



Structural, biosynthetic and serological interaction studies of the capsular polysaccharides from *Streptococcus pneumoniae*

Li, Chengxin

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Li, C. (2020). *Structural, biosynthetic and serological interaction studies of the capsular polysaccharides from Streptococcus pneumoniae*. Technical University of Denmark.

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.

**Structural, biosynthetic and serological interaction
studies of the capsular polysaccharides from
*Streptococcus pneumoniae***

PhD Thesis

Chengxin Li



Department of Chemistry

Technical University of Denmark

December 2020

PhD Supervisors:

Professor Jens Øllgaard Duus

Associate Professor Charlotte H. Gotfredsen

“Structure without function is a corpse; function without structure is a ghost.”

(Vogel and Wainwright, 1969)

Preface

This dissertation summarizes the work from my PhD studies at the Technical University of Denmark (DTU) from October 2017 to December 2020 under supervision of Professor Jens Øllgaard Duus and Associate Professor Charlotte H. Gotfredsen. A three months external stay was conducted at Research Center Borstel - Leibniz Lung Center in Dr. Katarzyna Duda's group. The project was financially supported by the China Scholarship Council (No.201708310119) and Oticon Fonden supported my external stay. The main works of my PhD project were performed at the DTU NMR center, supported by the Villum Foundation. In addition, the project is collaborating with SSI Diagnostica A/S, which provided pneumococcal polysaccharides and rabbit diagnostic antisera for studying.

The thesis consists of five chapters. The first chapter provides an introduction and background of the pathogen, *Streptococcus pneumoniae*, pneumococcal diseases, serotypes, the structure and biosynthesis of the capsular polysaccharides (CPS) and the motivation of the studied serogroups in this thesis. Chapter 2 introduces methods that were used in this study, including nuclear magnetic resonance (NMR) spectroscopy and other analytical methods for structural determination, bioinformatics tools for biosynthetic analysis and methods for serological interaction studies. Chapter 3 summarizes the results and discussions of CPS from serogroups 16 and 28, which is based on published paper 1 and paper 2 in the appendix. Chapter 4 is a part of the study of the discovery of a novel pneumococcal capsule type within serogroup 24. In Chapter 4, it focusses on the structural characterization of CPS from serogroup 24 and the novel serotype 24X. Chapter 5 explores the interactions between CPS and diagnostic antisera by diffusion NMR, NMR titration and AF4-MALS-dRI-UV.

The work presented in this thesis was performed by myself, except the general chemical analysis of CPS from serogroups 28 and 24 that was conducted by Dr. Katarzyna Duda at Research Center Borstel - Leibniz Lung Center.

The listed publications from 4 to 10 were from projects in collaboration with Andreas Eschenbacher, a former PhD student, from DTU chemical engineering. I was responsible for developing the NMR method for the quantitative analysis of bio-oil samples, which are not included in this thesis.

Acknowledgement

I would like to express my great appreciation to my supervisor Professor Jens Øllgaard Duus for giving me the opportunity to start my PhD at DTU chemistry and all the support during the last three years. I thank him for his trust, inspiration and encouraging me to explore new fields, which makes me a curious learner. I am truly grateful to Christian Kjeldsen for his guidance, inspiring discussions and very helpful and detailed feedback for my manuscripts and the thesis. It has been a great pleasure to work with him.

I would like to thank my co-supervisor Associate Professor Charlotte H. Gotfredsen for constructive guidance and discussions of my thesis. She helped me out when there were problems with the spectrometer and when I got confused in my project. I am genuinely thankful to Kasper Enemark-Rasmussen for his enormous support and training in NMR spectroscopy. His patience and excellent assistance in NMR acquisition and data analysis were of great help during my PhD study. I thank Sebastian Meier for useful discussions, encouragement, proofreading of my thesis and the humor that he brings to the working environment.

I would like to express my thanks to Dr. Katarzyna Duda and her group members at Research Center Borstel - Leibniz Lung Center for collaborations and for their professional training, patience and kindness during my external stay in Germany. I am thankful for Pernille Landsbo Elverdal and Ian Christian Skovsted at SSI diagnostica, Hillerød, for great collaboration of the project. They generously provided all the samples and supported me to do serological experiments in their lab. A big thank to David Tezé is for inspiring and fruitful discussions and great collaboration, which added the new perspective to the project.

I am grateful to Mariusz Kubus and Tao Jiang for instructing me of AF4-MALS analysis and helpful discussions. I thank David Frej Nielsen and other colleagues at the department for helpful technical and practical support. Kajta and Anne in the NMR group and other colleagues at DTU chemistry were thanked for a pleasant working environment. I acknowledge the DTU NMR center for all the instrumental access, the China Scholarship Council for supporting my PhD study and Oticon Fonden for supporting my external stay.

I would like to express my deep thanks to my parents for their continuous support and the love, freedom and trust that they have given to me. I want to thank my families and friends in China for always being supportive. I would also like to show my gratitude to my friends in Denmark and Germany, especially my friends at Østervold for all the hygge and memorable moments and for “training” me into Danish culture. It has been a wonderful experience for me to study and live in Denmark and I will always carry it with me.

Abstract

The rapid evolution of *Streptococcus pneumoniae* in response to clinical interventions, such as treatment with antibiotics and vaccination, has resulted in the change of prevalence in the carriage and invasive serotypes as well as frequent emergence of multidrug-resistant strains and novel serotypes. Thus, there are increasing needs in developing better diagnostic methods for the surveillance of *Streptococcus pneumoniae* and for gaining knowledge of non-vaccine serotypes for the development of future vaccines.

The capsular polysaccharides (CPS) are crucial virulence factors and the important antigen of *Streptococcus pneumoniae*. The structural elucidation of unknown pneumococcal CPS can provide insights into the function of CPS biosynthetic genes and enzymes, which is a significant addition to diagnosis based on gene sequencing methods. Moreover, the diagnostic antisera for pneumococcus are relying on serological reactivity of pneumococcal capsule, which also depends on CPS structures. Therefore, structure elucidation of pneumococcal CPS is fundamental for the development of better diagnostic methods and future pneumococcal vaccines. In this project, six previously unknown CPS structures of the pneumococcal serogroups 16, 24 and 28 (serotype 16F, 16A, 24A, 24B, 28F and 28A) and the newly discovered serotype 24X from serogroup 24 were thoroughly characterized by nuclear magnetic resonance (NMR) spectroscopy and other analytical methods.

A further CPS biosynthetic analysis to correlate CPS structures with biosynthetic enzymes revealed the presumable function of glycosyltransferases (GTs), acetyltransferases, phosphotransferases and polymerases. The functions of GT WcxN and WcxT from serogroup 16 were proposed for the first time. The proposed function of GT WcxN is in agreement with determined CPS structures of serotype 28F and 28A, which also possess GT WcxT. Moreover, the GT WciU in serotypes 28F and 28A showed different sugar donor specificity toward α -D-Glcp and α -D-GlcNAcp, respectively. The critical residue position that could be responsible for the substrate specificity of WciU was predicted through bioinformatics tools and protein 3D structure homology modeling.

In the serological study, the classic Quellung reaction and latex agglutination were applied to test the cross-reactions of related serotypes and the serological activities of partially degraded CPS. In addition, interactions between CPS and diagnostic antisera were studied by asymmetric flow field-flow fractionation coupled with multiangle light scattering, differential refractive index and ultraviolet detectors (AF4-MALS-dRI-UV), NMR titration and diffusion

ordered spectroscopy (DOSY), which provides alternative methods with potential to be used for the evaluation of diagnostic antisera and for studying CPS-antibody interactions.

Resumé

Den hurtige evolution af *Streptococcus pneumoniae* som følge af kliniske tiltag, såsom behandling med antibiotika og vaccination, har ført til ændring af prævalens i serotyper samt øget frekvens i fremkomsten af multiresistente stammer og nye serotyper. Derfor er der et stigende behov for udvikling af bedre diagnostiske værktøjer for at overvåge *Streptococcus pneumoniae* og øge forståelsen af serotyper der ikke er dækket af nuværende vacciner for at forbedre fremtidige vacciner.

Kapsel polysakkarider (CPS) er en essentiel virulensfaktor og et vigtig antigen af *Streptococcus pneumoniae*. Strukturoptæklaringen af ukendte CPS kan give indsigt i funktionerne af CPS relaterede biosyntetiske gener og enzymer, som er vigtige for diagnose baseret på gensekventerings metoder. Ydermere, diagnostiske antisera for pneumokok afhænger af den serologiske reaktivitet af pneumokok kapslen, som dermed også afhænger af CPS strukturen. Derfor er strukturoptæklaringen af CPS fundamental vigtig for udviklingen af nye diagnostiske værktøjer samt fremtidige vacciner. I dette projekt er seks serotyper med tidligere ukendte CPS strukturer fra serogruppe 16, 24 og 28 (serotype 16F, 16A, 24A, 24B, 28F og 28A) samt den nyopdagede serotype 24X fra serogruppe 24 blevet karakteriseret med kernemagnetisk resonans (NMR) spektroskopi samt med andre analysemetoder.

Dertil er der også lavet en CPS biosyntetisk analyse for at korrelere CPS struktur med biosyntesegener som afslører den formodede funktion af glykosyltransferaser, acetyltransferaser, phosphotransferaser og polymeraser. Funktionerne af glykosyltransferaserne WcxN og WcxT fra serogruppe 16 blev bestemt for første gang, og den tilordnede funktion stemmer overens med de karakteriserede strukturer af CPS fra serotype 28F og 28A. Ydermere viste glykosyltransferasen WciU fra serotype 28F og 28A forskellig donor specificitet mod henholdsvis α -D-Glc og α -D-GlcNAc, og den kritiske aminosyre position der kunne være ansvarlig for substratspecificitet i WciU blev forudsagt ved hjælp af bioinformatikværktøjer og 3D proteinstruktur modellering af homologe strukturer.

I serologistudiet benyttedes den klassiske Quellung reaktioner samt latexagglutination til at undersøge krydsreaktionerne af relaterede serotyper og serologisk aktivitet af delvist nedbrudt CPS. Derudover blev interaktionerne mellem CPS og diagnostisk antisera også undersøgt med asymmetric flow field-flow fractionation koblet med multiangle light scattering, differential refractive index og ultraviolet detektorer (AF4-MALS-dRI-UV), NMR titrering og diffusion ordered spectroscopy (DOSY), som udgør alternative metoder med

potentiale for at blive brugt til at evaluere diagnostisk antisera og studere CPS-antistof interaktioner.

Publications

1. Li C, Duda KA, Elverdal PL, Skovsted IC, Kjeldsen C, Duus JØ. 2019. Structural, Biosynthetic, and Serological Cross-Reactive Elucidation of Capsular Polysaccharides from *Streptococcus pneumoniae* Serogroup 16. *J Bacteriol* 201:1–13.
2. Li C, Duda KA, Elverdal PL, Skovsted IC, Kjeldsen C, Teze D, Duus JØ. 2021. Structural, biosynthetic and serological cross-reactive elucidation of capsular polysaccharides from *Streptococcus pneumoniae* serogroup 28. *Carbohydr Polym* 254:117323.
3. Maruhn K, Li C, Ganaie FA, Porambo R, Abeygunawardana C, Elverdal PL, Duda KA, Jens. Duus JØ, Van der Linden MP, Sheppard CL, Nahm MH. "Discovery of a novel pneumococcal capsule type within serogroup 24. (Manuscript in preparation).
4. Eschenbacher A, Saraeian A, Jensen PA, Shanks BH, Li C, Duus JØ, Smitshuysen TEL, Damsgaard CD, Hansen AB, Kling KI, Mentzel UV, Henriksen UB, Ahrenfeldt J, Jensen AD. 2020. Deoxygenation of wheat straw fast pyrolysis vapors over Na-Al₂O₃ catalyst for production of bio-oil with low acidity. *Chem Eng J* 394:124878.
5. Eschenbacher A, Saraeian A, Shanks BH, Jensen PA, Li C, Duus JØ, Hansen AB, Mentzel UV, Henriksen UB, Ahrenfeldt J, Jensen AD. 2020. Enhancing bio-oil quality and energy recovery by atmospheric hydrodeoxygenation of wheat straw pyrolysis vapors using Pt and Mo-based catalysts. *Sustain Energy Fuels* 4:1991–2008.
6. Eschenbacher A, Myrstad T, Bech N, Duus JØ, Li C, Jensen PA, Henriksen UB, Ahrenfeldt J, Mentzel UV, Jensen AD. 2020. Co-processing of wood and wheat straw derived pyrolysis oils with FCC feed—Product distribution and effect of deoxygenation. *Fuel* 260:116312.
7. Eschenbacher A, Jensen PA, Henriksen UB, Ahrenfeldt J, Jensen CD, Li C, Enemark-Rasmussen K, Duus JØ, Mentzel UV, Jensen AD. 2020. Catalytic upgrading of tars generated in a 100 kWth low temperature circulating fluidized bed gasifier for production of liquid bio-fuels in a polygeneration scheme. *Energy Convers Manag* 207:112538.
8. Eschenbacher A, Jensen PA, Henriksen UB, Ahrenfeldt J, Ndoni S, Li C, Duus JØ, Mentzel UV, Jensen AD. 2019. Catalytic deoxygenation of vapors obtained from ablative fast pyrolysis of wheat straw using mesoporous HZSM-5. *Fuel Process Technol* 194:106119.
9. Eschenbacher A, Jensen PA, Henriksen UB, Ahrenfeldt J, Li C, Duus JØ, Mentzel UV, Jensen AD. 2019. Deoxygenation of Wheat Straw Fast Pyrolysis Vapors using HZSM-5, Al₂O₃, HZSM-5/Al₂O₃ Extrudates, and Desilicated HZSM-5/Al₂O₃ Extrudates. *Energy & Fuels* 33:6405–6420.
10. Eschenbacher A, Jensen PA, Henriksen UB, Ahrenfeldt J, Li C, Duus JØ, Mentzel UV, Jensen AD. 2019. Impact of ZSM-5 Deactivation on Bio-Oil Quality during Upgrading of Straw Derived Pyrolysis Vapors. *Energy & Fuels* 33:397–412.

List of Abbreviations

| | |
|--------------------|--|
| 1D | One-dimensional |
| 2D | Two-dimensional |
| 3D | Three-dimensional |
| AATGalp | 2-acetamido-4-amino-2,4,6-trideoxy-galactopyranoside |
| Ab | Antibody |
| ACT | Artemis comparison tool |
| AF4 | Asymmetric flow field-flow fractionation |
| Ara-ol | Arabinitol |
| ax | Axial |
| BLAST | Basic local alignment search tool |
| CAZy | Carbohydrate-active enzymes database |
| CDC | Centers for disease control and prevention |
| CDP | Cytidine diphosphate |
| CDS | Coding sequence |
| Cho | Choline |
| CLIP | Clean in-phase |
| COSY | Correlation spectroscopy |
| CPS | Capsular polysaccharide |
| CRM197 | Cross-reactive material 197 |
| CWPS | Cell wall polysaccharide |
| dRI | Differential refractive index |
| De-OAc | De-O-acetylation |
| De-PO ₄ | Dephosphorylation |
| DOSY | Diffusion ordered spectroscopy |
| DQF | Double quantum filtered |
| EIA | Enzyme immunoassay |
| ELISA | Enzyme-linked immunosorbent assay |
| eq | Equatorial |
| <i>f</i> | Furanoside (for example Gal <i>f</i>) |
| FF | Force field |
| FID | Free induction decay |
| FucNAc | N-Acetyl Fucosamine |
| Gal | Galactose |
| GalA | Galacturonic acid |
| GalN | Galactosamine |
| GalNAc | N-acetyl galactosamine |

| | |
|------------|--|
| GC | Gas chromatography |
| GLC | Gas-liquid chromatography |
| Glc | Glucose |
| GlcA | Glucuronic acid |
| GlcN | Glucosamine |
| GlcNAc | <i>N</i> -acetyl glucosamine |
| Gro | Glycerol |
| Gro-1-P | Glycerol-1-phosphate |
| Gro-2-P | Glycerol-2-phosphate |
| GTs | Glycosyltransferases |
| Glc-T | Glucosyltransferase |
| GlcNAc-T | <i>N</i> -acetylglucosaminyltransferases |
| H2BC | Heteronuclear two-bond correlation |
| HGs | Homology groups |
| HMBC | Heteronuclear multiple-bond correlation |
| HSQC | Heteronuclear single quantum coherence |
| IgG | Immunoglobulin G |
| IPD | Invasive pneumococcal diseases |
| MALS | Multiangle light scattering |
| Man-ol-6-P | Mannitol-6-PO ₄ |
| Man | Mannose |
| ManNAc | <i>N</i> -acetyl mannosamine |
| Mn | Number-average molecular weight |
| MS | Mass spectrometry |
| Mw | Weight-average molecular weight |
| MSA | Multiple sequence alignment |
| NDP | Nucleoside diphosphate |
| NMR | Nuclear magnetic resonance |
| NOE | Nuclear Overhauser effect |
| NOESY | Nuclear Overhauser effect spectroscopy |
| NRD | Nucleotide recognition domain |
| OAc | O-Acetyl |
| <i>p</i> | Pyranoside |
| PBS | Phosphate-buffered saline |
| P-Cho | Choline-1-phosphate (phosphorylcholine) |
| PCR | Polymerase chain reaction |
| PCV | Protein-conjugate vaccine |

| | |
|-----------------|--|
| PFG | Pulsed-field gradients |
| PG | Peptidoglycan |
| PneNAc | N-acetylpneumosamine (2-acetamido-2,6-dideoxytalose) |
| PO ₄ | Phosphate |
| PPV | Pneumococcal polysaccharide vaccine |
| ppm | Parts per million |
| PS | Polysaccharides |
| QMQE | Global model quality estimation |
| QMEAN | Qualitative model energy analysis |
| RF | Radio-frequency |
| Rha | Rhamnose |
| Rib | Ribose |
| Rib-ol | Ribitol |
| RMS radius | Root mean square radius |
| Rn | Number-average mean square radius |
| ROE | Rotating Overhauser effect |
| ROESY | Rotating Overhauser effect spectroscopy |
| Rw | Weight-average mean square radius |
| SEC | Size exclusion chromatography |
| TOCSY | Total correlation spectroscopy |
| UV | Ultraviolet |
| UDP | Uridine diphosphate |
| Und-P | Undecaprenyl-phosphate |
| WGS | Whole genome sequence |

Table of content

| | |
|---|-------------|
| Preface | III |
| Acknowledgement | IV |
| Abstract | V |
| Resumé | VII |
| Publications | IX |
| List of Abbreviations | X |
| Table of content..... | XIII |
| Chapter 1 Introduction | 1 |
| 1.1 Streptococcus pneumoniae: diseases and virulence | 2 |
| 1.2 Serotype..... | 4 |
| 1.2.1 Serotype definition..... | 4 |
| 1.2.2 Serotyping..... | 5 |
| 1.3 Pneumococcal vaccines | 7 |
| 1.4 Serotype replacement..... | 9 |
| 1.5 CPS structures and biosynthesis | 10 |
| 1.5.1 CPS Chemical structures | 10 |
| 1.5.2 CPS biosynthesis..... | 11 |
| 1.6 Serogroups 16, 28 and 24 | 15 |
| Chapter 2 Methodology and experiments | 17 |
| 2.1 Structure elucidation of CPS | 17 |
| 2.1.1 NMR spectroscopy | 17 |
| 2.1.2 General chemical analysis for carbohydrates | 24 |
| 2.1.3 AF4-MALS-dRI-UV for characterizing CPS and CPS-antiserum interactions | 25 |
| 2.2 Biosynthetic analysis | 27 |
| 2.2.1 Bioinformatics | 27 |
| 2.2.2 Glycosyltransferases structure modeling | 29 |
| 2.3 Serological interaction study | 32 |
| 2.3.1 Quellung reaction..... | 32 |
| 2.3.2 Latex agglutination..... | 32 |
| 2.3.3 ¹ H NMR Titration for study CPS - antiserum interactions..... | 33 |
| 2.3.4 Diffusion NMR..... | 34 |

| | | |
|--|---|----|
| Chapter 3 | Structural, biosynthetic and serological cross-reactive studies of CPS from serogroups 16 and 28 | 36 |
| 3.1 | Structural elucidation of CPS repeat units from serogroups 16 and 28 | 36 |
| 3.1.1 | Molecular weight distribution measured by AF4-MALS-dRI | 36 |
| 3.1.2 | General chemical analysis of CPS | 38 |
| 3.1.3 | NMR analysis of serotype 16F, 16A, 28F and 28A CPS | 39 |
| 3.2 | Biosynthesis of serogroups 16 and 28 | 41 |
| 3.2.1 | The cps loci comparison of serotype 16F, 16A, 28F and 28A | 41 |
| 3.2.2 | Correlation of CPS structures with biosynthesis of serotype 16A, 16F, 28F and 28A | 43 |
| 3.2.3 | The donor specificity of WciU | 46 |
| 3.3 | serological cross-reactions of serogroups 16 and 28 | 51 |
| 3.3.1 | Serological cross-reactive analysis by Quellung reaction | 51 |
| 3.3.2 | Serological cross-reactive analysis by latex agglutination | 52 |
| Chapter 4 | A discovery of a new serotype in serogroup 24 | 53 |
| 4.1 | Structure analysis of CPS from serogroup 24 | 53 |
| 4.1.1 | General chemical analysis | 53 |
| 4.1.2 | NMR analysis of repeat unit structures of serotype 24B and 24A CPS | 54 |
| 4.1.3 | NMR analysis of repeat unit structure of serotype 24X CPS | 61 |
| Chapter 5 | Serological interactions study of CPS with diagnostic antisera | 65 |
| 5.1 | Study of interaction between 16A CPS and diagnostic antisera by DOSY | 66 |
| 5.2 | NMR Titration of 16A CPS with 16c antiserum | 69 |
| 5.2.2 | NMR Titration of 16A CPS with 16c antiserum | 69 |
| 5.3 | Characterization of 16c factor antiserum by AF4-MALS-dRI-UV | 73 |
| 5.3.1 | Evaluation of CPS binding antibodies in 16c factor antiserum | 74 |
| 5.4 | Summary | 78 |
| Conclusion and future prospective | | 79 |
| References | | 81 |
| Appendix | | 93 |

Chapter 1 Introduction

The fight against bacterial infections in humans and feedstock is still a major challenge in our society. The clinical interventions, such as applying antibiotics, have achieved great success and saved millions of lives. However, abusive use of antibiotics speeds up the process of bacteria evolution and emergence of antibiotic-resistant strains ¹⁻⁵. It has therefore become urgent to find new antibiotics or to develop alternative strategies for disease prevention and treatment, such as vaccines. Both vaccination and diagnosis of bacteria rely primarily on the immune recognition of bacterial surface polysaccharides (PS). Thus, it is of significance to get a better understanding of bacterial PS for the development of better diagnostic tools and vaccines.

Streptococcus pneumoniae has remained a major cause of morbidity and mortality globally among bacterial infectious disease, especially in infants and children. It is a highly recombinogenic bacterium, with genes encoding major antigens frequently affected ⁶. Rapid evolution of pneumococcus has been reported in response to clinical interventions, which has resulted in frequent emergence of multidrug-resistant pneumococcal strains, novel serotypes and serotype replacement of non-vaccine serotypes ^{1,7-9}. A clear understanding of pneumococcal virulence factors, especially the capsular polysaccharides (CPS), their biosynthesis and antigenic activities, and examining the current diagnostic techniques and vaccine strategies will allow for better surveillance and regulation of the pathogen and its diseases. Therefore, there are increased needs to learn more about the structure, biosynthesis and antigenic property of the main pneumococcal virulence factor CPS. *Streptococcus pneumoniae* is genetically well characterized with genetic analysis of *cps* loci from 90 serotypes and many genetic surveillance data available. However, there are still large gaps in the knowledge of CPS structures.

The knowledge of CPS structure is fundamental for understanding the relationships between genotype and phenotype. The structural characterization of unknown pneumococcal CPS can provide insights into the function of biosynthetic genes, which is significant for diagnosis that is based on gene sequence methods. Moreover, the immunological properties of PS are based on the chemical structures. Therefore, the elucidation of CPS structures can provide fundamental knowledge about the antigenic properties of bacterial PS. Consequently, structure elucidation of pneumococcal CPS would be useful for the development of better diagnostic methods and formulating future pneumococcal vaccines.

This PhD project focused on the structural characterization of previously unknown pneumococcal CPS by NMR spectroscopy and verified by classical analytical methods. A further correlation of determined CPS structures with *cps* genes to assign the function of related biosynthetic genes and enzymes was pursued. In addition, a serological interaction study of CPS with diagnostic antiserum was conducted in an attempt to explore the antigenic properties of CPS.

1.1 *Streptococcus pneumoniae*: diseases and virulence

Streptococcus pneumoniae (the pneumococcus) is an encapsulated, gram-positive, aerobic, continuously evolving human pathogen that commonly colonizes the human nasopharynx. The pneumococcal carriage rates range from 10 to 80% largely depending on the age of the population and is particularly higher in children. Pneumococcus can cause non-invasive diseases such as otitis media, and invasive pneumococcal diseases (IPD), like pneumonia, meningitis and sepsis ¹⁰. The mortality and morbidity of IPD are higher in children and the elderly. Pneumococcus was estimated to cause 735000 deaths in HIV-negative children, accounting for 8-12% of all deaths in children under five years old, especially in poor countries (Figure 1.1) ^{1,11-13}.

Disease Burden ► Pneumococcal Cases ► Number of Cases - Total

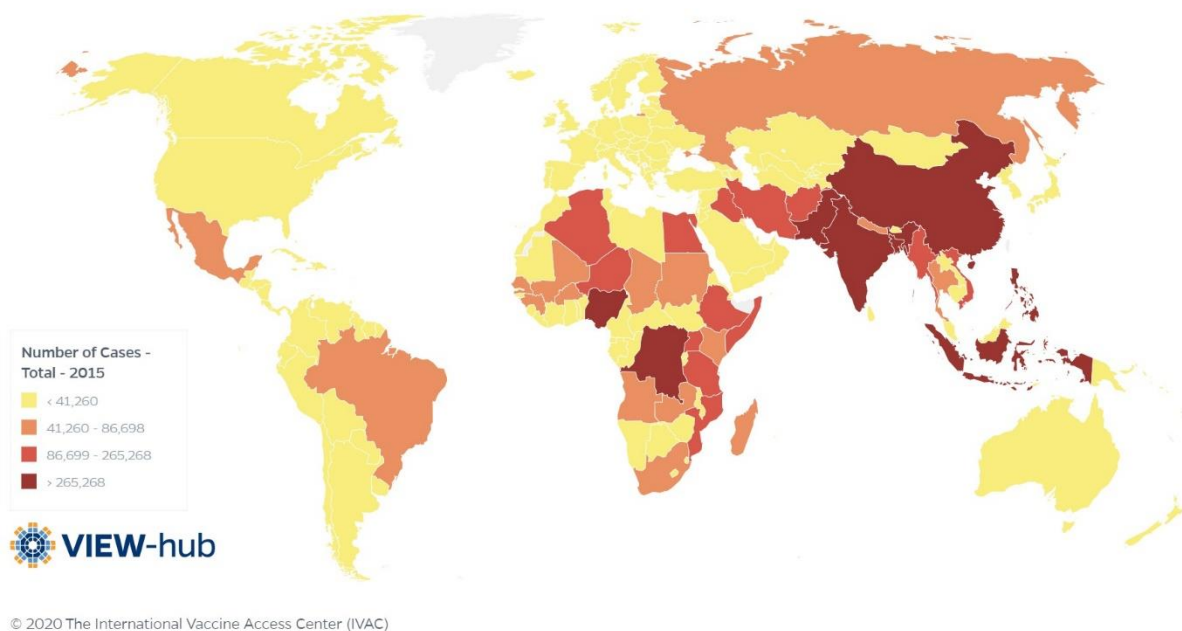


Figure 1.1 Total number of *Streptococcus pneumoniae* cases for a country in 2015. (International Vaccine Access Center (IVAC), Johns Hopkins Bloomberg School of Public

Health. Reprinted with permission from VIEW-hub, www.view-hub.org, modeled estimates from B. Wahl et al. ¹¹.)

Streptococcus pneumoniae displays a large selection of virulence factors including polysaccharides, surface proteins, excreted proteins and cytoplasmic proteins. These virulence factors help the bacteria in adherence, invasion of host tissues and evading the host immune defenses. Among all the known virulence factors, the surface components of pneumococcus play an important role in the pathogenicity of pneumococcus. There are two main layers outside of the plasma membrane, the cell wall and the capsule.

The cell wall consists of the peptidoglycan (PG) backbone, the cell wall polysaccharide (CWPS) and different surface proteins. CWPS is covalently linked to PG and is also referred to as teichoic acid or lipoteichoic acid with a lipid anchor on the polymer. There are three different structures of CWPS reported among all the pneumococcal serotypes (Table A1)
^{14,15}.

The capsule of pneumococcus, a thick polymer layer ranging from 200 to 400 nm (Figure 1.2), completely conceals the inner structures and has been recognized as the major virulence factor. The capsules are composed of high molecular weight polysaccharides, the capsular polysaccharides (CPS). The capsule plays an important role in providing resistance to phagocytosis by innate immune cells, preventing recognition by host receptors and complement factors, escaping from nasal mucus and neutrophil net traps and helping with adherence and colonization in the nasopharynx ^{10,16,17}. Most CPS have negative net charges, which is important for the pathogenesis of the capsule ¹⁸. The capsule interacts with negatively charged mucus and phagocytic cells, like macrophages, through electrostatic repulsion, which results in the reduction of mucociliary clearance of the bacteria ^{18,19}. The CPS are anchored on PG via a direct glycosidic bond to the 6-position of β -D-N-acetylglucosamine ²⁰. Unlike the CWPS that was mostly invariant among all pneumococcal isolates, the CPS showed great diversity in/of their chemical structures ¹⁵. The different CPS structures determine the serotype of the bacteria.

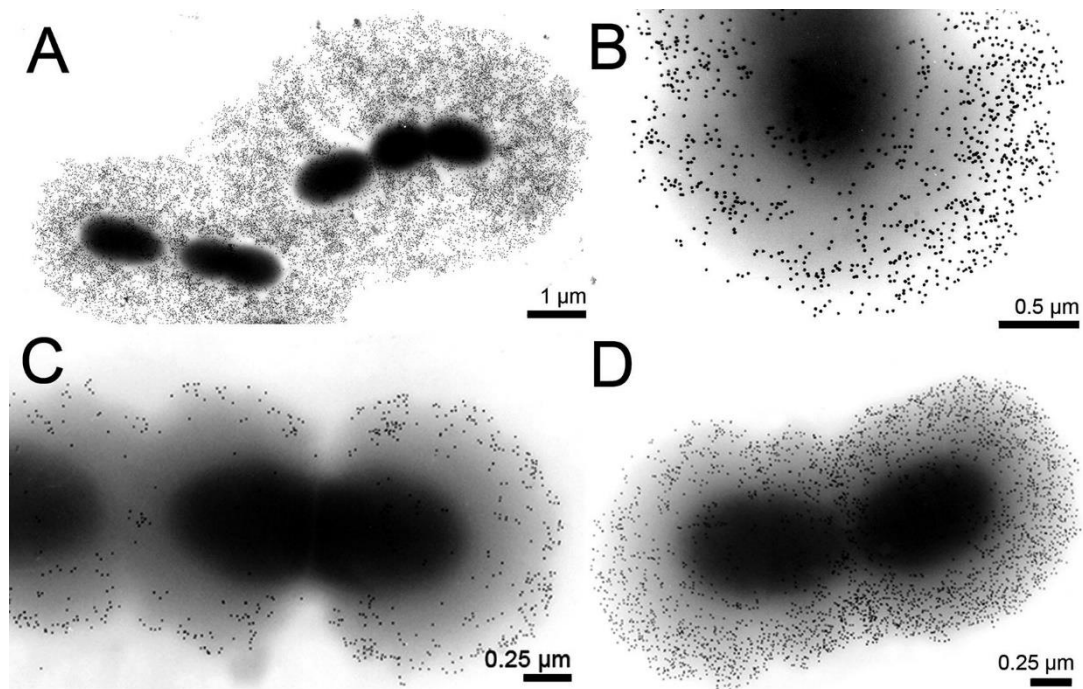


Figure 1.2 Visualization of pneumococcal CPS applying cationic gold-nanoparticles (small black dots surround the capsule). (A, B, D) Fixation of pneumococci in growth medium, pH shift to 3.0 with PBS, followed by incubation with 1:75 diluted stock solution of 15 nm cationic gold-nanoparticles or (D) 1: 200 diluted cationic gold-nanoparticles. Reprinted with permission from reference ²¹. Copyright © 2019, Springer Science Business Media, LLC, part of Springer Nature.

1.2 Serotype

1.2.1 Serotype definition

The pneumococci are classified into different serogroups and serotypes. A serotype is defined as pneumococcal strains producing a polysaccharide with unique chemical structure and serologic (immunologic) properties. A serogroup is defined to include serotypes that share many serologic properties (i.e., cross-reactive antibodies) ¹⁵. There are two different nomenclature systems for pneumococcal serotypes: the America system and the Danish system ²². The Danish system is based on cross-reactions between different types and the serologically cross-reactive types are assigned to a common serogroup. Then the individual serotypes within each serogroup are distinguished by the trailing letter, usually starting with “F” referring to “First” discovered in the serogroup. The American system numbers the serotypes sequentially according to the order of their discovery. The Danish nomenclature

system is widely accepted, as it classifies serotypes with similar antigenic properties ²³. As different pneumococcal serotypes have different epidemiology and different invasive potentials, identifying the serotype is critical for providing therapy and prevention of pneumococcal disease. The death of Danish Prince Valdemar in 1939 illustrated the importance of determination of serotypes when treating patients with serotype-specific antisera. His pneumonia was caused by a new serotype, at the time, in serogroup 9 and the serotype was named 9V after him. After Prince Valdemar's case, many serotyping methods were developed, including the precipitin test, agglutination test and the Quellung reaction ^{24–26}.

To date, there are around 100 different serotypes reported and each serotype has a unique CPS structure. However, with the increased understanding of pneumococcal CPS biosynthesis and recombination, additional serotypes were discovered and there might be more new serotypes to be discovered, still ^{27–31}. A new serotype should have a novel CPS structure and a stable genetic basis. Genetic variations are often not enough to produce a different capsule ^{32,33}. For example, the putative novel serotype 6E turned out to be a genetic variant of serotype 6B, which produced identical CPS ³⁴. Therefore, it is necessary to combine structural characterization with genotyping and serotyping when describing new serotypes.

Although the capsule can increase the virulence of pneumococci, only 20 to 30 serotypes out of near 100 serotypes show significant invasiveness. The ratio of IPD incidence and mortality are strongly associated with serotypes ^{9,35,36}. Moreover, different serotypes show different prevalence among different populations regarding to their countries and ages. It is likely that the interactions of pneumococci with host immune system vary with their capsule types. Therefore, the surveillance of serotype distribution in the given population is important for implementing effective pneumococcal vaccines.

1.2.2 Serotyping

The “gold standard” method for serotyping is the Quellung reaction (also known as Neufeld test), which uses specific rabbit antisera to detect the pneumococcal capsule type ³⁷. The Quellung reaction is an antigen-antibody reaction, which is normally performed using commercially available antisera. SSI Diagnostica A/S is the commercial source of diagnostic antisera to identify pneumococcal capsular types. Pool, group, type and factor antisera are used in a specific order during serotyping and the ultimate identification of the serotypes in the serogroup is carried out by serotyping with factor antisera. The combination of the results defines the serotype (<https://www.ssidiagnostica.com/neufeld-antisera/?cid=36>). A positive

reaction can be seen when antibodies in the serum binding to the capsule results in the visible capsule swelling phenomenon ³⁸. However, this method requires specific expertise and bacterial culture. Recently, SSI Diagnostica A/S has made a latex agglutination test available, known as ImmuLex™ Pneumotest Kit, which does not require a microscope for serotyping pneumococci. The rabbit diagnostic antisera are purified and coated onto blue latex particles. The positive reaction can be observed within 5-10 seconds after mixing equal quantities of antisera coated latex solution with bacterial culture. In recent decades, more molecular typing methods have emerged, such as serotyping based on the enzyme-linked immunosorbent assay (ELISA), a competitive enzyme immunoassay (EIA) and flow cytometry ^{39–42}.

Considering the extensive laboratory work when trying to serotype enormous pneumococcal clinic isolates, several “genetic serotyping” methods have been developed to identify capsule type from whole genome data based on polymerase chain reaction (PCR) techniques ^{26,43}. With increasing adoption of high-throughput next-generation sequencing for bacterial diagnosis, *in silico* serotyping has largely overtaken the standard serotyping methods and will most likely become the routine method of choice in the future, especially following the SARS-COVID19 pandemic. For example, the researchers at Public Health England developed an automated whole genome sequence (WGS)-based serotyping bioinformatics tool, PneumoCaT (Pneumococcal Capsule Typing) ⁴⁴. Although genetic serotyping has made serotyping simpler and available to laboratories worldwide, phenotypic methods are currently still necessary for pneumococcal serotyping. There are two reasons why the genetic serotyping is not a perfect predictor of phenotype ⁴⁵. Firstly, genetic mutations within capsule synthetic genes can alter specificity of encoded enzymes, thus producing a new serotype. The gap between serologically-determined diversity and genetically-determined diversity is widening. Secondly, additional genes located outside the CPS synthesis locus may also affect the final CPS structure. Therefore, it is necessary to get a better understanding of CPS biosynthetic genes and the specificity of the related enzymes to improve the accuracy of genetic methods for identification of serotypes.

1.3 Pneumococcal vaccines

The vaccines provide serotype-specific protection against pneumococcal infections. The introduction of pneumococcal vaccines is an important prevention of pneumococcal diseases among the population. Due to the huge burden of pneumococcal diseases on public health, it is necessary to develop and implement an effective vaccine to a given community or population. There are currently two types of pneumococcal vaccine: the pneumococcal polysaccharide vaccine (PPV) and the pneumococcal conjugate vaccines (PCV). The available pneumococcal vaccines and vaccines under development are summarized in Table 2. The first trial of whole-cell pneumococcal vaccine was coordinated by Sir Almroth Wright in South Africa from 1911 to 1912 ⁴⁶. Although Wright did not realize the serotype specificity at that time, his evaluation of the first pneumococcal vaccine opened the chapter of developing pneumococcal vaccine for the improvement of public health. Later, pneumococcal capsular polysaccharides were isolated by Alphonse Dochez and Oswald Avery in 1916 ⁴⁷. The pneumococcal capsule as a critical virulence factor and the immunogenicity of capsular polysaccharides were established. The first clinical trial of quadrivalent (serotype 1, 2, 5 and 7) pneumococcal polysaccharide vaccine (PPV) was conducted in 1944-1945, which demonstrated high efficacy in soldiers ⁴⁸. However, the PPV was not widely used in clinic, as the postwar clinicians preferred to use antibiotics, like penicillin, to treat patients with pneumonia. However, the high mortality and treatment failures of *Streptococcus pneumoniae* made people realize the necessity of implementing pneumococcal vaccines.

PPV

In 1983, the PPV23 vaccine was introduced, which included 23 serotypes covering about 87% of pneumococcal disease in the US ⁴⁹. The PPV23 contains 25 µg per serotype CPS antigen, which is currently still on the market (Table 2). However, there are some limitations of PPV23. The polysaccharides mainly induce a B-cell-dependent immune response, which showed poor immunogenicity in children younger than 2 years of age, probably due to their immature immune system ^{50,51}. Vaccination of adults with PPV23 requires a revaccination after 5-6 years and the subsequent doses of PPV23 appear to induce low antibody concentrations compared to the primary vaccination ⁵². Despite the relatively restricted efficacy of PPV23, it is clinically useful in adults and for those >65 years of age ^{53,54}. In the US, the Centers for Disease Control and Prevention (CDC) recommends PPV23 for all adults 65 years or older, people 2 through 64 years old with certain medical conditions and adults 19 through 64 years old who smoke cigarettes

(<https://www.cdc.gov/pneumococcal/vaccination.html>, December 30, 2020).

PCV

The lack of efficacy of PPV23 in infants <2 years of age motivated the development of polysaccharide conjugated vaccine (PCV). PCVs are produced by chemically conjugating capsular polysaccharides to a carrier protein. The PPV elicits T-cell independent immune response, while PCV induces T-cell dependent antibody response and generation of memory B-cell ⁵². PCVs showed high immunogenicity in young children less than 2 years old. The first licensed PCV was PCV7, which was introduced in the US in 2000. The capsular polysaccharides of PCV7 are conjugated to a highly immunogenic cross-reactive material 197 (CRM197), a non-toxic diphtheria toxoid protein. The conjugation method, the protein carrier, the ratio of carrier protein to CPS and the dose of CPS could affect the immunogenicity of PCV ⁵⁵. There are currently three PCVs available (Table 2). In the US, the CDC recommends PCV13 for all children younger than 2 years old and for persons 2 years or older with certain medical conditions. In addition, adults 65 years or older may discuss and decide, with their clinician, to receive PCV13 (<https://www.cdc.gov/pneumococcal/vaccination.html>, December 30, 2020).

Table 2 Current licensed pneumococcal vaccines and vaccines under development

| Vaccine (commercial name) | Valence: serotypes included | licensed year / Status |
|------------------------------|---|---------------------------|
| PPV23 (Pneumo 23) | 1, 2, 3, 4, 5, 6B, 7F, 8, 9N, 9V, 10A, 11A, 12F, 14, 15B, 17F, 18C, 19F, 19A, 20, 22F, 23F, 33F | 1983 |
| PVC7 (Pevnar) | 4, 6B, 9V, 14, 18C, 19F, 23F 6B, | 2000 |
| PCV10 (Synflorix) | PCV7 + 1, 5, 7F | 2008 |
| PCV13 (Pevnar 13) | PCV10 + 3, 6A, 19A | 2009 |
| PCV15 (Merck) | PCV13 + 22F, 33F | Phase 3 |
| PCV20 (Pfizer) | PCV13+ 8, 10A, 11A, 12F, 15BC, 22F, 33F | Phase 3 |
| PCV24 (Merck) | PCV13 + 2, 8, 9N, 10A, 11A, 12F, 15B, 17F, 20, 22F, 33F | Under development |

1.4 Serotype replacement

The success of PCV targeted for prevalent serotypes significantly reduced IPD, but resulted in rapid increase in some non-PCV13 serotypes and in the evolution of novel serotypes, which are compromising the benefits of current vaccines ^{1,11,12,56–60}. The introduction of PCVs has led to the emergence of the capsule-switch variants arising due to the vaccine derived selective pressures (Figure 1.3). To cover more serotypes and to adjust to the serotype replacement, multivalent PCVs that cover more serotypes are under development, such as PCV15, PCV20 and PCV24 ^{61,62}. The challenge is to maintain safety and immunogenicity while adding additional serotypes ⁶³. In addition, it is necessary to continue high-quality surveillance to monitor serotype distribution and emerging of novel serotypes to formulate future vaccines and implement better vaccine policies.

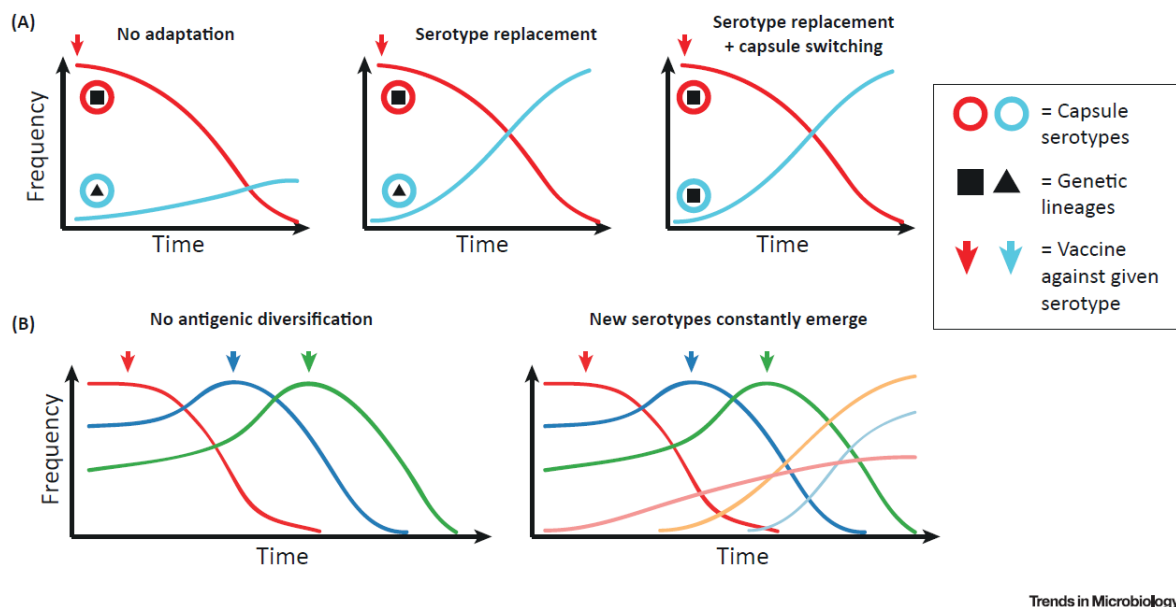


Figure 1.3 Mechanisms of bacterial adaptation against PCVs (A) Impact of PCVs on bacterial population structure. The latter two situations are frequently observed in *Streptococcus pneumoniae*. (B) Potential impact of antigenic diversification on multivalent vaccine strategies. On the left, it is assumed that serotypes do not diversify over time. On the right, new serotypes (light red, orange and light blue) constantly emerge at low frequencies and are selected for by the vaccine due to serotype replacement. Reprinted from an open access article ⁴⁵. Copyright © 2018 The Authors.

Pneumococcal capsule biosynthesis Locus *cps* was reported as evolutionary hotspot with potential to generate novel Serotypes by recombination ²⁹. As a consequence of faster evolution and serotype replacement of pneumococcus driven by vaccines, antibiotics and

environmental changes, the gap between serologically determined diversity and genetically determined diversity is widening. Furthermore, as genotyping gains popularity for diagnostic purposes, it is of significance to correlate genotype with phenotype. To fill up the gap, it is valuable to determine unknown capsule structures in order to correlate genotype with serotype for a better diagnosis.

1.5 CPS structures and biosynthesis

1.5.1 CPS Chemical structures

The polysaccharide diversity of pneumococcal capsules is a major challenge for eliminating pneumococcal disease ¹⁵. The CPS are high molecular weight carbohydrate polymers (polysaccharides), composed of oligosaccharide repeat units. The molecular weight of CPS ranges from 100 kDa to 2000 kDa, depending on the serotypes ^{64,65}. The heterogeneity of CPS is established through variation of monosaccharide composition and sequence of the repeat unit, type and configuration of the glycosidic linkages and substitutions. The CPS normally have negative net charges due to the presence of uronic acid, phosphate and pyruvate substitutions resulting in acidic polysaccharides, except for serotype 7F, 7A, 14, 33F and 37 that have no net charge ^{18,22}. The serotype 1 contains an equal positive and negative charge (zwitterionic charge). The repeat units of CPS usually contain 2 to 8 sugar residues (Figure 1.4). The monosaccharides that are most commonly present in CPS are D-Glcp, D-Galp, L-Rhap and D-GlcpNAc. Some serotypes contain D-GlcpA, D-ManpNAc, D-Manp, D-Galf, L-FucpNAc, D-Ribf, L-PnepNAc and sugar alcohols ¹⁵. CPS often contain labile groups, like O-Acetyl (OAc) and phosphate (PO₄), which can be removed during purification. The most unstable group is the OAc ester that often shows different locations and percentages. For example, the serotype 9V CPS has four OAc substituted sites with degrees of OAc varying from 0.1 %-100 % ⁶⁶. This partial O-Acetylation makes CPS even more heterogeneous. The OAc groups contribute to the configuration of CPS and play an important role in the antigenicity of CPS. CPS of more than half of the nearly 100 different serotypes contains PO₄. PO₄ can be found not only in the backbone of the polymer, but also in sidechains. Commonly, CPS from the same serogroup are highly similar with small differences in the structures. During the last three decades, many CPS structures have been determined. However, the structures of 18 serotypes are still unknown and the OAc substitutions of serotype 15F CPS are not assigned. To close the gap in CPS structural

information, we determined 7 CPS structures and one CPS from a newly discovered serotype.

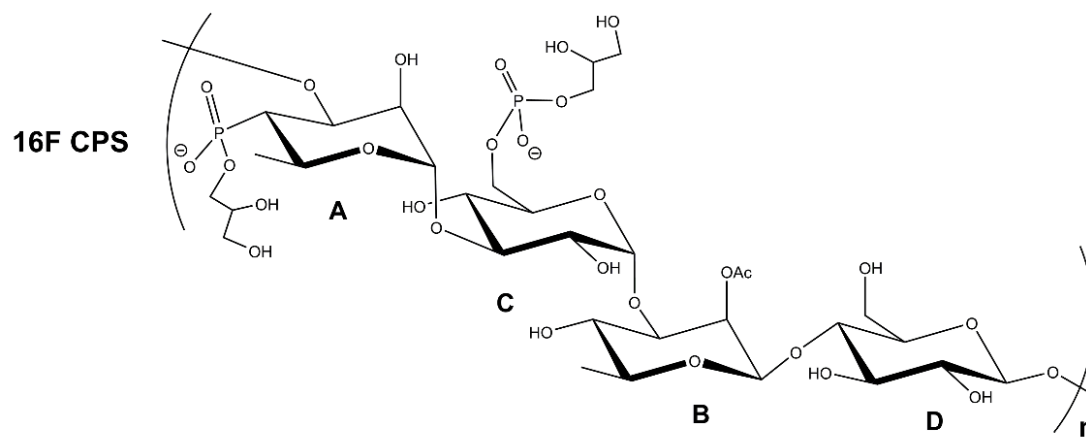


Figure 1.4 The repeat unit structure of serotype 16F CPS

Table 3 Serotypes with unknown CPS structures ¹⁵

| Serotype with unknown CPS structures | Structure information |
|---------------------------------------|---|
| 16F, 16A, 28F, 28A, 24A, 24B | Determined in this study |
| 21 | Constituents: Glc, Gal, and GlcN |
| 25F | Constituents: Glc, Rha, GlcN, Rib, and Rib-ol-P |
| 46 | Constituents: Gal, GalNAc, GlcNAc, and FucNAc |
| 12B, 22A, 25A, 36, 38, 40, 43, 44, 48 | None |

1.5.2 CPS biosynthesis

1.5.2.1 The CPS biosynthetic genes

The mechanisms of CPS synthesis in *Streptococcus pneumoniae* is primarily the Wzy dependent biosynthetic pathway, except for serotypes 3 and 37 that employ synthase dependent biosynthesis ^{67,68}. As the serotypes studied in this PhD thesis work are all using the Wzy-dependent pathway, we only introduce the Wzy-dependent biosynthesis. The genes that encode the CPS biosynthesis are in the *cps* locus, which is located in the pneumococcal

chromosome between *dexB* and *aliA* genes, with an average of 20,714 bp.^{67,69} The *wzg*, *wzh*, *wzd*, and *wze* genes (also known as *cps* ABCD) are regulatory genes that are highly conserved within all the serotypes. The downstream of *cps* locus encodes serotype specific biosynthetic genes, including the initial transferase genes, glycosyltransferase genes, the activated monosaccharide nucleotide sugar donor synthetic genes, flippase (*wzx*), polymerase (*wzy*) and other modifying enzyme genes, such as acetyl transferases and phosphotransferase (Figure 1.5).

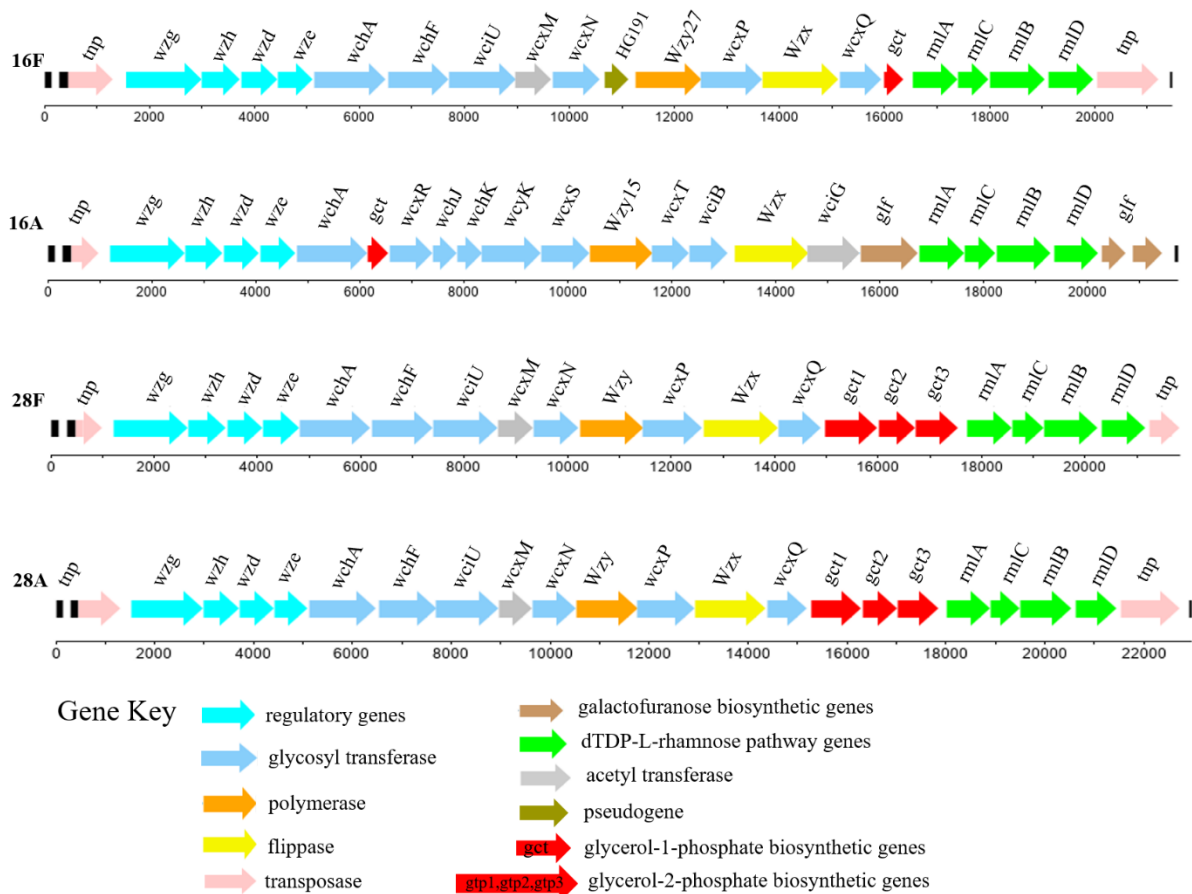
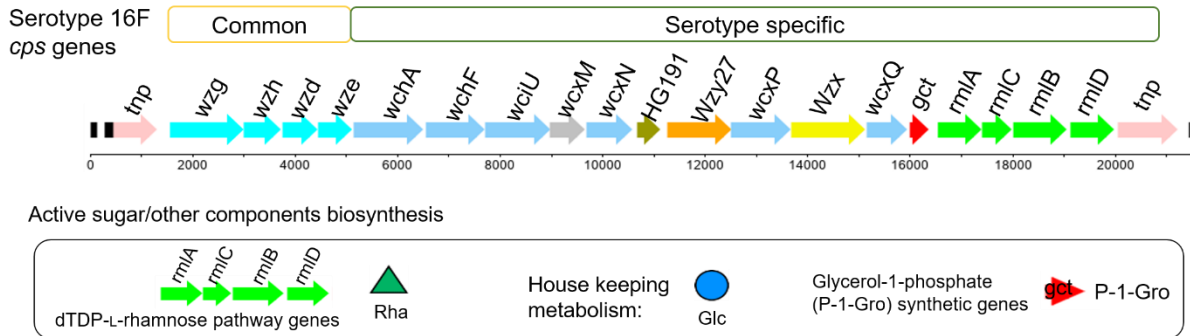


Figure 1.5 Serotype 16F, 16A, 28F and 28A cps loci

1.5.2.2 The Wzy dependent pathway

The Wzy dependent CPS biosynthesis includes synthesis of the precursors of sugars or other CPS components, transfer of an initial monosaccharide phosphate from a nucleotide diphosphate (UDP) sugar to a membrane-associated lipid carrier, sequential transfer of each sugar residue onto the lipid carrier by glycosyltransferases (GTs) to form the repeat unit. Then the repeat unit will subsequently be transported to the outer cytoplasmic membrane by

flippase (Wzx) and polymerized by the polymerase (Wzy). Finally, the CPS is exported and attached to the cell wall. Figure 1.6 is an example of Wzy-dependent biosynthesis of serotype 16F CPS.



Wzy-dependent CPS synthesis

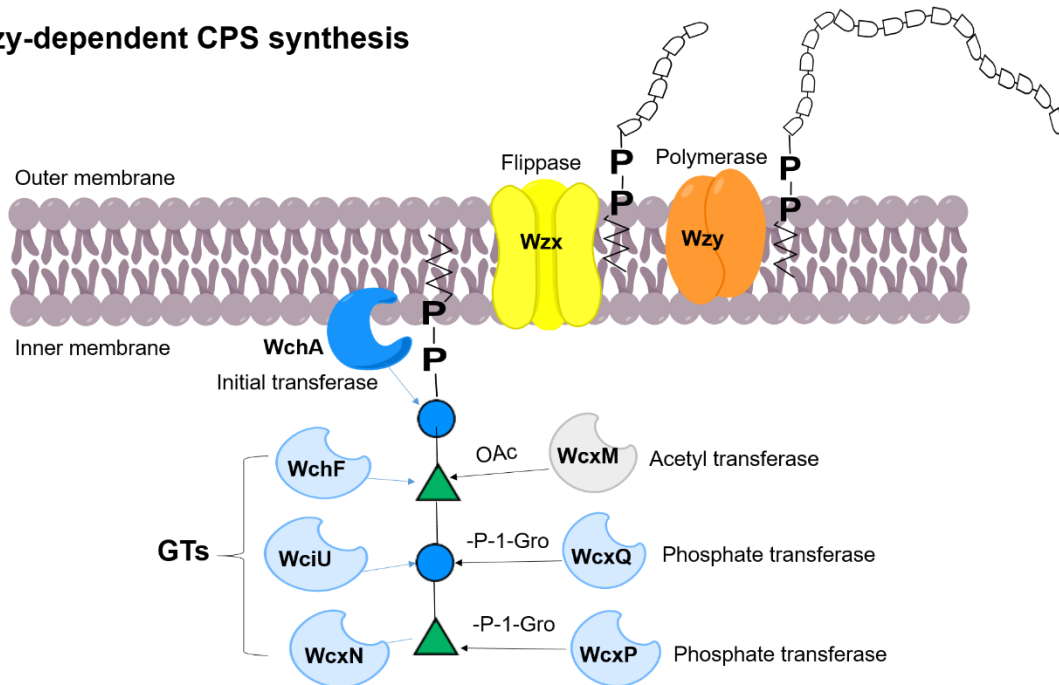


Figure 1.6 The Wzy-dependent biosynthesis of serotype 16F CPS.

For the biosynthesis of precursors for the sugar monomers and other components in the CPS, there are 7 out of 18 sugars available from housekeeping metabolic pathways⁶⁷. The non-housekeeping sugars are synthesized by the genes encoded in the *cps* locus. The precursors for the sugar alcohol phosphates, like arabinitol-1-PO₄ (Ara-ol-1-P), ribitol-1-PO₄ (Rib-ol-1-P) and mannitol-6-PO₄ (Man-ol-6-P) could be derived from respective sugar phosphates by two-step pathways. The synthesis of glycerol-1-phosphate (Gro-1-P) and glycerol-2-phosphate (Gro-2-P) are associated with *gct* gene and *gct1-3* genes respectively.

The precursor of choline-1-phosphate (Cho-P) is cytidine diphosphate (CDP)-Choline and can be produced during teichoic acid biosynthesis.

The synthesis of the repeat units starts with transferring the first UDP sugar to the C55 lipid undecaprenyl-phosphate (Und-P) by the initial transferase. There are mainly four initial transferase genes, *wchA*, *wciI*, *wcjG* and *wcjH*, present in all the known serotypes. Among the four initial transferases, the WchA is the most common one, which transfers UDP-Glcp to a lipid carrier ⁷⁰. After the initial transferase, the different GTs sequentially transfer sugar precursors to form the glycosidic linkages. Therefore, GTs are crucial in determining the sequence of CPS repeat units. The GT genes in *Streptococcus pneumoniae* are highly serotype specific. Understanding the specificity of GTs could help with the prediction of primary structures of CPS, which is of great value to correlate the genotypes with capsular types ⁷¹. Based on the determined CPS structures at that time and *cps* sequences of 90 serotypes, Andersen et al. assigned the functions of most GTs ⁷². Not surprisingly, the GTs form 92 homology groups (HGs), which is consistent with the diverse CPS structures.

In addition, more than half of the total serotype CPS structures contain phosphate groups, where a phosphodiester linked with two sugars or one sugar and one Cho / polyalcohol (Gro, Rib-ol, Ara-ol and Man-ol). The phosphodiester linkage is catalyzed by the sugar/polyalcohol phosphate transferase. Besides, acetyltransferases modify the repeat unit structure by adding OAc to sugars, which could happen during the repeat unit formation or after polymerization. There are often different variations of the degree of OAc and the mechanism for this variability is still not clear.

After the formation of oligosaccharide repeat units is completed, the repeat units are translocated to the outer membrane by the Wzx flippase and are polymerized at the reducing end to form the high molecular weight CPS. The Wzy polymerase determines the linkage between the initial sugar and the terminal residue and controls the polymer chain-length ⁷³. The *wzy* from 90 serotypes fall into 40 homology groups, which were commonly used as targets in PCR-based serotyping methods ^{26,43,44,67}. After polymerization, the CPS is transported to the cell wall and anchored at PG via a direct glycosidic bond to β -D-N-acetylglucosamine ²⁰.

The assignment of the functions of CPS biosynthetic genes and enzymes requires the correlation of accurately determined CPS structures with *cps* genes. Thus, it is valuable to determine these unknown CPS structures to facilitate characterization of function of above-mentioned biosynthetic enzymes.

1.6 Serogroups 16, 28 and 24

In serogroups 16, 24 and 28, there are currently seven different serotypes: 16F, 16A, 24F, 24A, 24B, 28F and 28A. None of them is included in current pneumococcal vaccines. However, the vaccine pressure could result in the change of prevalence in the carriage and invasive serotypes, which means these non-vaccine serotypes could become prevalent among the population.

In recent years, it has been reported that serotype 16F might be associated with IPD and increased case-fatality rates^{9,12,74–76}. Notably, in South Africa, serotype 16F has been frequently reported in recent years^{58,77}. One study proposed that serotype 16F could be included in a higher valence PCV⁷⁸. It is important to determine serotype 16F CPS structure in order to include it in a future vaccine. Serogroup 28 are genetically close to serotype 16F. As 16F become more prevalent and shows increased fatality risks, serotype 28F and 28A would also be potential prevalent types. Similarly in Africa, serotype 28F was found in both vaccinated and unvaccinated children^{79,80}.

Serotype 24F is often associated with penicillin-resistant strains and has been remarkably prevalent in many countries since 2013^{62,81}. In the UK and Israel, serotype 24F was the most common cause of non-PCV13 IPD in children < 5 years in 2013-2014, while 24F-caused IPD decreased in 2015-2016^{82,83}. In Japan, it was reported that serotype 24F was the most common non-vaccine type causing IPD in children in 2014 and a significant increase of 24F IPD since 2013⁸⁴. In Denmark, research also observed an increase of IPD due to serotype 24F since 2012 (Figure 1.7)⁸⁵. It was also found that serotype 24 is currently frequent in western European countries that use PCV13⁸⁶. However, that is not the case in the US.

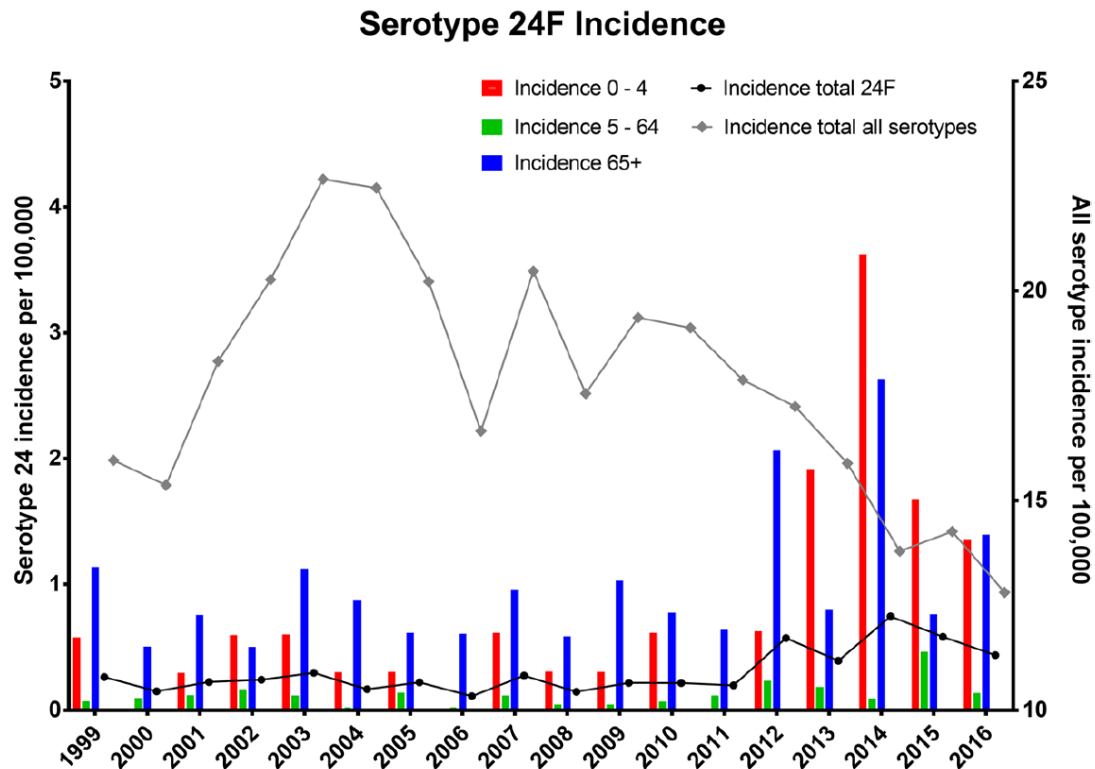


Figure 1.7 Serotype 24F incidence in Denmark since 1999. Reprinted from an open access article ⁸⁵. Copyright © 2018 The Authors.

The vaccine pressure altered serotype prevalence in the carriage and invasive serotypes, which resulted in non-vaccine serotypes 16F and 24F emerging after PCV. The efficacy of a given vaccine depends on the serotype carriage in the target population. Thus, high quality surveillance is necessary to monitor the epidemiology of pneumococcal serotypes to formulate future pneumococcal vaccines. In the meantime, it is important to gain an understanding of these prevalent non-vaccine serotypes. The CPS are vital for the virulence of the pathogen and antigenic properties of CPS are based on the structure. Consequently, a better understanding of the structure, biosynthesis and serology of the capsular polysaccharides can be of great importance towards developing future diagnostic tools and vaccines.

Chapter 2 Methodology and experiments

2.1 Structure elucidation of CPS

The primary structure of CPS is that of a high molecular weight polymer with an oligosaccharide repeat unit (Figure 1.4). The repeat unit structure was determined by the constitutions of monosaccharides and other components, sequence and linkage type of residues, the anomeric configuration of the glycosidic linkages and positions of substituents. NMR spectroscopy is a powerful technique for structure determination of carbohydrates^{87–89}. However, due to the complexity and heterogeneity of high molecular weight CPS, it is useful to combine classic analytical methods to assist in the NMR structure determination. In this study, NMR spectroscopy, combined with general chemical analysis, was applied to determine the repeat unit structure of CPS. Besides, the molecular weight distributions of CPS were measured by asymmetric flow field-flow fractionation coupled with multiangle light scattering, differential refractive index and ultraviolet detectors (AF4-MALS-dRI-UV), which can provide the information of the number of repeat unit in the polymer chain.

2.1.1 NMR spectroscopy

2.1.1.1 Basics of NMR spectroscopy

Some nuclei that contain positively charged protons and uncharged neutrons possess a physical property known as spin (I). If the number of neutrons plus the number of protons is odd, the nucleus has a half-integer spin (i.e. 1/2, 3/2, 5/2). If the number of neutrons and the number of protons are both odd, the nucleus has an integer spin (i.e. 1, 3, 5). If the number of neutrons and the number of protons are both even, the nucleus has spin 0. In quantum mechanics, a nuclear spin (I) can have $2I + 1$ possible Eigenstates. For example, ^1H , a spin 1/2 nucleus, has two possible Eigenstates. Nuclear spins act somewhat similar to a tiny bar magnet oriented along the rotation axis (Figure 2.1). When applying a magnetic field (B_0), the two states are not degenerate anymore but have different energy levels. The energy difference (ΔE) between the two states that depends on the strength of B_0 and the gyromagnetic ratio (γ) [2.1].

$$\Delta E = \frac{\gamma \hbar B_0}{2\pi} \quad [2.1]$$

Where h is Planck constant.

The spin angular momentum will precess around the magnetic field B_0 at the Larmor frequency (ω_0) [2.2].

$$\omega_0 = \gamma B_0 \quad [2.2]$$

There will be a slight excess of nuclei oriented with the magnetic field in the lower energy state, so the sum will yield a net magnetization (M_0 in Figure 2.1) along B_0 . When the sample is irradiated with a resonant radio frequency (RF) pulse perpendicular to B_0 , it can rotate the magnetization from its equilibrium position to a perpendicular plane, where receiver coil will record signals generated by oscillating magnetization. NMR experiments manipulate the magnetizations with different pulse programs to generate spectra carrying useful structural information.

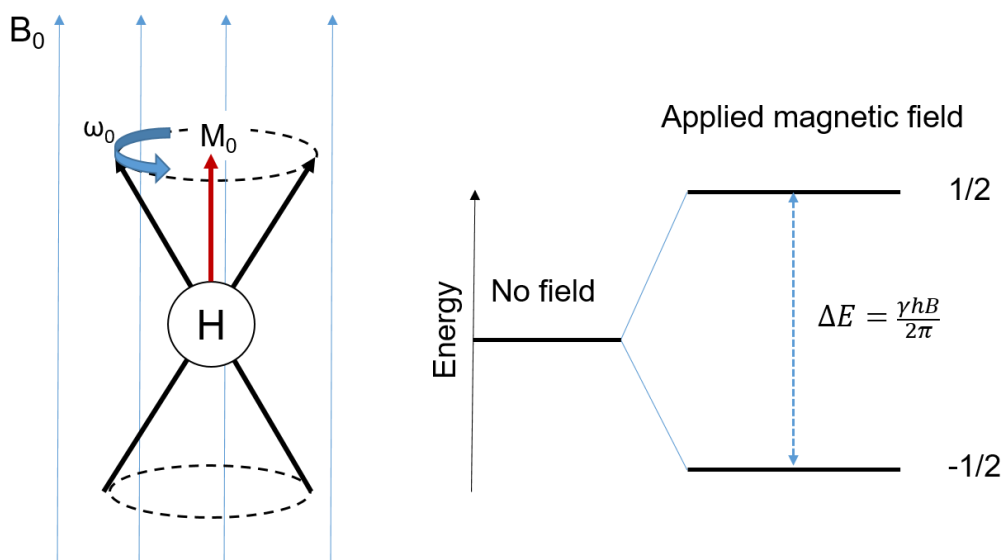
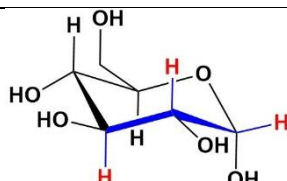
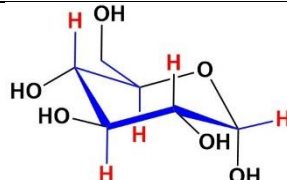
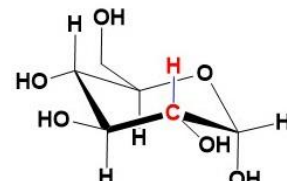
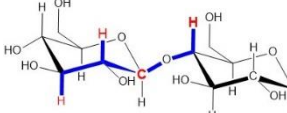
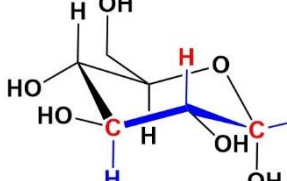
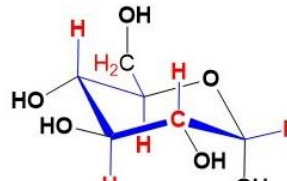
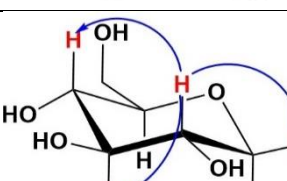


Figure 2.1 Schematic illustration of nuclear spin (^1H) in magnetic field B_0 and the energy level with and without magnetic field

One-dimensional (1D) and two-dimensional (2D) solution state NMR can provide information of the spin Larmor frequency (chemical shifts), spin-spin scalar coupling constants (J), correlations and spin relaxation. The scalar, or J -coupling, is a through-bond interaction, in which the spin of one nucleus perturbs the spins of neighboring nucleus. The J coupling constants are field-independent and the coupling over one (1J), two (2J) and three (3J) bonds dominate the fine structure of NMR spectra. Besides, the nuclear Overhauser effect (NOE) occurs when two nuclei in sufficiently close spatial proximity ($< 5\text{\AA}$), which can be used to measure the distance between them. Both J coupling constants and NOE are useful in identifying carbohydrate configuration (discussed in the next section).

A summary of 1D and 2D NMR experiments that were used for carbohydrate structure analysis is given in Table 2.1. In this section, applying NMR spectroscopy to study complex carbohydrates, such as CPS, will be introduced.

Table 2.1 Typical 2D NMR Techniques for structural characterization

| NMR technique | Correlations | Information | Example |
|--|--|--|---|
| Double quantum filter correlated spectroscopy (DQF-COSY) | Proton J coupling constants typically over 2, 3 or 4 bonds | Neighbor protons and $^3J_{H,H}$ |  |
| Total correlation spectroscopy (TOCSY) | Relayed proton within a coupled spin system | Protons within a same spin system |  |
| Heteronuclear single quantum coherence (HSQC) | One-bond heteronuclear correlation | Directly connect 1H and ^{13}C |  |
| Heteronuclear multi-bond correlation (HMBC) | Long-range heteronuclear couplings over 2-4 bonds | Cross spin system correlations |  |
| Heteronuclear 2 bond correlation (H2BC) | Heteronuclear couplings over 2 bonds relying on $^3J_{H,H}$ coupling between the protons on adjacent carbons | Neighbor carbons with protons |  |
| HSQC-TOCSY | Hybrid 1H - ^{13}C HSQC with a 1H -TOCSY experiments | Give correlations between a ^{13}C to its attached 1H and all 1H within the same spin system |  |
| Nuclear Overhauser effect spectroscopy (NOESY) | Through-space correlations | Protons < 5 Å |  |

2.1.1.2 Carbohydrate NMR

Carbohydrates have low chemical shift dispersion in ^1H and ^{13}C NMR compared to other biomolecules, which often lead to overlap, especially in ^1H or ^1H - ^1H correlation NMR experiments. Therefore, a combination of different 1D and 2D NMR experiments is needed to correctly deduce the structure. A strategy using different NMR experiments is provided to get monosaccharide compositions, glycosidic linkages, substitution patterns and conformational information of carbohydrates (Figure 2.2).

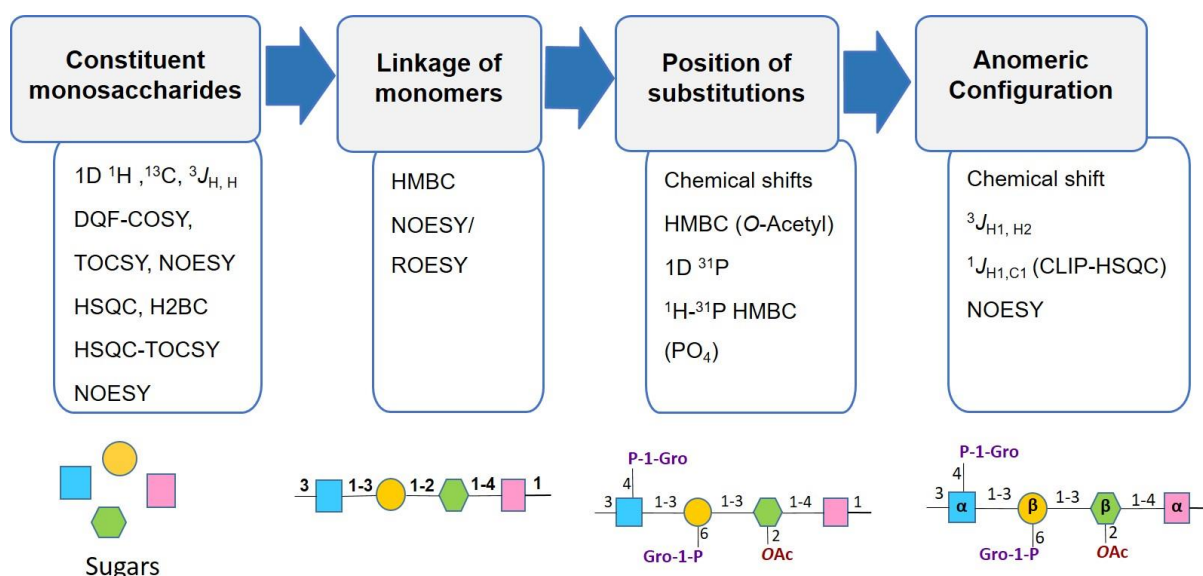


Figure 2.2 Graphical illustration of NMR for structural determination of carbohydrates

Constituent monosaccharides

The anomeric positions (1 position) of sugar residues are found in the chemical shift range 4.4 - 5.5 ppm for ^1H and 90 - 110 ppm for ^{13}C ⁸⁸. If 1D proton spectra showed overlap, it's recommended to use ^1H - ^{13}C HSQC spectra which are more sensitive than 1D ^{13}C spectra. Moreover, phase edited HSQC can also give information to distinguish CH_2 from CH and CH_3 . Starting from the anomeric signals, the number of sugar residues can be determined. Then following homonuclear TOCSY and DQF-COSY, each spin system corresponding to each sugar unit can be found. When experiencing overlap in the bulk region (3 - 4.2 ppm), H2BC and HSQC-TOCSY are useful to disperse signals in the carbon dimension and help with assignment of spin systems. Furthermore, to identify to which sugars individual spin systems belong, the $^3J_{\text{H,H}}$ pattern of the sugar ring protons can provide information of proton orientations. In pyranoses, the six-membered ring generally forms a chair conformation providing a classification of protons as axial (ax) or equatorial (eq). The 3J -coupling

constants of $H_{ax, ax}$ are >7 Hz for the 180° dihedral angles, while 3J -coupling constants of $H_{eq, eq}$ are smaller, about 2 Hz, and the 3J -coupling constants of $H_{ax, eq}$ are about 2 - 5 Hz. Moreover, the NOEs or ROEs are useful when small coupling constants are hard to observe. Furanose sugars are characterized by distinctive carbon chemical shifts^{87,89,90}. In general, the chemical shifts of C4 for aldofuranose and C5 for ketofuranose are in the region 80 - 85 ppm.

Linkage of monosaccharides

The chemical shifts of 1H and ^{13}C could indicate the glycosidic positions⁸⁹. The linkages between two sugar residues can be found through HMBC correlation from one sugar anomeric position to glycosidic position of another sugar (Figure 2.3). Besides, cross glycosidic NOEs could also give the information of linkage sites.

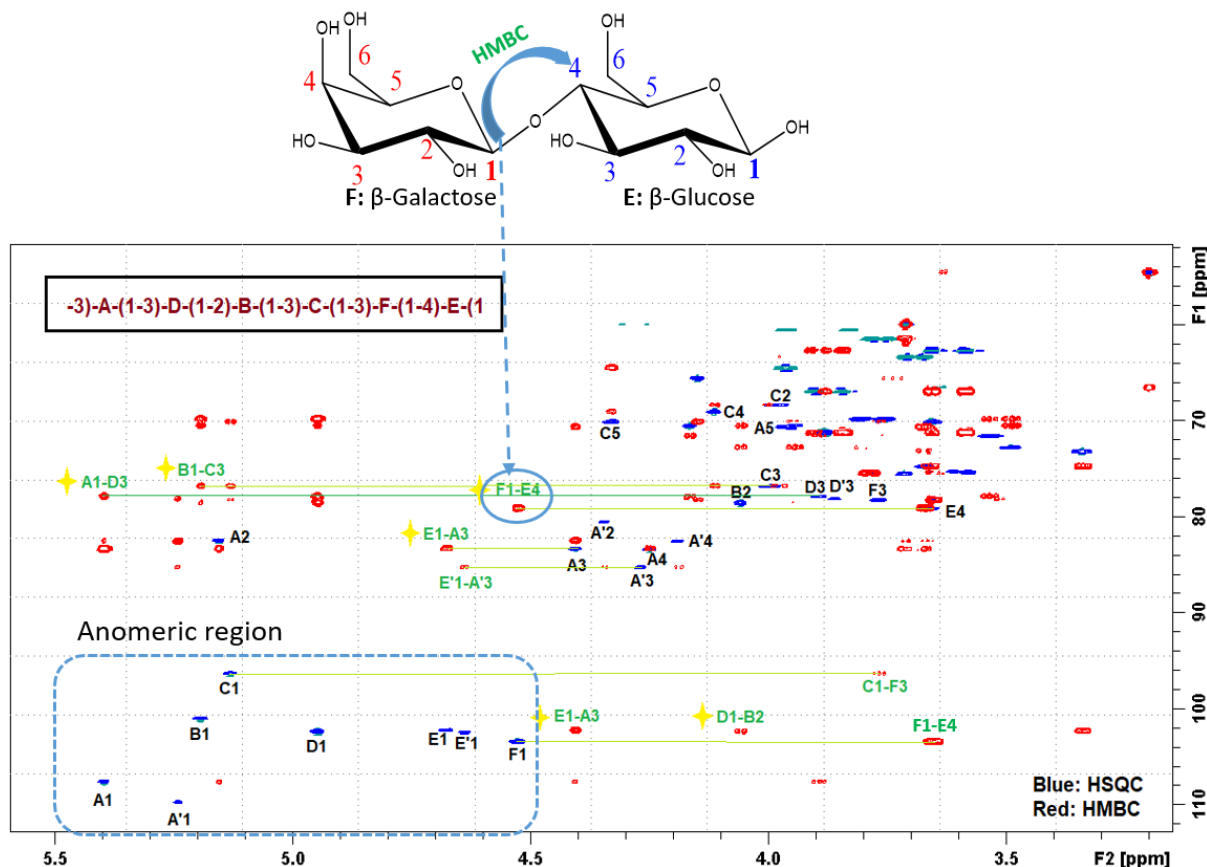


Figure 2.3 Expansion and overlap of 1H - ^{13}C HSQC (blue: +, light blue: -) and HMBC (red) spectra of 16A CPS.

Position of Substitutions

Bacterial polysaccharides often have O-acetylation and phosphorylation. Therefore, 1D ^{31}P spectra are useful to check if there are PO_4 groups in the polysaccharides structures. If there are PO_4 groups, ^1H - ^{31}P HMBC is useful to find the substituted position.

The O-Ac groups can result in high shifts of proton and HMBC correlations between carbonyl carbons with acetylated positions can be found to determine the acetylated sites. However, CPS sometimes are partially acetylated and have multiple acetylated sites, which makes the sample more heterogeneous and leads to complicated NMR spectra. Thus, acquiring NMR spectra of CPS before and after de-O-acetylation and comparing the changes in spectra would help to determine acetylation content and positions.

Anomeric configuration

In pyranose sugars, the α -anomeric positions normally have rather high chemical shifts at 4.8-5.8 ppm for ^1H and 95 - 103 ppm for ^{13}C , compared to β -anomeric protons at around 4.4-5.0 ppm for ^1H and 103 -110 ppm for ^{13}C ^{88,89,91}. Moreover, the scalar coupling constants are important for the determination of sugar anomeric configuration. For pyranose sugars, the $^3J_{\text{H}_1, \text{H}_2}$ can give information about the orientation of anomeric protons, thus to identify α or β configuration. However, when both H_1 and H_2 are at equatorial positions, the $^3J_{\text{H}_1, \text{H}_2}$ often are too small to determine. For example, the rhamnose has all small $^3J_{\text{H}_1, \text{H}_2}$. Therefore, $^1J_{\text{C}_1, \text{H}_1}$ would be more reliable for indication of anomeric configurations of pyranosides ⁹¹. A $^1J_{\text{C}_1, \text{H}_1}$ of ~170 Hz indicates the α configuration, while a $^1J_{\text{C}_1, \text{H}_1}$ of ~ 160 Hz indicates the β configuration ⁹². To get the $^1J_{\text{C}_1, \text{H}_1}$, clean in-phase (CLIP) -HSQC can be used ^{92,93}. However, the $^1J_{\text{C}_1, \text{H}_1}$ values of pentofuranosides are around 168-171 Hz for both α and β anomers. In this case, the comparison of chemical shifts with known assignments in the literatures would help to identify the anomeric configuration of furanosides ^{87,90,94}.

Sample preparations and NMR experiments

Purified pneumococcal CPS samples, as well as rabbit diagnostic antisera, were produced and provided by SSI Diagnostica A/S, Hillerød, Denmark. The serotype 24X CPS was purified by Sofie Slott when she did her bachelor project at SSI Diagnostica A/S. All the CPS samples, except serotype 16A CPS, were subjected to ultra-sonication to depolymerize the sample in order to get well-resolved NMR spectra. About 2 mg/mL CPS were subjected to sonication using LABSONIC® P ultrasonic homogenizer with 3 mm Probe for 2 hours under the ice bath, following by desalting and ultrafiltration (Amicon® Ultra-4, 3 kDa).

Dephosphorylation (de-PO₄) and de-O-acetylation (de-OAc) of CPS were performed using methods described by Richards and Perry⁹⁵. De-OAc CPS were generated by treating approximately 5-10 mg of each polysaccharide with 1 mL of a 0.2 M NaOH in D₂O solution at room temperature for 2 h, which was subsequently neutralized with 2M HCl, followed by desalting and removal of residuals using centrifugal filters (Amicon® Ultra-4, 3 kDa). Alternatively, 5-10 mg CPS were dissolved in 500 µl D₂O and 1-2 µL 5 M NaOH was added. Samples were incubated at room temperature for 1-2 hours. The de-OAc of CPS was performed inside or outside NMR tubes.

De-PO₄ of CPS was performed by treating 10 mg of each CPS sample in 2mL 1.0 M NaOH (adding a trace amount of sodium borohydride) at 100 °C in screw-capped glass tubes filled with nitrogen for 4 hours. Then the solution was cooled in an ice bath and neutralized with acetic acid, followed by desalting and ultrafiltration (Amicon® Ultra-4, 10 kDa).

The NMR experiments were acquired on a Bruker Avance III (799.90 MHz for ¹H and 201.14 MHz for ¹³C) equipped with a 5 mm TCI ¹H/(¹³C, ¹⁵N) cryoprobe, or on a Bruker AVANCE 600 MHz instrument with 5 mm SmartProbe BB(F)-H-D(¹⁵N-³¹P, ¹H, ¹⁹F).

All spectra of 16A CPS were recorded at 313 K. While serotype 16F, 28F and 28A CPS spectra were recorded at 333K or 323K. The assignments of CPS repeat units were performed using the following experiments: 1D ¹H and ³¹P and the 2D NMR spectra, including DQF-COSY, TOCSY with 80 ms mixing time, NOESY with 200-400 ms mixing time, HSQC, HMBC optimized for 8 Hz long range coupling constants, ¹H-¹³C HSQC-TOCSY with 80 ms mixing time, CLIP-HSQC and ¹H-³¹P HMBC. Experiments were conducted with standard gradient enhanced Bruker pulse sequences, CLIP-HSQC⁹³. All spectra of serogroup 24 CPS were recorded at 323 K. The assignments of serogroup 24 CPS repeat units were performed using similar experiments as for serogroups 16 and 28 CPS.

NMR spectra were processed on Topspin 3.6 or 4.07 (Bruker) using zero filling to twice the data size in the direct dimension and 4 times the data size in the indirect dimension. The one-dimensional ¹H spectra were processed with an exponential window function with a line broadening of 0.3 Hz for ¹H spectra. All 2D spectra were processed with shifted sine bell window functions in both dimensions. All spectra were referenced to residual cell wall polysaccharide phosphocholine CH₃ signals (¹H 3.2 ppm and ¹³C 54.5 ppm and the shielded ³¹P signal at 1.30 ppm)¹⁴.

2.1.2 General chemical analysis for carbohydrates

General chemical analysis, including monosaccharide composition, methylation, D / L configuration and phosphate content was conducted by classic gas chromatograph (GC) or GC-Mass spectroscopy (MS) methods in collaboration with Dr. Katarzyna Duda at Research Center Borstel - Leibniz Lung Center.

The monosaccharide composition was determined by GLC after hydrolysis with 2 M TFA at 120 °C, 2 h, reduction and peracetylation⁹⁶. The amino sugar was analysed by hydrolysis of the samples with 4M HCl at 100 °C for 20 h. Xylose was used as the internal standard. Analysis of samples was performed on a HP 5890 (series II) gas chromatograph with a flame ionization detector and a HP-5MS column (30 m × 0.25 mm × 0.25 µm film thickness, Agilent). The temperature in gas-liquid chromatography (GLC) started at 150°C for 3 min and increased with 3°C/min to 250°C, then increased to 320 °C for 10 min. The carrier gas was Helium at 70 kPa. A mixture of standard monosaccharides was used for sugar identification and xylose for quantification.

Methylation / linkage analysis was performed using Ciucanu-Kerek' s method⁹⁷. Because phosphate groups are negatively charged and could result in problems in later methylation, the CPS sample was treated with 48% HF for 48 h at 4 °C to remove the phosphate groups. As de-PO₄ CPS samples are still difficult to dissolve in DMSO, CPS samples were suspended in DMSO at 60 °C overnight, and then methylated with iodomethane⁹⁸. The product was hydrolyzed with 4 M TFA for 4 h at 100 °C and was reduced with NaBD₄, following peracetylation with acetic anhydride in pyridine. Partially methylated acetylated alditols were analyzed by GC-MS.

The absolute configuration was determined by butanolysis with 2 M HCl / (R)-2-Octanol or (S)-2-Butanol at 65 °C for 4 h and analyzed by GLC⁹⁹. The retention time was compared with sugar standards.

The content of total phosphate was determined by colorimetric analysis¹⁰⁰. CPS samples were subjected to persulfate digestion (digestion reagent: 30.6 mL conc. H₂SO₄, 6.7 mL 70% HClO₄ and 62.7 mL H₂O) at 100 °C first for 1 h and then at 165 °C for 2 h. Subsequently, the sample was cooled down to room temperature and freshly prepared color reagent was added (1 mL 1 M sodium acetate, 1 mL 2.5% ammonium molybdate solution, 7 mL H₂O, 1 mL freshly prepared ascorbic acid; 37 °C for 90 min). Na₂HPO₄ solutions were used as external standards. The absorbance of the ammonium phosphomolybdate complex was measured at λ = 820 nm (HELIOS BETA 9423 UVB 1002E spectrophotometer, Thermo Electron Ltd., Altrincham, Cheshire, UK).

2.1.3 AF4-MALS-dRI-UV for characterizing CPS and CPS-antiserum interactions

Asymmetrical flow field-flow fractionation (AF4) is a separating technique based on molecular diffusion properties¹⁰¹. AF4 has a wide range of capacity for separation of macromolecules and particles from a few nm up to about 100 μm in size, which is being increasingly used in purification of biopharmaceuticals and nanomedicines^{102–104}. One important advantage of AF4, compared to size exclusion chromatography (SEC), is that the separation of analytes is based on the interaction of the sample with a physical field rather than with a stationary phase. The separating device is a thin, flat channel with an ultrafiltration membrane in the bottom (Figure 2.4). The ultrafiltration membrane often is referred as the accumulation wall. A parabolic laminar flow is generated to elude the samples to the outlet that connects with different detectors. A cross flow is applied perpendicular to the forward channel flow, and is an important driving force for separation. The cross flow pushes the sample to the accumulation wall, while the Brownian motion of the molecules (the diffusion force) counteracts with the cross flow force and drives them toward the interior of the channel. The equilibrium of cross flow and diffusion force resolves components in the samples according to their inherent biophysical properties. Therefore, smaller molecules that diffuse faster are further away from the accumulation wall and will be eluted out earlier. While larger molecules diffuse slower and experience a longer elution time. AF4 is normally coupled with multiple downstream detectors such as multiangle light scattering (MALS), differential refractive index (dRI) and UV/Vis detectors, which can provide information of molecular size distributions.

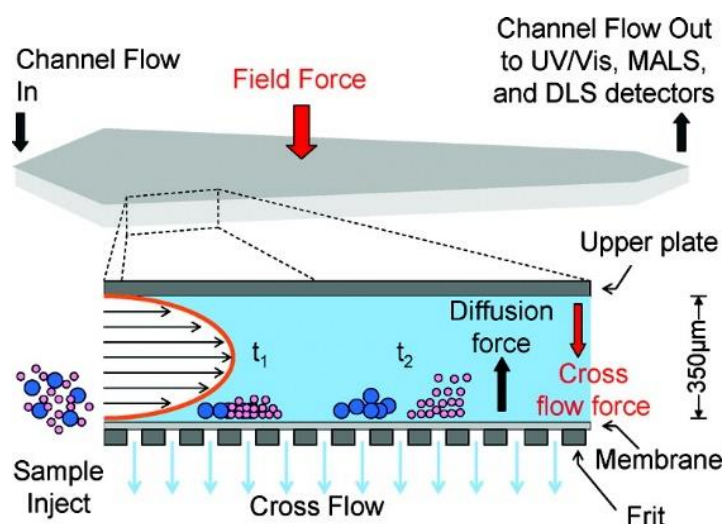


Figure 2.4 Schematic illustration of the separation process AT different elution times ($t_2 > t_1$) in an AF4 channel. Reprinted (adapted) with permission from¹⁰⁵. Copyright (2020) American Chemical Society.

The adjustable cross flow and tolerance to variable solvents or buffers make AF4 a flexible technique for studying different biological samples and nanoparticles^{102,103,105–109}. AF4 was applied to study proteins and antibodies in serum, plasma and whole¹¹⁰. The lack of a stationary phase in AF4 system can minimize of nonspecific interactions and structural deformation at the interface with the column packing in SEC, which is a good choice for characterization of macromolecule-macromolecule interactions, like CPS-antibody interactions.

In this project, the AF4-MALS-dRI-UV system was applied to characterize the molecular weight of CPS and was used to study CPS and antisera interactions. For AF4-MALS-dRI-UV analysis, CPS samples were dissolved in phosphate-buffered saline (PBS, pH 7.4) to 2.0 mg/mL. Antisera were diluted by adding 50 μ L into 500 μ L PBS. 1 mg CPS and 50 μ L antiserum were mixed in PBS to a final volume of 500 μ L for the study of interactions between CPS and antiserum. All the samples were filtered by Durapore 0.1 μ m PVDF membrane filters (Merck Millipore) before injection.

The AF4-MALS-dRI system consisted of a pump (Agilent 1260 Infinity II) with an online degasser. The separation was performed on an AF4 system (ECLIPSE; Wyatt Technology) using a polyethersulfone membrane with 10 kDa cut-off in a 17.5 cm separation channel with a 350 μ m spacer. A constant flow of 2.0 mL/min was applied for 3 min before injection. PBS was also used as the carrier solvent. 100 μ L sample was injected with a flow of 0.2 mL/min and then were eluted at 2.0 mL/min for 5 min. A constant detector flow of 0.5 mL/min was used during the separation process, which included 3 stages: (1) linear ramped flow from 2.0 to 0.7 mL/min in 22 min; (2) exponential ramped flow from 0.7 to 0.05 mL/min in 25 min; and (3) linear ramped flow from 0.05 to 0.0 mL/min in 3 min and 0.0 mL/min flow, held constant for 5 min. A TREOS II MALS detector (Wyatt Technology), a dRI detector, Optilab T-rEX (Wyatt Technology), and a UV detector were connected to the AF4 system. The UV detector was monitored at 280, 250, 220 nm. Data collection and processing were performed using the ASTRA software, V 7.1.4. For data analysis, the dn/dc of 0.130 mL/g for CPS concentration determination to obtain molecular weight⁶⁵ and a second-order Berry formalism for fitting light scattering data of CPS were employed¹¹¹. For analysis of proteins, the dn/dc of 0.185 mL/g and a second-order Debye formalism were applied for protein molecular weight determination. For the CPS-antibody complex, the average dn/dc of 0.16 mL/g and a second-order Zimm formalism was applied.

2.2 Biosynthetic analysis

2.2.1 Bioinformatics

Bioinformatics is a computer-assisted data management that allows us to gather, classify, analyze, and distribute large-scale biological information derived from sequencing and functional analysis projects to enable life science and novel drug discovery to progress much faster^{112,113}. Bioinformatics is an interdisciplinary field, which include DNA sequence assembly and annotation, sequence analysis, protein and RNA structural analysis, genomics, phylogenetics, metagenomics, and so on. In microbiology, bioinformatics is important and useful to integrate the information from all organisms and provide new insights into all aspects of microbiology.

In the study of *Streptococcus pneumoniae*, the genetic sequencing, assembly and annotation of the *cps* gene locus from 90 pneumococcal serotypes was done by Bentley⁶⁷. Genetic relation of *cps* loci and predicted functions of genes encoded in *cps* loci were further analyzed by Bentley's group^{72,114}. However, as limited CPS structures were known at that time, the functions of some CPS biosynthetic genes could not be proposed. Thus, here we want to compare *cps* loci of interest to related serotypes within and between the serogroups in correlation with CPS structures to assign the functions of related biosynthetic enzymes. Furthermore, in order to get a deeper understanding of what resulted in the different functions of GTs, protein homology modeling of GTs was performed to predict the key residues for the specificity of GTs. Therefore, we use bioinformatics tools 1) to compare gene and protein sequences and 2) to predict protein structure by homology modeling.

Bioinformatics tools are software programs that are designed for extracting meaningful information from biological databases to carry out sequence or structural analysis. Many useful bioinformatics tools are listed in the website <https://bio.tools/>. In this study, BLAST was used for nucleotide and protein sequence alignment and comparison. The phyre2 or SWISS-MODEL were used for protein homology modeling.

2.2.1.1 BLAST

BLAST (Basic Local Alignment Search Tool) is an algorithm and program provided by NCBI (National Center for Biotechnology Information) for searching and comparing primary sequences of DNA, protein or RNA in a target database¹¹⁵. BLAST can be used to find regions of similarity between two and multiple sequences and to generate sequence alignments. The NCBI BLAST homepage provides different BLAST tools to perform sequence alignment on the web (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). In a BLAST

search, one needs to enter a query nucleotide or protein sequence, which can be FASTA sequences or accession numbers. A published sequence accession number can be found in GenBank, which is the NIH genetic sequence database ¹¹⁶. There are mainly five different types of BLASTs available depending on the query sequences and the target databases (Table 2.2). The BLAST outputs include a table of sequence information with hits and scores, a graphic summary and alignments. The results of alignments can be downloaded as FASTA, GenBank, Hit Table and text for later use in visualization and comparison tools.

Table 2.2 Different BLAST programs

| BLAST Program | Query | Database |
|---------------|-----------------------|-----------------------|
| blastn | Nucleotide | Nucleotide |
| blastp | Protein | Protein |
| blastx | Translated Nucleotide | Protein |
| tblastx | Protein | Translated Nucleotide |
| tblastn | Translated Nucleotide | Translated Nucleotide |

In this study, blastn and blastx were mainly used for generating multiple DNA sequence alignments of pneumococcal *cps* gene clusters from different serotypes. Multiple protein sequence alignments of biosynthetic enzymes and pairwise protein sequence comparisons were conducted using BLASTp.

2.2.1.2 Multiple sequence comparison and visualization

In bioinformatics, multiple sequence alignment (MSA) is the simultaneous alignment of three or more nucleic acid or amino acid sequences to identify regions of similarity to find the potential relationship of function, structure, or evolution between the sequences. A summary of common MSA tools can be found in Table 2.3. To launch MSA tools can be via EMBL-EBI webpage (<https://www.ebi.ac.uk/Tools/msa/>). For short sequences, BLAST is fast and convenient for generating an alignment. Comparison and visualization of similarities between pneumococcal *cps* loci were performed by the Artemis comparison tool (ACT) ¹¹⁷. ACT is a Java based graphical DNA sequence comparison viewer, which allows interactive visualization of pairwise comparisons created using BLAST. Multiple sequence comparisons can be stacked in ACT. For the alignment of a specific gene or its product (enzyme) from related serotypes, JalView was used to visualize the multiple sequence alignment in residue level and produce comparison figures ¹¹⁸.

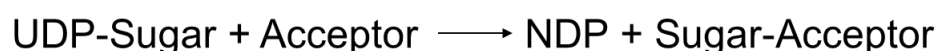
Table 2.3 A summary of MSA tools ^{119,120}

| MSA Tools | Suitability |
|------------------------------|---|
| T-Coffee ¹²¹ | Consistency-based MSA, small alignments |
| MUSCLE ¹²² | Accurate MSA tool, medium alignments |
| MAFFT ¹²³ | Medium-large alignments |
| Clustal Omega ¹²⁴ | Medium-large alignments |
| Kalign ¹²⁵ | Very fast MSA, large alignments |

2.2.2 Glycosyltransferases structure modeling

2.2.2.1 The glycosyltransferases (GTs)

The GTs are a group of enzymes that catalyze glycosidic bonds between the nucleoside diphosphate (NDP) sugar donor and the lipid-linked oligosaccharide precursor as the acceptor. The catalytic specificities of GTs include the acceptor, the donor sugar residues and the linkage type between them.



According to whether the GTs change the anomeric configurations of sugars or not, GTs can be classified as retaining and inverting enzymes, respectively. The Carbohydrate-Active enZymes Database (CAZy) groups the different GTs into families (EC 2.4.x.y) based on nucleotide diphospho-sugar, nucleotide monophospho-sugars and sugar phosphates and related proteins (<http://www.cazy.org/GlycosylTransferases.html>). GTs from same family are expected to have a similar three-dimensional fold. Although there are 111 GT families to date (12.30.2020) that are continuously updated, their catalytic domains fall into three different protein folds: GT-A, GT-B, and GT-C (Table 2.4) ^{126,127}. The GT-A fold enzymes only have one Rossmann-like fold, which is a domain that can bind nucleotides. The GT-B fold enzymes contain two Rossmann-like folded domains. The GT-C fold enzymes are transmembrane proteins and use mainly lipid-linked sugars as donors.

Table 2.4 Classification of GTs

| | Fold | Mechanism | Metal dependent | CAZy family |
|------|---------------------|-----------|-----------------|---|
| GT-A | 1 Rossmann fold | Inverting | Yes | 2,7,13,15,16,31,43,101 |
| | | | No | 14,29,42 |
| | | retaining | Yes | 6,8,15,24,27,34,44,55,62,64,78,81,88 |
| | | | No | 6 |
| GT-B | 2 Rossmann folds | Inverting | Yes | None |
| | | | No | 1,9,10,18,19,23,28,30,37,38,41,52,63,65,68,70,80,90,104 |
| | | retaining | Yes | None |
| | | | No | 3,4,5,20,35,72,99,107 |
| GT-C | Complex fold | Inverting | Yes | 66,83 |
| | | | No | None |
| | | retaining | Yes | None |
| | | | No | None |

2.2.2.2 Protein structure modeling

In bioinformatics, the three-dimensional (3D) structure of a protein may be constructed using experimentally determined structures of related protein family members as templates ¹²⁸. Even though the primary structure of GTs can be diverse, the 3D structure of GTs are rather conserved, especially if they are from a same CAZy family. Therefore, it is possible to predict the 3D structure of a GT based on its amino acid sequence. Some commonly used web-based comparative (or homology) protein structure modeling servers are SWISS-MODEL ^{129,130}, Phyre2 ¹³¹, I-TASSER ¹³² or RaptorX ¹³³. The overall 3D protein structure modeling procedure consists of similar steps and the differences in modeling accuracy between these tools are minor ¹³⁴. In paper 2, SWISS-MODEL was used for GT WciU 3D structure modeling. Models of WciU were built with Phyre2 in intensive mode.

The first step of comparative modeling is to submit a targeted protein sequence and search in the protein database for proteins that relate to the target sequence. The second step is to select a good template based on sequence identity and some quality estimations. The

sequence similarity should be above 25% in general. In the case of GT modeling, a known 3D structure of a GT from the same CAZy family should have a similar fold. The third step is to align the sequence of the target and the template. BLASTp search is sufficient to know which 3D-structures have high identity and coverage of the query. In some cases, multiple alignment algorithms, such as a Position-Specific Iterated BLAST (PSI-BLAST), T-coffee or MUSCLE are better suited to improve alignment quality. The fourth step is the model building, including backbone generation, loop modeling and side-chain modeling, which can be generated in automated mode in SWISS-MODEL or Phyre2 workplace. The fifth step is protein structure assessment and model quality estimation. The Swiss-MODEL uses the Global Model Quality Estimation (QMQE) and Qualitative Model Energy Analysis (QMEAN) as an absolute quality estimator based on different geometrical properties and provides both the entire structure and individual residue on the basis of one single model ^{135,136}. In Phyre 2, phyre Investigator can perform more in-depth analyses of model quality and potential function. The last step is to refine and optimize the model according to physics laws in computational 3D modeling. The energy minimization of a model is based on a force field (FF), which is a set of parameters and equations that describes the bond lengths and angles, dihedral angles, improper planes, electrostatics and van der Waals forces. There are different FF available for protein structure refinement. In this study, the energy minimization of the GT models was performed on the YASARA server ¹³⁷. Finally, the models were downloaded, visualized and compared in PyMol 2.0 for further analysis.

2.3 Serological interaction study

The serological interaction study contained mainly two parts. The first part was the study of serological cross-reactivity of CPS within related serotypes by rabbit diagnostic factor antisera. The factor antisera were obtained from rabbits immunized with pneumococci of specific serotype and only absorbed cross-reactive antibodies within the same serogroup. Therefore, the factor antisera could also show cross-reactions with related serotypes from other serogroups. Quellung reaction and latex agglutination were performed to check the cross-reactions. Another part was the study of the interactions between serotype 16A CPS and 16c antiserum in a semi-quantitative way by AF4-MALS-UV, NMR titration and diffusion NMR.

2.3.1 Quellung reaction

The Quellung reaction, also known as Neufeld test, is the gold standard method for serotyping *Streptococcus pneumoniae*. Anti-CPS antibodies in antiserum can react with the polysaccharide pneumococcal capsule, which leads to a precipitation reaction on the pneumococcal surface. Upon antibodies binding, it results in a change of refractive index of the capsule, which appears as a “swollen” capsule. This phenomenon can be observed under a phase contrast microscope.

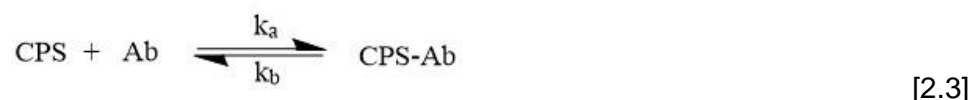
The procedure for the Quellung reaction is simple and fast. Equal amounts of diagnostic antiserum and pneumococcal culture were mixed on a slide and observed under a phase contrast microscope. However, the approach requires isolation and culture of pneumococci, which could be tedious and expensive for large numbers of clinic samples. The pneumococcal strains and rabbit diagnostic antisera in this project were produced by SSI diagnostica, Hillerød, and Quellung reaction was conducted in their laboratories.

2.3.2 Latex agglutination

Latex agglutination is also a CPS and antibody precipitation reaction. The Pneumotest-Latex assay was a simple and rapid latex agglutination for serotyping pneumococci and showed usefulness for direct serotyping pneumococci isolates in clinical samples with an agreement of 88.1% with a PCR-based reference method^{138,139}. The diagnostic antisera are coated on the latex particles. When antibodies binding with the pneumococcal capsule, the latex-antibody-pneumococci become big enough to be seen without microscopy. Therefore, it requires less expertise and is easier for identifying a positive reaction. In the study of serogroups 16 and 28 CPS, latex agglutination was performed by factor antiserum latex solution with partially degraded CPS to study the antigenic activity of de-OAc and de-PO₄ CPS.

2.3.3 ¹H NMR Titration for study CPS - antiserum interactions

When mixing CPS with antiserum, the antibodies (Ab) in the antiserum will react with CPS and form the CPS-Ab complex. Ab affinity is defined as strength of the binding interaction between antigen and antibody. Thermodynamic binding constant of pneumococcal CPS with monoclonal antibody was studied, which showed a weak affinity with K values in the range of 10⁶ to 10⁷ M⁻¹ ¹⁴⁰. The equilibrium between CPS antigen and Ab can be expressed as:



The affinity constant, K, depends on the association rate (k_a) and dissociation rate (k_b) [2.4].

$$K = \frac{k_a}{k_b} = \frac{[\text{CPS-Ab}]}{[\text{CPS}][\text{Ab}]} \quad [2.4]$$

The antibodies in diagnostic antisera are polyclonal antibodies, and thus one defined K cannot be determined. However, quantitative ¹H NMR titration of CPS with antiserum, based on the line broadening due to antibody binding, would give information of non-bound or non-affected CPS to probe the formation of CPS-Ab complex.

The antibody will likely not contribute NMR signals in normal solution NMR due to the fast spin relaxation in large proteins. While 16A CPS are more flexible compared to antibodies, which can give sharp NMR peaks. Upon addition of antiserum, free CPS can be quantified by ¹H NMR.

1.0 mg CPS were dissolved in 450 μL PBS (D₂O) and transferred into 5 mm NMR tubes. Then 5, 10, 20, 30, 40, 50 μL of 16c antiserum were added. After every addition of antiserum, the sample solution was thorough mixed and incubated at 310 K for 10 min. Subsequently, the acquisition of quantitative ¹H NMR spectra was started.

The inversion-recovery experiment was used to measure the T₁ spin relaxation time of ¹H in 16A CPS. The ¹H NMR spectra were acquired with d1 (relaxation delay) of 15 s and 32 scans at 310 K on a Bruker Avance III (799.90 MHz for ¹H and 201.14 MHz for ¹³C) equipped with a 5 mm TCI ¹H/(¹³C, ¹⁵N) cryoprobe. NMR data were processed on Topspin 4.07 (Bruker) using zero filling to twice the data size in the direct dimension and 4 times the data size in the indirect dimension. The ¹H spectra were processed with an exponential window function with a line broadening of 1 Hz. Baseline correction was conducted manually in Topspin. All the spectra were rescaled to the same values of ns, receiver gain and 90° pulse length for quantification. The integration of the interesting peaks was performed by the “multi_integ” command in Topspin.

2.3.4 Diffusion NMR

The interactions between CPS and antisera occur in a complex multicomponent system, which makes them rather challenging to observe. When adding antiserum into CPS solution, the overall diffusion behavior of CPS would be affected due to the interactions with antibodies. Thus, diffusion NMR methods were applied to study interactions of these macromolecules in solution.

Diffusion NMR, in particular Diffusion Ordered Spectroscopy (DOSY), measures the diffusion coefficients of solutes, providing an additional, size-dependent dimension to resolve different components in a mixture. Thus, DOSY is a useful tool for mixture analysis and binding studies¹⁴¹.

According to the Stokes-Einstein equation [2.5] for Brownian motion of a spherical molecule, the diffusion coefficient (D) depends on viscosity (η), temperature (T) and the hydrodynamic radius (r). The hydrodynamic radius is determined by molecular properties, such as size, shape, mass, charge as well as its surrounding environment, temperature and aggregation state.

$$D = \frac{kT}{6\pi\eta r} \quad [2.5]$$

In NMR, diffusion is measured using pulsed-field gradients (PFG) by an echo-based pulse sequence. The first part of the pulse sequence dephases spins by gradient pulses, which spatially labels each spin along the z-axis in the Cartesian coordinate system defined by the NMR instrument magnetic field along the z-axis. During a diffusion delay, displacement of the spins along the z-axis can occur. Finally, the signals are refocused by a gradient pulse that reverts the effect of the first gradient pulse for spins that have not moved along the z-axis. Greater displacement of spins along the z-axis due to translational diffusion results in fewer signals getting refocused. The decreased signal intensity or integral, therefore, results for faster diffusing species. By fitting the attenuated signal intensity to the Stejskal-Tanner equation [2.6], diffusion coefficients can be obtained for NMR signals according to the equation:

$$\frac{I}{I_0} = e^{-D \cdot \Delta (2\gamma \cdot s \cdot G \cdot \delta)^2} \quad [2.6]$$

where I is the signal intensity, I_0 is the initial signal intensity, D is diffusion coefficient, Δ is the diffusion delay, γ is the gyromagnetic ratio, s is gradient pulse shape; G is the applied gradient strength and δ is the duration of the gradient pulse applied.

DOSY, a 2D plot with NMR chemical shift on the x-axis and corresponding diffusion coefficients on the y-axis has been used to determine the average molecular weights of uncharged soluble polysaccharides and other macromolecules^{142–144}. Here, we apply the DOSY experiment to study the interactions between CPS and antisera at the atomic level in complex biofluids.

In this study, 10 mg of 16A CPS were dissolved in 1 ml D₂O as 16A CPS stock solution. For the interaction study, 1 mg 16A CPS was put into a 3 mm NMR tube and 5 to 50 μ l of antiserum were added. After thorough mixing, the sample solution was incubated at 310K for 15 minutes. Subsequently, the acquisition of NMR spectra was started.

Conventional ¹H NMR spectra and ¹H detected DOSY experiments were acquired at 310 K on a Bruker Avance III (799.90 MHz for ¹H and 201.14 MHz for ¹³C) equipped with a 5 mm TCI 1H/(¹³C, ¹⁵N) cryoprobe. DOSY experiments were performed using a stimulated echo sequence incorporating bipolar gradient pulses, two spoil gradients and a longitudinal eddy current delay with water presaturation (ledbpgppr2s). The diffusion time (d20) was set to 800 ms and the gradient pulse duration (p30) was 1800 μ s. The gradient strength was logarithmically incremented in at least 16 steps from 2% up to 95% of the maximum gradient strength (56 G/cm). After Fourier transformation, phase correction and baseline correction in Topspin 4.0.7, the diffusion dimension of the 2D DOSY spectrum was processed and fitted using the Bruker Dynamics Center software (version 2.6).

Chapter 3 Structural, biosynthetic and serological cross-reactive studies of CPS from serogroups 16 and 28

3.1 Structural elucidation of CPS repeat units from serogroups 16 and 28

The structure determination of pneumococcal serotype 16F, 16A, 28F and 28A CPS are mainly based on NMR spectroscopy and was validated by general chemical analysis. A summary of CPS structural information of serotype 16F, 16A, 28F and 28A are discussed in this section and the full NMR assignments can be found in paper 1 and paper 2 in the appendix.

3.1.1 Molecular weight distribution measured by AF4-MALS-dRI

The native serotype 16A CPS gave good NMR spectra, while serotype 16F, 28F and 28A CPS gave broad lines in ^1H NMR due to relatively short transverse relaxation times (T_2) of high molecular weight polysaccharides or the heterogeneity of CPS samples. This indicated that 16F, 28F and 28A CPS might have a higher molecular weight or that they are more heterogeneous than 16A CPS. To validate this assumption, AF4-MALS-dRI was used to measure the molecular weight distribution of these four CPS (Figure 3.1). The average molecular weights and sizes of CPS from serogroups 16 and 28 are summarized in Table 3.1.

Both weight average molecular weight (M_w) and number-average molecular weight (M_n) of 16A CPS are lower than the other three CPS, while 16A CPS has a higher polydispersity (M_w/M_n) that indicates a broader molecular weight distribution of 16A CPS. Moreover, 16A CPS also have a smaller root mean square radius (RMS radius). The results are in agreement with NMR line width. According to NMR determined 16A CPS structure, there was 70% O-Acetylation at the 2- β -D-Galf residue (Figure 3.3). In AF4-dRI, 16A CPS had a shoulder peak at 10 - 23 min, with a mass ratio of about 30% based on dRI peak area. This suggests it might be from 30% of non-O-acetylated CPS.

The M_w of 16F, 28F and 28A were around 653.7 kDa, 540.5 kDa and 421.9 kDa, respectively, with a similar polydispersity of around 1.2. While 16A CPS had a relatively smaller M_w of about 273.1 kDa with a higher polydispersity of around 1.35. Interestingly, 16F CPS had relatively higher M_w than 28F and 28A CPS, but showed a smaller RMS radius than 28F CPS and similar RMS radius with 28A CPS (Figure 3.1). This would be an indication of a relatively compact conformation of 16F CPS compared to 28F and 28A CPS. It is consistent with CPS repeat unit structures (Figure 3.3), as 16F CPS has a linear repeat unit, while the repeat unit of 28F and 28A CPS have the α -L-Rhap-[4-P-2-Gro] as a side

chain. In addition, 28F CPS showed a broader peak and later elution time than 16F and 28A, which could be for a stronger intra-molecular interaction. Besides, 28F CPS had a small peak at about 50 min, which was from aggregates. The small peaks at around 12 min in dRI were from the impurity, CWPS.

Table 3.1 The molecular weight, polydispersity and RMS radius of serotype 16F, 16A, 28F and 28A CPS

| Sample name | Mn (kDa) | Mw (kDa) | Polydispersity (Mw/Mn) | Rn (nm) | Rw (nm) |
|-------------|------------|------------|------------------------|----------|----------|
| 16F CPS | 563.1±10.8 | 653.7±11.8 | 1.16±0.03 | 40.1±4.2 | 46.4±3.4 |
| 16A CPS | 202.6±1.7 | 273.1±1.8 | 1.35±0.01 | 25.3±3.1 | 29.2±2.3 |
| 28F CPS | 476.9±11.5 | 540.5±9.1 | 1.13±0.03 | 51.3±3.9 | 55.5±2.9 |
| 28A CPS | 363.6±3.4 | 421.9±3.3 | 1.16±0.01 | 41.7±0.9 | 45.1±0.8 |

Mn: number-average molecular weight; Mw: weight-average molecular weight; Rn: number-average mean square radius; Rw: weight-average mean square radius.

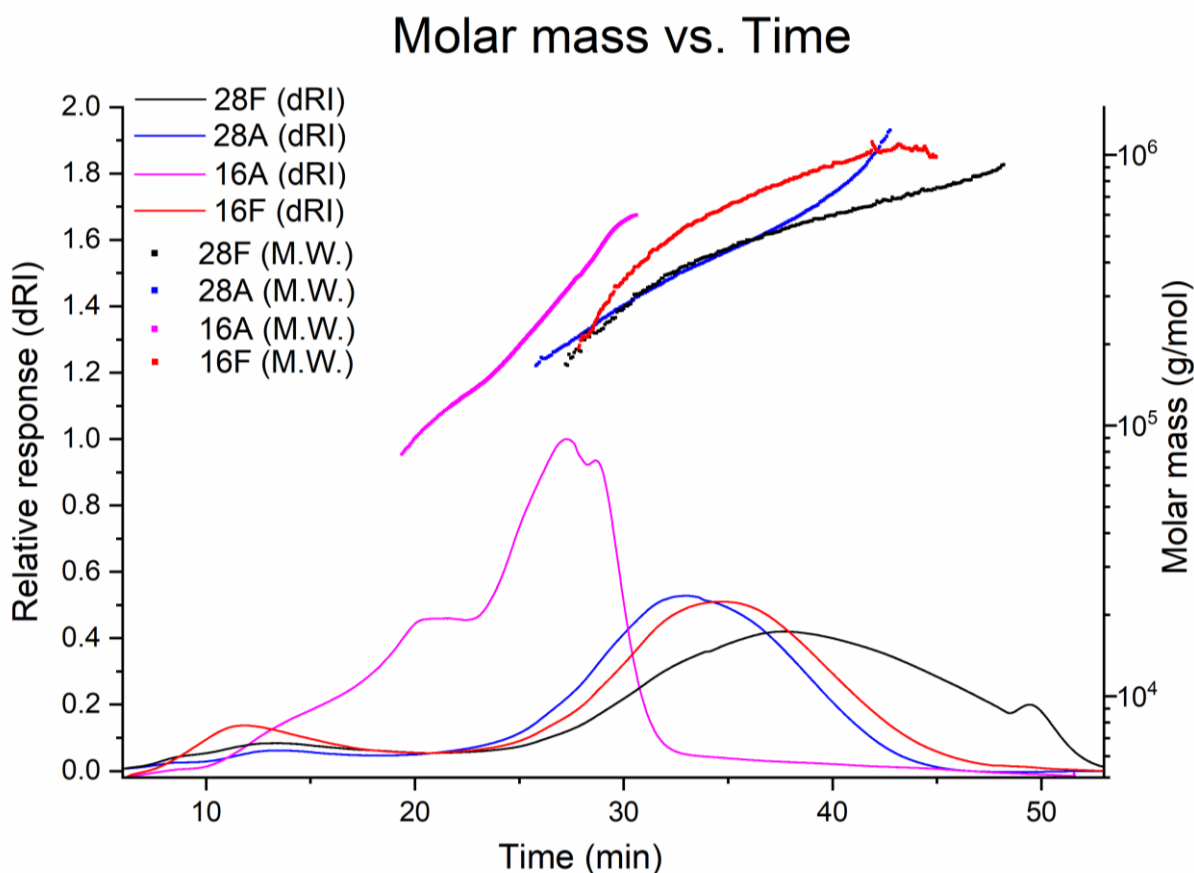


Figure 3.1 Molar mass distribution (dots) and differential refractive index response (solid lines) of serotype 16F (red), 16A (pink), 28F (black) and 28A (blue) CPS

3.1.2 General chemical analysis of CPS

The results of monosaccharides composition, phosphate content and methylation analysis of CPS from serotype 16F, 16A, 28F and 28A are summarized in Table 3.2. 16A CPS had a different monosaccharides profile, as it contained three D-Gal, while the other three CPS did not have D-Gal in the repeat units. The 16F and 28F CPS have similar monosaccharide and phosphate composition, but different linkage types. The linkage types of 16F indicated that the repeat unit structure was linear. While the presence of 3, 4-Glc in 28F and 3, 4-GlcNAc in 28A CPS suggested that these two sugar residues are 3,4-substituted by other sugar residues, as the samples were subjected to dephosphorylation before the linkage analysis. The 28A CPS showed a higher amount of PO₄ which probably originate from a higher level of CWPS impurities in the sample. The hydrolysis of the glycosidic bonds of 2-acetamido sugars can give rise to severe problems, due to a competition between the hydrolysis of the N-acetyl linkage, yielding an inhibiting protonated amino group at C2, and the hydrolysis of the glycosidic linkage. Therefore, the 2-amino sugars are not quantitative. However, the signal in GC can still indicate the amount of amino sugars.

Table 3.2 Results of monosaccharides compositional analysis, phosphate, methylation and absolute configuration analysis

| Serotype | L-Rha | D-Glc | D-Gal | D-GlcN | PO ₄ | Linkage type |
|----------|-----------|-------|-------|--------|-----------------|---|
| | (nmol/mg) | | | | | |
| 16F | 1545 | 1276 | - | + | 1770 | 2-Rha, 4-Glc, 3-Glc |
| Ratio | 1.2 | 1.0 | - | - | 1.4 | |
| 16A | 1455 | 686 | 2005 | - | 942 | 2-Rha, 3-Rha, 4-Glc, 3-Glc, 3-Galf, 2,3-Gal |
| Ratio | 2.1 | 1.0 | 2.9 | - | 1.4 | |
| 28F | 1605 | 1521 | - | - | 1686 | t-Rha, 3-Rha, 4-Glc, 3,4-Glc |
| Ratio | 1.1 | 1.0 | - | - | 1.1 | |
| 28A | 1011 | 610 | - | +++ | 1505 | t-Rha, 3-Rha, 4-Glc, 3,4-GlcNAc |
| Ratio | 1.7 | 1.0 | - | 1 | 2.5 | |

3.1.3 NMR analysis of serotype 16F, 16A, 28F and 28A CPS

The assignment of native 16A CPS by NMR was relatively straightforward at the beginning, as it gave sharp NMR peaks and contained less CWPS impurities (Figure 3.2). However, 16A CPS presented partial O-acetylation at the 2- β -D-Galf, which resulted in two different chemical shifts with 70% and 30% relative intensity for the positions close to O-acetylation site. Therefore, a de-O-acetylation (de-OAc) is needed to simplify NMR spectra and confirm the position of O-acetylation. Full assignment of native, de-OAc and de-PO₄ 16A CPS can be found in the appendix (paper 1).

As the native 16F, 28F and 28A CPS all gave broad lines in NMR, even at 333 K (Figure 3.2), they were all subjected to ultra-sonication to depolymerize the samples. The ¹H spectra of native 16F and 28F were highly similar, because they had the same monosaccharide compositions. Furthermore, the intensities of the phosphorylcholine (P-Cho) peaks were extremely high in 28F and 28A CPS, while P-Cho in the 16F sample had similar intensity as the other CWPS signals. This indicated that 28F and 28A CPS might contain P-Cho in their CPS repeat unit. Full assignment of native, de-OAc and de-PO₄ 16F and 16A CPS can be found in paper 1. We did not acquire NMR data of de-PO₄ 28F and 28A CPS, as they were hardly dissolvable in D₂O. Detailed assignment of native and de-OAc 28F and 28A CPS can be found in paper 2.

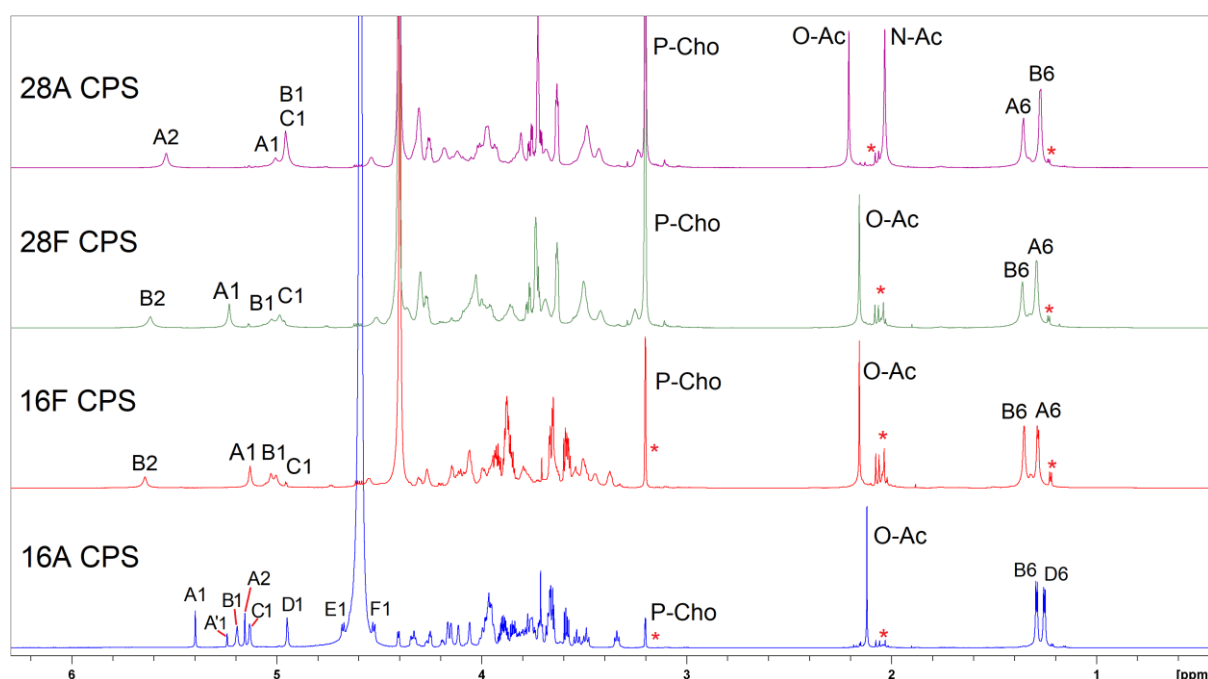


Figure 3.2 Overlapped ¹H spectra of native 16A (blue, 313k), 16F (red, 333k), 28F (green, 333k), and 28A (purple, 333k) CPS. The peaks from CWPS impurity are marked with the red “*”.

The repeat unit structures of 16A, 16F, 28F and 28A CPS are shown below.

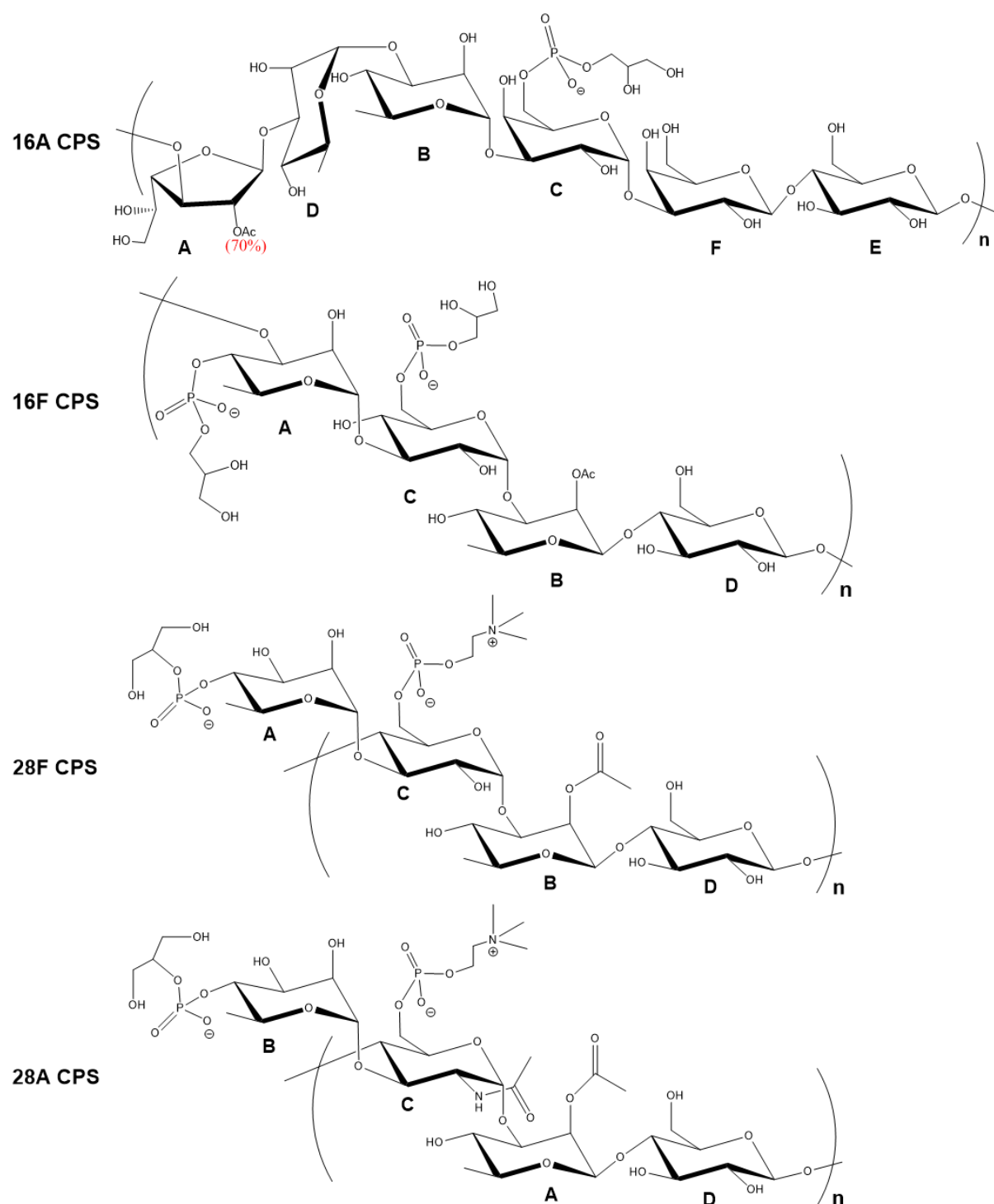


Figure 3.3 The repeat units structures of 16A, 16F, 28F and 28A CPS

3.2 Biosynthesis of serogroups 16 and 28

As serogroup 16 is closely related to serogroup 28, they are discussed together for comparison.

3.2.1 The *cps* loci comparison of serotype 16F, 16A, 28F and 28A

Biosynthesis of pneumococcal serotype 16F, 16A, 28F and 28A CPS use the Wzy-dependent mechanism. The blastn alignment results of *cps* sequences of serotype 16A, 28F and 28A are shown in Table 3.3. The *cps* genes showed more than 90% nucleotide identity within serotype 16F, 28F and 28A with around 80% query cover. Serotype 16F, 28F and 28A *cps* loci constitute a subcluster in cluster analysis with conservative threshold^{114,145}. In contrary, serotype 16A *cps* locus has little in common with serotype 16F. The sequence alignment of 16F and 16A *cps* showed a low query cover (45%), as their coding sequence (CDS) did not share many genes, and particularly GT-encoding genes are different. On the other hand, serotype 16F, 28F and 28A have most genes in common except the glycerol-phosphate synthetic genes, which is *gct* in 16F but is *gct1*, *gct2* and *gct3* in 28F and 28A (Figure 3.4). In addition, the polymerase gene *wzy* in serogroup 28 falls into a different homology group compared to serotype 16F. A further pairwise comparison of amino acid sequences of each CPS biosynthetic enzyme in 16F, 28F and 28A is summarized in Table 3.4. A list of accession numbers of protein sequences of enzymes encoded in *cps* are summarized in the appendix (Table A2). A detailed discussion of serogroup 16 and serogroup 28 biosynthetic genes, their functions and related serotypes can be found in paper 1 and paper 2 in the appendix.

Table 3.3 The BLAST results of *cps* nucleotide sequence of serotype 16F (query sequence), 16A, 28F and 28A

| Strain and serotype | Accession | <i>cps</i> locus DNA length (bp) | Total score | Query cover | Percentage identity |
|--|------------|-------------------------------------|----------------|----------------|------------------------|
| <i>Streptococcus pneumoniae</i> strain nr. 34361 (serotype 16F) | CR931668.1 | 21481 | | | |
| <i>Streptococcus pneumoniae</i> strain R105 (serotype 16A) | CR931667.1 | 21730 | 16827 | 45% | 96.7% |
| <i>Streptococcus pneumoniae</i> strain 34372 (serotype 28F) | CR931693.2 | 21839 | 25119 | 79% | 92.4% |
| <i>Streptococcus pneumoniae</i> strain 1982/45 (serotype 28A) | CR931692.1 | 22978 | 26751 | 85% | 98.0% |

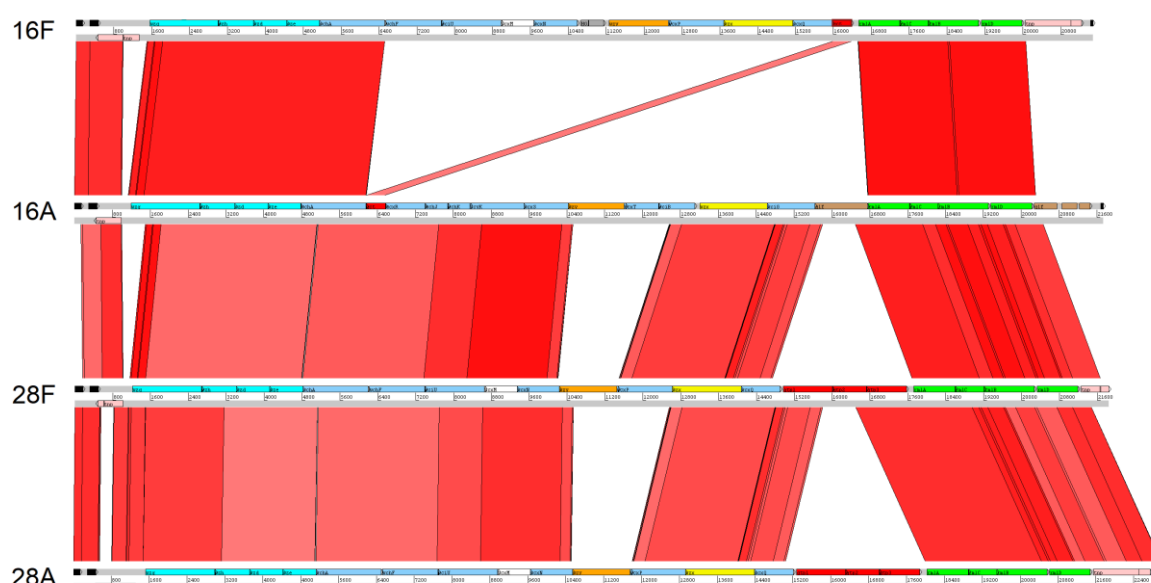


Figure 3.4 Comparison of *cps* gene loci of serotype 16F (reference sequence), 16A, 28F and 28A in ACT with score cutoff 200 and identity cutoff 60%.

Table 3.4 Pairwise comparison of synthetic enzymes in serotype 28F with 28A and 16F

| Enzyme name | Putative function | Percentage identity (%) | |
|-------------|--|-------------------------|--------------|
| | | serotyp 28A | serotype 16F |
| Wzg | Regulatory enzymes | 99.6 | 99.2 |
| Wzh | | 99.6 | 85.6 |
| Wzd | | 99.1 | 76.1 |
| Wze | | 100.0 | 78.8 |
| WchA | Initial transferase | 99.6 | 71.4 |
| WchF | GT | 99.5 | 94.6 |
| WciU | GT | 94.1 | 98.4 |
| WcxM | Acetyl transferase | 100.0 | 99.6 |
| WcxN | GT | 100.0 | 94.1 |
| Wzy | Polymerase | 99.5 | 26.5 |
| WcxP | phosphotransferase | 100.0 | 86.3 |
| Wzx | Flipase | 100.0 | 94.0 |
| WcxQ | phosphotransferase | 100.0 | 90.8 |
| Gtp1 | Glycerol-2-phosphate synthetic enzymes | 99.7 | / |
| Gtp2 | | 97.0 | |
| Gtp3 | | 98.9 | |
| Gtp (16F) | Glycerol-1-phosphate synthetic enzymes | / | - |
| RmlA | Rhamnose synthetic enzymes | 100.0 | 99.3 |
| RmlC | | 96.5 | 97.5 |
| RmlB | | 99.4 | 99.4 |
| RmlB | | 98.9 | 98.9 |

3.2.2 Correlation of CPS structures with biosynthesis of serotype 16A, 16F, 28F and 28A

Based on published CPS structures in literatures and determined structures of 16F, 16A, 28F and 28A CPS, the catalytic linkages of related enzymes were assigned (Figure 3.5).

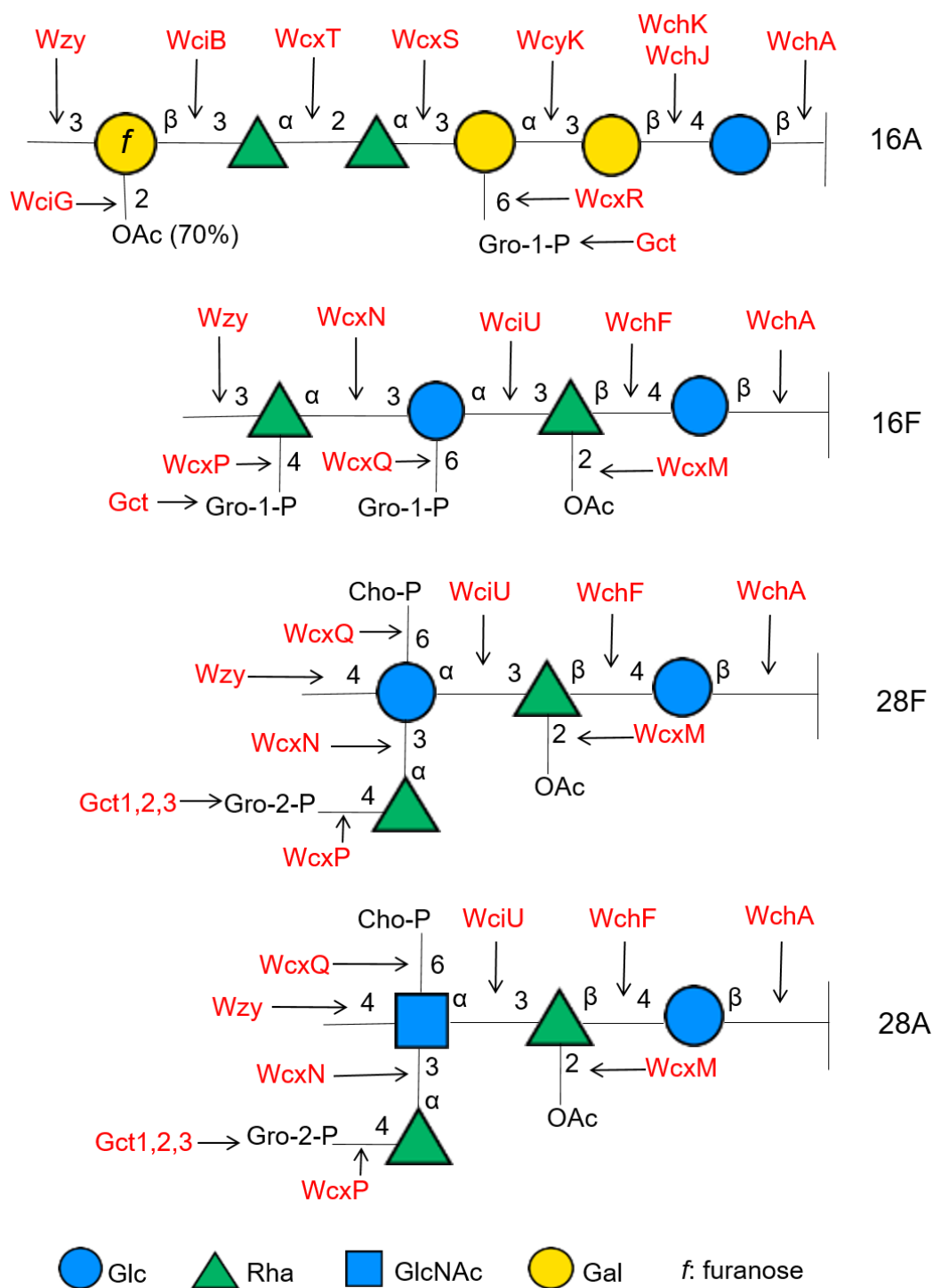


Figure 3.5 Assignment of the function of biosynthetic enzymes (red letters) in serotype 16A, 16F, 28F and 28A.

Biosynthesis of CPS repeat units from all the four serotypes starts with transferring the first Glc by initial transferase WchA. Next in 16A, the addition of β -D-Galp onto the 4-position of β -D-Glcp is transferred by the GT WchK together with a GT enhancer WchJ⁷². The GT WcyK, which are also present in serogroup 11, added α -D-Galp to 3- β -D-Galp. The WcxS catalyzes the linkage of α -L-Rhap-(1-3)- α -D-Galp in serotype 45, which is in agreement with the assignment in 16A. The GT WciB is putative β -1-3-galactofuranosyl transferase, found in 25 serotypes including serotype 16A^{67,72}. Therefore, by the process of elimination, the WcxT was assigned to catalyze the glycosidic bond of α -L-Rhap-(1-2)- α -L-Rhap. The *wcxT* gene is unique in serotype 16A, which could be used as a target for genetic serotyping. Wzy of 16A was assigned to polymerize the repeat unit at β -D-Glcp-(1-3)- β -D-Galf. The phosphotransferase WcxR transferred Gro-1-P to the 6-position of α -D-Galp. The acetyl transferase WciG is responsible for partially adding the acetyl groups to 2- β -D-Galf.

Serotype 16F, 28F and 28A have three common GTs (WchF, WciU and WcxN) for synthesis of the CPS repeat units. WchF (also known as Cps2T) was elucidated to be a β -1-4 rhamnosyltransferase that adds the second sugar to the repeat unit¹⁴⁶. The WciU is also present in serogroup 18, which was assigned as adding the α -D-Glcp (serotype 18F, 18B and 18C) or α -D-GlcpNAc (serotype 18A) to the 3-position of β -L-Rhap⁷². Similarly, the GT WciU showed different sugar donor specificity in serotype 28A compared to serotype 16F and 28F (Table 3.5)¹⁴⁷. The WciU in serotype 16F, 18F, 18B, 18C and 28F are glucosyltransferases, which are specific for adding α -D-Glcp to the 3-position of β -L-Rhap. in serotype 18A and 28A WciU transfers α -D-GlcpNAc to O3 of β -L-Rhap. While WciU in serotype 18A and 28A transfer α -D-GlcpNAc to 3- β -L-Rhap. The GT WcxN was assigned to add α -L-Rhap to the 3-position of α -D-Glcp in serotype 16F, which is in agreement with the function of WcxN in 28F. However, WcxN in serotype 28A transfer α -L-Rhap to the 3-position of α -D-GlcpNAc. This means that WcxN could have a broader specificity of the acceptor in the catalytic process.

As there are two phosphate groups present in serotype 16F, 28F and 28A, two putative phosphotransferase WcxP and WcxQ are responsible for adding the phosphate substitutions. The putative function of WcxP and WcxQ are summarized in Table 3.6. However, further experiments are needed to validate the putative functions of these phosphotransferases. In addition, the acetyl transferase WcxM in serotype 16F, 28F and 28A catalyzes the O-acetylation at the 2-position of β -L-Rhap, which was also found in serotype 18F.

Table 3.5 A summary of pneumococcal serotypes with WciU, their protein sequence identity and catalytic donor, linkage and acceptor ¹⁴⁷

| Serotype with WciU | Percentage identity (refer to WciU of 28F) | Donor | Linkage | Acceptor |
|--------------------|--|-----------|----------------|-----------------|
| 28F | 100.0 % | | | |
| 18F | 99.1 % | D-Glcp | α (1-3) | |
| 16F | 98.6 % | | | |
| 18B | 92.6 % | | | β -L-Rhap |
| 18C | 92.6 % | D-Glcp | α (1-3) | |
| 18A | 89.0 % | D-GlcNAcp | α (1-3) | |
| 28A | 94.1 % | D-GlcNAcp | α (1-3) | |

Table 3.6 Comparison of the phosphotransferase WcxP and WcxQ in serotype 16F, 28F and 28A

| Enzyme | Putative function | Serotype | AA Length | Identity* (%) | Donor | Acceptor |
|--------|--------------------------------|----------|-----------|---------------|---------|------------------------|
| WcxP | glycerol phosphotransferase | 28F | 386 | 100 | Gro-2-P | 4- α -L-Rhap |
| | | 28A | 386 | 100 | Gro-2-P | 4- α -L-Rhap |
| | | 16F | 388 | 86.3 | Gro-1-P | 4- α -L-Rhap |
| WcxQ | LicD-family phosphotransferase | 28F | 270 | 100 | Cho-P | 6- α -D-Glcp |
| | | 28A | 270 | 100 | Cho-P | 6- α -D-GlcpNAc |
| | | 16F | 271 | 90.7 | Gro-1-P | 6- α -D-Glcp |

*The percentage identity was from blastp alignment using WcxP or WcxQ in 28F as the query sequence.

3.2.3 The donor specificity of WciU

The phylogenetic analysis indicated that the split between serotype 28F and 28A was due to the recombination at the *wciU* gene¹⁴⁵. As discussed in 3.2.1, WciU transfers either D-Glcp or D-GlcpNAc to the 3-position of β -L-Rhap. Therefore, the differences in *wciU* gene of serotype 28F and 28A lead to different CPS structures and thus the *wciU* gene is crucial for distinguishing serotype 28F from serotype 28A. A better understanding of the key residues that result in the donor specificity of WciU would be important for genetically differentiating them. Both UDP-Glcp and UDP-GlcpNAc are available in cytoplasm from housekeeping metabolic pathways. Consequently, the specificity of WciU in serotype 28F and 28A should result from the different protein structures. The WciU belongs to the retaining GT4 enzyme in CAZy family with a GT-B fold and is present in seven serotypes among all the pneumococcal serotypes^{67,114,148} (Table 3.5). Two Pfam domains are found in 28F WciU (UniProtKB: Q4K002): PF13439 (position 16-137) and PF00534 (position 323-376). Besides, the conserved EX₇E motif is often found to be critical for the nucleotide recognition domain (NRD) of the retaining GTs and the fourth position of the EX₇E motif (two Glu separated by 7 aa) is usually a glycine for the Glc-transferase (Glc-T)^{149–152}. There are EX₇E motifs found in these seven WciU sequences (Figure 3.6). Only EX₇E (348-356) is within the predicted domain (position 323-376). The multiple sequence alignment of seven WciU GTs with four GT4 family N-acetylglucosaminyltransferases (GlcNAc-Ts) show that the residues from position 319 to 376, especially around the EX₇E motifs, are relatively conserved compared to the other parts in the whole sequences (Figure 3.6, Figure A1). MSA of WciU from all seven serotypes shows that there are 18 aa out of 423 aa, which are the same in 28A and 18A, but different from those in 16F, 28F, 18F, 18B and 18C (Figure 3.6). Among these 18 residues, the ones in positions 329, 342, 343 and 346 are within the predicted GT domain, which could be critical for the donor substrate specificity.

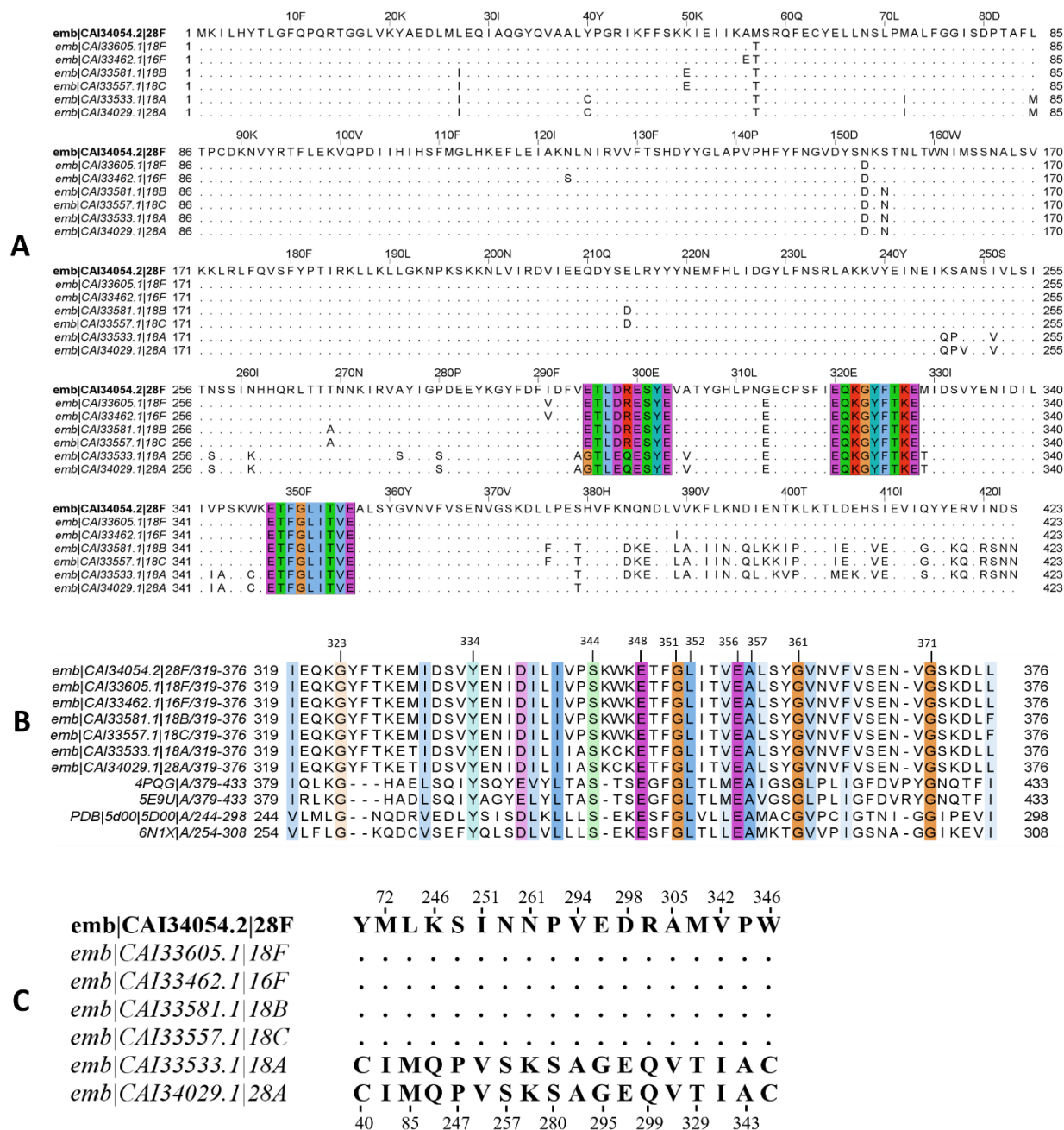


Figure 3.6 (A) Multiple sequence alignment of WciU from serotype 16F, 18F, 18A, 18B, 18C, 28F and 28A. EX₇E motifs are colored by Clustalx. (B) Multiple sequence alignment of WciU from serotype 16F, 18F, 18A, 18B, 18C, 28F and 28A with protein sequences of GT4 GlcNAc-Ts (4PQG.pdb, 5E9U.pdb, 5D00.pdb and 6N1X.pdb). Alignment is colored by Clustalx with 50% conservation increased visibility. (C) The 18 residues that are identical in 28F, 16F, 18F, 18B and 18C, but are different in the corresponding sites of 18A and 28A which are identical.

3.2.3.1 Protein structure models of WciU

In order to identify the critical amino acid residues that are responsible for the substrate selectivity, 3D protein models of WciU from 28F and 28A were built by Phyre2 (Figure 3.7). The templates used for building WciU models are summarized in Table A3 and Table A4.

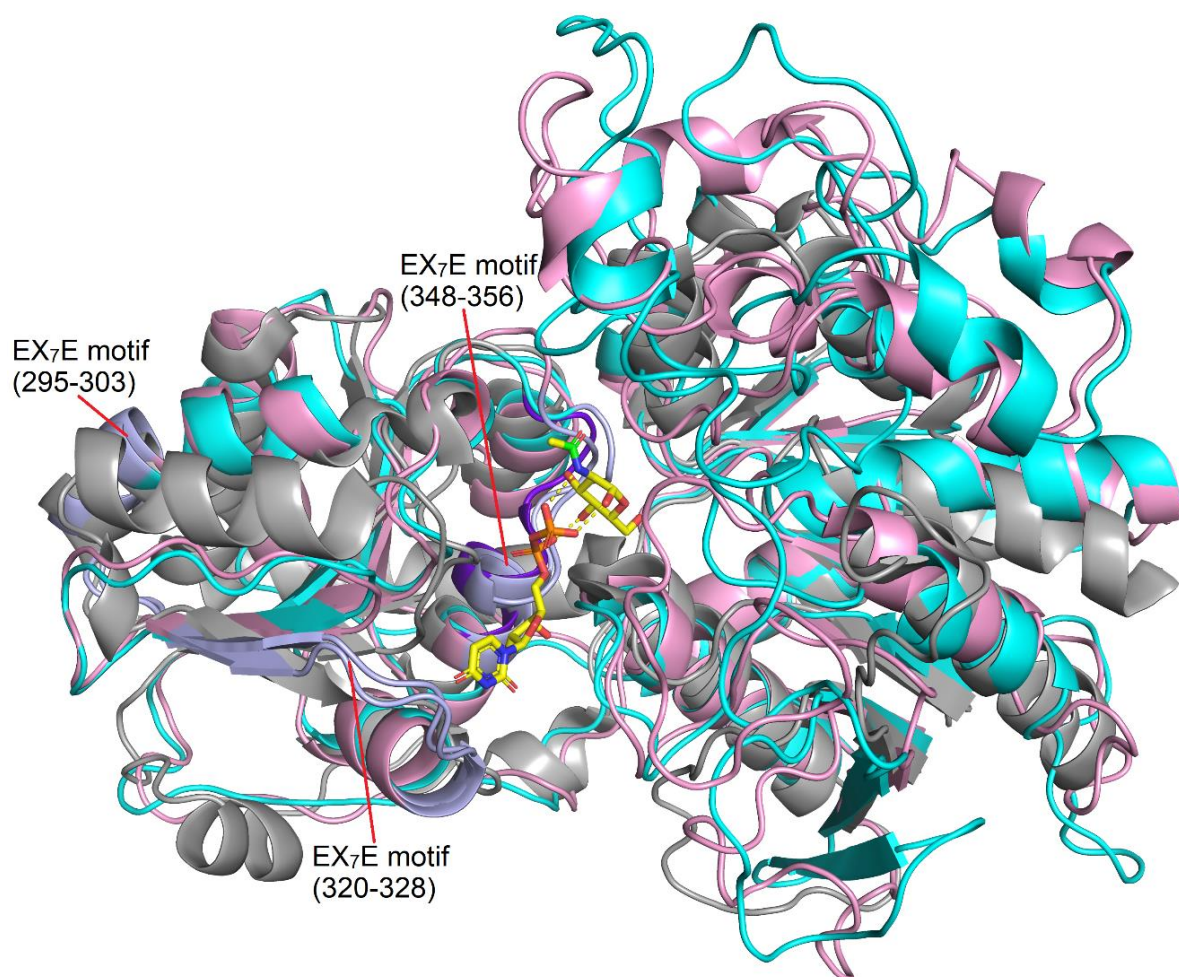


Figure 3.7 Superimposed protein models of WciU of 28F (pink, light purple highlights the three EX₇E motifs), 28A (cyan, light purple highlights the three EX₇E motifs) and protein structures of *N*-acetylglucosaminyltransferases from *S. aureus* (6N1X.pdb, grey, dark purple highlights the EX₇E motif). The UDP-GlcNAc (yellow stick) is flanked by the EX₇E motif of positions 348-356 (dark purple). The executive RMSD between the superimposed models with the GlcNAc-T (6N1X.pdb) are < 3.0 Å².

Among the three EX₇E motifs in WciU, only the one at position 348-356 is close to the UDP-GlcNAc ligand. Moreover, the EX₇E (348-356) motifs of 28F and 28A WciU models (light purple) align very well with the EX₇E motif of GlcNAc-T (dark purple, 6N1X.pdb), which is

flanking the UDP-GlcNAc ligand site. This is consistent with the predicted domain and the essential role of EX₇E motif for retaining GT4 catalytic activity.

Furthermore, by comparison with the sequences and the structures of distant GT4 *N*-acetylglucosaminyltransferases from *S. aureus* (6N1X.pdb), *S. gordonii* (5E9U.pdb) and *S. pneumoniae* (4PQG.pdb), the residue in position 346 seems to be the one in the closest vicinity (~5 Å to the C_α) of the 2 acetamido group among the non-identical 18 residues between the two isoforms of WciU (Figure 3.8). It is also consistent that a change from a bulky Trp in 28F WciU into a smaller Cys in 28A WciU could allow more space for the acetamido group to bind into the pocket. Subsequently, a donor substrate specificity of WciU shifts from a Glc (28F) to a GlcNAc (28A). The second closest residue (342) to the 2-acetamido group is located around 10 Å away with a rather small change (Ile versus Val), and deemed unlikely to affect the specificity.

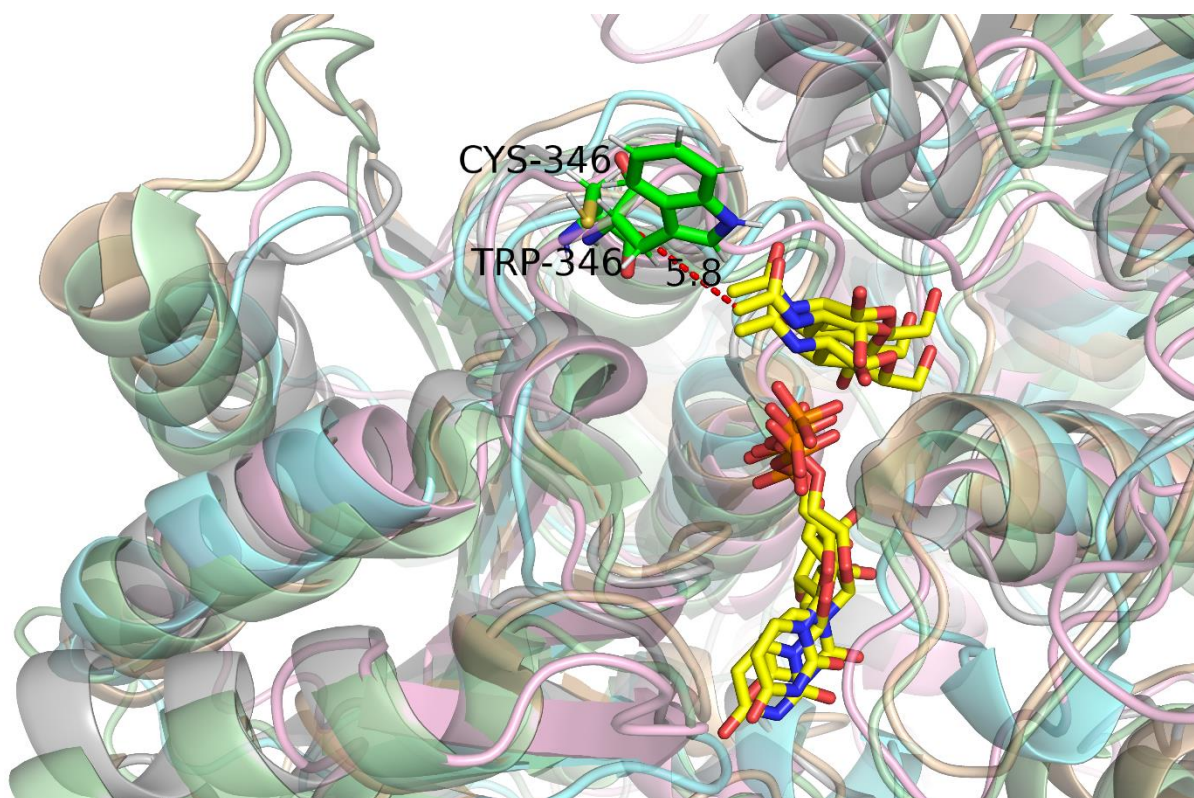


Figure 3.8 Superimposed protein models of WciU of 28F (pink), 28A (cyan) and protein structures of GlcNAc-T from *S. aureus* (grey, 6N1X.pdb), *S. gordonii* (wheat, 5E9U.pdb) and *S. pneumoniae* (light green, 4PQG.pdb). The UDP-GlcNAc ligands from GlcNAc-Ts are all shown in yellow. The residues Trp346 (28F WciU) and Cys346 (28F WciU) are highlighted with green sticks. The dashed red line indicates that the distance from the 2-acetamido group to the TRP-346 is 5.8 Å.

While the orientation and exact location of the side-chain residues from *in silico* WciU models are not reliable, the overall position of the residue, especially in the conserved domain, is most likely correct. Therefore, we could draw a preliminary conclusion that the residue 346 is critical for the donor specificity of WciU. In addition, a similar donor specificity shifting from Glc towards GlcNAc of a retaining GT4 WciY from *E. coli* O107 were created by mutating a bulky Arg194 to a Cys¹⁴⁹. The superimposed models of WciY from *E. coli* O107 with WciU in 28A showed that the C194 that was responsible for the specificity shift was in the close vicinity of C346 (Figure 3.9). However, experimental validations are required to confirm the predicted key 346 residue.

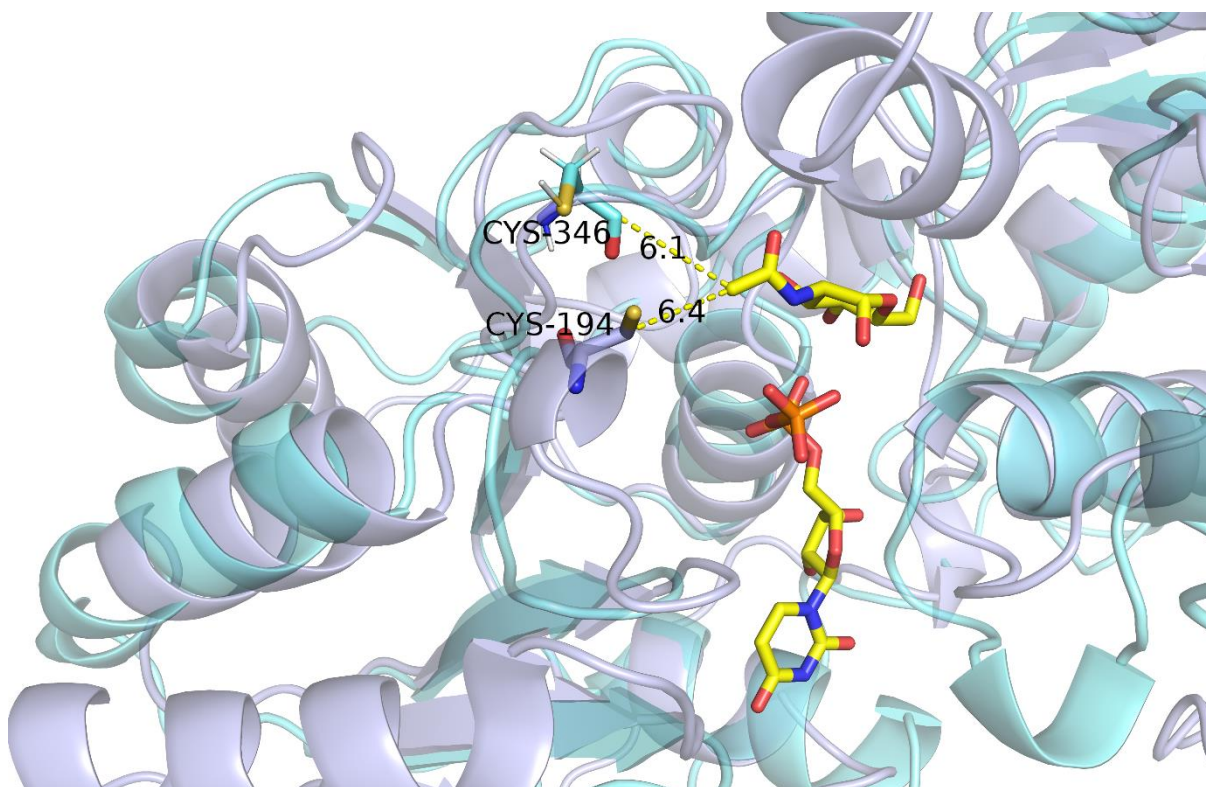


Figure 3.9 Superimposed protein models of WciU of 28A (cyan) and the GlcNAc-T WciY (light purple) from *E. coli* O107. The UDP-GlcNAc ligand are shown in yellow. The yellow dash lines indicate that the distances from the 2-acetamido group to the Cys346 (6.1 Å) and Cys194 (6.4 Å).

3.3 serological cross-reactions of serogroups 16 and 28

3.3.1 Serological cross-reactive analysis by Quellung reaction

When using Quellung reaction for serotyping *pneumococcus*, serogroup 16 first reacts with pool antiserum D and was identified by group 16 antisera. Then, serotype 16F and 16A was further distinguished by factor antisera 16b and 16c, respectively. The factor antisera were made from immunized rabbits with specific serotype pneumococcus, while cross-reactive antibodies of other serotypes within the serogroup were absorbed in the factor antisera. Therefore, we could use factor antisera to check the cross-reaction with other serotypes outside of the serogroup.

Serotype 16F cross-reacted with factor antiserum 28b (Table 3.7), which was used for serotyping 28F. While serotype 16A showed negative Quellung reactions with factor antiserum 28b. This was in agreement with the structural similarity between 28F and 16F CPS. The positive Quellung reaction between serotype 16F and factor serum 11d was observed before ¹¹⁴. Factor antiserum 11c, which reacted with serotype 11A, 11C and 11D, was reported to bind to P-1-Gro and OAc groups ¹⁵³. Thus, we tested cross-reaction of serogroup 16 with serogroup 11 and serogroup 28. 16F and 16A cross-reacted with factor antiserum 11c were observed (Table 3.7), which indicated that they shared similar antigenic determinants. The CPS structures of serotype 11A, 11C and 11D also consist of a linear oligosaccharide with P-1-Gro and O-acetyl substitutes. Negative Quellung reactions between serotype 28F and factor antiserum 11c suggested that factor antiserum 11c might be specific to P-1-Gro and not P-2-Gro. Moreover, it was found that the O-acetylation in serogroup 11 CPS was necessary for their antigenicity ^{153,154}. The cross-reactions of serogroup 16 and factor antiserum 11c supported that factor 11c epitope was associated with P-1-Gro and OAc groups. Additionally, serotype 28F and 28A cross-reacted with factor antiserum 23d that was used for serotyping 23B. The common part of CPS structures of serotype 28F, 28A and 23B was a trisaccharide backbone with P-2-Gro on the external surface, which might be the reason for the cross-reactions ^{147,155}.

Table 3.7 Cross-reactions between serogroup 16 and serogroup 28 using Quellung reaction

| Serotype | Factor sera | | | | | | | | | | |
|----------|-------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | 16b | 16c | 28b | 28c | 11b | 11c | 11f | 11g | 23b | 23c | 23d |
| 16F | +++ | - | +++ | - | - | +++ | - | - | - | - | - |
| 16A | - | +++ | - | - | - | +++ | - | - | - | - | - |
| 28F | - | - | +++ | - | - | - | - | - | - | - | +++ |
| 28A | - | - | - | +++ | - | - | - | - | - | - | +++ |

3.3.2 Serological cross-reactive analysis by latex agglutination

In order to understand the antigenic properties of serogroups 16 and 28 CPS, we tested agglutination reactions of the de-OAc and de-PO₄ 16F, 16A, 28F and 28A CPS with latex coated factor antisera (Table 3.8, Figure A2-A3). Interestingly, the de-OAc and de-PO₄ 16A CPS can still react with 16c latex solution with reduced intensity, which indicated that there were other epitopes in 16A CPS for antibodies in 16c antiserum binding. While the de-OAc and de-PO₄ 16A and 16F CPS all showed negative reaction with 11c latex solution, which confirmed the importance of OAc groups for 11c antiserum binding.

According to the determined CPS structures of 28F and 28A, α -D-Glcp in 28F CPS and α -D-GlcpNAc in 28A CPS are the critical epitopes for factor antiserum 28b and 28c binding, respectively. The reduced cross-reactions between the de-OAc 28F / 28A CPS and 23d latex solution showed that OAc groups would affect the binding affinity of 23d antiserum. While the removal of PO₄ groups from 28F and 28A CPS can result in negative agglutination. It showed the PO₄ group was necessary for antiserum 23d binding. A detailed discussion can be found in the appendix (Paper 2).

Table 3.8 Cross-reactions factor antisera with de-OAc and de-PO₄ 16F, 16A, 28F and 28A CPS using latex agglutination

| Samples | Factor antisera Latex solution | | | | | |
|----------------------------|--------------------------------|-----|-----|-----|-----|-----|
| | 16b | 16c | 28b | 28c | 11c | 23d |
| 16F native CPS | +++ | - | +++ | - | +++ | / |
| 16F de-OAc CPS | - | - | - | - | - | / |
| 16F de-PO ₄ CPS | - | - | - | - | - | / |
| 16A native CPS | - | +++ | - | - | +++ | / |
| 16A de-OAc CPS | - | ++ | - | - | - | / |
| 16A de-PO ₄ CPS | - | + | - | - | - | / |
| 28F native CPS | / | / | +++ | - | / | +++ |
| 28F de-OAc CPS | / | / | - | - | / | + |
| 28F de-PO ₄ CPS | / | / | - | - | / | - |
| 28A native CPS | / | / | - | +++ | / | +++ |
| 28A de-OAc CPS | / | / | - | - | / | + |
| 28A de-PO ₄ CPS | / | / | - | - | / | - |

“+++” : positive agglutination reaction; “++” : medium positive agglutination reaction; “+” : weakly positive agglutination reaction; “-” : negative agglutination reaction; “/” : not tested.

Chapter 4 A discovery of a new serotype in serogroup 24

There are three different serotypes in serogroup 24, serotype 24F, 24A and 24B, none of which are included in current vaccines. As a part of the pneumococcal IPD surveillance strategy, an unknown serotype with antigenic disparity among several pneumococcal isolates belonging to serogroup 24. The isolates displayed a serological profile which does not correspond to any known serotypes in serogroup 24, hence, were provisionally labelled as serotype “24X” (Table 4.1). The 24X strain reacts with both 24d and 24e factor antisera, which indicate 24X produce a new capsule. In serogroup 24, only serotype 24F CPS structure was previously known¹⁵⁶. In order to get more knowledge of serogroup 24 and to confirm the novelty of 24X, additional CPS structures from serogroup 24 (24F, 24A, 24B) were studied.

Table 4.1 Quellung reaction of serogroup 24 with diagnostic antisera

| Serotype | 24c | 24d | 24e |
|----------|-----|-----|-----|
| 24F | - | + | - |
| 24A | + | + | - |
| 24B | - | - | + |
| 24X | - | + | + |

4.1 Structure analysis of CPS from serogroup 24

4.1.1 General chemical analysis

The results of monosaccharides composition, phosphate content and methylation analysis of CPS from serotype 24F, 24A, 24B and 24X are summarized in Table 4.2. According to the NMR determined serogroup 24 CPS structures, CPS of all the four serotypes had the same backbones, $-4)-\beta\text{-Glc}p\text{NAc}-(1-4)-\beta\text{-Rhap}-(1-4)-\beta\text{-Glc}p-(1$. However, the sidechains of these four CPS varied a lot. 24F and 24B CPS have the same $-1-\alpha\text{-Rhap}-(1-4)-\beta\text{-Rib}f$ sidechain, but have the different phosphate sidechains, which are $\text{PO}_4-1\text{-Ara-ol}$ and $\text{PO}_4-1\text{-Rib-ol}$, respectively. While 24A CPS has the $-1-\alpha\text{-Rhap}$ without 4-position substituted $\beta\text{-Rib}f$ in the sidechain. Moreover, 24A CPS has two phosphate substitutions, the P-Cho and the $\text{PO}_4-1\text{-Ara-ol}$. However, general chemical analysis indicated that the 24F, 24B and 24X samples were not very homogenous, especially the sidechains were not stoichiometric. The ratio of Rib and Ara showed a lower ratio than expected. However, this could also be a result of the

hydrolysis methods having potentially partially destroyed the sugar alcohols. Before the methylation analysis, all the CPS samples were subjected to de-phosphorylation by HF at 4 °C. The terminal Rhamnose (t-Rha) was found in all serogroup 24 CPS. However, based on NMR determined structures, t-Rha should only be present in 24A CPS. Thus, it might be generated by the purification procedures or hydrolysis methods that inadvertently removed some Ribf sidechain. The 24F and 24X CPS samples contained high percentage of CWPS, which contains at least two phosphate groups in the repeat unit ¹⁴. Consequently, the PO₄ amounts in 24F and 24X CPS were not reliable.

Table 4.2 Results of monosaccharides compositional analysis as well as phosphate and methylation analysis

| Serotype | Rha | Glc | Rib | Ara | GlcN | PO ₄ | Linkage type |
|----------|-----------|-----|-----|-----|------|-----------------|---|
| | (nmol/mg) | | | | | | |
| 24F | 1203 | 535 | 325 | 440 | +++ | 1014 | t-Rib <i>f</i> , t-Rib <i>p</i> , 4-Rha, 4-Glc, 3,4-GlcNAc, t-Rha |
| Ratio | 2.2 | 1.0 | 0.6 | 0.8 | 1 | 1.9 | |
| 24A | 1304 | 585 | 0 | 509 | +++ | 1514 | t-Rha, 4-Rha, 4-Glc, 3,4-GlcNAc |
| Ratio | 2.2 | 1.0 | 0.0 | 0.9 | 1 | 2.6 | |
| 24B | 1485 | 635 | 818 | 0 | +++ | 895 | t-Rib <i>f</i> , t-Rib <i>p</i> , 4-Rha, 4-Glc, 3,4-GlcNAc, t-Rha |
| Ratio | 2.3 | 1.0 | 1.3 | 0 | 1 | 1.4 | |
| 24X | 862 | 433 | 293 | 264 | +++ | 1128 | t-Rib <i>f</i> , t-Rib <i>p</i> , 4-Rha, 4-Glc, 3,4-GlcNAc, t-Rha |
| Ratio | 2.0 | 1.0 | 0.7 | 0.6 | 1 | 2.6 | |

4.1.2 NMR analysis of repeat unit structures of serotype 24B and 24A CPS

The native CPS from serogroup 24 all gave broad lines due to fast relaxation of macromolecules or the heterogeneous nature of the samples. Slightly depolymerizing CPS by ultra-sonication and acquisition at a higher temperature (323K) can improve NMR line shapes by increasing molecular tumbling rate. The repeat unit structures of serotype 24A, 24B and 24X CPS were determined (Figure 4.1).

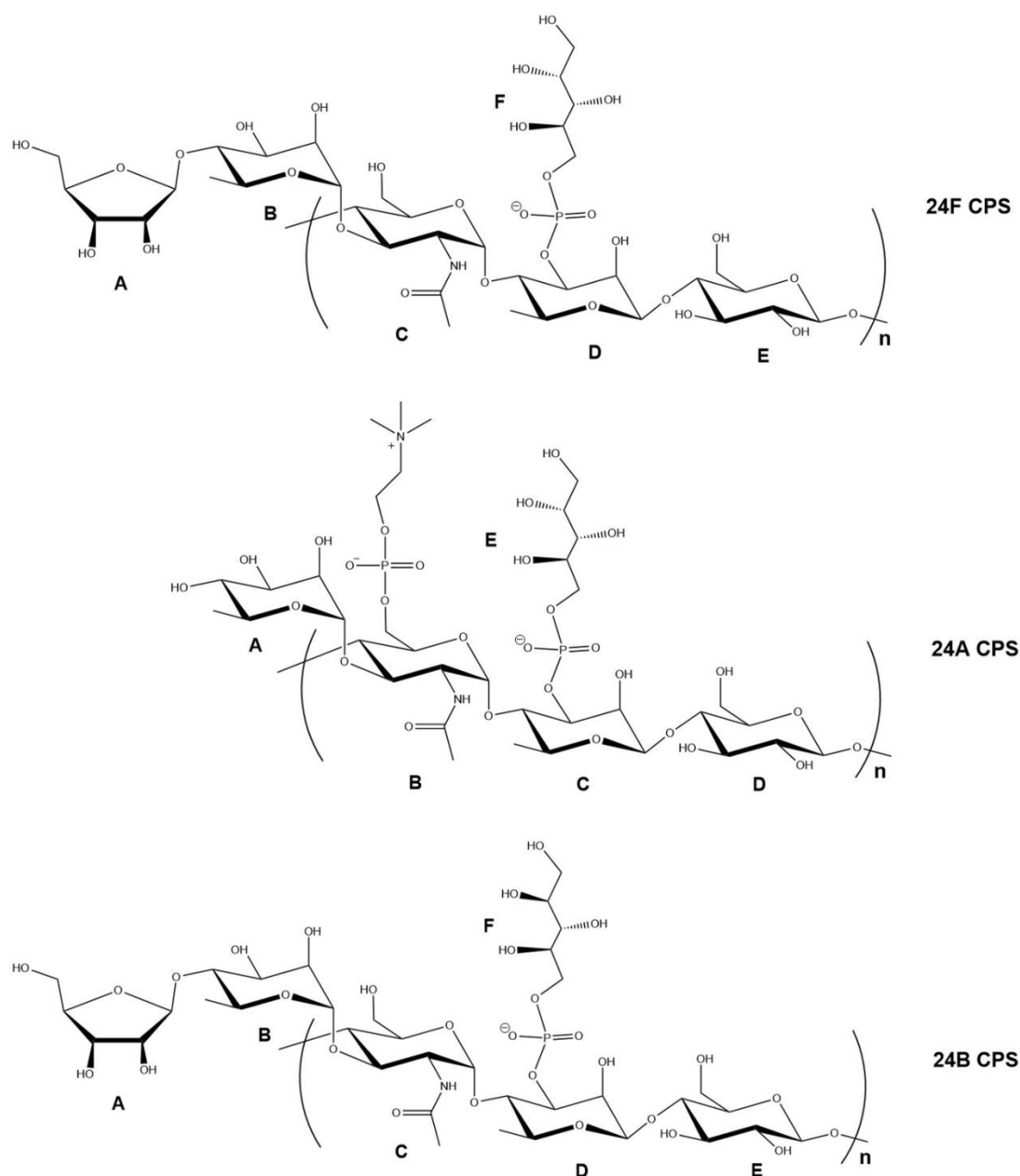


Figure 4.1 Repeat unit structures of serotype 24F, 24A and 24B CPS

As the CWPS peaks were dominant in 1D ^1H spectra of serogroup 24 CPS (Figure A4), 2D NMR experiments were more useful to identify each monosaccharides component. According to HSQC spectrum, 24B CPS also consisted of five different anomeric signals, which were almost identical to 24F CPS, labelled as **A**, **B**, **C**, **D** and **E** in order of descending ^1H chemical shift (Figure 4.2). Utilizing DQF-COSY, TOCSY, HSQC-TOCSY and H2BC, spin systems were assigned which corresponded to each sugar or polyalcohol residue. Combined with monosaccharide composition analysis, residues **A**, **B**, **C**, **D** and **E** were identified as Ribf, Rhap, GlcpNAc, Rhap and Glcp, respectively. In order to determine the

anomeric configurations of pyranoside sugar residues, $^1J_{H1, C1}$ of each sugar residue was measured by CLIP-HSQC. Residue **B**, with the $^1J_{H1, C1}$ of about 170Hz, was identified as α -Rhap, while **C**, **D** and **E**, with $^1J_{H1, C1}$ between 160-163Hz, were β -configuration. As the residue **A** was a furanoside sugar, the $^1J_{H1, C1}$ for α - and β -configuration are both around 168-171 Hz^{89,157}. Residue **A**, with a chemical shift of about 108 ppm, was identified as a β -Ribf based on chemical shift data in the literature^{90,158–161}.

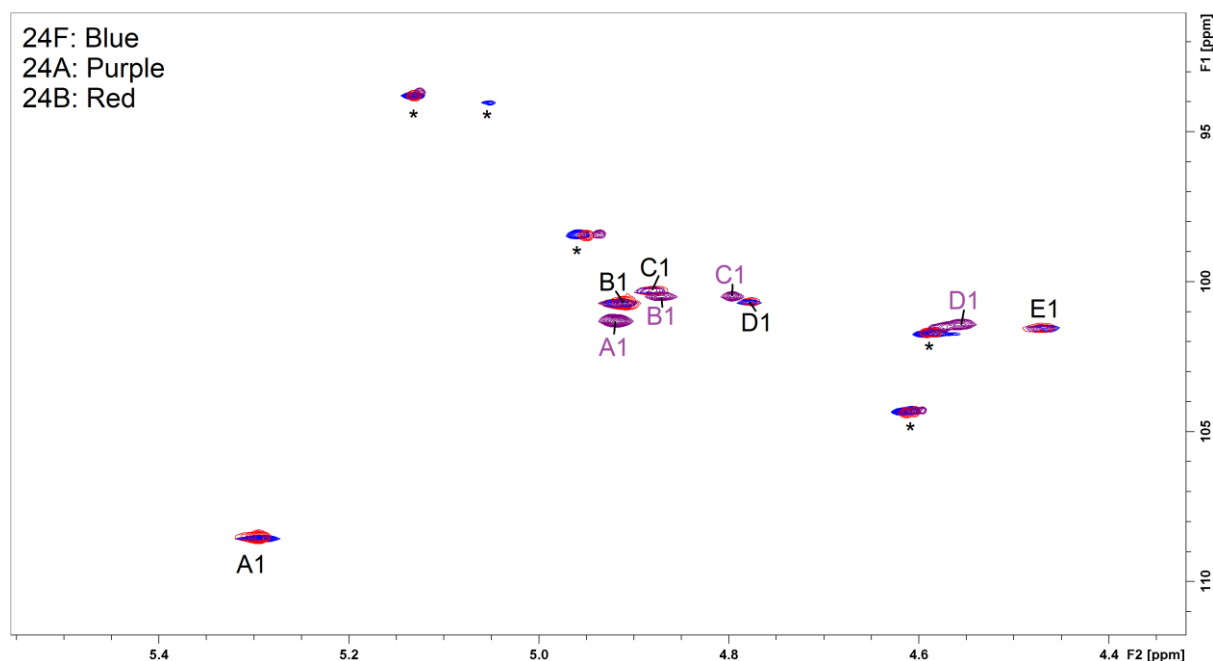


Figure 4.2 Expansion and overlay of HSQC spectra of 24F (blue), 24A (purple) and 24B (red) CPS. The black A1-E1 labels correspond to the anomeric position of 24F and 24B in table 4.3. The purple A1-D1 labels correspond to the anomeric position of 24A in table 4.3. The CWPS impurity peaks are marked with *.

Besides, monosaccharide compositions showed that there were near two molar Rib (ribose or ribitol) in the repeat unit of 24B CPS, whereas there were one Ara (arabinose or arabinitol) and one Rib (ribose or ribitol) in 24F CPS. The ribitol (Rib-ol) of 24B CPS consisted of another spin system labeled as **F**, which differentiated 24B CPS from 24F CPS in HSQC spectra (Figure 4.3).

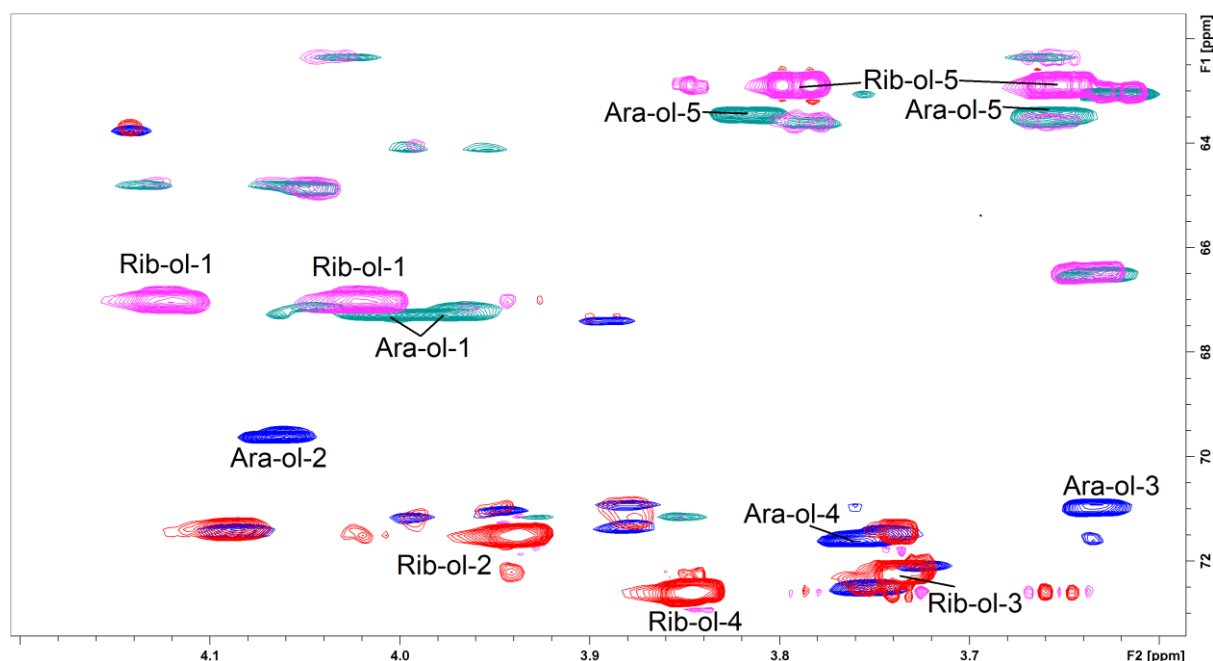


Figure 4.3 Expansion and overlay of HSQC spectra of 24F (blue: + and light blue: -) and 24B (red: + and pink: -) CPS

Furthermore, the linkages between residues were determined by HMBC and NOESY (Figure 4.4). The anomeric position of **A** had HMBC correlations with **B4**, while the anomeric position of **B** had correlations with **C3**. HMBC cross peaks between **C1** and **D4**, between **D1** and **E4** and between **E1** and **C4** were found. Therefore, the order of residues in 24B CPS repeat unit was: 4)-[**A**-(1-4)-**B**-(1-3)]-**C**-(1-4)-**D**-(1-4)-**E**-(1-). In addition, according to ^{31}P and ^1H - ^{31}P HMBC spectra, there was one phosphate group, correlated with the 1-position of the Rib-ol **F** and the 3-position of the β -Rhap **C** (Figure 4.5). Based on the above analysis, the repeat unit of serotype 24B CPS was determined to be: -4)[β -Ribf(1-4)- α -Rhap-(1-3)]- β -Glc_pNAc-(1-4)- β -Rhap-[3-P-1-Rib-ol] (1-4)- β -Glc_p-(1- (Figure 4.1). Full chemical shift assignment of serotype 24B is shown in Table 4.3.

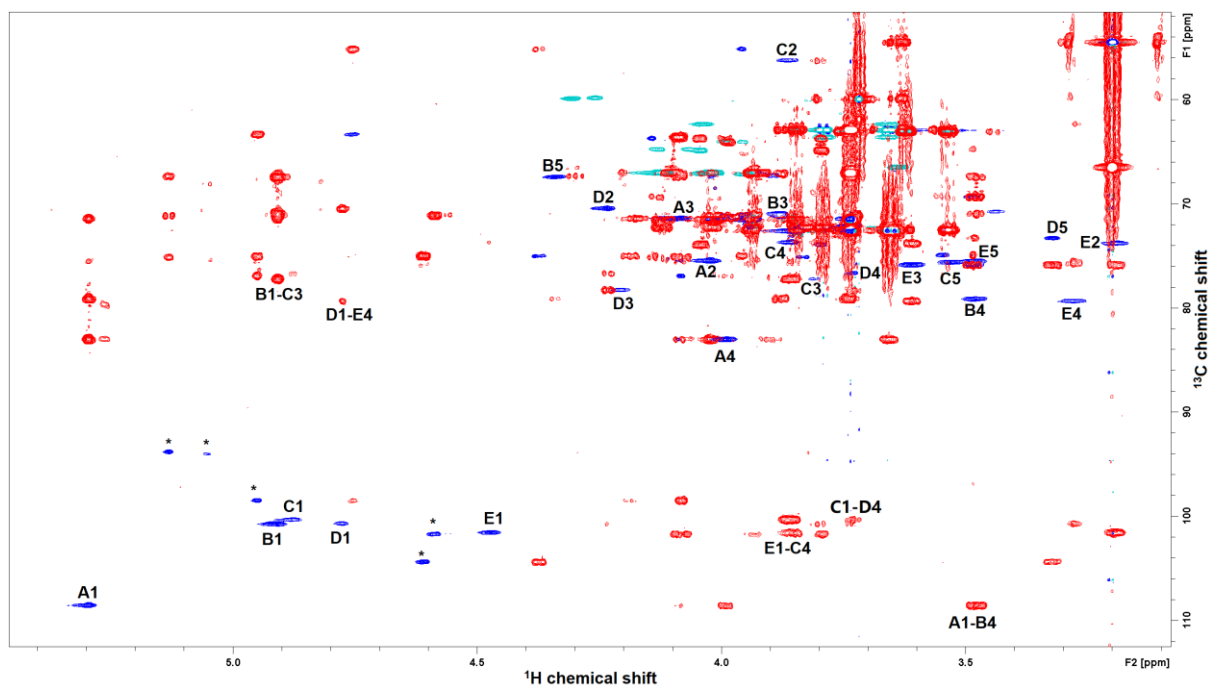


Figure 4.4 Expansion and overlay of ^1H - ^{13}C HSQC (blue: +, light blue: -) and HMBC (red) spectra of 24B CPS.

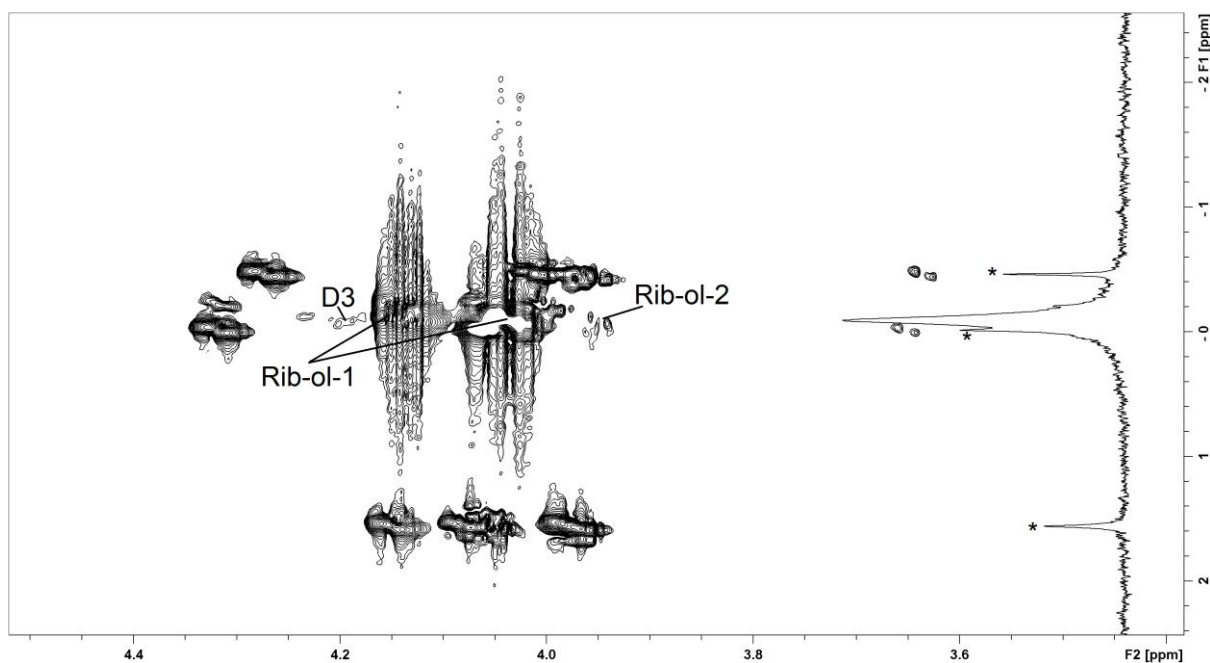


Figure 4.5 Expansion and overlay of ^1H - ^{31}P HMBC and ^{31}P spectra of 24B CPS. The P-Chol peaks from CWPS impurity marked with *.

Table 4.3 Chemical shift assignment of serotype 24F, 24A, 24B and 24X CPS

| Serotype | Residue | 1 | 2 | 3 | 4 | 5 | 6 | NAc | PO ₄ |
|----------|-----------------|-----------------|-------|-------|-------|-------------|-----------------|-------|-----------------|
| 24F | β -Ribf | 5.292 | 4.029 | 4.086 | 3.993 | 3.778/3.658 | | | |
| | A | 108.56 | 75.49 | 71.39 | 83.11 | 63.65 | | | |
| | α -Rhap | 4.916 | 3.744 | 3.879 | 3.483 | 4.328 | 1.236 | | |
| | B | 100.70 | 71.47 | 70.92 | 79.24 | 67.40 | 17.56 | | |
| | β -GlcNAc | 4.885 | 3.859 | 3.800 | 3.857 | 3.531 | ND | 2.063 | |
| | C | 100.35 | 56.26 | 77.25 | 73.85 | 75.61 | ND | 22.84 | |
| | β -Rhap | 4.779 | 4.231 | 4.207 | 3.727 | 3.434 | 1.326 | | |
| | D | 100.70 | 70.45 | 78.31 | 76.66 | 70.80 | 17.68 | | |
| | β -Glc | 4.469 | 3.192 | 3.613 | 3.291 | 3.478 | 3.661/ 4.022 | | |
| | E | 101.52 | 73.85 | 75.96 | 79.24 | 75.49 | 62.36 | | |
| 24A | Ara-ol | 4.009/ 3.978 | 4.064 | 3.632 | 3.760 | 3.819/3.659 | | | -0.276 |
| | F | 67.28 | 69.63 | 70.92 | 71.62 | 63.41 | | | |
| | α -Rhap | 4.919 | 3.750 | 3.720 | 3.366 | 4.347 | 1.204 | | |
| | A | 101.31 | 71.00 | 70.47 | 72.23 | 68.94 | 17.10 | | |
| | β -GlcNAc | 4.873 | 3.873 | 3.787 | 3.914 | 3.657 | ND | 2.075 | -0.471 |
| | B | 100.49 | 55.93 | 78.45 | 73.28 | 74.10 | ND | 22.87 | |
| | β -Rhap | 4.797 | 4.215 | 4.203 | 3.697 | 3.406 | 1.292 | | |
| | C | 100.49 | 70.12 | 78.09 | 76.92 | 70.70 | 17.57 | | |
| | β -Glc | 4.557 | 3.188 | 3.645 | 3.501 | 3.459 | 3.792/ 3.920 | | |
| | D | 101.43 | 73.99 | 75.63 | 77.39 | 75.16 | 61.55 | | |
| 24B | Ara-ol | 3.976 | 4.061 | 3.612 | 3.745 | 3.804/3.819 | | | -0.305 |
| | E | 67.07 | 69.41 | 70.70 | 71.29 | 63.31 | | | |
| | β -Ribf | 5.295 | 4.030 | 4.087 | 3.997 | 3.779/3.663 | | | |
| | A | 108.56 | 75.49 | 71.39 | 83.11 | 63.65 | | | |
| | α -Rhap | 4.915 | 3.744 | 3.885 | 3.482 | 4.327 | 1.237 | | |
| | B | 100.70 | 71.49 | 70.92 | 79.24 | 67.40 | 17.56 | | |
| | β -GlcNAc | 4.886 | 3.865 | 3.799 | 3.858 | 3.533 | ND | 2.062 | |
| | C | 100.35 | 56.26 | 77.25 | 73.85 | 75.61 | ND | 22.84 | |
| | β -Rhap | 4.783 | 4.239 | 4.215 | 3.728 | 3.438 | 1.327 | | |
| | D | 100.70 | 70.45 | 78.31 | 76.66 | 70.80 | 17.68 | | |
| 24X | β -Glc | 4.470 | 3.189 | 3.616 | 3.292 | 3.480 | 4.035/ 3.663 | | |
| | E | 101.53 | 73.85 | 75.96 | 79.24 | 75.49 | 62.36 | | |
| | Rib-ol | 4.127/ 4.025 | 3.944 | 3.737 | 3.852 | 3.803/3.660 | | | -0.045 |
| | F | 67.05 | 71.62 | 72.33 | 72.68 | 62.94 | | | |
| | | | | | | | | | |

The structures of 24A CPS were determined in the same way as 24B CPS. 24A CPS was similar to 24F CPS, except lacking β -Ribf in the side chain. Therefore, there were only four anomeric signals from 24A CPS in HSQC spectrum (Figure 4.6). The monosaccharide composition and methylation analysis of 24A CPS also confirmed the missing of β -Ribf residue. In addition, ^{31}P and ^1H - ^{31}P HMBC spectra showed that 24A CPS had two phosphate groups. One of them correlated with 1-position of Ara-ol and 3-position of β -Rhap, another corresponded to P-Cho (Figure 4.7). The P-Cho substitution site was determined by excluding the possibility of sitting on other positions, since the phosphate will affect the chemical shift of the substitution position. Therefore, the P-Cho should be linked to the 6-position of β -GlcNAc. Consequently, the repeat unit of serotype 24A CPS was determined to be: $-4)[\alpha\text{-Rhap-(1-3)]}\text{-}\beta\text{-GlcNAc-[6-P-Cho]-(1-4)-}\beta\text{-Rhap-[3-P-1-Ara-ol] (1-4)-}\beta\text{-GlcP-(1-}$ (Figure 4.1).

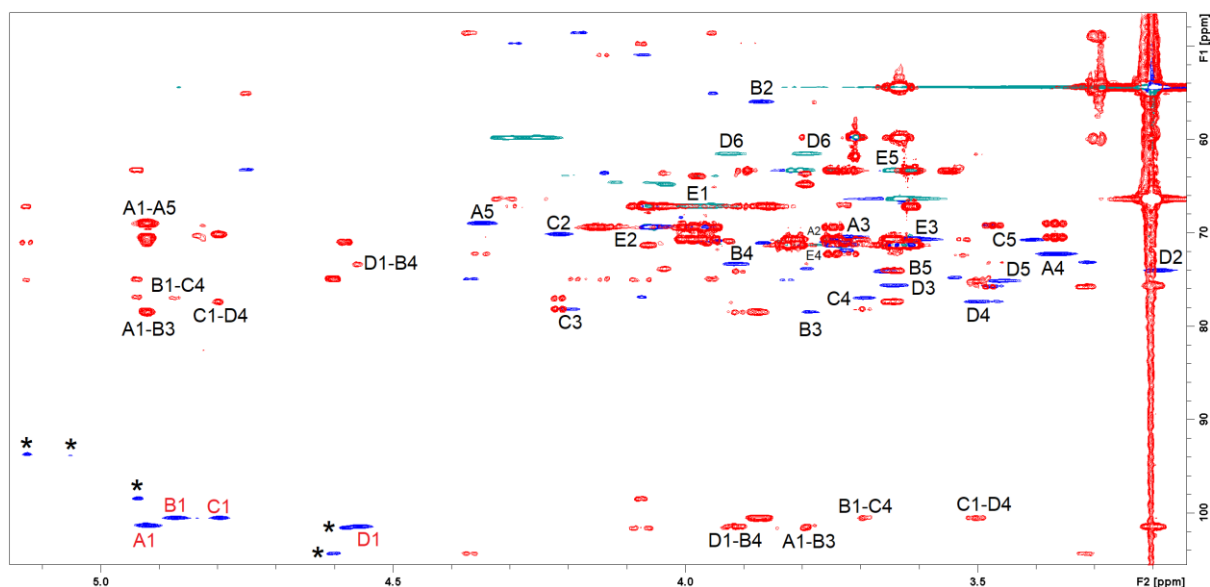


Figure 4.6 Expansion and overlap of ^1H - ^{13}C HSQC (blue: +, light blue: -) and HMBC (red) spectra of 24A CPS. The CWPS impurity peaks marked with *.

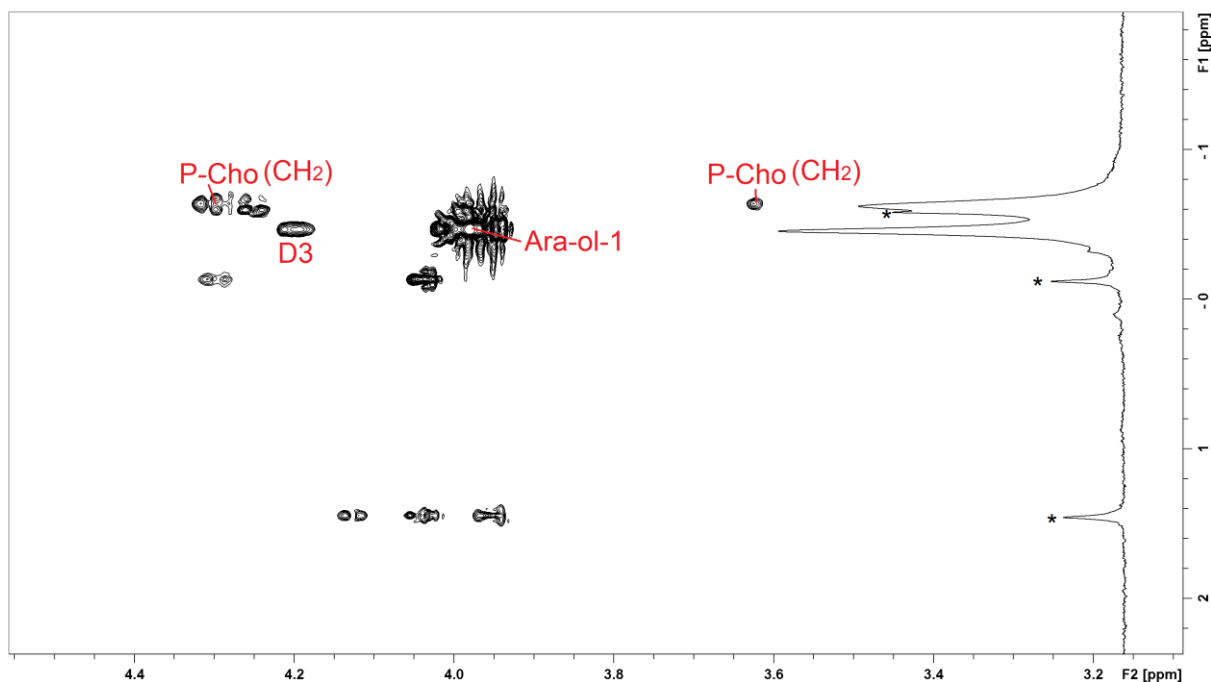


Figure 4.7 Superimposed ^1H - ^{31}P HMBC and ^{31}P spectra of serotype 24A CPS. The CWPS impurity peaks marked with *.

4.1.3 NMR analysis of repeat unit structure of serotype 24X CPS

The HSQC spectrum of 24X CPS showed peaks that were mostly identical to those from 24F and 24B, while it contained peaks from both Rib-ol and Ara-ol (Figure 4.8). Besides, ^{31}P spectrum of 24X CPS, regardless of peaks from CWPS, also showed two phosphors, which showed highly similar chemical shifts to the phosphor of 24A and 24F, respectively (Figure 4.9 (A), Figure 4.10). Thus, the repeat unit structure of 24X CPS had both $\text{PO}_4\text{-1-Ara-ol}$ and $\text{PO}_4\text{-1-Rib-ol}$. The monosaccharide composition analysis also supported it. Therefore, it indicated that the repeat unit 24X CPS should contain both repeat units of 24F and 24B CPS, in a ratio approximately (3.5: 1), by measuring integrals from ^{31}P NMR and HSQC. The repeat unit structure is shown in Figure 4.9 (B). Full assignment of 24X CPS is given in Table 4.4.

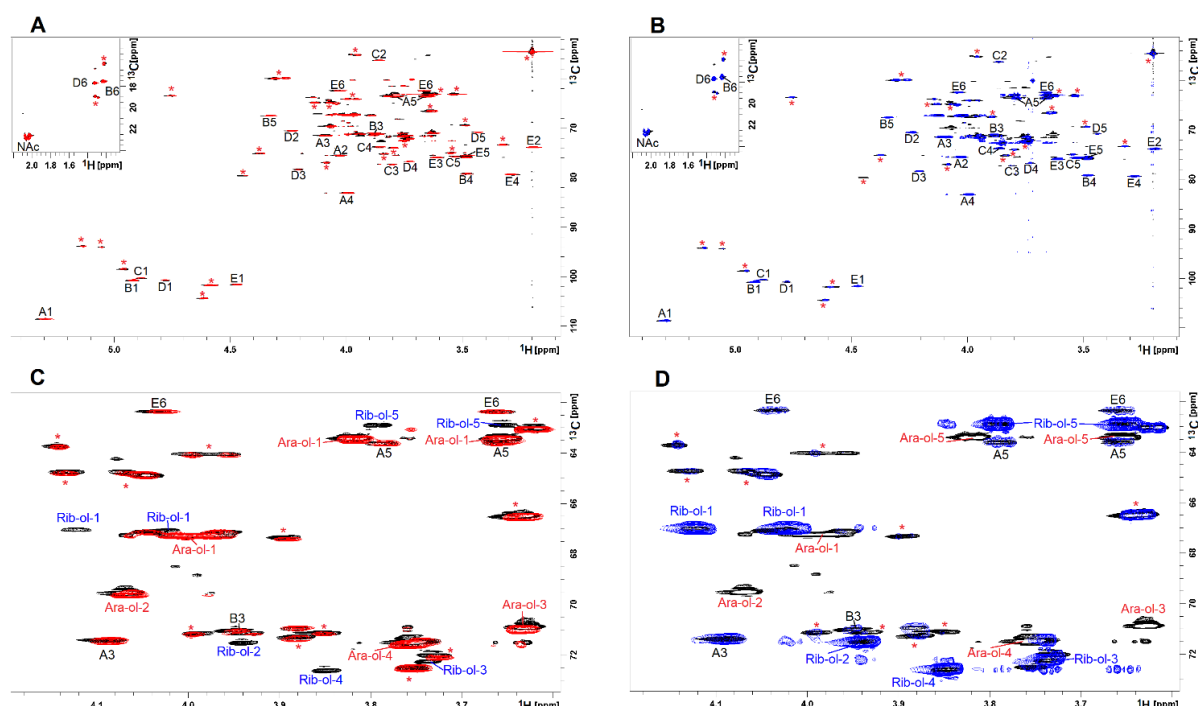


Figure 4.8 (A): Superimposed HSQC spectra of 24X (Black) and 24F (Red); **(B):** Superimposed HSQC spectra of 24X (Black) and 24B (Blue); **(C):** Expansion of **A** showing Ara-ol and Rib-ol peaks region **(D):** Expansion of **B** showing Ara-ol and Rib-ol peaks region. Impurities peaks from CWPS and free sugar alcohols were marked with red *.

Table 4.4 Chemical shift assignment of serotype 24X CPS

| Serotype | Residue | 1 | 2 | 3 | 4 | 5 | 6 | NAc | PO4 |
|----------|-----------------|-------------|-------|-------|-------|-----------------|-----------------|-------|--------|
| 24X | β -Ribf | 5.296 | 4.025 | 4.089 | 3.992 | 3.781/ 3.659 | | | |
| | A | 108.57 | 75.49 | 71.39 | 83.00 | 63.58 | | | |
| | α -Rhap | 4.913 | 3.740 | 3.885 | 3.482 | 4.342 | 1.237 | | |
| | B | 100.77 | 71.45 | 70.92 | 79.13 | 67.40 | 17.50 | | |
| | β -GlcNAc | 4.882 | 3.863 | 3.807 | 3.861 | 3.535 | ND | 2.06 | |
| | C | 100.30 | 56.20 | 77.20 | 73.69 | 75.61 | ND | 22.83 | |
| | β -Rhap | 4.776 | 4.232 | 4.204 | 3.731 | 3.440 | 1.327 | | |
| | D | 100.71 | 70.45 | 78.25 | 76.67 | 70.74 | 17.67 | | |
| | β -Glc | 4.472 | 3.200 | 3.609 | 3.279 | 3.480 | 4.038/ 3.658 | | |
| | E | 101.53 | 73.79 | 75.85 | 79.31 | 75.49 | 62.36 | | |
| | Ara-ol | 4.003/3.975 | 4.067 | 3.631 | 3.757 | 3.819/ 3.655 | | | -0.274 |
| | F1 | 67.28 | 69.51 | 70.86 | 71.51 | 63.42 | | | |
| | Rib-ol | 4.121/4.022 | 3.940 | 3.735 | 3.847 | 3.799/ 3.653 | | | -0.041 |
| | F2 | 67.04 | 71.51 | 72.21 | 72.62 | 62.88 | | | |

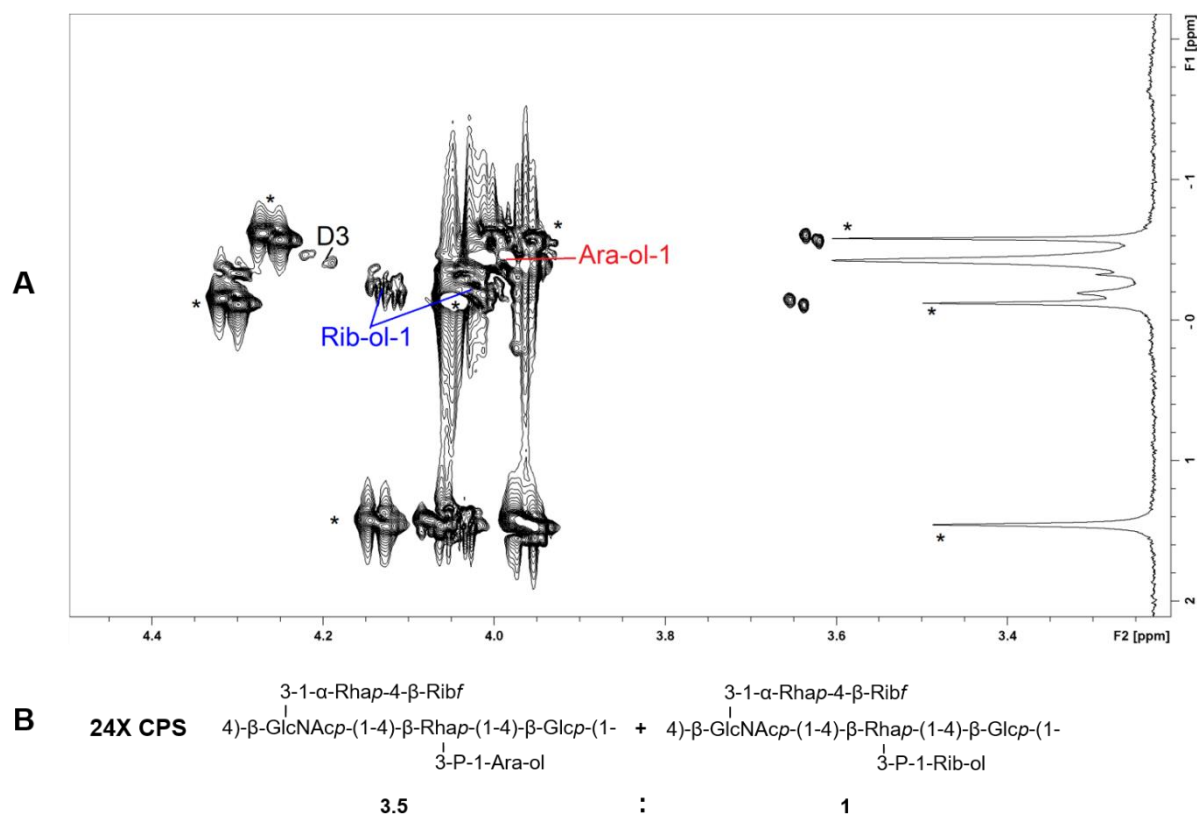


Figure 4.9 (A) Superimposed ^1H - ^{31}P HMBC and ^{31}P spectra of serotype 24X CPS. The CWPS impurity peaks marked with *. **(B)** Repeat unit structures of serotype 24X CPS.

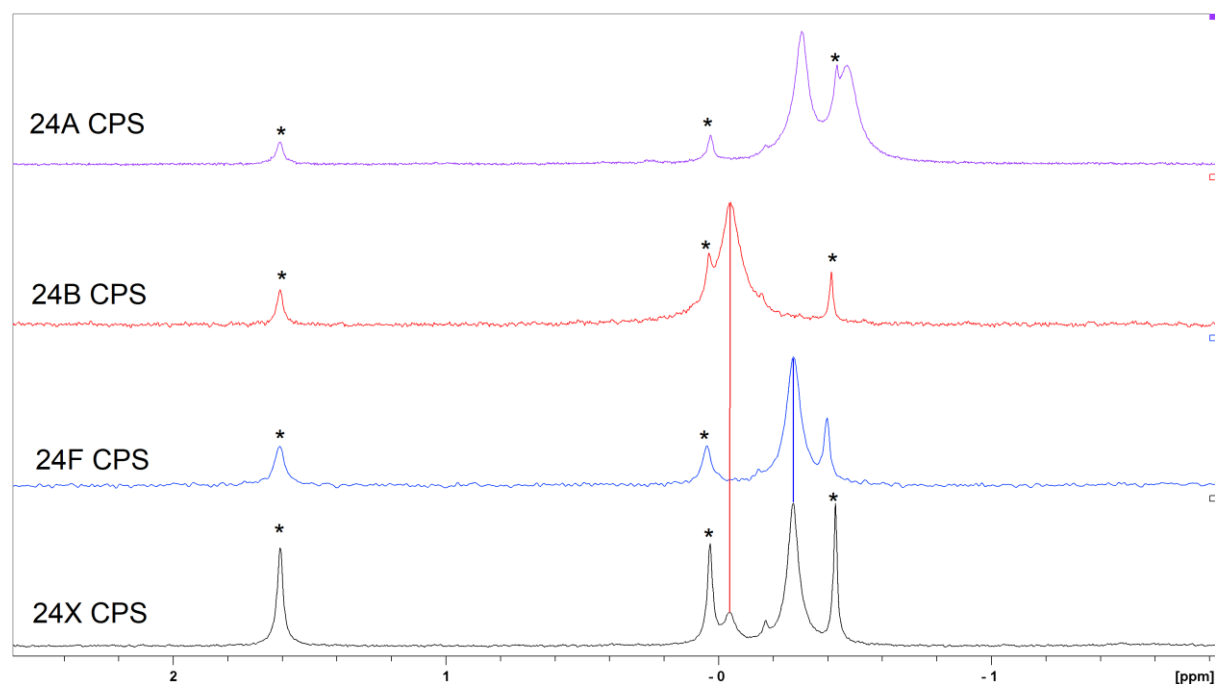


Figure 4.10 Overlapped ^{31}P spectra of 24X (black), 24F (blue), 24B (red) and 24A (purple) CPS. The CWPS impurity peaks marked with *.

4.2 Conclusion

The CPS structures of serogroup 24 were thoroughly elucidated by NMR and general chemical analysis, which provided fundamental knowledge of serogroup 24. The CPS structures of serogroup 24 have the same backbones and the main differences among them are the sidechains. The repeat unit of 24A CPS has the same P-Ara-ol as 24F, but lack of the β -Ribf at the 4-position of α -Rhap. The only difference in the repeat unit of 24B CPS, compared to 24F CPS, is the P-Rib-ol sidechain instead of the P-Ara-ol in 24F CPS.

Serotype 24X CPS contains both repeat unit of 24F and 24B. These findings could explain the reactivity of serotype 24X with both factor antisera 24d and 24e. Serotype 24X capsule has a distinct serology and a repeating unit structure, which qualify it as a new serotype. The genetic basis for its existence along with the previously unexamined related serotypes 24F and 24B and 24A was investigated by collaborators, which is not included in this thesis. According to Danish naming system, 24X will be named as serotype 24C and be included as the 4th member of the serogroup 24.

Chapter 5 Serological interactions study of CPS with diagnostic antisera

The interactions between CPS and diagnostic antisera belong to macromolecule-macromolecule (CPS-antibody) interactions in a biological mixture (serum). Moreover, as CPS have multiple binding sites and as the antibodies are bivalent, CPS and antibodies can form cross-linked complexes, which makes the interaction quite challenging to study at the molecular level. When antibodies bind to CPS, CPS might show a different diffusion behavior compared to CPS alone without adding antiserum. Therefore, at the beginning of this study, diffusion NMR was employed in an attempt to find out if there were any observable differences in diffusion coefficients of CPS upon antibodies binding. The interactions between 16A CPS and rabbit diagnostic antisera were studied in this Chapter for as native 16A CPS gave good NMR spectra. The 16c factor antiserum reacts with 16A CPS and different batches of 16c factor antisera used in this study are summarized in Table 5.1. Interestingly, we observed different diffusion coefficients between 16A CPS with and without OAc when adding 16c-0 antiserum. However, when mixing 16A CPS with 16c-1 and 16c-2 antiserum, no such different diffusion behaviors were observed.

Table 5.1 Information of 16c antiserum used in this study

| Antiserum name | Antiserum production | Production date |
|----------------|--|-----------------|
| 16c-0 | Commercial 16c factor antiserum (made from rabbit that was immunized with serotype 16A pneumococci and removal of antibodies that cross-react with serotype 16F CPS) | 2018-04-16 |
| 16c-1 | Raw 16A antiserum, which was made from rabbit that was immunized with serotype 16A pneumococci | 2020-05-27 |
| 16c-2 | Same as 16c-0 | 2020-05-27 |

One possible reason could be that the different batches of 16c antisera have different antibodies that possess different binding affinities or preferences, such as binding to different epitopes of CPS. When the antibodies bind strongly to the epitope on the CPS, the CPS-antibody bound parts and nearby residues could become too rigid to yield detectable NMR signals. Consequently, such rigid parts of CPS would disappear into the spectral noise. In other words, the effect of CPS NMR line broadening due to antibody binding resulted in a

decrease of integration of CPS signals. This effect implies that the non-bound parts of CPS could be quantified and used to assess the affinity of antiserum^{162,163}. Therefore, NMR titration was used to study the decreased signal integral of CPS in an attempt to probe non-affected CPS during the titration. In addition, AF4-MALS-dRI-UV was applied to characterize the diagnostic antisera and CPS-antibody complex.

5.1 Study of interaction between 16A CPS and diagnostic antisera by DOSY

The 16A CPS contains 70% partial O-Acetylation (OAc) at the 2- β -Gal β residue (Figure 5.1). Before the addition of any antisera, DOSY of 16A CPS showed mainly two distinct diffusion coefficients (D) (Figure 5.1). The peaks from 16A CPS aligned at almost the same diffusion level with a D of about 1.6×10^{-11} . The choline peak from CWPS had a higher D of about 3.1×10^{-11} . Hence, the CWPS diffused faster compared to 16A CPS, which was consistent with smaller Mw and sharper NMR linewidths of the CWPS.

16A CPS repeat unit structure:

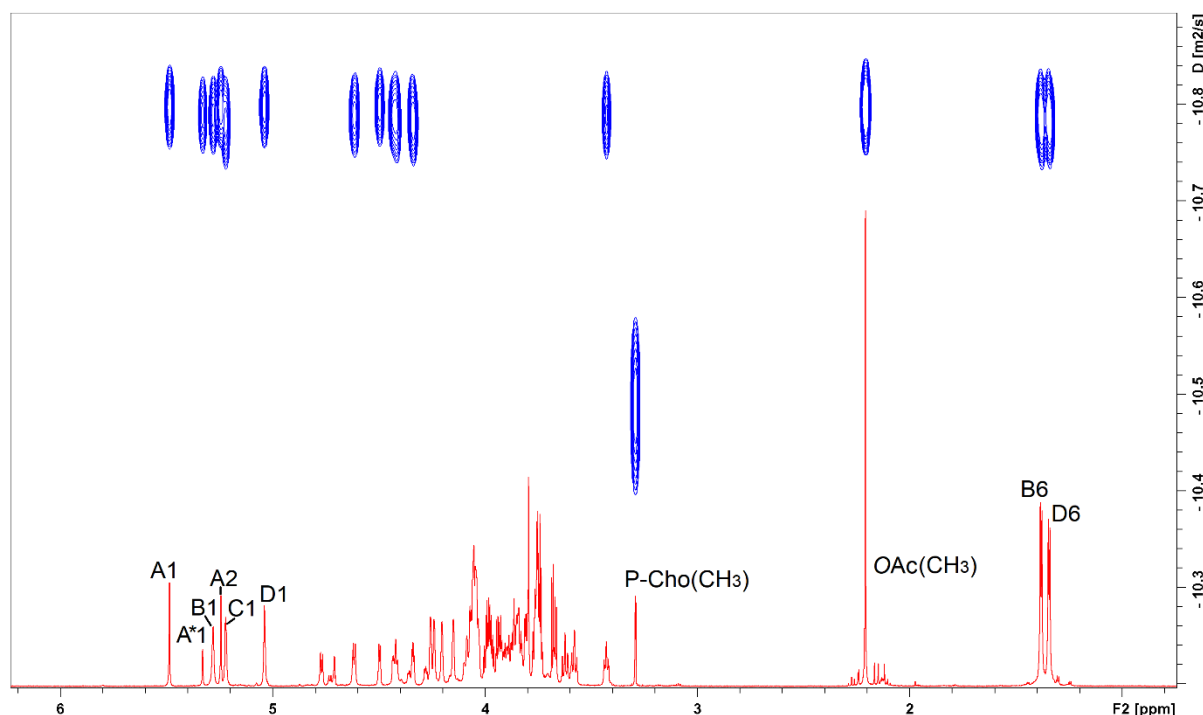
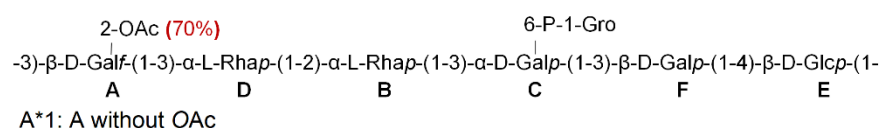


Figure 5.1 Repeat unit structure and DOSY spectrum of 16A CPS with 16c factor antiserum

Interestingly, when adding 16c-0 antiserum (the first batch of 16c factor antiserum, Table 5.1), the β -Gal residue of 16A CPS with and without O-Acetylation showed different diffusion behaviors (Figure 5.2). The peaks from 2-OAc- β -Gal residue showed smaller D of about 9.0×10^{-12} . However, D of the anomeric peak of β -Gal without OAc, A*1, was about 1.5×10^{-11} . The other residues showed D values between A and A*. This observation indicated an observable effect resulting in reduced diffusion especially for acetylated species when antibodies in 16c-0 antiserum bound to 16A CPS.

In addition, the 16b factor antiserum was used as a negative control. In 16b factor antiserum, the 16A CPS specific antibodies were removed from the serum so that 16A CPS do not react with 16b factor antiserum. Not surprisingly, when adding 16b factor antiserum, there were no such distinct diffusion behaviors within each residue of 16A CPS (Figure 5.2). Therefore, the 2-OAc- β -Gal residue of 16A CPS could be important for antibody in 16c factor antiserum binding. The different diffusion behavior suggested that there might be two different polymer chains, one with O-Acetylation modification and another without O-Acetylation. This observation was consistent with the AF4-MALS-dRI measurement of 16A CPS, which showed two parts in dRI after AF4 separation (Chapter 3.1.1). However, further separation of these two CPS polymers is needed to validate the assumption.

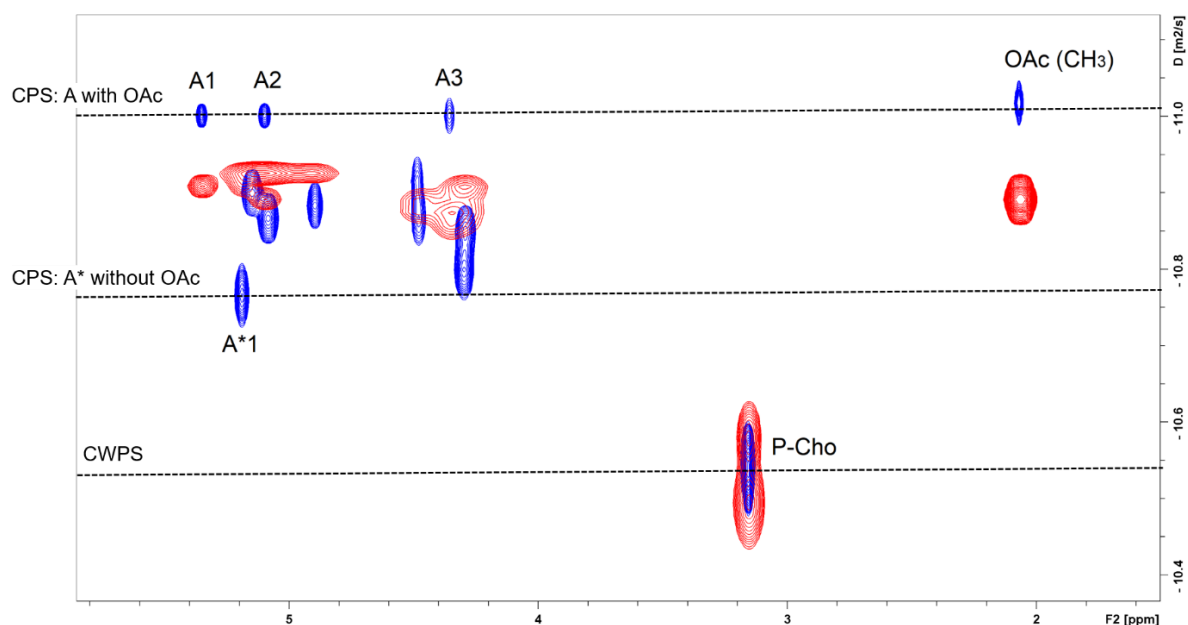


Figure 5.2 DOSY of 16A CPS with 16c antiserum (blue) and 16A CPS with 16b factor antiserum (red)

As mentioned in chapter 2, DOSY is sensitive to temperature, solvents, convolution and other factors, which causes some limitations when applying it to study complex systems. Thus, the temperature, pH, salt and the instrumental parameters all need to be well controlled, and the same applies to data processing and fitting to get the D values. The components in rabbit diagnostic antisera could vary a lot among different batches. The 16c-0 antiserum used in DOSY study above was from the first batch that was provided by SSI Diagnostica A/S. The interactions of 16A CPS with the other two 16c factor antisera (16c-1 and 16c-2) did not show the different diffusion patterns that were observed in 16A CPS with 16c-0 antiserum. Upon addition of 16c-1 and 16c-2 antiserum, slightly larger diffusion coefficients for signals from 16A CPS were measured than 16A CPS alone. This observation indicated that the non-binding CPS were either smaller in size than binding CPS, or that the non-binding became had higher diffusion rates due to a dilution effect. Due to the complexity of antiserum, the changes in local dynamics of CPS could be different when adding different batches of antiserum. Further studies would be needed to understand the batch differences and the usefulness of diffusion NMR in understanding antibody affinity and concomitant structural changes.

5.2 NMR Titration of 16A CPS with 16c antiserum

5.2.1 T_1 and T_2 measurements of 16A CPS

Before NMR titration, it is necessary to know the T_1 of CPS in order to set long enough interscan relaxation delays (d_1), the relaxation delay for quantitative NMR acquisition. The T_1 and T_2 of non-overlapping peaks are shown in Table 5.2. T_1 of 16A CPS ranges from 1.07 to 3.04 s. Therefore, the delay between each scan should be set at least $5 \times T_1$, approximately 15 s.

Table 5.2 T_1 and T_2 of anomeric and methyl protons of 16A CPS

| Position | A1 | A'1 | B1 | A2 | C1 | D1 | Ac(CH ₃) | Rha (6-CH ₃) |
|------------|------|------|------|------|------|------|----------------------|-----------------------------|
| T_1 (s) | 2.59 | 2.53 | 2.25 | 2.76 | 2.48 | 2.44 | 3.04 | 1.07/1.14 |
| T_2 (ms) | 128 | 148 | 60.3 | 107 | 66.8 | 85 | 263 | 66 |

5.2.2 NMR Titration of 16A CPS with 16c antiserum

The NMR titration of 16A CPS with 16c antisera was based on the size limitation of solution NMR. One of the assumptions for NMR titration is that, when CPS specific antibody bound to CPS, the antibody binding parts of CPS and the nearby repeat units, but not the whole CPS chain, will be affected and tumbled too slow to give NMR signals. The affinity of CPS specific antibody had previously been studied through thermodynamic methods and they found that an average of 12 monosaccharides in CPS bind per monoclonal antibody ¹⁴⁰. As mentioned before, when antibodies in the antisera bind with CPS, the slower tumbling rate of the parts of CPS affected by binding of antibodies would vanish from the spectra and result in a decrease of signal integral. The integral of the anomeric peak from 3- β -Gal β -2-OAc (A1) was chosen for calculating the unaffected 16A CPS, as this signal did not overlap with peaks from components of antisera (Figure 5.3, Figure A5).

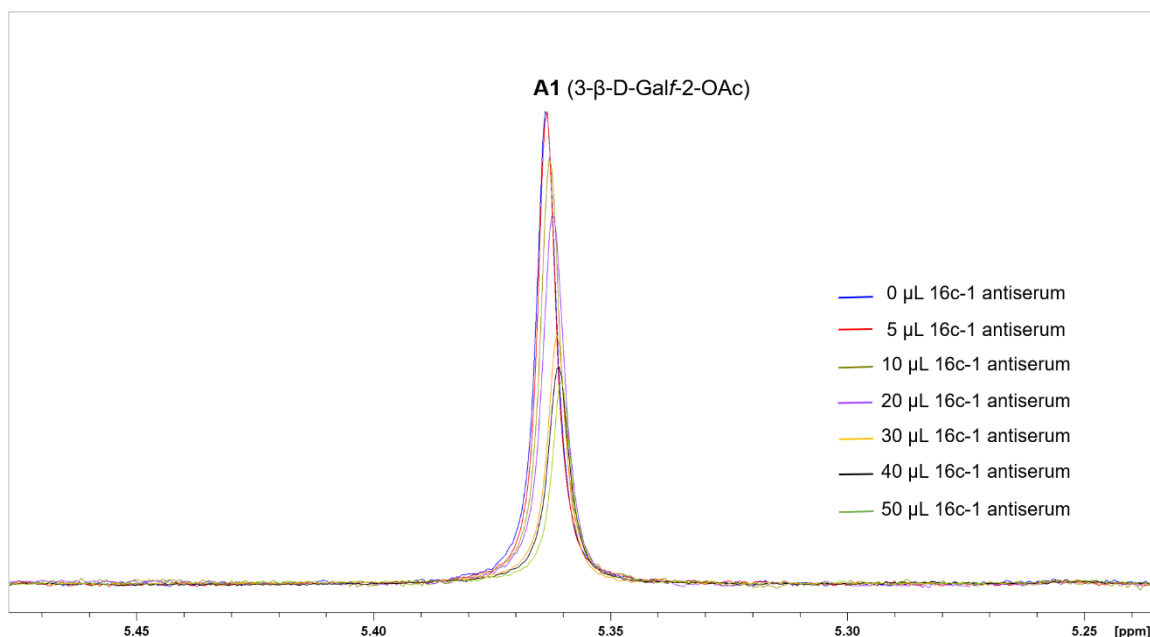


Figure 5.3 Expansion of ^1H NMR titration of 16A CPS with 16c-1 antiserum

NMR titration curves of 16A CPS with 16c antisera are shown in Figure 5.4. The concentration of 16A CPS was based on the integral of A1. When titrating 16A CPS with 16c-1 antiserum, the integral of A1 was slowly decreasing at the beginning. However, there was a sharply decreasing signal upon 16c-1 antiserum addition at volume up to between 20 μL and 30 μL . After adding 30 μL of 16c-1 antiserum, the decrease in signal became less pronounced and was comparable to the dilution rate. For CPS titration with 16c-2 antiserum, a slight decrease at the beginning was observed. Similar to 16c-1 antiserum, a more dramatic decline of 16A CPS signal when adding 16c-1 antiserum between 20 μL and 30 μL was observed. However, when adding more than 30 μL 16c-2 antiserum, the integral of A1 peak increased back.

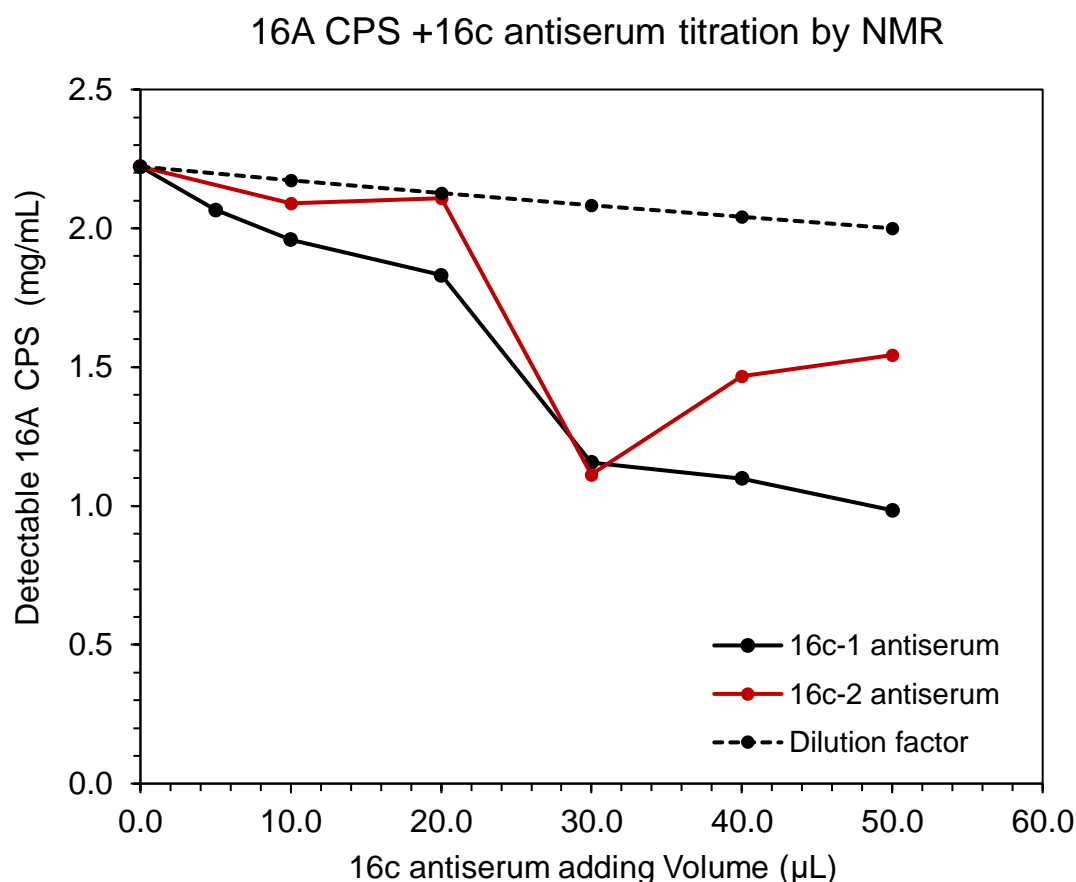


Figure 5.4 NMR titration of 16A CPS with 16c-1 and 16c-2 antiserum

One possible explanation of the NMR titration of 16A CPS and 16c antiserum can be based on the antigen-antibody reactions, similar to agglutination reactions (Figure 5.5) ¹⁶⁴. At the beginning, the 16A CPS antigen was present in excess. When adding antibodies in CPS solution, each antibody can bind with one CPS epitope. The cross-link was not formed with substoichiometric amounts. However, when the added amount of antibody reached equivalence, cross-linked antibody-antigen complexes appeared to have formed. The cross-linked antibody-antigen complexes had a slower tumbling rate in the solution or even became visible (precipitation). Due to the size limitations of solution NMR, the NMR signals of 16A CPS that were formed in the cross-linked complex will all disappear from the spectra. Therefore, the sharply descending region would be the equivalence zone, where cross-linked CPS-antibody complexes were formed. After the equivalence zone, the excess antibody could break the cross-linked CPS-antibody complex and form soluble immune complexes. Therefore, the observable 16A CPS signal was increased when adding more antiserum beyond 30 μL. The reason that free 16A CPS in 16c-1 antiserum titration did not increase could be that the antibodies could not break the cross-linked CPS-Antibody

complex. It could also be that the different antibodies in the two antisera showed different affinities or a variation of other components in the antiserum could affect the reaction of CPS and antibodies.

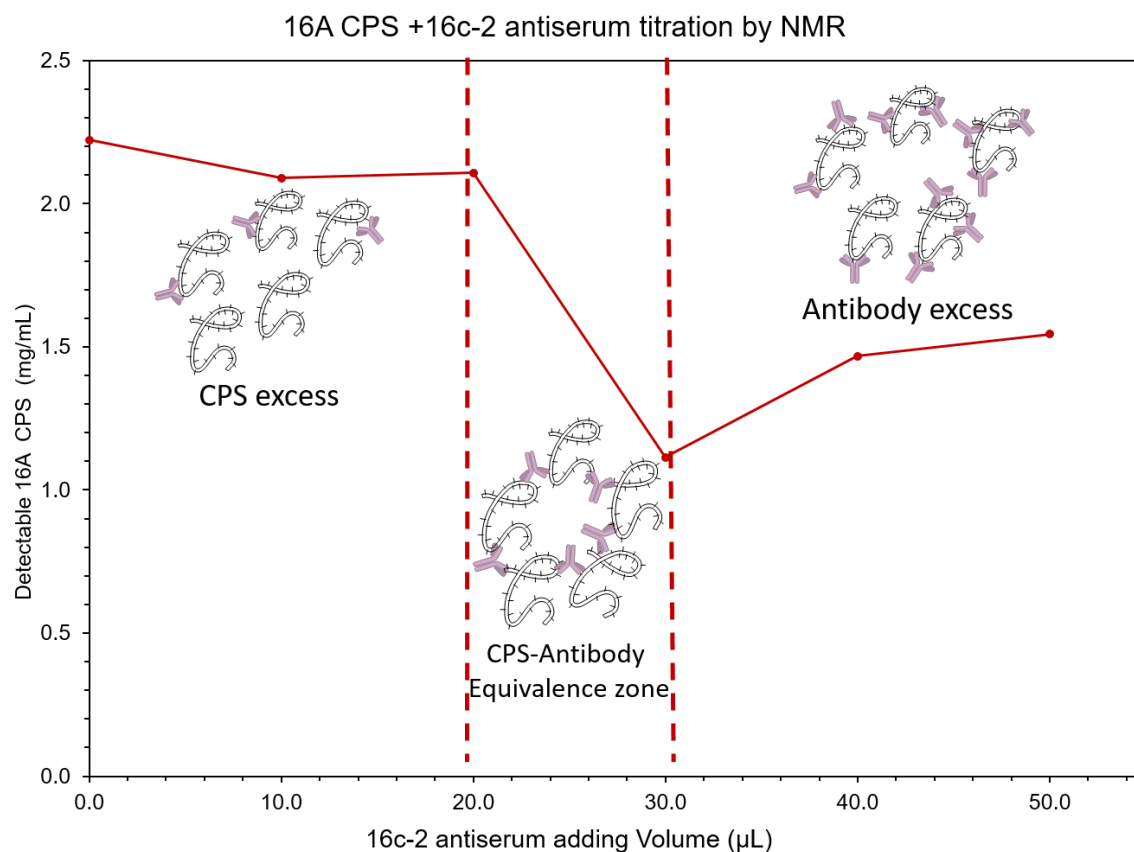


Figure 5.5 Graphic elucidation of NMR titration of 16A CPS with 16c-2 antiserum

5.3 Characterization of 16c factor antiserum by AF4-MALS-dRI-UV

16c-1 antiserum is the raw antiserum, which was obtained from rabbits that were immunized with serotype 16A pneumococci (Table 5.1). 16c-2 antiserum was obtained in a similar way as 16c-1 antiserum, but with an additional absorptive step to remove the cross-reactive antibodies with serotype 16F pneumococci. The 16c-1 antiserum was diluted 10-fold with phosphate-buffered saline (PBS) prior to AF4 separation. According to literature, the two most abundant components in serum are the serum albumin and immunoglobulins^{110,165}. The 16c-1 antiserum sample showed mainly three peaks in AF4-MALS-dRI-UV (Figure 5.6) (Table 5.3). The biggest peak at around 11 min with the Mw of about 64 kDa corresponded to the rabbit serum albumin. The peak with a maximum at 16 min corresponds to immunoglobulin G (IgG, the main antibody) with the Mw of about 158 kDa. The unknown small peak at 19 min with the Mw of about 775 kDa could be other larger protein, like alpha-2-macroglobulin (720 kDa, ~3 mg/mL) or IgM (950 kDa, ~1 mg/mL).

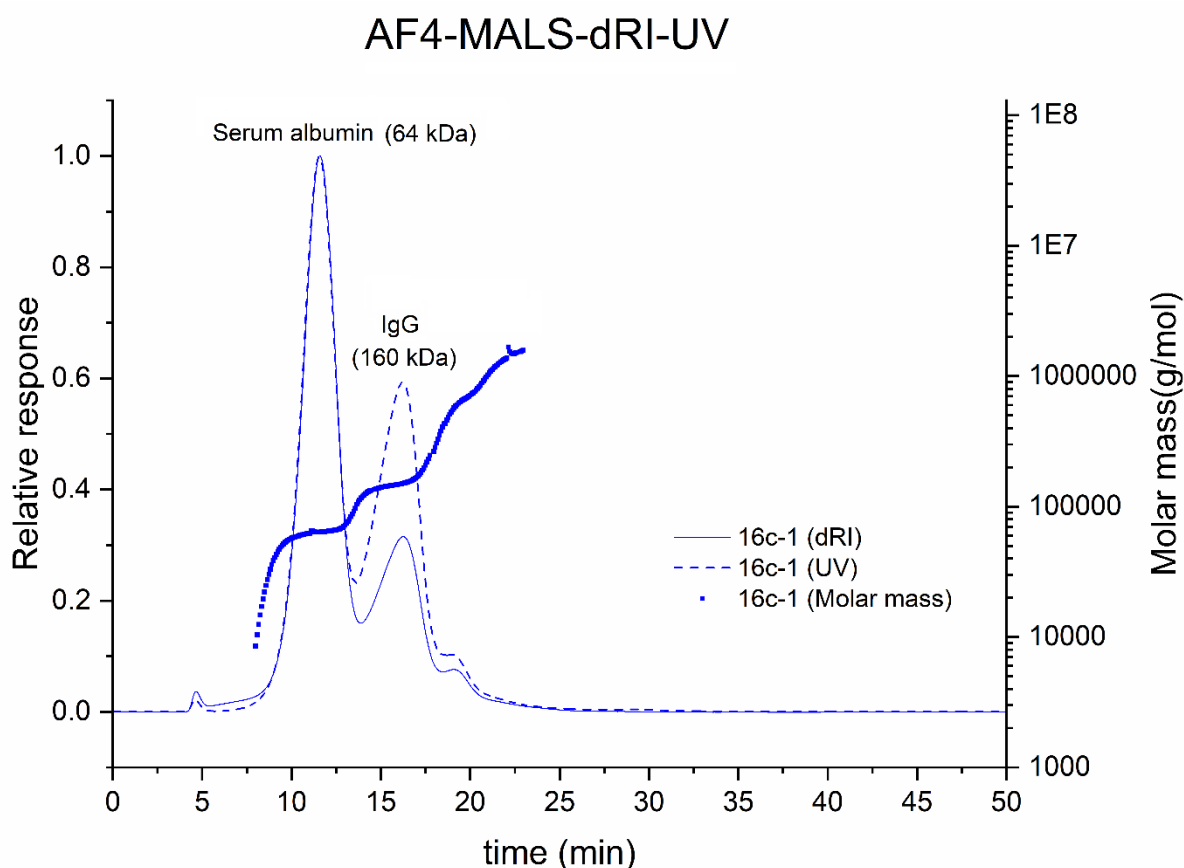


Figure 5.6 The components of 16c-1 antiserum and their size distributions measured by AF4-MALS-dRI-UV

Table 5.3 The components and Mw of 16c antiserum measured by AF4-MALS-dRI-UV

| Component | 16c-1 antiserum | | 16c-2 antiserum | |
|----------------|-----------------|------------------------|-----------------|------------------------|
| | Mw (kDa) | Polydispersity (Mw/Mn) | Mw (kDa) | Polydispersity (Mw/Mn) |
| Serum albumin | 63.1 ± 0.2 | 1.13 ± 0.04 | 64.0 ± 0.3 | 1.02 ± 0.008 |
| IgG | 157.3 ± 0.5 | 1.05 ± 0.005 | 157.8 ± 0.6 | 1.05 ± 0.006 |
| Other Proteins | 774.9 ± 2.9 | 1.13 ± 0.006 | 822.0 ± 3.2 | 1.21 ± 0.007 |

5.3.1 Evaluation of CPS binding antibodies in 16c factor antiserum

In order to analyze serum proteins quantitatively, bovine serum albumin (BSA) was used as standards for quantification of albumin in 16c antiserum. The AF4 analysis of BSA standard solutions was using identical AF4 methods. A calibration curve of BSA was made for calculating the rabbit serum albumin in 16c antiserum (Figure 5.7). The concentration of serum albumin, IgG and other proteins are summarized in Table 5.4. The concentrations of rabbit serum albumin in 16c-1 and 16c-2 antiserum were both around 41 mg/mL, which were within the range of mammal serum albumin level ^{110,165}. The concentration of IgG was calculated with a UV extinction coefficient at $\epsilon_{280\text{ nm}} = 1.380 \text{ [cm}^{-1}(\text{mg/mL})^{-1}]$ compared to BSA $\epsilon_{280\text{ nm}} = 0.667 \text{ [cm}^{-1}(\text{mg/mL})^{-1}]$. The concentration of the last other proteins was calculated assuming a UV extinction coefficient at $\lambda 280\text{ nm}$ (1.0), which was not reliable and just for reference.

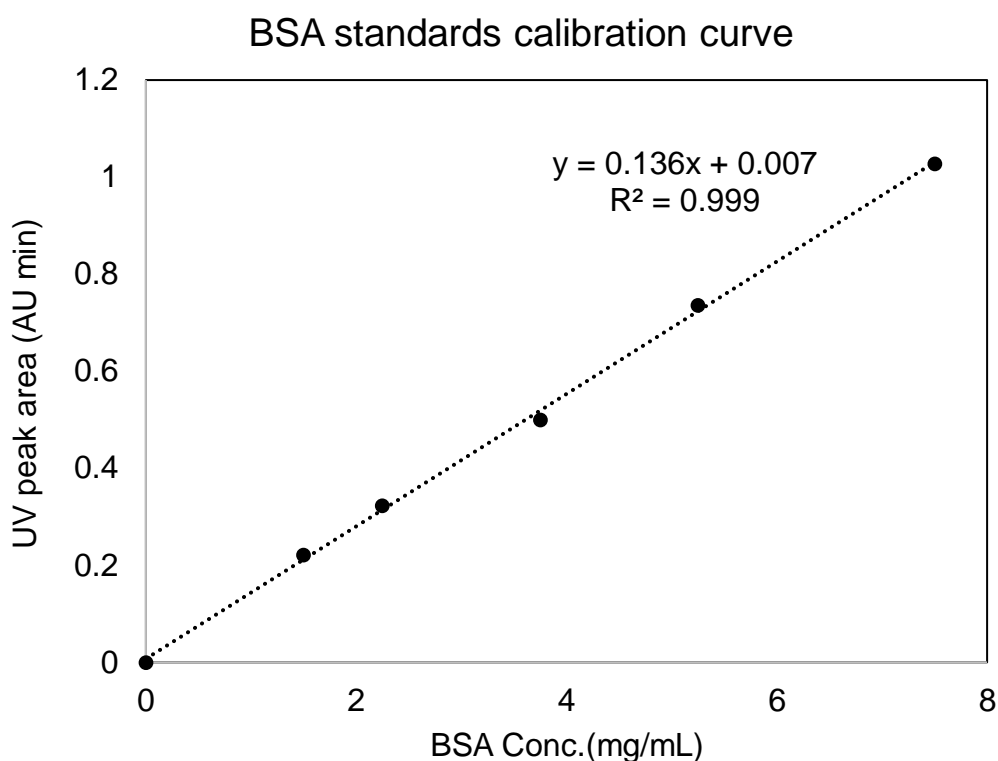


Figure 5.7 The BSA standards calibration curve by AF4-UV

The 16A CPS solution showed almost no UV response at λ 280 nm (Figure 5.8). When mixing 16A CPS with 16c-1 antiserum, the overall UV profile changed dramatically (Figure 5.8). The concentrations of 16A CPS and 16c-1 antiserum used in AF4-MALS-dRI-UV were the same as the last NMR titration point, in which were 1mg 16A CPS with 50 μ L 16c antiserum in 500 μ L PBS buffer. In the mixture of 16A CPS and 16c antiserum (16A + 16c-1), the peak at around 12 min corresponds to serum albumin with a size of about 64 kDa. The IgG peak at around 16 min largely reduced in the 16A + 16c-1 mixture compared to the solution containing only 16c-1 antiserum, which indicated that IgG was interacting with 16A CPS. The 16A CPS peak shifted to a longer retention time in 16A + 16c-1 antiserum mixture and dRI peak area of 16A CPS increased. Notably, 16A CPS peak region showed strong UV absorption compared to 16A CPS alone with almost no UV signal at 280 nm. The fact that the UV peak area of IgG decreased significantly, while the 16A CPS peak region gained UV absorption indicated that IgG was thus binding onto 16A CPS.

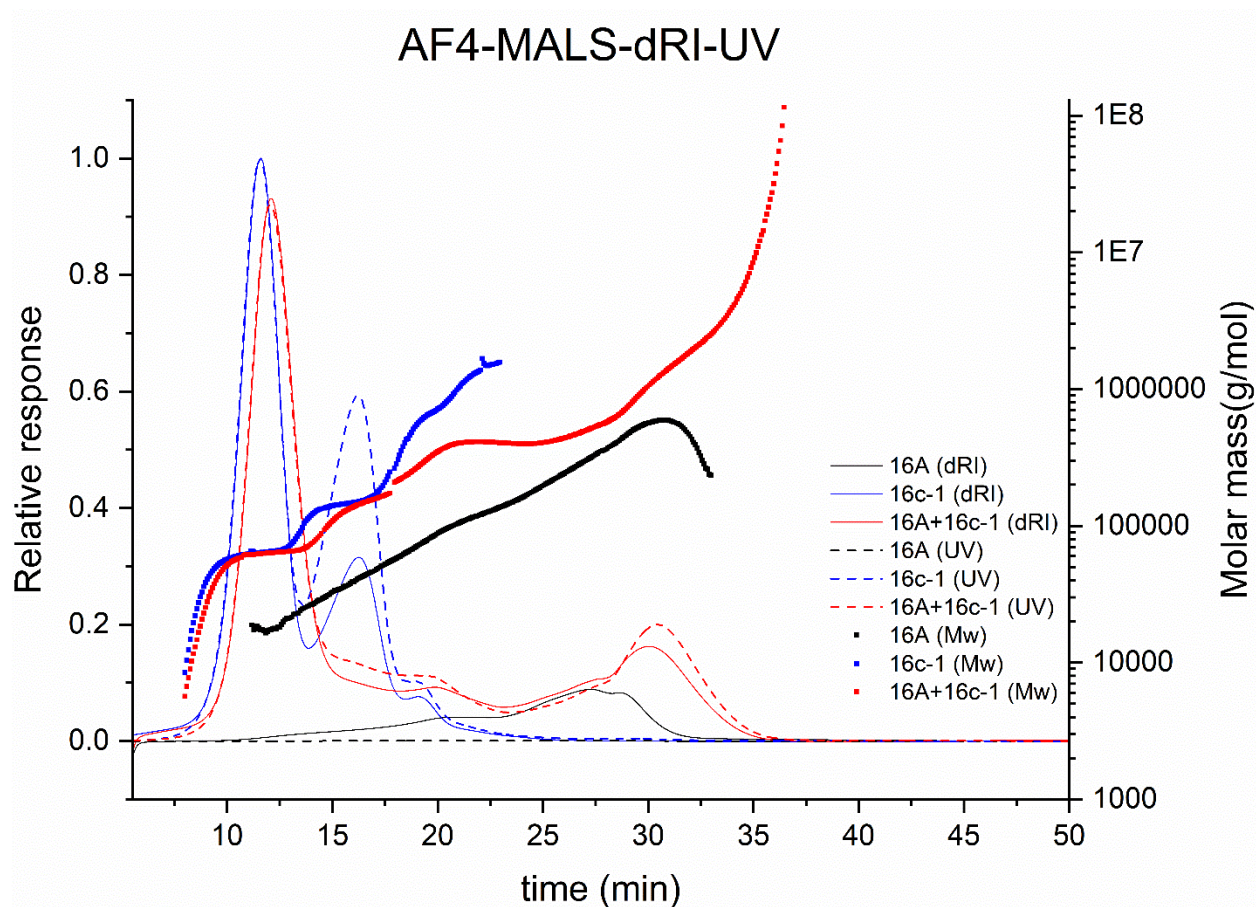


Fig. 5.8 AF4-MALS-dRI-UV results of 16A CPS (black), 16c-1 antiserum (blue) and mixture of 16A CPS with 16c-1 antiserum (red).

As IgG are bivalent and CPS are multivalent, the CPS-IgG can form the cross-linked complex. Therefore it was not surprisingly, that the molar mass of the 16A + 16c mixture showed a high polydispersity at the CPS-IgG complex peak. In 16A + 16c-1, 16A CPS-IgG complex had Mw of about 1141 kDa with a polydispersity of about 1.6. In 16A + 16c-2, 16A CPS-IgG complex had a Mw of about 950 kDa with a polydispersity of about 1.5. As the Mw of 16A CPS and IgG was about 273 kDa and 157 kDa, respectively, the CPS-IgG complex could contain multiple IgG or CPS polymer chains, which was consistent with NMR titration. According to NMR titration, when adding 50 μ L antiserum, the antibodies excess the equivalence zone. Therefore, one CPS chain could bind with multiple antibodies and cross-linked CPS-IgG complex could also exist. The concentration of IgG that bound to 16A CPS (CPS binding IgG) can be calculated from the area of UV in the CPS-IgG complex region. The percentage of CPS binding IgG was calculated by dividing the concentration of CPS binding IgG to the concentration of IgG in the antiserum alone solution. The results of concentrations of serum albumin, IgG and CPS binding IgG are shown in Table 5.4. The

concentrations of IgG and CPS binding IgG in 16c-2 antiserum were slightly higher than IgG in 16c-1 antiserum. However, the percentage of CPS binding IgG relative to total IgG in 16c-1 antiserum was slightly higher than the corresponding fraction in 16c-2 antiserum.

Therefore, AF4-MALS-dRI-UV shows the potential for evaluation of antisera activity and quantitative analysis of CPS binding antibodies in antiserum.

Table 5.4 Concentration of rabbit serum albumin, IgG, other proteins and CPS binding IgG in diagnostic antisera calculated by UV response

| Sample | Serum albumin (mg/mL) | IgG (mg/mL) | CPS binding IgG (mg/mL) | CPS binding IgG in total IgG (%) |
|---------------------------|-----------------------|-------------|-------------------------|----------------------------------|
| 16c-1 antiserum | 41.6 | 14.3 | - | 67.1 |
| 16A CPS + 16c-1 antiserum | 42.7 | 3.0 | 9.6 | |
| 16c-2 antiserum | 41.1 | 15.7 | - | 65.6 |
| 16A CPS + 16c-2 antiserum | 41.7 | 3.1 | 10.3 | |

5.4 Summary

Application of diffusion NMR for studying the interactions of 16A CPS with 16c-0 antiserum showed that the β -Galf residue of 16A CPS with and without O-Acetylation displayed different diffusion behaviors, which indicated that the O-Acetyled β -Galf residue was important for antibodies in 16c-0 antirum binding. In NMR titration of 16A CPS with 16c antiserum, the reduced intensity or integral of CPS peaks implied the binding of antibodies on to CPS. Through quantifying the unaffected CPS by ^1H NMR during the titration, the equivalence zone in CPS-antibody reactions can be found. Moreover, AF4-MALS-dRI-UV was applied to study antisera and CPS-antibody complex. The molecular weight distribution of CPS-antibody complex was consistent with last NMR titration point where the antibodies were excess.

The preliminary studies in this chapter were performed to provide alternative methods for studying the interactions between CPS and rabbit diagnostic antisera. However, as the antiserum is a complicated mixture and binding modes can be complex. The study of CPS-antibody interactions in a complicated matrix by diffusion NMR in the current setup has too many variables, which become challenging to predict the effects of these factors in the evaluation of the results. Moreover, since the small molecular components in antiserum vary from batches to batches and they could overlap with CPS peaks in NMR, it would be better to use monoclonal antibodies to study their interaction with CPS in the future.

Conclusion and future prospective

The capsule of *S. pneumoniae* plays vital roles in both environmental survival and pathogenesis. The structural characterization of seven unknown pneumococcal CPS filled the gap of chemical structures within pneumococcal serogroups 16, 24 and 28. Moreover, a new serotype 24X was discovered and the structural determination of its CPS confirmed the novelty. As CPS structures are determined, biosynthesis and serological activity can be better understood, which has essential implications in antibacterial capsule-based vaccine design and aids molecular epidemiological tracking.

The correlation of CPS structures with biosynthesis provided a better understanding of related CPS biosynthetic enzymes. After the structures of 16F and 16A were determined, the functions of GTs WcxN and WcxT were predicted for the first time. The putative function of WcxN was further supported by CPS structures of serotype 28F and 28A. Moreover, WcxN showed a broad acceptor specificity in serogroup 28, which can transfer α -l-Rhap to 3-position of α -d-Glcp or α -d-GlcNAcp. Besides, the different sugar donor specificity of GT WciU resulted in the CPS structural difference in serotype 28F and 28A. The 3D structural modeling of WciU and comparison with published GT structures in the database revealed the critical residue (Trp in the Glc specific WciU, Cys in the GlcNAc one) at position 346 that could be responsible for the donor specificity of WciU.

Furthermore, the cross-reactions of serogroups 16 and 28 within these two serogroups and other related serogroups were studied by Quellung reaction and Latex agglutination. Serological reactivity test of diagnostic antisera with partially degraded CPS revealed the importance of OAc and phosphate substitutions in CPS antigenic activities. In addition, the interactions between CPS and diagnostic antisera were further studied by diffusion NMR, NMR titration and AF4-MALS-dRI-UV. Application of diffusion NMR shows the potential to study the interactions of CPS with antibodies at the molecular level, which could be used for identifying important functional epitopes and for understanding the structural basis of immune recognition of CPS. The ^1H NMR titration proved useful to study unaffected CPS in the CPS-antiserum mixture and to find the equivalence zone in CPS-antibody reaction. In addition, AF4-MALS-dRI-UV could be a useful tool for evaluation of antisera activities, which could be applied for quality control of diagnostic antisera and analysis of specific antibodies in the antisera.

In the future, other unknown pneumococcal CPS structures still need to be determined to prepare for the serotype replacement and changes in IPD and prevalent serotypes. Further study to correlate the CPS structure with biosynthesis are important to help understand the relationship between genotype and capsule type, which requires a deeper understanding of the genetic basis of related biosynthetic enzymes. To continue the study of WciU, the genetic experiments can be conducted to verify the predicted key residue³⁴⁶ that results in the donor sugar specificity.

The relationship between CPS structures and their immunological activities should be further investigated to design better CPS based pneumococcal vaccines. As NMR titration focuses on the change in CPS concentration and AF4-MALS-dRI-UV could provide information of antibody and CPS-antibody complex, the combination of NMR titration with a titration measured in AF4-MALS-dRI-UV would be useful for understanding the interactions between CPS and diagnostic antisera. Alternatively, starting with the study of interactions of a monoclonal antibody with CPS can simplify the system, which might enable us to map binding epitopes by applying other NMR techniques, such as saturation transfer difference (STD) NMR^{166,167}.

References

1. Croucher, N. J. *et al.* Rapid pneumococcal evolution in response to clinical interventions. *Science* (80-.). **331**, 430–434 (2011).
2. Lázár, V. *et al.* Bacterial evolution of antibiotic hypersensitivity. *Mol. Syst. Biol.* **9**, 700 (2013).
3. Gillings, M. R. & Stokes, H. W. Are humans increasing bacterial evolvability? *Trends Ecol. Evol.* **27**, 346–352 (2012).
4. Henriques-Normark, B. & Normark, S. Bacterial vaccines and antibiotic resistance. *Ups. J. Med. Sci.* **119**, 205–208 (2014).
5. Fridman, O., Goldberg, A., Ronin, I., Shores, N. & Balaban, N. Q. Optimization of lag time underlies antibiotic tolerance in evolved bacterial populations. *Nature* **513**, 418–421 (2014).
6. Chaguza, C., Cornick, J. E. & Everett, D. B. Mechanisms and impact of genetic recombination in the evolution of *Streptococcus pneumoniae*. *Comput. Struct. Biotechnol. J.* **13**, 241–247 (2015).
7. Stacevičienė, I. *et al.* Antibiotic resistance of *Streptococcus pneumoniae*, isolated from nasopharynx of preschool children with acute respiratory tract infection in Lithuania. *BMC Infect. Dis.* **16**, 216 (2016).
8. Manenzhe, R. I. *et al.* Nasopharyngeal Carriage of Antimicrobial-Resistant Pneumococci in an Intensively Sampled South African Birth Cohort. *Front. Microbiol.* **10**, 1–10 (2019).
9. Harboe, Z. B. *et al.* Pneumococcal Serotypes and Mortality following Invasive Pneumococcal Disease: A Population-Based Cohort Study. *PLoS Med.* **6**, e1000081 (2009).
10. Henriques-Normark, B. & Tuomanen, E. I. The Pneumococcus: Epidemiology, Microbiology, and Pathogenesis. *Cold Spring Harb. Perspect. Med.* **3**, a010215–a010215 (2013).
11. Wahl, B. *et al.* Burden of *Streptococcus pneumoniae* and *Haemophilus influenzae* type b disease in children in the era of conjugate vaccines: global, regional, and national estimates for 2000–15. *Lancet Glob. Heal.* **6**, e744–e757 (2018).
12. Grabenstein, J. D. & Musey, L. K. Differences in serious clinical outcomes of infection caused by specific pneumococcal serotypes among adults. *Vaccine* **32**, 2399–2405 (2014).
13. O'Brien, K. L. *et al.* Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. *Lancet* **374**, 893–902 (2009).
14. Skovsted, I. C. *et al.* 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**, 6490–6500

- (2007).
15. Geno, K. A. *et al.* Pneumococcal capsules and their types: Past, present, and future. *Clin. Microbiol. Rev.* **28**, 871–899 (2015).
 16. Brooks, L. R. K. & Mias, G. I. Streptococcus pneumoniae's Virulence and Host Immunity: Aging, Diagnostics, and Prevention. *Front. Immunol.* **9**, 1366 (2018).
 17. Alonsodevelasco, E., Verheul, A. F. M., Verhoef, J. & Snippe, H. Streptococcus pneumoniae: Virulence factors, pathogenesis, and vaccines. *Microbiol. Rev.* **59**, 591–603 (1995).
 18. Li, Y., Weinberger, D. M., Thompson, C. M., Trzciński, K. & Lipsitch, M. Surface charge of Streptococcus pneumoniae predicts serotype distribution. *Infect. Immun.* **81**, 4519–4524 (2013).
 19. Nelson, A. L. *et al.* Capsule Enhances Pneumococcal Colonization by Limiting Mucus-Mediated Clearance. *Infect. Immun.* **75**, 83–90 (2007).
 20. Larson, T. R. & Yother, J. Streptococcus pneumoniae capsular polysaccharide is linked to peptidoglycan via a direct glycosidic bond to β -D- N -acetylglucosamine . *Proceedings of the National Academy of Sciences* **114**, 5695–5700 (2017).
 21. Hammerschmidt, S. & Rohde, M. Electron Microscopy to Study the Fine Structure of the Pneumococcal Cell. in *Methods in Molecular Biology* **1968**, 13–33 (2019).
 22. Kamerling, J. P. Pneumococcal Polysaccharides: A Chemical View. in *Streptococcus pneumoniae: Molecular Biology & Mechanisms of Disease* (2000).
 23. Kauffmann, F., Lund, E. & Eddy, B. E. Proposal for a change in the nomenclature of Diplococcus pneumoniae and a comparison of the Danish and American type designations. *Int Bull Bacteriol Nomencl. Taxon.* 1960; 10 31–40. **10**, 31–40 (1960).
 24. Austrian, R. 3. The Quellung Reaction: A Neglected Microbiologic Technique. in *Life with the Pneumococcus* (University of Pennsylvania Press, 1985). doi:10.9783/9781512800135-005
 25. Satzke, C. *et al.* Standard method for detecting upper respiratory carriage of Streptococcus pneumoniae: Updated recommendations from the World Health Organization Pneumococcal Carriage Working Group. *Vaccine* **32**, 165–179 (2013).
 26. Jauneikaite, E. *et al.* Current methods for capsular typing of Streptococcus pneumoniae. *J. Microbiol. Methods* **113**, 41–49 (2015).
 27. Ganaie, F. *et al.* A New Pneumococcal Capsule Type, 10D, is the 100th Serotype and Has a Large cps Fragment from an Oral Streptococcus. *MBio* **11**, 1–15 (2020).
 28. Chewapreecha, C. *et al.* Frequent recombination of pneumococcal capsule highlights future risks of emergence of novel serotypes. *Mol. Biol. Evol.* 1–30 (2017). doi:http://dx.doi.org/10.1101/098335.
 29. Mostowy, R. J. *et al.* Pneumococcal Capsule Synthesis Locus cps as Evolutionary Hotspot with Potential to Generate Novel Serotypes by Recombination. *Mol. Biol. Evol.* **34**, 2537–2554 (2017).
 30. Kjeldsen, C. *et al.* Discovery and description of a new serogroup 7 Streptococcus

- pneumoniae serotype, 7D, and structural analysis of 7C and 7D. *Carbohydr. Res.* **463**, 24–31 (2018).
31. Oliver, M. B., Van Der Linden, M. P. G., Küntzel, S. A., Saad, J. S. & Nahm, M. H. Discovery of streptococcus pneumoniae serotype 6 variants with glycosyltransferases synthesizing two differing repeating units. *J. Biol. Chem.* **288**, 25976–25985 (2013).
 32. Manna, S. *et al.* A novel genetic variant of Streptococcus pneumoniae serotype 11A discovered in Fiji. *Clin. Microbiol. Infect.* **24**, 428.e1-428.e7 (2018).
 33. van Tonder, A. J. *et al.* Putatively novel serotypes and the potential for reduced vaccine effectiveness: capsular locus diversity revealed among 5405 pneumococcal genomes. *Microb. Genomics* **2**, 000090 (2016).
 34. Burton, R. L., Geno, K. A., Saad, J. S. & Nahm, M. H. Pneumococcus with the “6E”. **54**, 967–971 (2016).
 35. Sleeman, K. L. *et al.* Capsular serotype-specific attack rates and duration of carriage of Streptococcus pneumoniae in a population of children. *J. Infect. Dis.* **194**, 682–688 (2006).
 36. Rodrigo, C. & Lim, W. S. The Relevance of Pneumococcal Serotypes. *Curr. Infect. Dis. Rep.* **16**, 403 (2014).
 37. Konradsen, H. B. Validation of serotyping of Streptococcus pneumoniae in Europe. *Vaccine* **23**, 1368–1373 (2005).
 38. Habib, M., Porter, B. D. & Satzke, C. Capsular serotyping of Streptococcus pneumoniae using the quellung reaction. *J. Vis. Exp.* 1–4 (2014). doi:10.3791/51208
 39. Lankinen, K. S. *et al.* Type-specific enzyme immunoassay for detection of pneumococcal capsular polysaccharide antigens in nasopharyngeal specimens. *J. Microbiol. Methods* **56**, 193–199 (2004).
 40. Schaffner, A., Michel-Harder, C. & Yeginsoy, S. Detection of Capsular Polysaccharide in Serum for the Diagnosis of Pneumococcal Pneumonia: Clinical and Experimental Evaluation. *J. Infect. Dis.* **163**, 1094–1102 (1991).
 41. Leeming, J. P., Cartwright, K., Morris, R., Martin, S. A. & Smith, M. D. Diagnosis of Invasive Pneumococcal Infection by Serotype-Specific Urinary Antigen Detection. *J. Clin. Microbiol.* **43**, 4972–4976 (2005).
 42. Sheppard, C. L., Harrison, T. G., Smith, M. D. & George, R. C. Development of a sensitive, multiplexed immunoassay using xMAP beads for detection of serotype-specific Streptococcus pneumoniae antigen in urine samples. *J. Med. Microbiol.* **60**, 49–55 (2011).
 43. Rayner, R. E., Savill, J., Hafner, L. M. & Huygens, F. Genotyping Streptococcus pneumoniae. *Future Microbiol.* **10**, 653–664 (2015).
 44. Kapatai, G. *et al.* Whole genome sequencing of Streptococcus pneumoniae: Development, evaluation and verification of targets for serogroup and serotype prediction using an automated pipeline. *PeerJ* **2016**, (2016).
 45. Mostowy, R. J. & Holt, K. E. Diversity-Generating Machines: Genetics of Bacterial

- Sugar-Coating. *Trends Microbiol.* **26**, 1008–1021 (2018).
46. Wright, A., Morgan, W., Colebrook, L. & Dodgson, R. W. Observations on prophylactic inoculation against pneumococcus infections, and on the results which have been achieved by it. *Lancet* **10**, 87–95 (1914).
 47. Dochez, A. R. & Avery, O. T. The elaboration of specific soluble substance by pneumococcus during growth. *J. Exp. Med.* (1917). doi:10.1084/jem.26.4.477
 48. Grabenstein, J. D. & Klugman, K. P. A century of pneumococcal vaccination research in humans. *Clin. Microbiol. Infect.* **18**, 15–24 (2012).
 49. Robbins, J. B. *et al.* Considerations for formulating the second-generation pneumococcal capsular polysaccharide vaccine with emphasis on the cross-reactive types within groups. *J. Infect. Dis.* **148**, 1136–1159 (1983).
 50. Stein, K. E. Thymus-Independent and Thymus-Dependent Responses to Polysaccharide Antigens. *J. Infect. Dis.* **165**, S49–S52 (1992).
 51. Pletz, M. W., Maus, U., Krug, N., Welte, T. & Lode, H. Pneumococcal vaccines: mechanism of action, impact on epidemiology and adaption of the species. *Int. J. Antimicrob. Agents* **32**, 199–206 (2008).
 52. Clutterbuck, E. A. *et al.* Pneumococcal conjugate and plain polysaccharide vaccines have divergent effects on antigen-specific B cells. *J. Infect. Dis.* **205**, 1408–1416 (2012).
 53. Christenson, B., Lundbergh, P., Hedlund, J. & Örtqvist, Å. Effects of a large-scale intervention with influenza and 23-valent pneumococcal vaccines in adults aged 65 years or older: A prospective study. *Lancet* **357**, 1008–1011 (2001).
 54. Suzuki, M. *et al.* Serotype-specific effectiveness of 23-valent pneumococcal polysaccharide vaccine against pneumococcal pneumonia in adults aged 65 years or older: a multicentre, prospective, test-negative design study. *Lancet Infect. Dis.* **17**, 313–321 (2017).
 55. Poolman, J. T., Peeters, C. C. & Van Den Dobbelsteen, G. P. The history of pneumococcal conjugate vaccine development: Dose selection. *Expert Rev. Vaccines* **12**, 1379–1394 (2013).
 56. Ladhani, S. N. *et al.* Rapid increase in non-vaccine serotypes causing invasive pneumococcal disease in England and Wales, 2000–17: a prospective national observational cohort study. *Lancet Infect. Dis.* **18**, 441–451 (2018).
 57. Massora, S. *et al.* Invasive disease potential of *Streptococcus pneumoniae* serotypes before and after 10-valent pneumococcal conjugate vaccine introduction in a rural area, southern Mozambique. *Vaccine* **37**, 7470–7477 (2019).
 58. van Tonder, A. J. *et al.* Putative novel cps loci in a large global collection of pneumococci. *Microb. Genomics* **5**, (2019).
 59. Varon, E. & Cohen, R. Novel insights into pneumococcal lineages in the vaccine era. *Lancet Infect. Dis.* **19**, 679–681 (2019).
 60. Lewnard, J. A. & Hanage, W. P. Making sense of differences in pneumococcal

- serotype replacement. *Lancet Infect. Dis.* **19**, e213–e220 (2019).
61. Stacey, H. L. *et al.* Safety and immunogenicity of 15-valent pneumococcal conjugate vaccine (PCV-15) compared to PCV-13 in healthy older adults. *Hum. Vaccines Immunother.* **15**, 530–539 (2019).
 62. Klugman, K. P. & Rodgers, G. L. Time for a third-generation pneumococcal conjugate vaccine. *Lancet Infect. Dis.* **3099**, 20–21 (2020).
 63. Kwambana-Adams, B. A., Mulholland, E. K. & Satzke, C. State-of-the-art in the pneumococcal field: Proceedings of the 11th International Symposium on Pneumococci and Pneumococcal Diseases (ISPPD-11). *Pneumonia* **12**, 1–14 (2020).
 64. Hadidi, M., Buckley, J. J. & Zydney, A. L. Effects of solution conditions on characteristics and size exclusion chromatography of pneumococcal polysaccharides and conjugate vaccines. *Carbohydr. Polym.* **152**, 12–18 (2016).
 65. Bednar, B. & Hennessey, J. P. Molecular size analysis of capsular polysaccharide preparations from *Streptococcus pneumoniae*. *Carbohydr. Res.* **243**, 115–130 (1993).
 66. Calix, J. J., Saad, J. S., Brady, A. M. & Nahm, M. H. Structural characterization of *Streptococcus pneumoniae* serotype 9A capsule polysaccharide reveals role of glycosyl 6-O-acetyltransferase *wcjE* in serotype 9V capsule biosynthesis and immunogenicity. *J. Biol. Chem.* **287**, 13996–14003 (2012).
 67. Bentley, S. D. *et al.* Genetic Analysis of the Capsular Biosynthetic Locus from All 90 Pneumococcal Serotypes. *PLoS Genet.* **2**, e31 (2006).
 68. Yother, J. Capsules of *streptococcus pneumoniae* and other bacteria: Paradigms for polysaccharide biosynthesis and regulation. *Annu. Rev. Microbiol.* **65**, 563–581 (2011).
 69. Tettelin, H. *et al.* Genomics, Genetic Variation, and Regions of Differences. in *Streptococcus Pneumoniae* 81–107 (Elsevier, 2015). doi:10.1016/B978-0-12-410530-0.00005-3
 70. Pelosi, L., Boumedienne, M., Saksouk, N., Geiselmann, J. & Geremia, R. A. The glucosyl-1-phosphate transferase *WchA* (Cap8E) primes the capsular polysaccharide repeat unit biosynthesis of *Streptococcus pneumoniae* serotype 8. *Biochem. Biophys. Res. Commun.* **327**, 857–865 (2005).
 71. Schmid, J., Heider, D., Wendel, N. J., Sperl, N. & Sieber, V. Bacterial Glycosyltransferases: Challenges and Opportunities of a Highly Diverse Enzyme Class Toward Tailoring Natural Products. *Front. Microbiol.* **7**, (2016).
 72. Aanensen, D. M., Mavroidi, A., Bentley, S. D., Reeves, P. R. & Spratt, B. G. Predicted functions and linkage specificities of the products of the *Streptococcus pneumoniae* capsular biosynthetic loci. *J. Bacteriol.* **189**, 7856–7876 (2007).
 73. Whitfield, C. Polymerases: Glycan chain-length control. *Nat. Chem. Biol.* **6**, 403–404 (2010).
 74. Jansen, A. G. S. C. *et al.* Invasive Pneumococcal Disease among Adults: Associations among Serotypes, Disease Characteristics, and Outcome. *Clin. Infect. Dis.* **49**, e23–e29 (2009).

75. Horácio, A. N., Diamantino-Miranda, J., Aguiar, S. I., Ramirez, M. & Melo-Cristino, J. The Majority of Adult Pneumococcal Invasive Infections in Portugal Are Still Potentially Vaccine Preventable in Spite of Significant Declines of Serotypes 1 and 5. *PLoS One* **8**, 1–9 (2013).
76. Luján, M. *et al.* Influence of pneumococcal serotype group on outcome in adults with bacteraemic pneumonia. *Eur. Respir. J.* **36**, 1073–1079 (2010).
77. Dube, F. S. *et al.* Longitudinal characterization of nasopharyngeal colonization with *Streptococcus pneumoniae* in a South African birth cohort post 13-valent pneumococcal conjugate vaccine implementation. *Sci. Rep.* **8**, 1–9 (2018).
78. Marsh, R. L. *et al.* Molecular characterisation of pneumococcal serotype 16F: Established predominant carriage and otitis media serotype in the 7vPCV era. *Vaccine* **25**, 2434–2436 (2007).
79. Walekhwa, M., Muturi, M., Gunturu, R., Kenya, E. & Kabera, B. *Streptococcus pneumoniae* Serotype Epidemiology among PCV-10 Vaccinated and Unvaccinated Children at Gertrude's Children's Hospital, Nairobi County: A Cross-Sectional Study. *F1000Research* **7**, 879 (2018).
80. Ramdani-Bougoussa, N. *et al.* Evolution of antimicrobial resistance and serotype distribution of *Streptococcus pneumoniae* isolated from children with invasive and noninvasive pneumococcal diseases in Algeria from 2005 to 2012. *New Microbes New Infect.* **6**, 42–48 (2015).
81. Hultén, K. G. The changing epidemiology of pneumococcal diseases. *Lancet Infect. Dis.* **18**, 929–930 (2018).
82. Ben-Shimol, S. *et al.* Comparative incidence dynamics and serotypes of meningitis, bacteremic pneumonia and other-IPD in young children in the PCV era: Insights from Israeli surveillance studies. *Vaccine* **36**, 5477–5484 (2018).
83. Waight, P. A. *et al.* Effect of the 13-valent pneumococcal conjugate vaccine on invasive pneumococcal disease in England and Wales 4 years after its introduction: an observational cohort study. *Lancet Infect. Dis.* **15**, 535–543 (2015).
84. Nakano, S. *et al.* Serotypes, antimicrobial susceptibility, and molecular epidemiology of invasive and non-invasive *Streptococcus pneumoniae* isolates in paediatric patients after the introduction of 13-valent conjugate vaccine in a nationwide surveillance study conducted in. *Vaccine* **34**, 67–76 (2016).
85. Kavalari, I. D., Fuursted, K., Krogfelt, K. A. & Slotved, H. C. Molecular characterization and epidemiology of *Streptococcus pneumoniae* serotype 24F in Denmark. *Sci. Rep.* **9**, 1–9 (2019).
86. Tin Tin Htar, M., Christopoulou, D. & Schmitt, H. J. Pneumococcal serotype evolution in Western Europe. *BMC Infect. Dis.* **15**, 1–10 (2015).
87. Beier, R. C., Mundy, B. P. & Strobel, G. A. Assignment of anomeric configuration and identification of carbohydrate residues by ¹³C nmr. 1. Galacto- and glucopyranosides and furanosides. *Can. J. Chem.* **58**, 2800–2804 (1980).
88. Duus, J., Gotfredsen, C. H. & Bock, K. Carbohydrate structural determination by NMR

- spectroscopy: modern methods and limitations. *Chem. Rev.* **100**, 4589–4614 (2000).
89. Agrawal, P. K. NMR Spectroscopy in the structural elucidation of oligosaccharides and glycosides. *Phytochemistry* **31**, 3307–3330 (1992).
 90. Beier, R. C. & Mundy, B. P. Assignment of Anomeric Configuration and Identification of Carbohydrate Residues by ¹³C NMR: Arabino- and Ribopyranosides and Furancsides. *J. Carbohydr. Chem.* **3**, 253–266 (1984).
 91. Bubb, W. A. NMR spectroscopy in the study of carbohydrates: Characterizing the structural complexity. *Concepts Magn. Reson. Part A Bridg. Educ. Res.* **19**, 1–19 (2003).
 92. Bock, K. & Pedersen, C. A study of ¹³CH coupling constants in hexopyranoses. *J. Chem. Soc. Perkin Trans. 2* 293 (1974). doi:10.1039/p29740000293
 93. Enthart, A., Freudenberger, J. C., Furrer, J., Kessler, H. & Luy, B. The CLIP/CLAP-HSQC: Pure absorptive spectra for the measurement of one-bond couplings. *J. Magn. Reson.* **192**, 314–322 (2008).
 94. Hopley, P., Howarth, O. & Ibbett, R. N. ¹H and ¹³C NMR Shifts for Aldopyranose and Aldofuranose Monosaccharides: Conformational Analysis and Solvent Dependence. *Magn. Reson. Chem.* **34**, 755–760 (1996).
 95. Richards, J. C. & Perry, M. B. Structure of the specific capsular polysaccharide of *Streptococcus pneumonia* type 23F (American type 23). *Biochem Cell Biol* **66**, 758–771 (1987).
 96. Duda, K. A., Petersen, S. & Holst, O. Structural characterization of the lipoteichoic acid isolated from *Staphylococcus sciuri* W620. *Carbohydr. Res.* **430**, 44–47 (2016).
 97. Ciucanu, I. & Kerek, F. A simple and rapid method for the permethylation of carbohydrates. *Carbohydr. Res.* **131**, 209–217 (1984).
 98. Sims, I. M., Carnachan, S. M., Bell, T. J. & Hinkley, S. F. R. Methylation analysis of polysaccharides: Technical advice. *Carbohydr. Polym.* **188**, 1–7 (2018).
 99. Gerwig, G. J., Kamerling, J. P. & Vliegenthart, J. F. G. Determination of the absolute configuration of monosaccharides in complex carbohydrates by capillary g.l.c. *Carbohydr. Res.* **77**, 1–7 (1979).
 100. Lowry, O. H., Roberts, N. R., Leiner, K. Y., Wu, M. L. & Farr, A. L. The quantitative histochemistry of brain. I. Chemical methods. *J. Biol. Chem.* **207**, 1–17 (1954).
 101. Wahlund, K. G. & Giddings, J. C. Properties of an Asymmetrical Flow Field-Flow Fractionation Channel Having One Permeable Wall. *Anal. Chem.* **59**, 1332–1339 (1987).
 102. Hu, Y., Crist, R. M. & Clogston, J. D. The utility of asymmetric flow field-flow fractionation for preclinical characterization of nanomedicines. *Anal. Bioanal. Chem.* **412**, 425–438 (2020).
 103. Zhang, H. & Lyden, D. Asymmetric-flow field-flow fractionation technology for exomere and small extracellular vesicle separation and characterization. *Nat. Protoc.* **14**, 1027–1053 (2019).

104. Fraunhofer, W. & Winter, G. The use of asymmetrical flow field-flow fractionation in pharmaceuticals and biopharmaceuticals. *Eur. J. Pharm. Biopharm.* **58**, 369–383 (2004).
105. Ma, P. L., Buschmann, M. D. & Winnik, F. M. One-step analysis of DNA/chitosan complexes by field-flow fractionation reveals particle size and free chitosan content. *Biomacromolecules* **11**, 549–554 (2010).
106. Sitar, S. *et al.* Size Characterization and Quantification of Exosomes by Asymmetrical-Flow Field-Flow Fractionation. *Anal. Chem.* **87**, 9225–9233 (2015).
107. Mudalige, T. K., Qu, H., Van Haute, D., Ansar, S. M. & Linder, S. W. Capillary electrophoresis and asymmetric flow field-flow fractionation for size-based separation of engineered metallic nanoparticles: A critical comparative review. *TrAC - Trends Anal. Chem.* **106**, 202–212 (2018).
108. Pitkänen, L. & Striegel, A. M. Polysaccharide characterization by hollow-fiber flow field-flow fractionation with on-line multi-angle static light scattering and differential refractometry. *J. Chromatogr. A* **1380**, 146–155 (2015).
109. Guo, P., Li, Y., An, J., Shen, S. & Dou, H. Study on structure-function of starch by asymmetrical flow field-flow fractionation coupled with multiple detectors: A review. *Carbohydr. Polym.* **226**, (2019).
110. Leeman, M., Choi, J., Hansson, S., Storm, M. U. & Nilsson, L. Proteins and antibodies in serum, plasma, and whole blood—size characterization using asymmetrical flow field-flow fractionation (AF4). *Anal. Bioanal. Chem.* **410**, 4867–4873 (2018).
111. Andersson, M., Wittgren, B. & Wahlund, K. G. Accuracy in multiangle light scattering measurements for molar mass and radius estimations. Model calculations and experiments. *Anal. Chem.* **75**, 4279–4291 (2003).
112. Bansal, A. K. Bioinformatics in microbial biotechnology - A mini review. *Microb. Cell Fact.* **4**, 1–11 (2005).
113. Trends, I. Bioinformatics. **17**, (1999).
114. Mavroidi, A. *et al.* Genetic relatedness of the *Streptococcus pneumoniae* capsular biosynthetic loci. *J. Bacteriol.* **189**, 7841–7855 (2007).
115. Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. A Greedy Algorithm for Aligning DNA Sequences. *J. Comput. Biol.* **7**, 203–214 (2000).
116. Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. GenBank. *Nucleic Acids Res.* **44**, D67–D72 (2016).
117. Carver, T. J. *et al.* ACT: The Artemis comparison tool. *Bioinformatics* **21**, 3422–3423 (2005).
118. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
119. Daugelaite, J., O' Driscoll, A. & Sleator, R. D. An Overview of Multiple Sequence Alignments and Cloud Computing in Bioinformatics. *ISRN Biomath.* **2013**, 1–14 (2013).

120. Pervez, M. T. *et al.* Evaluating the Accuracy and Efficiency of Multiple Sequence Alignment Methods. *Evol. Bioinforma.* **10**, EBO.S19199 (2014).
121. Di Tommaso, P. *et al.* T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* **39**, W13–W17 (2011).
122. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* (2004). doi:10.1093/nar/gkh340
123. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* (2013). doi:10.1093/molbev/mst010
124. Sievers, F. & Higgins, D. G. Clustal Omega. *Curr. Protoc. Bioinforma.* **48**, (2014).
125. Lassmann, T., Frings, O. & Sonnhammer, E. L. L. Kalign2: High-performance multiple alignment of protein and nucleotide sequences allowing external features. *Nucleic Acids Res.* (2009). doi:10.1093/nar/gkn1006
126. Moremen, K. W. & Haltiwanger, R. S. Emerging structural insights into glycosyltransferase-mediated synthesis of glycans. *Nat. Chem. Biol.* **15**, 853–864 (2019).
127. Clarke, B. R. *et al.* A bifunctional O-antigen polymerase structure reveals a new glycosyltransferase family. *Nat. Chem. Biol.* **16**, 450–457 (2020).
128. Fiser, A. Template-based protein structure modeling. *Methods Mol. Biol.* **673**, 73–94 (2010).
129. Bordoli, L. *et al.* Protein structure homology modeling using SWISS-MODEL workspace. *Nat. Protoc.* **4**, 1–13 (2009).
130. Bordoli, L. & Schwede, T. Automated protein structure modeling with swiss-model workspace and the protein model portal. *Methods Mol. Biol.* **857**, 107–136 (2012).
131. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
132. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: A unified platform for automated protein structure and function prediction. *Nat. Protoc.* (2010). doi:10.1038/nprot.2010.5
133. Källberg, M. *et al.* Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* **7**, 1511–1522 (2012).
134. Moult, J., Fidelis, K., Kryshtafovych, A., Schwede, T. & Tramontano, A. Critical assessment of methods of protein structure prediction (CASP) - round x. *Proteins Struct. Funct. Bioinforma.* (2014). doi:10.1002/prot.24452
135. Benkert, P., Biasini, M. & Schwede, T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **27**, 343–350 (2011).
136. Studer, G. *et al.* QMEANDisCo-distance constraints applied on model quality estimation. *Bioinformatics* **36**, 1765–1771 (2020).

137. Krieger, E. *et al.* Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins Struct. Funct. Bioinforma.* **77**, 114–122 (2009).
138. Kuch, A. *et al.* Usefulness of Pneumotest-Latex for direct serotyping of *Streptococcus pneumoniae* isolates in clinical samples. *J. Clin. Microbiol.* **52**, 2647–2649 (2014).
139. Slotved, H. C., Kaltoft, M., Skovsted, I. C., Kern, M. B. & Espersen, F. Simple, rapid latex agglutination test for serotyping of pneumococci (Pneumotest-Latex). *J. Clin. Microbiol.* **42**, 2518–2522 (2004).
140. Harris, S. L. & Fernsten, P. Thermodynamics and density of binding of a panel of antibodies to high-molecular-weight capsular polysaccharides. *Clin. Vaccine Immunol.* **16**, 37–42 (2009).
141. Johnson, C. S. Diffusion ordered nuclear magnetic resonance spectroscopy: Principles and applications. *Prog. Nucl. Magn. Reson. Spectrosc.* **34**, 203–256 (1999).
142. Ferrage, F., Zoonens, M., Warschawski, D. E., Popot, J. L. & Bodenhausen, G. Slow diffusion of macromolecular assemblies by a new pulsed field gradient NMR method. *J. Am. Chem. Soc.* **125**, 2541–2545 (2003).
143. Viel, S., Capitani, D., Mannina, L. & Segre, A. Diffusion-ordered NMR spectroscopy: A versatile tool for the molecular weight determination of uncharged polysaccharides. *Biomacromolecules* **4**, 1843–1847 (2003).
144. Groves, P. Diffusion ordered spectroscopy (DOSY) as applied to polymers. *Polym. Chem.* **8**, 6700–6708 (2017).
145. Mostowy, R. J. *et al.* Pneumococcal Capsule Synthesis Locus cps as Evolutionary Hotspot with Potential to Generate Novel Serotypes by Recombination. *Mol. Biol. Evol.* **34**, 2537–2554 (2017).
146. James, D. B. A. & Yother, J. Genetic and Biochemical Characterizations of Enzymes Involved in *Streptococcus pneumoniae* Serotype 2 Capsule Synthesis Demonstrate that Cps2T (WchF) Catalyzes the Committed Step by Addition of β 1-4 Rhamnose, the Second Sugar Residue in the Repeat Unit. *J. Bacteriol.* **194**, 6479–6489 (2012).
147. Li, C. *et al.* Structural, biosynthetic and serological cross-reactive elucidation of capsular polysaccharides from *Streptococcus pneumoniae* serogroup 28. *Carbohydr. Polym.* **254**, 117323 (2021).
148. Li, C. *et al.* Structural, Biosynthetic, and Serological Cross-Reactive Elucidation of Capsular Polysaccharides from *Streptococcus pneumoniae* Serogroup 16. *J. Bacteriol.* **201**, 1–13 (2019).
149. Kocev, A. *et al.* The wclY gene of *Escherichia coli* serotype O117 encodes an α 1,4-glucosyltransferase with strict acceptor specificity but broad donor specificity. *Glycobiology* **00**, 1–12 (2020).
150. Claus, H., Stummeyer, K., Batzilla, J., Mühlenhoff, M. & Vogel, U. Amino acid 310 determines the donor substrate specificity of serogroup W-135 and y capsule polymerases of *Neisseria meningitidis*. *Mol. Microbiol.* **71**, 960–971 (2009).

151. Kostova, Z. *et al.* Comparative importance in vivo of conserved glutamate residues in the EX7E motif retaining glycosyltransferase Gpi3p, the UDP-GlcNAc-binding subunit of the first enzyme in glycosylphosphatidylinositol assembly. *Eur. J. Biochem.* **270**, 4507–4514 (2003).
152. Azurmendi, H. F. *et al.* Chemical structure and genetic organization of the E. coli O6:K15 capsular polysaccharide. *Sci. Rep.* **10**, 1–12 (2020).
153. Calix, J. J., Nahm, M. H. & Zartler, E. R. Elucidation of structural and antigenic properties of pneumococcal serotype 11A, 11B, 11C, and 11F polysaccharide capsules. *J. Bacteriol.* **193**, 5271–5278 (2011).
154. Calix, J. J., Brady, A. M., Du, V. Y., Saad, J. S. & Nahma, M. H. Spectrum of pneumococcal serotype 11A variants results from incomplete loss of capsule O-acetylation. *J. Clin. Microbiol.* **52**, 758–765 (2014).
155. Ravenscroft, N. *et al.* Genetic and structural elucidation of capsular polysaccharides from *Streptococcus pneumoniae* serotype 23A and 23B, and comparison to serotype 23F. *Carbohydr. Res.* **450**, 19–29 (2017).
156. PORAMBO, Richard, J., ABEYGUNAWARDANA, Chitrananda, MUSEY, Luwy Kavuka, KOSINSKI, Michael, J., CUI, Yadong Adam, MCHUGH, Patrick, KONIETZKO, J. PNEUMOCOCCAL POLYSACCHARIDES AND THEIR USE IN IMMUNOGENIC POLYSACCHARIDE-CARRIER PROTEIN CONJUGATES. (2019).
157. Cloran, F., Carmichael, I. & Serianni, A. S. ¹³C–¹H and ¹³C–¹³C Spin Coupling Behavior in Aldofuranosyl Rings from Density Functional Theory. *J. Phys. Chem. A* **103**, 3783–3795 (1999).
158. Calix, J. J., Nahm, M. H. & Zartler, E. R. Elucidation of structural and antigenic properties of pneumococcal serotype 11A, 11B, 11C, and 11F polysaccharide capsules. *J. Bacteriol.* **193**, 5271–5278 (2011).
159. Beynon, L. M., Richards, J. C., Perry, M. B. & Kniskern, P. J. Antigenic and structural relationships within group 19 *Streptococcus pneumoniae* : chemical characterization of the specific capsular polysaccharides of types 19B and 19C . *Can. J. Chem.* **70**, 218–232 (1992).
160. Agrawal, P. K. NMR Spectroscopy in the structural elucidation of oligosaccharides and glycosides. *Phytochemistry* **31**, 3307–3330 (1992).
161. Kjeldsen, C. *et al.* Discovery and description of a new serogroup 7 *Streptococcus pneumoniae* serotype, 7D, and structural analysis of 7C and 7D. *Carbohydr. Res.* **463**, 24–31 (2018).
162. Cameron, K. S. & Fielding, L. NMR diffusion spectroscopy as a measure of host-guest complex association constants and as a probe of complex size. *J. Org. Chem.* **66**, 6891–6895 (2001).
163. Delius, J., Frank, O. & Hofmann, T. Label-free quantitative ¹H NMR spectroscopy to study low-affinity ligand–protein interactions in solution: A contribution to the mechanism of polyphenol-mediated astringency. *PLoS One* **12**, 1–14 (2017).
164. Alhabbab, R. Y. Precipitation and Agglutination Reactions. in *Basic Serological*

Testing 23–30 (2018). doi:10.1007/978-3-319-77694-1_3

165. Dati, F. *et al.* Technical Reports. *Clin. Chem. Lab. Med.* **34**, (1996).
166. Monaco, S., Tailford, L. E., Juge, N. & Angulo, J. Differential Epitope Mapping by STD NMR Spectroscopy To Reveal the Nature of Protein-Ligand Contacts. *Angew. Chemie Int. Ed.* **56**, 15289–15293 (2017).
167. Walpole, S., Monaco, S., Nepravishta, R. & Angulo, J. STD NMR as a Technique for Ligand Screening and Structural Studies. in *Methods in Enzymology* **615**, 423–451 (Elsevier Inc., 2019).
168. Vialle, S., Sepulcri, P., Dubayle, J. & Talaga, P. The teichoic acid (C-polysaccharide) synthesized by *Streptococcus pneumoniae* serotype 5 has a specific structure. *Carbohydr. Res.* (2005). doi:10.1016/j.carres.2004.10.021

Appendix

Table A1. CWPS structures

| CWPS | Structure | reference |
|------|---|-----------|
| 1 | $\begin{array}{c} \text{P-Cho} \\ \\ 6 \\ \\ \text{-6)-}\beta\text{-D-Glcp-(1-3)-}\alpha\text{-AATGalp-(1-4)-}\alpha\text{-D-GalpNAc-(1-3)-}\beta\text{-D-GalpNAc-(1-1)-D-Rib-ol-5-P-} \end{array}$ | 14 |
| 2 | $\begin{array}{cc} \text{P-Cho} & \text{P-Cho} \\ & \\ 6 & 6 \\ & \\ \text{-6)-}\beta\text{-D-Glcp-(1-3)-}\alpha\text{-AATGalp-(1-4)-}\alpha\text{-D-GalpNAc-(1-3)-}\beta\text{-D-GalpNAc-(1-1)-D-Rib-ol-5-P-} \end{array}$ | 14 |
| 3 | $\begin{array}{cc} \text{P-Cho} & \text{P-Cho} \\ & \\ 6 & 6 \\ & \\ \text{-6)-}\beta\text{-D-Galp-(1-3)-}\alpha\text{-AATGalp-(1-4)-}\alpha\text{-D-GalpNAc-(1-3)-}\beta\text{-D-GalpNAc-(1-1)-D-Rib-ol-5-P-} \end{array}$ | 168 |

*AATGal: 2-acetamido-4-amino-2,4,6-trideoxy-D-galactose;

Table A2. A Summary of amino acid (AA) lengths, accession numbers of biosynthetic enzymes from serotype 28F, 28A and 16F.

| Enzyme name | 16F | | 28F | | 28A | |
|-------------|-----------|---------------|-----------|---------------|-----------|---------------|
| | AA Length | Accession Nr. | AA length | Accession Nr. | AA length | Accession Nr. |
| Wzg | 481 | CAI33456.1 | 484 | CAI34048.2 | 484 | CAI34023.1 |
| Wzh | 243 | CAI33457.1 | 243 | CAI34049.2 | 243 | CAI34024.1 |
| Wzd | 230 | CAI33458.1 | 231 | CAI34050.2 | 231 | CAI34025.1 |
| Wze | 227 | CAI33459.1 | 229 | CAI34051.2 | 229 | CAI34026.1 |
| WchA | 455 | CAI33460.1 | 455 | CAI34052.2 | 455 | CAI34027.1 |
| WchF | 390 | CAI33461.1 | 390 | CAI34053.2 | 390 | CAI34028.1 |
| WciU | 423 | CAI33462.1 | 423 | CAI34054.2 | 423 | CAI34029.1 |
| WcxM | 228 | CAI33463.1 | 228 | CAI34055.2 | 228 | CAI34030.1 |
| WcxN | 301 | CAI33464.1 | 292 | CAI34056.2 | 292 | CAI34031.1 |
| Wzy | 420 | CAI33466.1 | 410 | CAI34057.2 | 410 | CAI34032.1 |
| WcxP | 388 | CAI33467.1 | 386 | CAI34058.2 | 386 | CAI34033.1 |
| Wzx | 483 | CAI33468.1 | 483 | CAI34059.2 | 483 | CAI34034.1 |
| WcxQ | 271 | CAI33469.1 | 270 | CAI34060.2 | 270 | CAI34035.1 |
| Gtp1 | | | 342 | CAI34061.2 | 342 | CAI34036.1 |
| Gtp2 | | | 234 | CAI34062.2 | 234 | CAI34037.1 |
| Gtp3 | | | 277 | CAI34063.2 | 283 | CAI34038.1 |
| Gtp | 130 | CAI33470.1 | | | | |
| RmlA | 289 | CAI33471.1 | 289 | CAI34064.2 | 289 | CAI34039.1 |
| RmlC | 197 | CAI33472.1 | 202 | CAI34065.2 | 197 | CAI34040.1 |
| RmlB | 349 | CAI33473.1 | 349 | CAI34066.2 | 349 | CAI34041.1 |
| RmlD | 283 | CAI33474.1 | 283 | CAI34067.2 | 283 | CAI34042.1 |

Table A3. Templates used for modeling of WciU of 28A in Phyre2 (93% of residues modeled at >90% confidence, 29 residues were modeled by ab initio.)

| Phyre2 Code | Confidence of homologous | PDB Code | Enzyme Name | Source Organism |
|-------------|--------------------------|----------|--|-----------------------------|
| c2r60A | 100.0% | 2r60 | Apo Sucrose Phosphate Synthase | Halothermothrix orenii |
| c6gnfC | 100.0% | 6gnf | Starch Synthase | Cyanobacterium sp. CLg1 |
| c6gneB | 100.0% | 6gne | Starch Synthase 4 catalytic domain | Arabidopsis thaliana |
| c3c4vB | 100.0% | 3c4v | the retaining glycosyltransferase MshA | Corynebacterium glutamicum |
| d2busa1 | 100.0% | 2bis | glycogen synthase | Pyrococcus abyssi |
| c3vufA | 100.0% | 3vuf | Starch Synthase I Catalytic Domain | Oryza sativa Japonica Group |

Table A4. Templates used for modeling of WciU of 28A in Phyre2 (97% of residues modeled at >90% confidence, 13 residues were modeled by ab initio.)

| Phyre2 Code | Confidence of homologous | PDB Code | Enzyme Name | Source Organism |
|-------------|--------------------------|----------|--|-----------------------------|
| c2r60A | 100.0% | 2r60 | Apo Sucrose Phosphate Synthase | Halothermothrix orenii |
| c6gnfC | 100.0% | 6gnf | Starch Synthase | Cyanobacterium sp. CLg1 |
| c6gneB | 100.0% | 6gne | Starch Synthase 4 catalytic domain | Arabidopsis thaliana |
| c3c4vB | 100.0% | 3c4v | the retaining glycosyltransferase MshA | Corynebacterium glutamicum |
| d2qzsA | 100.0% | 2qzs | Glycogen Synthase | Escherichia coli |
| c3vufA | 100.0% | 3vuf | Starch Synthase I Catalytic Domain | Oryza sativa Japonica Group |

emb|CA134054|228F1-423 1 --- MKILHYTGFG - QPORTGGLVKYAEIDLML --- EQI AAG --- YQVAALYP --- GR I KF --- FSKKIE I K - AMSRQFECYELLNSLP - -MALF 75
emb|CA133605|118F1-423 1 --- MKILHYTGFG - QPORTGGLVKYAEIDLML --- EQI AAG --- YQVAALYP --- GR I KF --- FSKKIE I K - ATSRQFECYELLNSLP - -MALF 75
emb|CA133462|118F1-423 1 --- MKILHYTGFG - QPORTGGLVKYAEIDLML --- EQI AAG --- YQVAALYP --- GR I KF --- FSKKIE I K - ETSRQFECYELLNSLP - -MALF 75
emb|CA133581|118B1-423 1 --- MKILHYTGFG - QPORTGGLVKYAEIDLMI --- EQI AAG --- YQVAALYP --- GR I KF --- FSKKIE I K - ATSRQFECYELLNSLP - -MALF 75
emb|CA133557|118C1-423 1 --- MKILHYTGFG - QPORTGGLVKYAEIDLMI --- EQI AAG --- YQVAALYP --- GR I KF --- FSKKIE I K - ATSRQFECYELLNSLP - -MALF 75
emb|CA133533|118A1-423 1 --- MKILHYTGFG - QPORTGGLVKYAEIDLMI --- EQI AAG --- YQVAALCP --- GR I KF --- FSKKIE I K - ATSRQFECYELLNSLP - -IALF 75
emb|CA134029|128A1-423 1 --- MKILHYTGFG - QPORTGGLVKYAEIDLMI --- EQI AAG --- YQVAALCP --- GR I KF --- FSKKIE I K - ATSRQFECYELLNSLP - -IALF 75
4PQG/A2-503 2 HHMTIYINLGI - QWASSG - VEYQAAYRAGVFRK - -LNLSSKF I F TDI LADNIOHLTANIGFDNDGVILWYHFDI K IAPTSTVTDVLLAYFG - -GEESH - -REKNGKLRV - - -FFDQDQFVTC - -115
5E9U/A1-503 1 --- STYVINLGI - QWASSG - VEYQAAYRAQILRRIQQPAKIFIM - -DM LADNIOHLTENIGFDEE I WLYNYFTDI K IAPTSTVTDVLLAYFG - -GOPER - -SEKEGKLV - R - - -YFPQDQDQ ITC - -115
2V1T/A2-371 2 --- IVA - FCYI - KYFP - -FGQLDRDFMR I AS - -TVAARG - -HHVRVYTQSW - -GDGP - -KAFEL I - - -52
PDB|5000|5000/A2-362 2 --- RKL - -K IGI TCYPS - VGGSG I IATELGK - -TVAARG - -HHVRVYTQSW - -GDGP - -KAFEL I - - -52
6N1X/A3-373 3 --- FQG - HMK IGI TCYPS - VGGSG I IATELGK - -TVAARG - -HHVRVYTQSW - -GDGP - -KAFEL I - - -52
emb|CA134054|228F1-423 76 GG SDPTAFLTPCDKNVYRTFLEKVPDI I H I H S F M G L H K F E I A K N L - - N I R V V F T S H D Y G L A - - - - - P V P H F Y F N G V D - - - - - Y S N K S T N L T W N I M S S N A L S V K K L R L F - Q V S F Y P 182
emb|CA133605|118F1-423 76 GG SDPTAFLTPCDKNVYRTFLEKVPDI I H I H S F M G L H K F E I A K N L - - N I R V V F T S H D Y G L A - - - - - P V P H F Y F N G V D - - - - - Y S D K S T N L T W N I M S S N A L S V K K L R L F - Q V S F Y P 182
emb|CA133462|118F1-423 76 GG SDPTAFLTPCDKNVYRTFLEKVPDI I H I H S F M G L H K F E I A K S L - - N I R V V F T S H D Y G L A - - - - - P V P H F Y F N G V D - - - - - Y S D K S T N L T W N I M S S N A L S V K K L R L F - Q V S F Y P 182
emb|CA133581|118B1-423 76 GG SDPTAFLTPCDKNVYRTFLEKVPDI I H I H S F M G L H K F E I A K N L - - N I R V V F T S H D Y G L A - - - - - P V P H F Y F N G V D - - - - - Y S D K N T L T W N I M S S N A L S V K K L R L F - Q V S F Y P 182
emb|CA133557|118C1-423 76 GG SDPTAFLTPCDKNVYRTFLEKVPDI I H I H S F M G L H K F E I A K N L - - N I R V V F T S H D Y G L A - - - - - P V P H F Y F N G V D - - - - - Y S D K N T L T W N I M S S N A L S V K K L R L F - Q V S F Y P 182
emb|CA133533|118A1-423 76 GG SDPTAFLTPCDKNVYRTFLEKVPDI I H I H S F M G L H K F E I A K N L - - N I R V V F T S H D Y G L A - - - - - P V P H F Y F N G V D - - - - - Y S D K N T L T W N I M S S N A L S V K K L R L F - Q V S F Y P 182
emb|CA134029|128A1-423 76 GG SDPTAFMTPCDKNVYRTFLEKVPDI I H I H S F M G L H K F E I A K N L - - N I R V V F T S H D Y G L A - - - - - P V P H F Y F N G V D - - - - - Y S D K N T L T W N I M S S N A L S V K K L R L F - Q V S F Y P 182
4PQG/A2-503 116 --- YLVENKDLVQH - -AEY - - - - - V F K G - - - N L I R K D Y F S Y - - - - - R Y C S E Y F A P K D N V A V L Q R T F Y N E D G - T P V D I L W N Q G K - - E E V Y H F K D I F Y G 196
5E9U/A1-503 116 --- YLREQDDFVEH - -VEY - - - - - V S R G - - - R L I R K D Y F S Y - - - - - R Y A S E Y F A P H N D A A T L Y Q R R F Y H E D G - S V A Y D M L I E D G Q - - E K I Y R F D R I F Y S 196
2V1T/A2-371 53 --- --- - -QV - - P V K S H T H G R N A E Y Y A W Q N H L K E H P A D R V G F N K M P G L D V Y F A - - - - - A D V C Y A E K V A Q E R G F L Y R L S R Y R 122
PDB|5000|5000/A2-362 76 --- --- - -VAERENLD I H A H Y A L - P H A V C A Y L A K O M L K R N I G I V T T L H G T - - - - - --- - -117
6N1X/A3-373 79 --- --- - -V I K E Y D L D L H M H Y A V - P H A I C G - - - - - --- - -100
emb|CA134054|228F1-423 183 --- T I R K L L K L G K N P - K S K N L V I R D V - - - - - I E E D V Y S E L R Y Y N E M F H L I - D - - - G Y L F N S R L - A K K V Y E I N E - I K S A N S I V L S I T N S S I N H H Q R L T T T N N K I R V A Y - I G P D E E Y K G Y F 288
emb|CA133605|118F1-423 183 --- T I R K L L K L G K N P - K S K N L V I R D V - - - - - I E E D V Y S E L R Y Y N E M F H L I - D - - - G Y L F N S R L - A K K V Y E I N E - I K S A N S I V L S I T N S S I N H H Q R L T T T N N K I R V A Y - I G P D E E Y K G Y F 288
emb|CA133462|118F1-423 183 --- T I R K L L K L G K N P - K S K N L V I R D V - - - - - I E E D V Y S E L R Y Y N E M F H L I - D - - - G Y L F N S R L - A K K V Y E I N E - I K S A N S I V L S I T N S S I N H H Q R L T T T N N K I R V A Y - I G P D E E Y K G Y F 288
emb|CA133581|118B1-423 183 --- T I R K L L K L G K N P - K S K N L V I R D V - - - - - I E E D V Y S D L R Y Y N E M F H L I - D - - - G Y L F N S R L - A K K V Y E I N E - I K S A N S I V L S I T N S S I N H H Q R L T T T N N K I R V A Y - I G P D E E Y K G Y F 288
emb|CA133557|118C1-423 183 --- T I R K L L K L G K N P - K S K N L V I R D V - - - - - I E E D V Y S D L R Y Y N E M F H L I - D - - - G Y L F N S R L - A K K V Y E I N E - I K S A N S I V L S I T N S S I N H H Q R L T T T N N K I R V A Y - I G P D E E Y K G Y F 288
emb|CA133533|118A1-423 183 --- T I R K L L K L G K N P - K S K N L V I R D V - - - - - I E E D V Y S E L R Y Y N E M F H L I - D - - - G Y L F N S R L - A K K V Y E I N E - I Q P A N S V L S I T S S S I K H H Q R L T T T N N K I R V S Y - I G D E E Y K G Y F 288
emb|CA134029|128A1-423 183 --- T I R K L L K L G K N P - K S K N L V I R D V - - - - - I E E D V Y S E L R Y Y N E M F H L I - D - - - G Y L F N S R L - A K K V Y E I N E - I Q P A N S V L S I T S S S I K H H Q R L T T T N N K I R V A Y - I G D E E Y K G Y F 288
4PQG/A2-503 197 KQAFVRAFMSLNLN - -KSDVL I D R E T G I G Q V F E E A Q T A H L A V V H A E H Y S E N A T N E D V I L W N Y D Y Q F T N A - D K V D F I V S T O R - Q N E V L Q - - - - - E Q F A K Y T 294
5E9U/A1-503 197 K A E L V R Y F L Q C L Q - - - - - A D D V I L D R E T G I G Q V F E E S Q A K L G V V H A E H F S E N A S S D V I L W N Y D Y Q F T N A - D K V D F I V A T E A - Q K R I L E - - - - - Q O F Q H Y S 294
2V1T/A2-371 123 H Y A A F E R A T F E Q S T L K M M E T - - - - - P E S L R H - N T L F V V G O Q K P R - K F E A L A E K L G V R S N V H F F S - - G R N D V S E L M A A A D L L H P A - Y O E A A G I V L L E A I T A G L P V L T T A V - C B A Y A H I A 310
PDB|5000|5000/A2-362 118 --- - -D I T V L G Y D P - S - L K D - L R F A - - - - - I E S S D R V T A - - - - - V S S A L - A E A T Y D L I K P E K K I E T I Y N F I K N T A A I K E K H G I L P D E K V - V I H - Y S N F R K V R Q 202
6N1X/A3-373 101 --- - -I L A R - E M S - G D K I M T T - - - - - L H G T I T V L G D H S L Q G A I K F G I E K S D I V T S V S - K S L A Q T H E - I I E T N K E I I P I Y N F - V R E N E F P T K H N T A L K S Q G I A P D E K V L I H V 202
emb|CA134054|228F1-423 289 DF DFVETL - - - - - D R E - S Y - E V A - - T Y G H - - - - - L P N E C P S F I E Q K G Y F T K M I D S V Y E N I D I L I V P S K W K E T F G L I T V E A L S Y G V N V F V S E N - V S K D L P 377
emb|CA133605|118F1-423 289 DF DFVDFVETL - - - - - D R E - S Y - E V A - - T Y G H - - - - - L P N E C P S F I E Q K G Y F T K M I D S V Y E N I D I L I V P S K W K E T F G L I T V E A L S Y G V N V F V S E N - V S K D L P 377
emb|CA133462|118F1-423 289 DF DFVDFVETL - - - - - D R E - S Y - E V A - - T Y G H - - - - - L P N E C P S F I E Q K G Y F T K M I D S V Y E N I D I L I V P S K W K E T F G L I T V E A L S Y G V N V F V S E N - V S K D L P 377
emb|CA133581|118B1-423 289 DF DFVETL - - - - - D R E - S Y - E V A - - T Y G H - - - - - L P N E C P S F I E Q K G Y F T K M I D S V Y E N I D I L I V P S K W K E T F G L I T V E A L S Y G V N V F V S E N - V S K D L P 377
emb|CA133557|118C1-423 289 DF DFVETL - - - - - D R E - S Y - E V A - - T Y G H - - - - - L P N E C P S F I E Q K G Y F T K M I D S V Y E N I D I L I V P S K W K E T F G L I T V E A L S Y G V N V F V S E N - V S K D L P 377
emb|CA133533|118A1-423 289 DF DFAGT - - - - - E Q E - S Y - E V V - - T Y G H - - - - - L P N E C P S F I E Q K G Y F T K E T I D S V Y E N I D I L I A S K C K E T F G L I T V E A L S Y G V N V F V S E N - V S K D L P 377
emb|CA134029|128A1-423 289 DF DFAGT - - - - - E Q E - S Y - E V V - - T Y G H - - - - - L P N E C P S F I E Q K G Y F T K E T I D S V Y E N I D I L I A S K C K E T F G L I T V E A L S Y G V N V F V S E N - V S K D L P 377
4PQG/A2-503 286 QHQP I V T I P V G S I D S L T D S G R K P F S L I T A S R L A E K H I D W L V A V I E A H K E L P E L T F D - I Y G S G E E D K L R R I I E A G A Q D Y I R L K G - - H A D L S Q I Y A G Y E L U T A S - T S S G F G L L M E A G S S L P I I G F D V R Y G N O T F I D 434
5E9U/A1-503 285 DQOP I A T I P V G S I D S L T D S G R K P F S M I T A S R L A E K H I D W L V A V I E A H K E L P E L T F D - I Y G S G E E D K L R R I I E A G A Q D Y I R L K G - - H A D L S Q I Y A G Y E L U T A S - T S S G F G L L M E A G S S L P I I G F D V R Y G N O T F I D 434
2V1T/A2-371 213 R S I E A L A S - - - - - P E S L R H - N T L F V V G O Q K P R - K F E A L A E K L G V R S N V H F F S - - G R N D V S E L M A A A D L L H P A - Y O E A A G I V L L E A I T A G L P V L T T A V - C B A Y A H I A 310
PDB|5000|5000/A2-362 203 D V I R V R N I - - - - - A - - - - - G K T A - K L L - L V E D G P E K S T A E L I R Y G L E D Q L M L G - - N O D R V E D L Y S I S D L K L L S - E K S F G L L L E A M K T G V P I G S N A - G B I K E V I K 299
6N1X/A3-373 203 S N F R Q V K R - - - - - D T I I E T F A K V R E K I P S - K L I - - L L G D G P E L V P M R Q L K E L N E E D V L F L G - - K D C V S E F Y Q L S D L V L L S - E K S F G L L L E A M K T G V P I G S N A - G B I K E V I K 309
emb|CA134054|228F1-423 378 ESHV - -FKN - - - - - Q N D - L V V K F L K N D I - - - - - E N T K L K - - - - - T I D - - - - - E H S I E V I Q - Y Y - - - - - E R V I N D - - - - - 423
emb|CA133605|118F1-423 378 ESHV - -FKN - - - - - Q N D - L V V K F L K N D I - - - - - E N T K L K - - - - - T I D - - - - - E H S I E V I Q - Y Y - - - - - E R V I N D - - - - - 423
emb|CA133462|118F1-423 378 ESHV - -FKN - - - - - Q N D - L I V K F L K N D I - - - - - E N T K L K - - - - - T I D - - - - - E H S I E V I Q - Y Y - - - - - E R V I N D - - - - - 423
emb|CA133581|118B1-423 378 ETHV - -FKD - - - - - K E D - L L A K I I N N Q L - - - - - K K I P L K - - - - - T I E - - - - - E H V E E I V Q - Y Y - - - - - K Q V R S N - - - - - 423
emb|CA133557|118C1-423 378 ETHV - -FKD - - - - - K E D - L L A K I I N N Q L - - - - - K K I P L K - - - - - T I E - - - - - E H V E E I V Q - Y Y - - - - - K Q V R S N - - - - - 423
emb|CA133533|118A1-423 378 ETHV - -FKD - - - - - K E D - L L A K I I N N Q L - - - - - K K I P L K - - - - - T I E - - - - - E H V E E I V Q - Y Y - - - - - K Q V R S N - - - - - 423
emb|CA134029|128A1-423 378 ETHV - -FKN - - - - - Q N D - L V V K F L K N D I - - - - - E N T K L K - - - - - T I D - - - - - E H S I E V I Q - Y Y - - - - - E R V I N D - - - - - 423
4PQG/A2-503 435 DQONQ I P S - - - - - S D H V E D Q I K Q A Y A A K I C Q L Y O E N R L E - - - - - A M R - - - - - A Y S Y O A E - G F L T K E I L E K W K Y V E E V L H - - - - - 503
5E9U/A1-503 435 DQONQ I P S - - - - - S D H V E D Q I K Q A Y A A K I C Q L Y O E N R L E - - - - - A M R - - - - - A Y S Y O A E - G F L T K E I L E K W K Y V E E V L H - - - - - 503
2V1T/A2-371 311 D A N C O T V I A E P - - - - - F S Q E Q - L N E V L R K A L - - - - - T Q S P L R M A W A E N A R H Y A D T Q D L Y S L P - - - - - E K A A D I I T - G Q - - - - - 371
PDB|5000|5000/A2-362 300 H N V S G F L V D G D V T A A T A R A M S I L E D Q L S N R F - - - - - T K A A I E - - - - - M L E M E F - - - - - S S K K I V S - Q Y - - - - - E Q I Y A D L A E - - - - - 362
6N1X/A3-373 310 H G E T G F V D G D C D S A S Y A I R L L E D K V L Y N K L - - - - - Q N K M L A - - - - - D J A E R F G S E L I T Q D Y E Y Y - - - - - Q K M L N - - - - - 373

Figure A1 Full multiple sequence alignment of WciU from serotype 16F, 18F, 18A, 18B, 18C, 28F and 28A with protein sequences of GT4 GlcNAc-Ts (4PQG.pdb, 5E9U.pdb, 5D00.pdb and 6N1X.pdb). Alignment are colored by Clustalx with 50% conservation increased visibility.

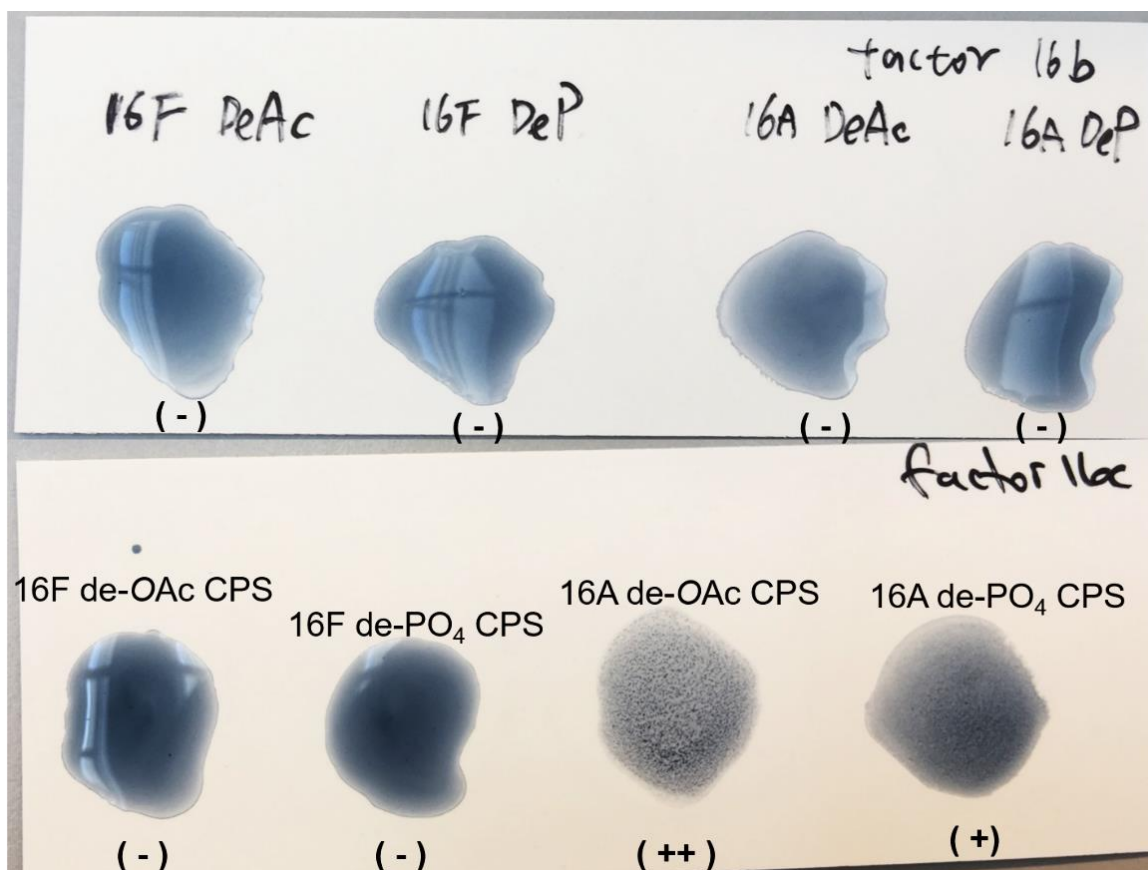


Figure A2 Latex agglutination test results of factor antiserum 16b and 16c with de-OAc and de-PO₄ 16F and 16A CPS

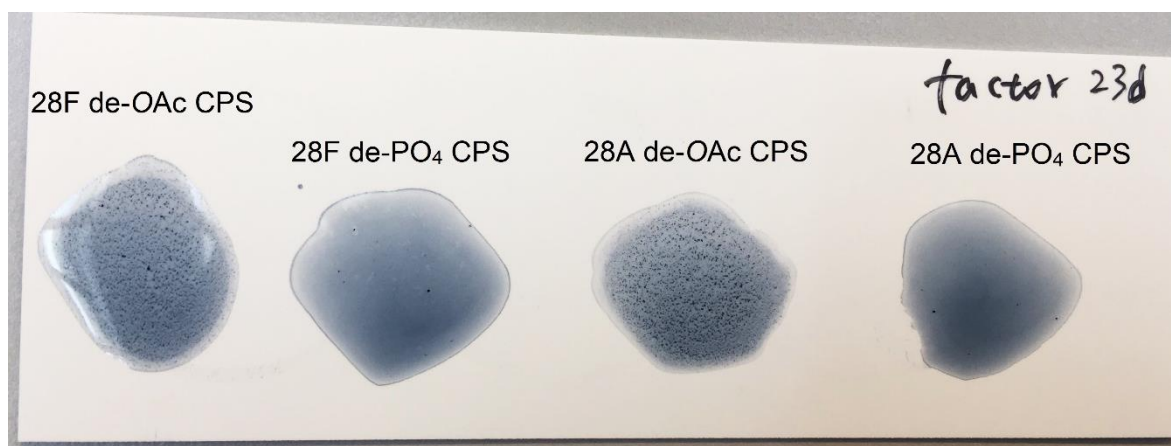


Figure A3 Latex agglutination test results of factor antiserum 23d with de-OAc and de-PO₄ 28F and 28A CPS

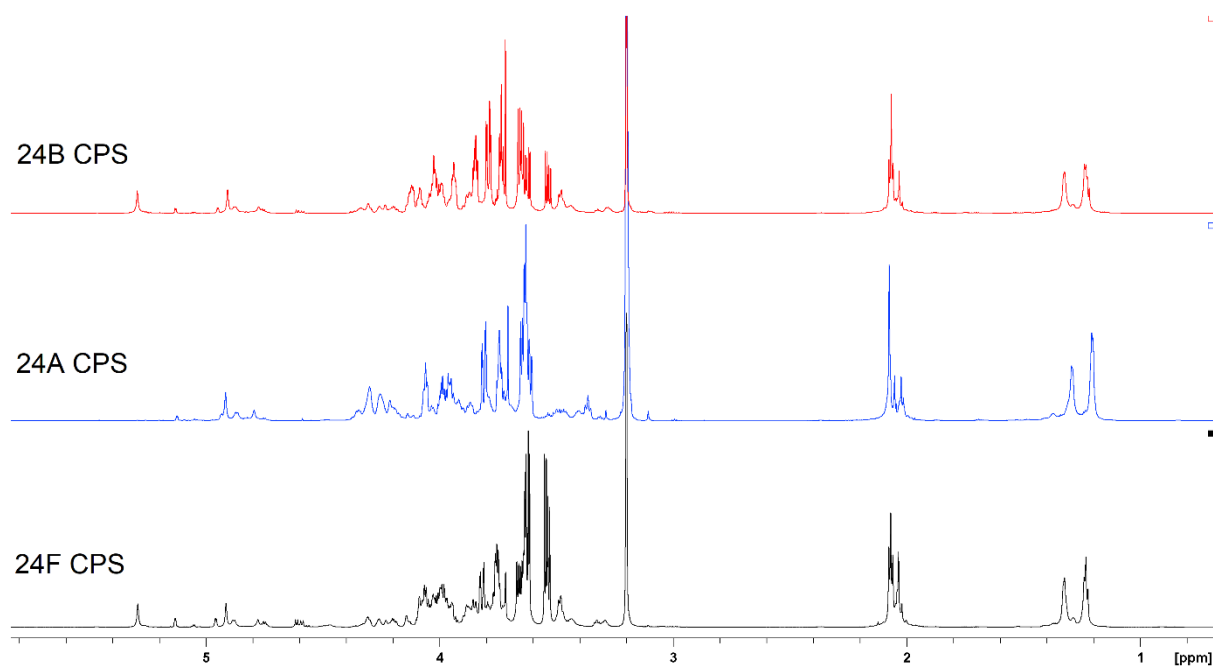


Figure A4 Overlapped ^1H spectra of 24F (black), 24A (blue), 24B (red) CPS.

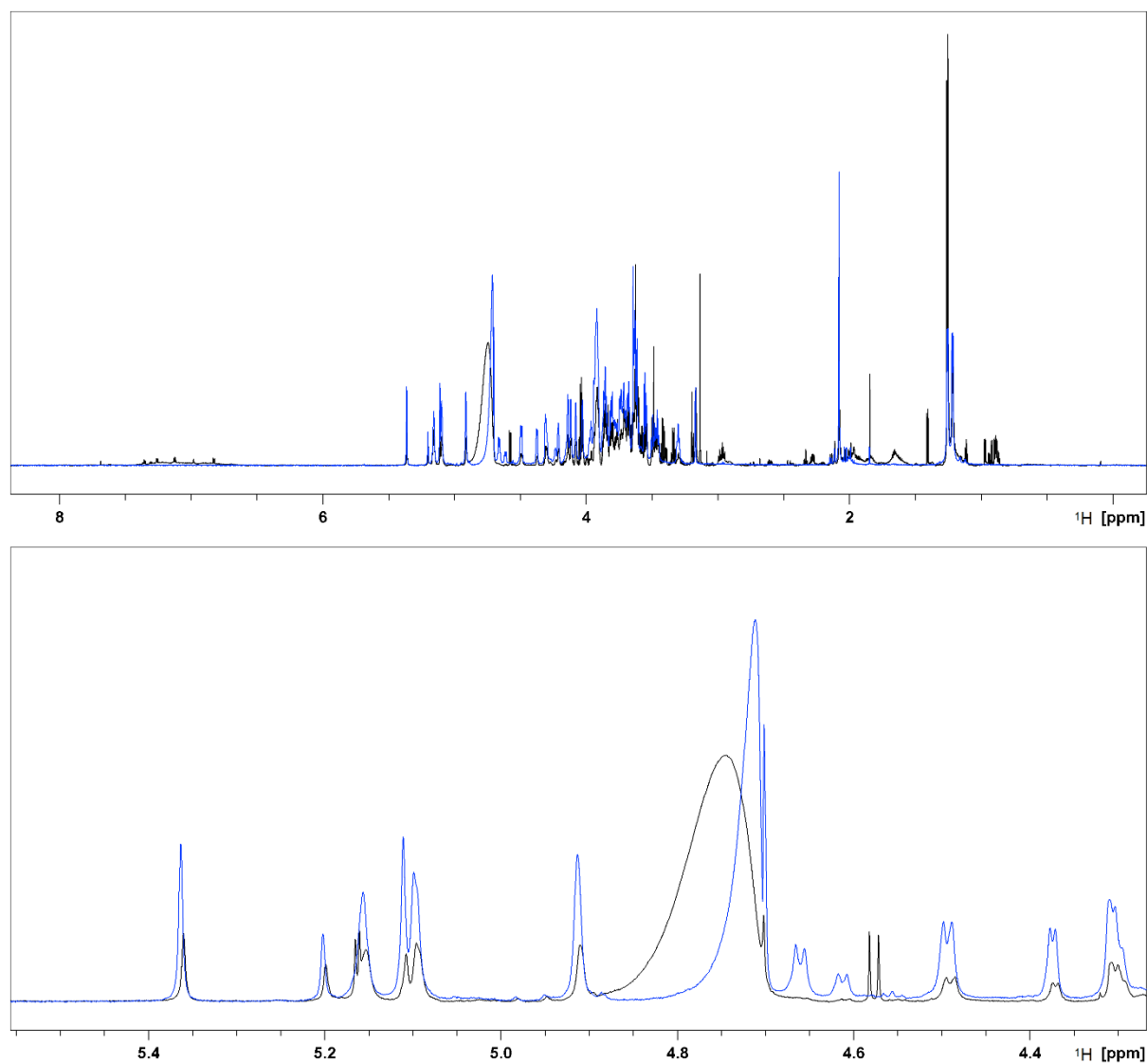
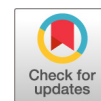


Figure A5 Overlapped ^1H spectra of 16A CPS with 0 μL 16c-1 antiserum (blue) and 50 μL 16c-1 antiserum 24A (black).

Paper 1

Li C, Duda KA, Elverdal PL, Skovsted IC, Kjeldsen C, Duus JØ. 2019.
Structural, biosynthetic, and serological cross-reactive elucidation of capsular polysaccharides from *Streptococcus pneumoniae* serogroup 16. J Bacteriol 201:1–13.



Structural, Biosynthetic, and Serological Cross-Reactive Elucidation of Capsular Polysaccharides from *Streptococcus pneumoniae* Serogroup 16

Chengxin Li,^a Katarzyna A. Duda,^b Pernille L. Elverdal,^c Ian C. Skovsted,^c Christian Kjeldsen,^a Jens Ø. Duus^a

^aDepartment of Chemistry, Technical University of Denmark, Kongens Lyngby, Denmark

^bJunior Research Group of Allergobiocchemistry, Research Center Borstel, Leibniz Lungenzentrum, Airway Research Center North (ARCN), German Center for Lung Research (DZL), Borstel, Germany

^cSSI Diagnostica A/S, Hilleroed, Denmark

ABSTRACT Capsular polysaccharides (CPS) are crucial virulence factors of *Streptococcus pneumoniae*. The previously unknown CPS structures of the pneumococcal serogroup 16 (serotypes 16F and 16A) were thoroughly elucidated by nuclear magnetic resonance (NMR) spectroscopy and verified by chemical analysis. The following repeat unit structures were determined: 16F, -3)- α -L-Rhap-[4-P-1-Gro]-(1-3)- α -D-Glcp-[(6-P-1)-Gro]-(1-3)- β -L-Rhap-[2-OAc]-(1-4)- β -D-Glcp-(1-; 16A, -3)- β -D-Galp-[2-OAc (70%)]-(1-3)- α -L-Rhap-(1-2)- α -L-Rhap-(1-3)- α -D-Galp-[(6-P-1)-Gro]-(1-3)- β -D-Galp-(1-4)- β -D-Glcp-(1- (OAc, O-acetyl substitution; P-1-Gro, glycerol-1-phosphate substitution). A further analysis of CPS biosynthesis of serotypes 16F and 16A, in conjunction with published *cps* gene bioinformatics analysis and structures of related serotypes, revealed presumable specific function of glycosyltransferase, acetyltransferase, phosphotransferase, and polymerase. The functions of glycosyltransferases WcxN and WcxT were proposed for the first time, and they were assigned to catalyze linkage of α -L-Rhap-(1-3)- α -D-Glcp and α -L-Rhap-(1-2)- α -L-Rhap, respectively. Furthermore, since serotype 16F was genetically close to serogroup 28, cross-reactions between serogroup 16 and serogroup 28 were studied using diagnostic antisera, which provided further understanding of antigenic properties of CPS and diagnostic antisera. Interestingly, serotype 16F cross-reacted with factor antisera 28b and 11c. Meanwhile, serotype 16A cross-reacted with factor antiserum 11c.

IMPORTANCE The vaccine pressure against *Streptococcus pneumoniae* could result in a change of prevalence in carriage and invasive serotypes. As such, it is necessary to monitor the distribution to achieve successful vaccination of the population, and similarly, it is important to increase the knowledge of even the currently less prevalent serotypes. The CPS are vital for the virulence of the pathogen, and antigenic properties of CPS are based on the structure. Consequently, a better understanding of the structure, biosynthesis, and serology of the capsular polysaccharides can be of great importance toward developing future diagnostic tools and vaccines.

KEYWORDS NMR spectroscopy, *Streptococcus pneumoniae*, capsular polysaccharides, serotypes 16F and 16A, structural characterization

Streptococcus pneumoniae is an important encapsulated Gram-positive human pathogen with 93 distinct serotypes reported (1, 2). The capsular polysaccharide (CPS), forming the hydrated outer layer of *S. pneumoniae*, protects the pathogen from the host immune system, which is the basis for successful vaccines against infections caused by *S. pneumoniae*. Based on the importance of CPS, it has been extensively studied from basic structure and genetic biosynthesis to immunology. Determination of

Citation Li C, Duda KA, Elverdal PL, Skovsted IC, Kjeldsen C, Duus JO. 2019. Structural, biosynthetic, and serological cross-reactive elucidation of capsular polysaccharides from *Streptococcus pneumoniae* serogroup 16. J Bacteriol 201:e00453-19. <https://doi.org/10.1128/JB.00453-19>.

Editor Michael J. Federle, University of Illinois at Chicago

Copyright © 2019 American Society for Microbiology. All Rights Reserved.

Address correspondence to Jens Ø. Duus, jduus@kemi.dtu.dk.

Received 4 July 2019

Accepted 25 July 2019

Accepted manuscript posted online 5 August 2019

Published 20 September 2019

unknown pneumococcal CPS structures is important to understand the function of biosynthetic genes and enzymes and the relationships between structure and antigenicity. The elucidation of a genetic basis for all capsular types has been completed by sequencing the corresponding DNA stretches (3), while the structural analysis of the corresponding capsular polysaccharide has remained incomplete.

Serotype 16F is not included in any current vaccine; however, in recent years, it was reported as a potential cause of invasive pneumococcal disease (IPD) and increased case fatality rates in several studies (4–8). The efficacy of a given vaccine is inherently dependent on the serotype distribution in the target population for the serotype-specific protection of pneumococcal vaccines. Thus, it is important to monitor the epidemiology of pneumococcal serotypes to formulate future pneumococcal vaccines. One study proposed 16F as a candidate for higher-valency pneumococcal conjugate vaccines (9). However, the CPS structures of serotypes 16F and 16A were both unknown, which consequently resulted in ambiguity of the function of related biosynthetic genes and enzymes. Here, we report the repeat unit structure of serotype 16F and 16A CPS based on nuclear magnetic resonance (NMR) and chemical analysis, together with biosynthetic analysis to correlate CPS structure with the presence of related enzymes. Assignments of specific functions of glycosyltransferases (GT), acetyltransferase, phosphotransferases, and polymerases are then proposed. In addition, cross-reactions with rabbit antisera between serogroup 16 with genetically related serotypes were studied to gain an understanding of CPS from serological perspectives.

RESULTS AND DISCUSSION

NMR structural analyses of 16F and 16A CPS repeat units were conducted mainly in two steps: assignment of individual monosaccharides or residues and connecting the identified structural motifs to yield the repeat unit structure (10).

Assignment of individual spin systems was based on one-dimensional (1D) ^1H NMR and 2D double quantum filter correlated spectroscopy (DQF-COSY) spectra, total correlation spectroscopy (TOCSY) spectra, multiplicity-edited ^1H - ^{13}C heteronuclear single quantum coherence (HSQC) spectra, and HSQC-TOCSY spectra. In addition, heteronuclear 2-bond correlation (H2BC) spectra and heteronuclear multibond correlation (HMBC) spectra were used to confirm spin systems, especially in some overlapping regions. Glycosidic bonds then were identified using ^1H - ^{13}C HMBC spectra and nuclear Overhauser effect spectroscopy (NOESY) spectra, as were O-acetylated positions. Phosphodiester linkages were determined by ^1H - ^{31}P HMBC. Clean in-phase (CLIP)-HSQC (11) was used to determine the anomeric configuration of pyranosides (12). In addition, NMR analysis of the de-O-acetylated (de-OAc) CPS and dephosphorylated (de-P-Gro) CPS was conducted to simplify spectra and confirm the position of substitutions.

The 16F CPS NMR spectra gave broad lines because of short transverse relaxation times (T_2) due to slow molecular tumbling expected from high-molecular-weight polysaccharides, but this could also be partly due to the heterogeneous nature of the sample. Applying higher temperature (333 K) and slightly depolymerizing CPS by sonication can increase the molecular tumbling rate in order to improve line shapes. Moreover, removal of OAc groups and glycerol-1-phosphate substitution (P-1-Gro) substituents also led to narrower lines, as the sample became more homogeneous. There was a relatively strong phosphocholine signal at 3.2/54.5 ppm and other small signals from cell wall polysaccharides (CWPS) (13) in both 16F and 16A CPS NMR spectra.

NMR assignments for serotype 16F CPS repeat unit. According to 1D ^1H and HSQC spectra of 16F CPS, there were four different anomeric signals corresponding to four sugar residues, labeled A, B, C, and D in order of descending ^1H chemical shift (Fig. 1). Furthermore, there also was a downfield signal at 5.634 ppm corresponding to an OAc substitution as well as three methyl signals, two of which corresponded to the 6 position of 6-deoxy sugars as well as one corresponding to an OAc group. Residues A and B were both identified as rhamnopyranosides, as they had methyl groups at their 6 position and $^3J_{\text{H}, \text{H}}$ coupling patterns (see Table S1 in the supplemental material)

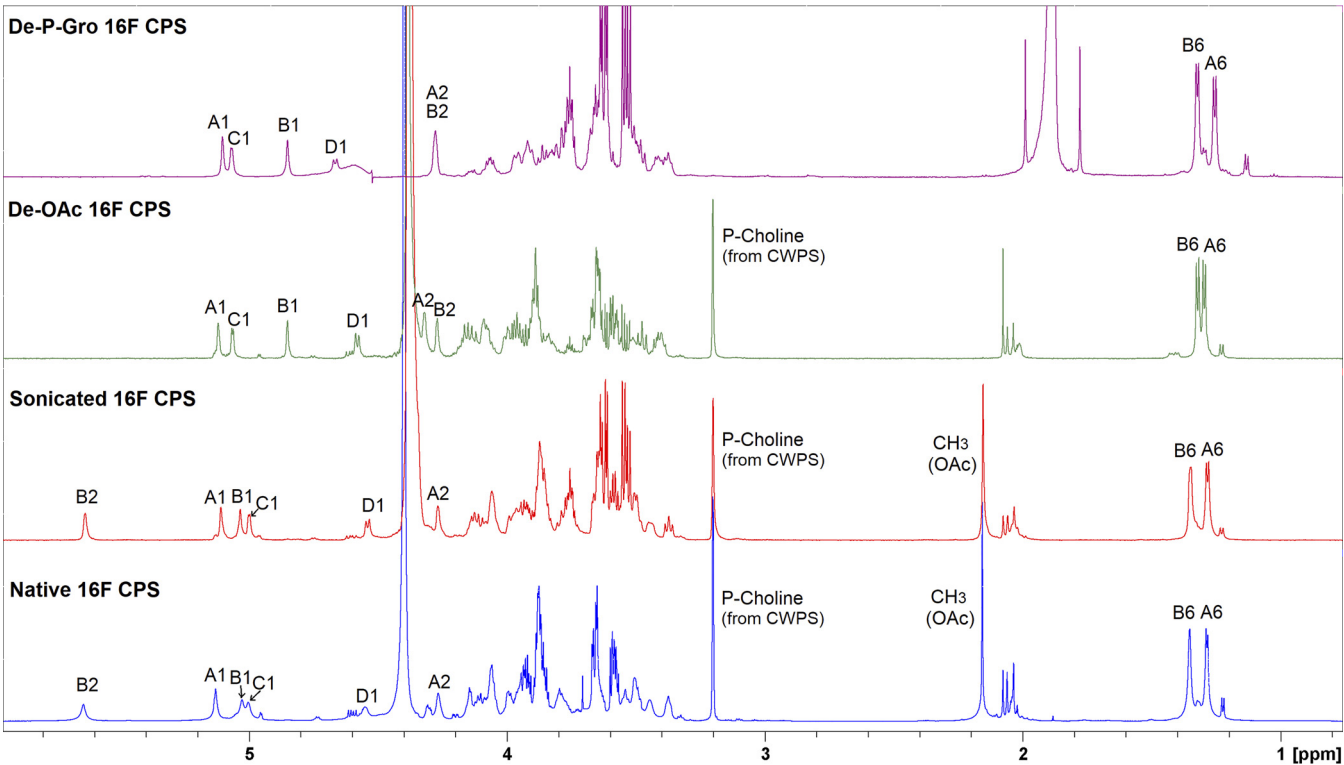


FIG 1 Expansion of ¹H spectra of native 16F CPS, ultrasonicated 16F CPS, de-OAc 16F CPS, and de-P-Gro 16F CPS.

corresponding to 6-deoxy mannopyranoside configurations. Residue C was determined to be an α -glucopyranose according to $^3J_{H,H}$ coupling pattern. D was determined to be a β -glucopyranose, as it only showed large $^3J_{H,H}$ coupling constants. The $^3J_{H3,H4}$ constants were difficult to measure because of the overlap in both ¹H and DQF-COSY spectra. Based on $^1J_{H1,C1}$ measured from CLIP-HSQC, A and C, with $^1J_{H1,C1}$ of about 170 Hz, were identified as being in α configuration, while B and D, with $^1J_{H1,C1}$ of about 160 Hz, were in β configuration (12). The monosaccharide composition analysis showed that 16F CPS were composed of rhamnose and glucose at a 1:1 ratio (Table 1), which was consistent with the results of the NMR analysis.

After the monosaccharide units were identified, HMBC and NOESY were used to find the linkage of these monosaccharides (Table 2 and Fig. S1). The anomeric position of A had HMBC and NOESY correlations with C3, while the anomeric position of C had correlations with B3. Similarly, there were HMBC and NOESY cross peaks between B1 and D4 and between D1 and A3. Therefore, these four sugar residues make up the backbone of the repeat unit in the order 3)-A-(1-3)-C-(1-3)-B-(1-4)-D-(1-. Furthermore, according to ³¹P and ¹H-³¹P HMBC spectra, there were two phosphate groups attached to the backbone. One connected to the 4 position of the α -rhamnose A and another attached to the 6 position of the α -glucose C. The phosphorylated positions were confirmed by NMR analysis of de-P-Gro 16F CPS, as the chemical shifts of A4 (4.110/75.72 ppm) and C6 (4.058/64.70 ppm) in sonicated CPS moved upfield to 3.604/71.38 ppm and 3.772/3.779/60.78 ppm in de-P-Gro 16F CPS, respec-

TABLE 1 Results of monosaccharide compositional analysis and phosphate, methylation, and absolute configuration analysis

| Serotype | Amt [nmol/mg (ratio)] of: | | | | Linkage types | L and D configurations |
|----------|---------------------------|-------------|-------------|-----------------|--|------------------------|
| | Rha | Glc | Gal | PO ₄ | | |
| 16F CPS | 1,545 (1.2) | 1,276 (1.0) | | 1,770 (1.4) | 2-Linked Rha, 4-linked Glc, 3-linked Glc | L-Rha, D-Glc |
| 16A CPS | 1,455 (2.1) | 686 (1.0) | 2,005 (2.9) | 942 (1.4) | 1,2-Linked Rha, 1,3-linked Rha, 4-linked Glc, 3-linked Glc, 1,3-linked Gal, 1,2,3-linked Gal | L-Rha, D-Glc, D-Gal |

TABLE 2 ^1H and ^{13}C NMR chemical shift assignment of sonicated 16F CPS, de-OAc 16F CPS, and de-P-Gro 16F CPS repeat units as well as important HMBC/NOE correlations

| | Value(s) (δ , ppm) for position: | | | | | | | |
|--------------------------|--|----------------|----------------------------|----------------|----------------|----------------------|----------------------------------|----------|
| Sample and residue | 1 | 2 | 3 | 4 | 5 | 6 | Substituents | HMBC/NOE |
| Sonicated 16F CPS | | | | | | | | |
| 3)- α -Rhap-(1-A) | 5.107 100.69 | 4.268 70.56 | 3.964 80.99 | 4.110 75.72 | 4.128 68.33 | 1.286 17.34 | | A1-C3 |
| 3)- β -Rhap-(1-B) | 5.037 99.63 | 5.634 68.92 | 3.866 75.25 | 3.504 71.15 | 3.506 72.67 | 1.350 17.22 | | B1-D4 |
| | | | | | | | OAc, 2.154/20.74; C=O, 174.07 | B2 |
| 3)- α -Glc-(1-C) | 5.000 95.18 | 3.657 71.91 | 3.788 79.94 | 3.543 67.98 | 3.985 71.15 | 4.058 64.70 | | C1-B3 |
| 4)- β -Glc-(1-D) | 4.536 104.56 | 3.373 73.61 | 3.580 75.48 | 3.654 77.68 | 3.447 74.66 | 3.755/3.860 61.13 | | D1-A3 |
| P ₁ -1-Gro | 3.921/3.855 66.92 | 3.874 71.26 | 3.666/3.646/3.588 62.70 | | | | | |
| P ₁ | | | | | | | 0.6031 | A4 |
| P ₂ -1-Gro | 3.874 67.04 | 3.874 71.26 | 3.666/3.646/3.588 62.70 | | | | | |
| P ₂ | | | | | | | 1.2222 | C6 |
| De-OAc 16F | | | | | | | | |
| 3)- α -Rhap-(1-A) | 5.116 100.81 | 4.320 70.44 | 4.010 80.88 | 4.150 76.19 | 4.164 68.33 | 1.292 17.22 | | A1-C3 |
| 3)- β -Rhap-(1-B) | 4.850 100.92 | 4.269 67.75 | 3.670 78.30 | 3.473 70.79 | 3.417 72.67 | 1.324 17.22 | | B1-D4 |
| 3)- α -Glc-(1-C) | 5.063 96.00 | 3.700 72.20 | 3.869 80.06 | 3.575 67.98 | 4.088 71.26 | 4.117/4.094 64.93 | | C1-B3 |
| 4)- β -Glc-(1-D) | 4.583 104.56 | 3.397 73.73 | 3.653 75.60 | 3.653 77.24 | 3.508 74.90 | 3.836/3.887 61.30 | | D1-A3 |
| P ₁ -1-Gro | 3.959/3.910 67.28 | 3.889 71.15 | 3.666/3.598 62.71 | | | | | |
| P ₁ | | | | | | | 0.2338 | A4 |
| P ₂ -1-Gro | 3.924 67.39 | 3.889 71.15 | 3.647/3.590 62.71 | | | | | |
| P ₂ | | | | | | | 0.8944 | C6 |
| De-P-Gro 16F | | | | | | | | |
| 3)- α -Rhap-(1-A) | 5.104 101.16 | 4.281 70.44 | 3.920 80.64 | 3.604 71.38 | 4.064 69.15 | 1.251 16.99 | | A1-C3 |
| 3)- β -Rhap-(1-B) | 4.850 100.92 | 4.275 67.63 | 3.659 78.06 | 3.465 70.68 | 3.411 72.67 | 1.32 17.23 | | B1-D4 |
| 3)- α -Glc-(1-C) | 5.067 95.88 | 3.662 72.2 | 3.850 80.17 | 3.508 68.33 | 3.955 72.32 | 3.772/3.790 60.78 | | C1-B3 |
| 4)- β -Glc-(1-D) | 4.659 104.09 | 3.374 73.84 | 3.637 75.84 | 3.637 77.12 | 3.508 74.90 | 3.811/3.903 61.18 | | D1-A3 |

tively. The β -rhamnose B was 2-OAc substituted according to HMBC correlations between carbonyl of the OAc group and B2 proton, which could explain an extremely downfield proton chemical shift of B2 at 5.634 ppm. Moreover, in the NMR spectra of de-OAc 16F CPS, the signals from the acetyl group disappeared and the B2 proton signal moved upfield from 5.634 ppm to 4.269 ppm. Furthermore, there were three additional sharp and intense signals, two CH_2 groups and one CH group, in sonicated CPS and de-P-Gro CPS, which arose from free glycerol. These signals had chemical shifts and line shapes almost identical to those detected from a reference sample of glycerol in D_2O (Fig. S2) and did not show any correlations to other units. The free glycerol could be generated during strong sonication. Therefore, one needs to pay extra attention when using strong sonication to depolymerize large polysaccharides.

Based on the above-described analysis, the repeat unit of serotype 16F CPS was determined to be a tetrasaccharide: -3)- α -L-Rhap-[4-P-1-Gro]-(1-3)- α -D-Glc-[6-P-1-Gro]-(1-3)- β -L-Rhap-[2-OAc]-(1-4)- β -D-Glc-(1- (Fig. 2). Full assignments of the 16F CPS, de-OAc 16F CPS, and de-P-Gro 16F CPS repeat units are given in Table 2.

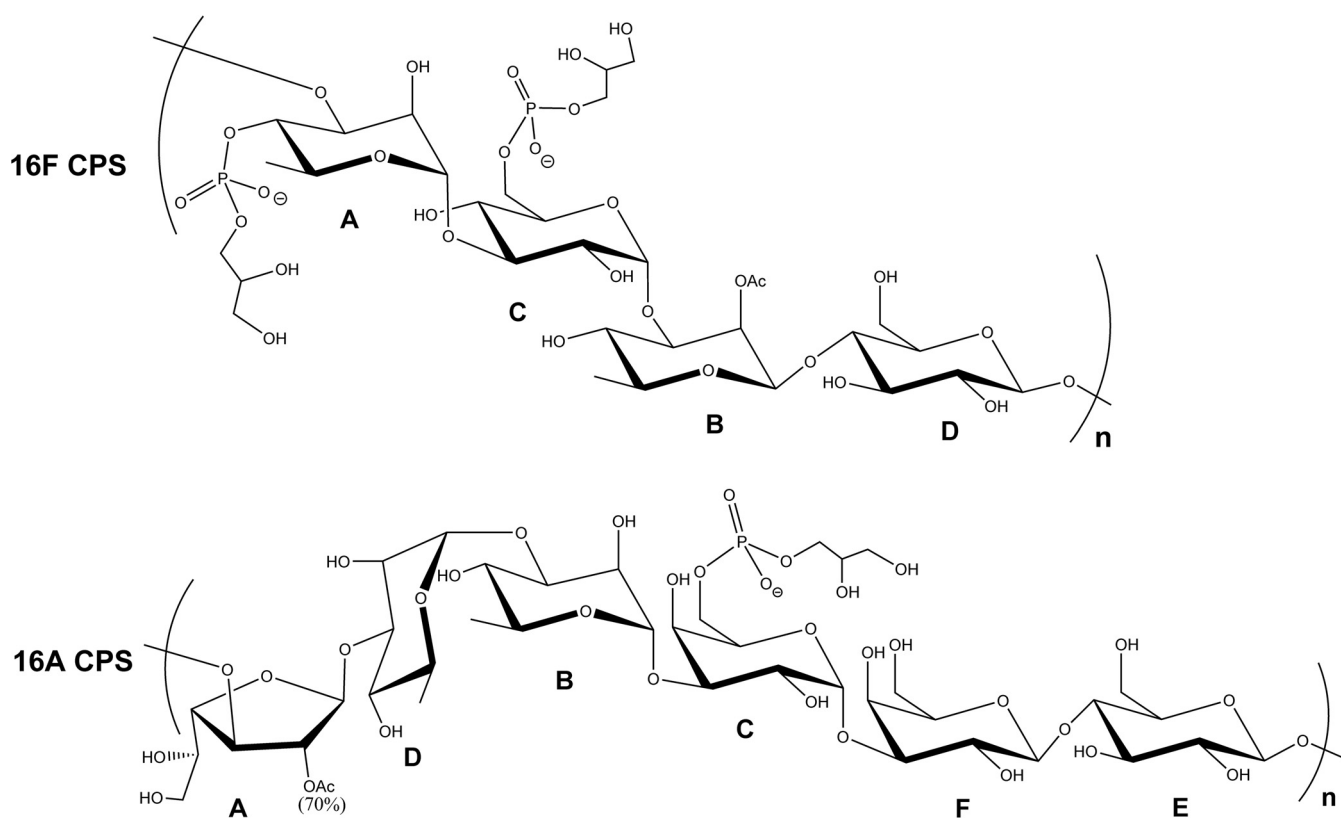


FIG 2 Repeat unit structures of 16F and 16A CPS.

NMR assignments for serotype 16A CPS repeat unit. 1D ^1H and HSQC spectra of 16A CPS contained eight different anomeric signals. However, the intensity of these anomeric signals varied, and this turned out to be due to partial 2-OAc substitution at a single position (Fig. 3). Moreover, NMR analysis of de-O-acetylated 16A CPS showed only six anomeric signals, which indicated that 2-OAc substitution affected nearby anomeric signals. The monosaccharide units were labeled A, B, C, D, E, and F in order of descending ^1H chemical shift. As A was only partially 2-OAc substituted, A without 2-OAc substitution was labeled A'. Similarly, the anomeric position of E was affected by the acetylation of A, and as such the corresponding monosaccharide unit in the approximately 30% nonacetylated repeat unit was labeled E'.

Furthermore, three methyl groups were observed, two of which had almost identical chemical shift corresponding to the methyl of 6-deoxy sugars, and the third one was the methyl of an acetyl group.

A was identified as a galactose, approximately 70% of which was 2-OAc substituted according to integration of ^1H NMR anomeric signals. The $^3J_{\text{H}, \text{H}}$ coupling constants of A did not show a regular pattern, unlike pyranose sugar, and based on the chemical shifts, A was identified as a β -galactofuranose (14, 15). This 2-OAc substitution of A would explain the downfield-shifted proton chemical shift of A2 at 5.155 ppm. The de-OAc 16A CPS spectra became far simpler than those of native 16A CPS, as signals caused by the partial O-acetylation disappeared. The anomeric region of de-OAc 16A CPS showed only six signals with almost identical integrals in the ^1H spectrum. The chemical shift of A2 moved upfield from 5.155/82.45 ppm to 4.342/80.45 ppm (Table 3). B and D were both identified as rhamnopyranosides because of the methyl groups at their 6 positions and the $^3J_{\text{H}, \text{H}}$ coupling constants that corresponded to 6-deoxy mannopyranose configurations. C and F were identified as α - and β -galactopyranosides, respectively, according to their $^3J_{\text{H}, \text{H}}$ coupling constants (Table S1) and chemical shifts. The 6 position of C was connected to a glycerol-1-phosphate substi-

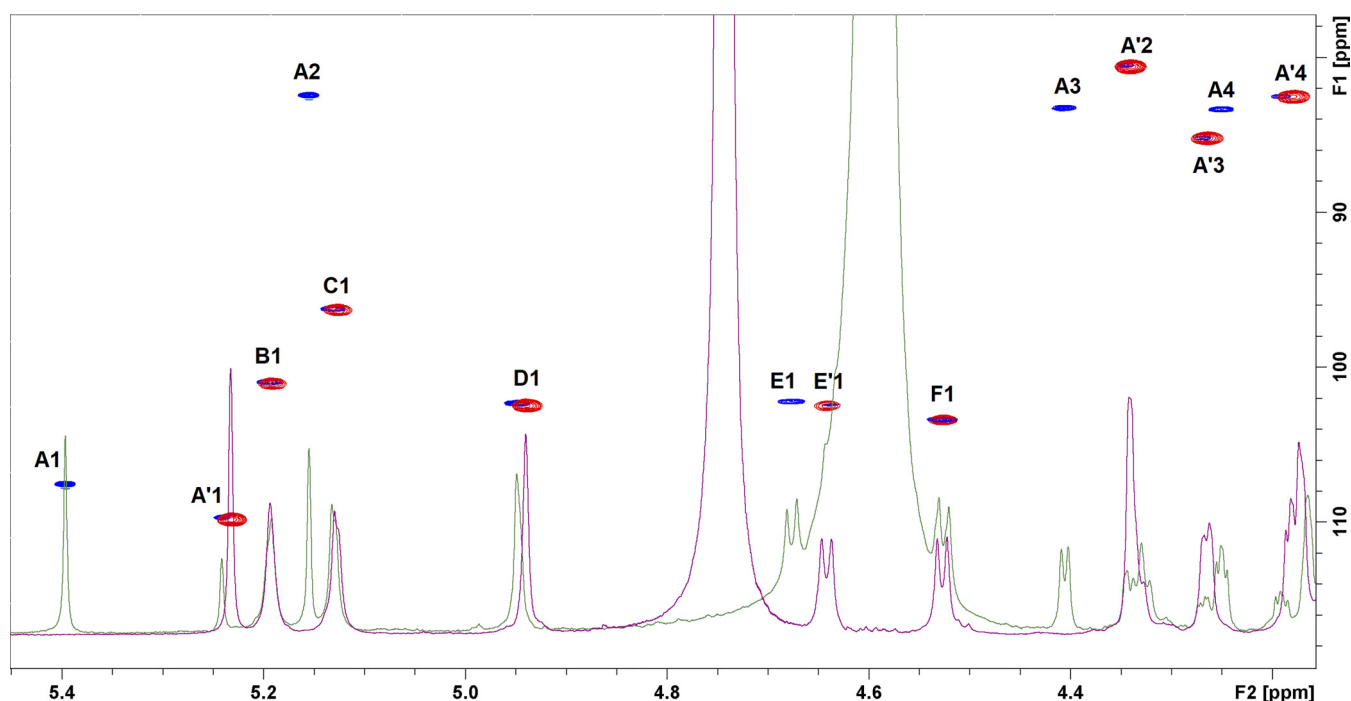


FIG 3 Expansion of ^1H and HSQC spectrum of native 16A CPS (green and blue) at 313 K and de-OAc 16A CPS (purple and red) at 298 K.

tution according to ^1H - ^{31}P HMBC, which can explain the relatively downfield chemical shift of 3.960/64.46 ppm at C6 versus 3.704/61.24 ppm in de-P-Gro 16A CPS. E was determined to be a β -glucopyranose, as it only had large $^3J_{\text{H}_1, \text{H}}$ coupling constants. Based on the $^1J_{\text{H}_1, \text{C}_1}$ coupling constants of approximately 170 Hz, measured from CLIP-HSQC, the two rhamnosides B and D were identified as having α configuration. Additionally, CLIP-HSQC was also used to confirm the anomeric configurations of C, E, and F with $^1J_{\text{H}_1, \text{C}_1}$ coupling constants of approximately 170, 160, and 160 Hz, respectively. The anomeric configurations are in agreement with $^3J_{\text{H}_1, \text{H}_2}$ coupling constants and the chemical shifts in general. NMR analysis of the monosaccharide repeat units were in agreement with the results of the monosaccharide composition analysis: rhamnose-galactose-glucose in an approximately 2:3:1 ratio (Table 1). Furthermore, in the HMBC and NOESY spectra (Fig. S3), cross peaks of A1 with D3, D1 with B2, B1 with C3, C1 with F3, F1 with E4, and E1 with A3 revealed linkages of these monosaccharide units. Thus, these six sugars constitute a linear backbone of the 16A repeat unit in the sequence 3)-A-(1-3)-D-(1-2)-B-(1-3)-C-(1-3)-F-(1-4)-E-(1-. Thus, the repeat units of serotype 16A CPS were determined to be a hexasaccharide with the following structure: -3)- β -D-Galp-[2-OAc (70%)]-(1-3)- α -L-Rhap-(1-2)- α -L-Rhap-(1-3)- α -D-Galp-[(6-P-1)-Gro]-(1-3)- β -D-Galp-(1-4)- β -D-Glcp-(1- (Fig. 2). The full assignment of the 16A CPS, de-OAc, and de-P-Gro 16A CPS repeat unit is given in Table 3.

The ^{13}C chemical shift assignments of 16F and 16A CPS also supported glycosidic linkage sites determined by HMBC/NOESY, as all substituted positions had downfield chemical shifts from the same unsubstituted position (16). The assigned chemical shifts of de-P-Gro 16F and 16A CPS are in good agreement with those calculated by CASPER (17).

Genetic elucidation of serotype 16F and 16A CPS biosynthesis. Biosynthesis of pneumococcal serotype 16F and 16A CPS uses the Wzy-dependent mechanism. A summary of 16F and 16A CPS biosynthetic genes, their functions, and related serotypes is shown in Table 4, which is based on published genetic sequence data, bioinformatics analysis, and CPS structures (1, 3, 18, 19). The *cps* loci of 16F and 16A did not show high similarity (Fig. 4). They have similar conserved regulatory genes *cpsABCD* (*wzg*, *wzh*, *wzd*, and *wze*), the initial glucosyl-1-phosphate transferase gene (*wchA*) (20), and the

TABLE 3 ^1H and ^{13}C NMR chemical shift assignment of native 16A CPS, de-OAc 16A, and de-P-Gro CPS repeat units as well as important HMBC/NOE correlations

| | Value(s) (δ , ppm) for position: | | | | | | | |
|---|--|----------------|----------------------|----------------|----------------|----------------------|----------------------------------|----------|
| Sample and residue | 1 | 2 | 3 | 4 | 5 | 6 | Substituents | HMBC/NOE |
| Native 16A CPS | | | | | | | | |
| 3)- β -Gal f -2-OAc-(1-A (70%)) | 5.395 107.55 | 5.155 82.45 | 4.407 83.27 | 4.250 83.27 | 3.966 70.6 | 3.706/3.674 63.27 | OAc, 2.119/20.75; C=O, 173.45 | A1-D3 |
| 3)- β -Gal f -(1-A' (30%)) | 5.2406 109.66 | 4.345 80.45 | 4.267 85.14 | 4.191 82.45 | 3.952 70.83 | 3.706/3.674 63.27 | | A'1-D'3 |
| 2)- α -Rhap-(1-B | 5.192 100.98 | 4.057 78.46 | 3.949 70.48 | 3.489 72.71 | 3.805 69.78 | 1.29 17.23 | | B1-C3 |
| 3)- α -Gal p -(1-C | 5.131 96.29 | 3.981 68.37 | 3.993 76.70 | 4.113 68.96 | 4.328 70.01 | 3.960 64.44 | | C1-F3 |
| 3)- α -Rhap-(1-D | 4.947 102.27 | 4.164 70.48 | 3.895 77.87 | 3.533 71.54 | 3.753 69.78 | 1.254 17.23 | | D1-B2 |
| D'3 | | | 3.864/78.11 | | | | | |
| 4)- β -Glc p -(1-E | 4.673 102.15 | 3.340 73.18 | 3.668 74.70 | 3.665 79.04 | 3.58 75.29 | 3.962/3.830 60.45 | | E1-A3 |
| E' | 4.637/102.39 | | | | | | | |
| 3)- β -Gal p -(1-F | 4.524 103.33 | 3.653 70.01 | 3.771 78.22 | 4.147 65.56 | 3.714 75.53 | 3.744/3.775 61.39 | | F1-E4 |
| Gro-1-P | 3.901/3.845 66.79 | 3.881 71.19 | 3.647/3.590 62.57 | | | | 0.9144 | |
| P | | | | | | | | C6 |
| De-OAc 16A CPS | | | | | | | | |
| 3)- β -Gal f -(1-A | 5.230 109.79 | 4.342 80.54 | 4.264 85.20 | 4.178 82.46 | 3.943 70.73 | 3.708/3.656 63.34 | | A1-D3 |
| 2)- α -Rhap-(1-B | 5.191 101.11 | 4.057 78.61 | 3.943 70.73 | 3.493 72.66 | 3.802 69.93 | 1.287 17.21 | | B1-C3 |
| 3)- α -Gal p -(1-C | 5.124 96.29 | 3.967 68.32 | 3.998 76.84 | 4.111 68.96 | 4.334 70.09 | 3.951 64.46 | | C1-F3 |
| 3)- α -Rhap-(1-D | 4.941 102.39 | 4.170 70.57 | 3.857 78.13 | 3.513 71.54 | 3.752 69.77 | 1.252 17.21 | | D1-B2 |
| 4)- β -Glc p -(1-E | 4.639 102.39 | 3.329 73.14 | 3.665 74.75 | 3.669 78.93 | 3.611 75.39 | 3.963 60.43 | | E1-A3 |
| 3)- β -Gal p -(1-F | 4.526 103.36 | 3.642 70.09 | 3.767 78.13 | 4.147 65.59 | 3.712 75.55 | 3.744/3.775 61.41 | | F1-E4 |
| Gro-1-P | 3.650/3.583 62.54 | 3.877 71.21 | 3.896/3.842 66.88 | | | | 0.7042 | |
| P | | | | | | | | |
| De-P-Gro 16A CPS | | | | | | | | |
| 3)- β -Gal f -(1-A | 5.230 109.79 | 4.342 80.47 | 4.269 85.05 | 4.178 82.41 | 3.936 70.86 | 3.702/3.662 63.29 | | A1-D3 |
| 2)- α -Rhap-(1-B | 5.184 101.05 | 4.059 78.42 | 3.936 70.39 | 3.494 72.62 | 3.790 69.80 | 1.281 17.15 | | B1-C3 |
| 3)- α -Gal p -(1-C | 5.137 95.89 | 3.976 68.34 | 3.973 77.01 | 4.059 69.39 | 4.188 71.33 | 3.704 61.24 | | C1-F3 |
| 3)- α -Rhap-(1-D | 4.941 102.34 | 4.169 70.45 | 3.851 78.13 | 3.513 71.50 | 3.748 69.80 | 1.251 17.26 | | D1-B2 |
| 4)- β -Glc p -(1-E | 4.643 102.28 | 3.330 73.14 | 3.665 74.67 | 3.660 78.89 | 3.614 75.31 | 3.807/3.797 60.48 | | E1-A3 |
| 3)- β -Gal p -(1-F | 4.504 103.34 | 3.662 69.98 | 3.775 77.66 | 4.161 65.29 | 3.699 75.55 | 3.780/3.733 61.47 | | F1-E4 |

D-TDP L-rhamnose pathway genes (*rmIA*, *rmIC*, *rmIB*, and *rmID*). Genes for biosynthesis of glycerol-1-phosphate (*gct*) were present in both 16F and 16A with 77% identities in BLASTp alignment (21). The presence of these similar genes indicated that 16F and 16A CPS both contain glucose, rhamnose, and glycerol-1-phosphate in their repeat unit structures. However, genes for synthesis of galactofuranose (*glf*) were present in serotype 16A but not in 16F. Moreover, they have significantly different glycosyltransferase (GT) genes and oligosaccharide repeat unit polymerases (Wzy). Wzy27 in 16F and

TABLE 4 Serotype 16F and 16A CPS biosynthetic genes, their predicted functions or products, structures they catalyzed, and related serotypes containing the same genes

| Gene ^a | No. of occurrences of gene | Structure | | Serotype(s) containing the gene ^d | | |
|--------------------------|----------------------------|---|--|--|------------------|--|
| | | Product or function | Donor | Linkage | Acceptor | CPS structure known |
| <i>wchF</i> | 27 | GT | L-Rhap | β 1-4 | β -D-Glcp | 2, 7F, 7A, 7B, 17F, 18F, 18A, 18B, 18C, 22F, 23F, 27, 32F, 32A, 16F |
| <i>wciU</i> | 7 | GT | D-GlcpNAC | α 1-3 | β -L-Rhap | 18A |
| <i>wcxN</i> | 3 | GT | D-Glcp | α 1-3 | β -L-Rhap | 18F, 18B, 18C, 16F |
| <i>wcxM</i> | 4 | GT | L-Rhap | α 1-3 | α -D-Glcp | 16F |
| <i>wzy27^b</i> | 1 | Acetyl transferase | β -L-Rhap-2-OAc | | | 16F |
| <i>gct</i> | 13 | Polymerase | β -D-Glcp-(1,3)- α -L-Rhap | | | 16F |
| | | Glycerol-1-phosphate biosynthetic protein | P-1-Gro | | | 11A, 11C, 11D, 11E, 18F, 18A, 18B, 18C, 45, 16F , 16A |
| <i>wcxP</i> | 3 | Sugar/polyalcohol phosphate transferase | Rib-ol-1-P | | | 11B, 11F (truncated <i>gct</i>) |
| <i>wcxQ</i> | 3 | Sugar/polyalcohol phosphate transferase | D-Glcp-6-P-1-Gro/ L-Rhap-4-P-1-Gro ^c | | | 16F |
| <i>wchK</i> | 13 | GT | D-Galp | β 1-4 | β -D-Glcp | 11F, 11A, 11B, 11C, 11D, 11E, 13, 14, 15F, 15A, 15B, 15C, 16A |
| <i>wchJ</i> | 13 | GT enhancer | D-Galp | α 1-3 | β -D-Galp | 11F, 11A, 11B, 11C, 11D, 11E, 16A |
| <i>wcyK</i> | 7 | GT | L-Rhap | α 1-3 | α -D-Galp | 45, 16A |
| <i>wcxS</i> | 2 | GT | L-Rhap | α 1-2 | α -L-Rhap | 16A |
| <i>wcxT</i> | 1 | Hypothetical protein | D-Galp | β 1-3 | β -D-Galp | 10F, 10A, 10B, 10C, 29, 39, 47F, 47A |
| <i>wciB</i> | 25 | GT | | | β -D-Glcp | 17A, 20A, 20B, 33F, 33A, 35F, 35A, 35B, 35C, 41F, 41A, 42 |
| | | | | | α -D-Glcp | 34 |
| | | | | | α -L-Rhap | 16A |
| <i>wzy15^b</i> | | Polymerase | β -D-Glcp-(1,3)- β -D-Galp | | | 37 (synthase pathway) |
| <i>wciG</i> | 16 | Acetyl transferase | β -D-Galp-2-OAc | | | 16A |
| | | | β -D-Galp-5,6-OAc | | | 20A, 20B, 33F, 33A, 33B, 33D, 35F, 35A, 35B, 47F, 16A |
| | | | β -D-Glcp-2,3-OAc | | | 13 |
| | | | β -D-Galp-5,6-OAc | | | 42 (containing <i>wciG</i> and <i>wciE</i>) |
| | | | NO OAc | | | 35C (containing <i>wciG</i> and <i>wciE</i>), 10F, 10C |
| <i>wcxR</i> | 1 | Sugar/polyalcohol phosphate transferase | Gro | 1-P-4 | α -D-Galp | 16A |

^aRegulatory genes (*wzg*, *wzh*, *wzd*, and *wze*) and initial transferase (*WchA*), L-rhamnose, and galactofuranose synthetic genes are not included.^b*wzy27* and *wzy15* belong to Wzy homology groups 27 and 15, respectively.^cThe specific functions of two sugar/polyalcohol phosphate transferases (*WcxP* and *WcxQ*) cannot be distinguished.^dBoldface highlights serotype 16F and 16A.

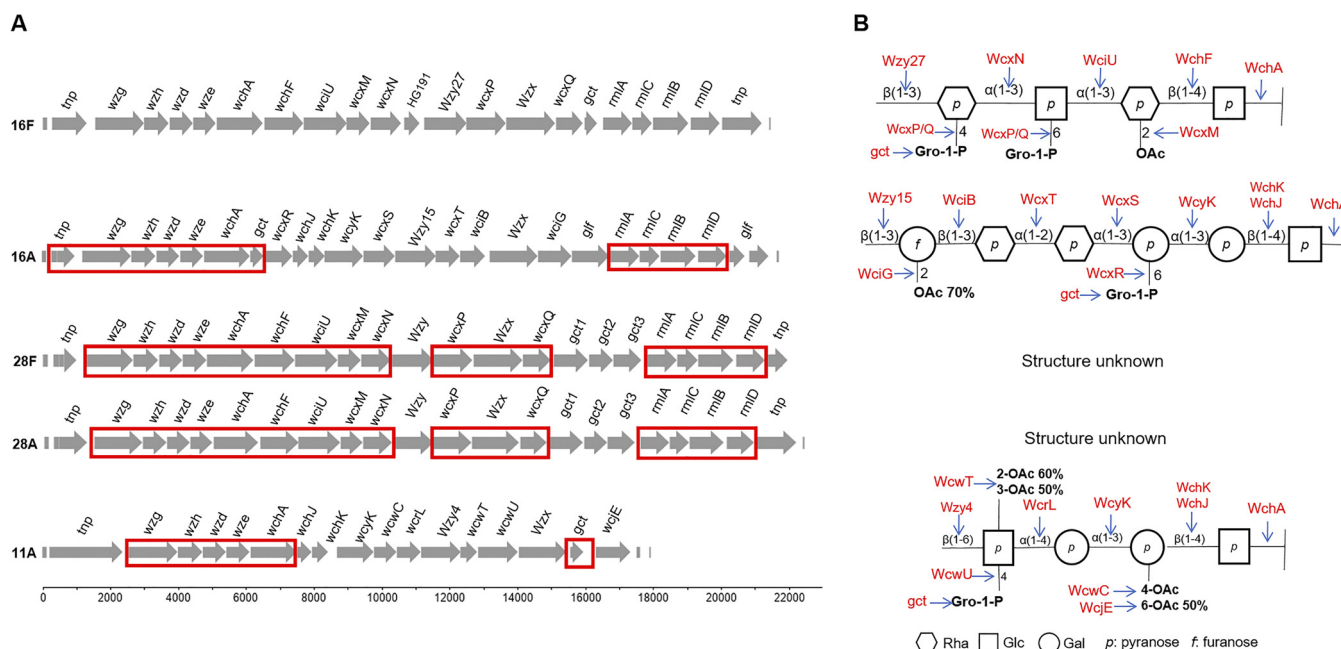


FIG 4 Comparison of serogroup 16 CPS biosynthetic gene loci and repeat unit structures with other serotypes. (A) *cps* gene loci of *S. pneumoniae* serogroup 16 and related serotypes. Red frames highlight genes highly similar to the 16F *cps* gene locus. (B) CPS repeat unit structures and biosynthetic enzymes (red) that are responsible for corresponding structures, indicated by arrows.

Wzy15 in 16A showed only 28% identities in BLASTp comparison. According to 16F and 16A CPS structures, Wzy27 is assigned to polymerize at β -D-Glcp-(1-3)- α -L-Rhap, while Wzy15 is assigned to polymerize at β -D-Glcp-(1-3)- β -D-Galf.

The *cps* locus of 16F contained three GT genes. The GT WchF gene has been shown to encode the β -1-4 rhamnosyltransferase for addition of the second sugar (22). The GT gene (*wciU*) in the 16F *cps* locus, also present in serogroup 18, added α -D-Glcp or α -D-GlcNAc to the 3 position of β -L-Rhap (19). Therefore, the only unassigned GT (WcxN) should be responsible for addition of α -L-Rhap to the 3 position of α -D-Glcp, based on the 16F CPS structure we determined in this paper. The putative glycerol phosphotransferase (WcxP) and putative LicD family phosphotransferase (WcxQ) were responsible for addition of Gro-1-P substitution to the 4 position and 6 position of α -D-Glcp and α -L-Rhap, respectively, but it was hard to distinguish which specific catalyzed linkage either one of them was responsible for. WcxM, a putative acetyltransferase, catalyzed O-acetylation at the 2 position of β -L-Rhap, which was consistent with serotype 18F (1, 23).

The serotype 16A *cps* locus had five GT genes. The GT (WchK) that is found in 13 different serotypes was assigned as β -1-4-galactosyltransferase, and WchJ was suggested to enhance activity of WchK (18, 24). WcyK, also found in serogroup 11 (25), catalyzed linkage of α -D-Galp-(1-3)- β -D-Galp. WcxS catalyzed the linkage of α -L-Rhap-(1-3)- α -D-Galp, which was shown for serotype 45 (19). WciB, present in 25 pneumococcal serotypes (Table 4), was assigned as β -1-3-galactofuranosyl transferase (26, 27) (Table 4). Thus, the only unassigned GT (WcxT) was assigned, by process of elimination, as being responsible for catalyzing the linkage of α -L-Rhap-(1-2)- α -L-Rhap. WciG, a putative acetyltransferase, catalyzed 2-O-acetylation at β -D-Galf (26–29). WcxR is a putative LicD family phosphotransferase that was assigned to catalyze phosphorylation at the 6 position of α -D-Galp. The analysis of the biosynthesis of 16F and 16A CPS was in an agreement with determined CPS repeat unit structure and functions of related genes reported in the literature.

Compared to other serogroups, the 16F *cps* locus showed high similarity to serotypes 28A and 28F, and it was in the same subcluster of cluster 2 (18, 23) (Fig. 4). In addition to containing the same conserved regulatory genes, they have highly similar

TABLE 5 Cross-reactions between serogroup 16 and serogroup 28 using Quellung reaction

| Serotype | Reaction to factor serum: | | | | | | | | | | |
|----------|---------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | 16b | 16c | 28b | 28c | 11b | 11c | 11f | 11g | 23b | 23c | 23d |
| 16F | +++ | — | +++ | — | — | +++ | — | — | — | — | — |
| 16A | — | +++ | — | — | — | +++ | — | — | — | — | — |
| 28F | — | — | +++ | — | — | — | — | — | — | — | +++ |
| 28A | — | — | — | +++ | — | — | — | — | — | — | +++ |

WchA, GTs, acetyltransferase (WcxM), and flippase (Wzx). Main differences in 28F and 28A *cps* loci were glycerol-2-phosphate biosynthetic genes (*gtp1*, *gtp2*, and *gtp3*) compared to the glycerol-1-phosphate gene (*gct*) in 16F (Fig. 4). Thus, the repeat unit structure of 16F CPS should have similarities to the 28F and 28A CPS structures, which unfortunately are both unknown.

Serological cross-reactions of diagnostic antisera between serogroup 16 and other serotypes. When serotyping *S. pneumoniae*, serogroup 16 reacts with pool antiserum D and was identified by group antiserum 16. Furthermore, factor antisera 16b and 16c were used to determine serotypes 16F and 16A (Table 5). Positive Quellung reactions were observed between serotypes 16F and 16A and factor antiserum 11c. Factor antiserum 11c reacted with serotypes 11A, 11C, and 11D and is reported to bind to P-Gro and OAc groups (25). Moreover, the cross-reaction between serotype 16F and factor serum 11d was previously observed (18). These cross-reactions indicated similar antigenic properties among them. However, the *cps* gene locus of serotype 11A does not show high similarity to the 16F *cps* locus apart from regulatory genes *wchA* and *gct* (Fig. 4), while 16A shared some similar GTs (WchJ, WchK, and WcyK) with serogroup 11. The common parts of the 16F CPS repeat unit structure, compared with that of 11A, are that they consist of a linear tetrasaccharide backbone with glycerol-1-phosphate branches and O-acetyl substitution. O-acetylation was reported to be crucial to the antigenicity of serogroup 11 (25, 30). The cross-reactions between group 16 and factor antiserum 11c confirmed that the factor 11c epitope was associated with P-1-Gro and OAc groups. Serotype 16F cross-reacted with factor antiserum 28b, which was used for serotyping 28F. Thus, it can be predicted that the CPS structure of 28F is similar to that of 16F, as they have highly similar *cps* genes and common serological activity. No cross-reaction was observed between factor antiserum 11c and serotype 28F, which has genes to express P-2-Gro synthase, indicating that factor antiserum 11c reacts specifically to P-1-Gro and not P-2-Gro. Additionally, serotypes 28F and 28A reacted with factor antiserum 23d, which is used for serotyping 23B. CPS structures of group 23 all contain P-2-Gro (31). However, factor antiserum 23d only reacts with serotype 23B of serogroup 23. This suggests that there are similar structural properties other than that for P-2-Gro between serogroup 28 and serotype 23B. However, these assumptions about serogroup 28 need further structural information to be confirmed.

Summary of results. The CPS repeat unit structures of serotypes 16F and 16A were elucidated using NMR spectroscopy and chemical analysis (Fig. 2), which provide the first comprehensive understanding of serogroup 16 CPS. Furthermore, based on the determined CPS structures, a comparison analysis of CPS biosynthesis of serotype 16F and 16A with related serotypes that share the same biosynthetic genes revealed functions of previously unknown glycosyltransferases WcxN and WcxT (Table 4). Finally, the serogroup 16 CPS were subjected to serological studies to determine any relevant cross-reactions with antisera from the genetically similar serogroup 28, where it was shown that the CPS of serotypes 16F and 28F have similar antigenic properties, as they both react with factor serum 28b, which is used for 28F serotyping. Thus, it could be predicted that the CPS structure of serotype 28F is similar to that of serotype 16F, since they have highly similar *cps* genes and serological behavior.

MATERIALS AND METHODS

Sample preparation. Purified pneumococcal capsular polysaccharides (CPS) from serotypes 16A and 16F, as well as rabbit antisera, were produced by SSI Diagnostica, Hillerød, Denmark. The polysaccharide samples were dissolved in D₂O (99.9%; Sigma) to concentrations of approximately 1.5%, wt/vol (10 mg in 600 μ l).

Ultrasonicated 16F CPS were obtained by sonication of the CPS (10 mg in 5 ml 2 M KCl) using a Labsonic P ultrasonic homogenizer with a 3-mm probe for 2 h under an ice bath, after which it was desalted and concentrated using centrifugal filters (Amicon Ultra-4, 3 kDa).

De-OAc and de-P-Gro procedures were performed using methods described by Richards and Perry (32). De-OAc removed only acetyl groups, while de-P-Gro removed both acetyl and glycerophosphate groups. De-OAc CPS were obtained by treating approximately 5 to 10 mg of each polysaccharide with 1 ml of a solution of 0.2 M NaOH in D₂O at room temperature for 2 h, which was subsequently neutralized with 2 M HCl, followed by desalting and removal of residuals using centrifugal filters (Amicon Ultra-4, 3 kDa). Dephosphorylated CPS were generated by treating approximately 10 mg of each CPS in 2 ml of 1 M NaOH, containing a trace amount of sodium borohydride, at 100°C for 4 h in screw-cap glass tubes filled with nitrogen. The solutions then were cooled in an ice bath and neutralized with acetic acid, followed by desalting and removal of residuals using centrifugal filters (Amicon Ultra-4, 10 kDa). Finally, these samples were transferred to 5- or 3-mm NMR tubes.

NMR spectroscopy. NMR spectroscopy was used to determine repeat unit structures of serotype 16A and 16F CPS. The NMR experiments were carried out on a Bruker Avance III (799.90 MHz for ¹H and 201.14 MHz for ¹³C) equipped with a 5-mm TCI 1H/(¹³C, ¹⁵N) cryoprobe or a Bruker AVANCE 600-MHz instrument with 5-mm SmartProbe BB(F)-H-D(¹⁵N-³¹P, ¹H, ¹⁹F).

All spectra of 16A CPS were recorded at 313 K, while 16F CPS spectra were recorded at 333 K or 323 K. The assignments of 16A and 16F CPS repeat units were performed using the following experiments: 1D ¹H and ³¹P NMR and the 2D NMR spectra, including DQF-COSY, TOCSY with 80-ms mixing time, NOESY with 200- to 400-ms mixing time, HSQC, HMBC optimized for 8-Hz long-range coupling constants, ¹H-¹³C HSQC-TOCSY with 80-ms mixing time, CLIP-HSQC, and ¹H-³¹P HMBC. Experiments were conducted with standard gradient-enhanced Bruker pulse sequences, except for those of CLIP-HSQC (11).

NMR spectra were processed with Topspin 3.5 (Bruker) using extensive zero filling in all dimensions. The one-dimensional ¹H spectra were processed with an exponential window function with a line broadening of 0.3 Hz for ¹H spectra. All 2D spectra were processed with shifted sine bell window functions in both dimensions. All spectra were referenced to residual cell wall polysaccharide phosphocholine signals (¹H, 3.2 ppm; ¹³C, 54.5 ppm; shielded ³¹P signal, 1.30 ppm) (13).

General chemical analysis. The monosaccharide composition was determined by gas chromatography (GC) after hydrolysis (33) (2 M trifluoroacetic acid [TFA] at 120°C, 2 h), addition of xylose as an internal standard, reduction (NaBH₄, 16 h in the dark), and peracetylation (acetic anhydride and pyridine [1:1], 85°C, 10 min, twice). GC analysis was performed on an HP 5890 (series II) gas chromatograph with a flame-ionization detector and a column (30 m by 0.25 mm; 0.25- μ m film thickness; Agilent) of 5% phenylmethylsiloxane (HP-5MS); helium was used as the carrier gas (70 kPa). The temperature in gas-liquid chromatography (GLC) was 150°C for 3 min and then increased at 3°C/min to 320°C. A mixture of standard monosaccharides was used for sugar identification and xylose for quantification.

Methylation was performed using Ciucanu-Kerek's method (34). Prior to this, the removal of phosphate from CPS samples was done by using 48% HF for 48 h at 4°C, followed by removal of HF by evaporation under N₂ while suspended in dimethyl sulfoxide for 16 h at 60°C and then methylated with iodomethane in the presence of NaOH. The product was hydrolyzed with 4 M TFA for 4 h at 100°C, reduced with NaBD₄, and peracetylated with acetic anhydride in pyridine. Partially methylated and acetylated alditols were analyzed by GC mass spectrometry.

The content of total phosphate was determined by colorimetric analysis according to Lowry et al. (35). The samples were subjected to digestion reagent (30.6 ml H₂SO₄, 6.7 ml 70% HClO₄, and 62.7 ml H₂O) at 100°C for 1 h and then at 165°C for 2 h. Subsequently, it was cooled down to room temperature and freshly prepared color reagent was added (1 ml 1 M sodium acetate, 1 ml 2.5% ammonium molybdate solution, 7 ml H₂O, 1 ml freshly prepared ascorbic acid; 37°C for 90 min). The extinction of the ammonium phosphomolybdate complex was measured at a wavelength of 820 nm (Helios Beta 9423 UVB 1002E spectrophotometer; Thermo Electron Ltd., Altrincham, Cheshire, UK). For quantification, 5 mM Na₂HPO₄ solution was used as an external standard.

The absolute configuration of Rha and Glc was determined by GLC (36) by comparison with authentic standards of the acetylated O-[(R)-2-Octyl] glycoside after methanolysis (0.5 M HCl-methanol [MeOH], 85°C, 3.5 h), butanolysis [2 M HCl-(R)-2-octanol, 65°C, 4 h], and peracetylation (85°C, 10 min). The absolute configuration of Gal was determined by GLC by comparison with authentic standards of the acetylated O-[(S)-2-butyl] glycoside after methanolysis (2 M HCl-MeOH, 85°C, 45 min), butanolysis [2 M HCl-(S)-2-butanol; 65°C, 4 h], and peracetylation (85°C, 10 min).

Genetic analysis of serogroup 16. The published *cps* locus sequences (GenBank accession numbers of serotypes 16F, 16A, 28F, 28A, and 11A were CR931668.1, CR931667.1, CR931693.2, CR931692.1, and CR931653.1, respectively) have been downloaded from GenBank (<https://www.ncbi.nlm.nih.gov/nucleotide>). Pairwise protein sequence identity was assessed using BLASTp. Comparison of similarities of the gene products (TBLASTX) of the *cps* biosynthetic loci was performed by the Artemis comparison tool (ACT) (37).

The Quellung reaction was performed at SSI Diagnostica using factor antisera (SSI Diagnostica A/S, Hillerød, Denmark). For the Quellung reaction, bacteria were grown on a horse blood agar plate (SSI Diagnostica A/S, Hillerød, Denmark) overnight. Two to 4 μ l of culture was added onto a glass slide and

mixed with the same amount of factor antisera. The mixture was then observed for capsule swelling reaction in a phase-contrast microscope.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/JB.00453-19>.

SUPPLEMENTAL FILE 1, PDF file, 0.8 MB.

ACKNOWLEDGMENTS

The 800-MHz NMR spectra were recorded at the NMR Center-DTU, supported by the Villum Foundation. C.L. appreciates the financial support of the China Scholarship Council (Ph.D. scholarship no. 201708310119) and Oticon Fonden for financial support during her external stay at the Research Center Borstel.

We thank Kasper Enemark-Rasmussen for excellent technical assistance with NMR instrument setup, Katharina Jakob (Research Center Borstel) for general chemical analysis, and Katrine Helander Pedersen for serological testing.

REFERENCES

- Geno KA, Gilbert GL, Song JY, Skovsted IC, Klugman KP, Jones C, Konradsen HB, Nahm MH. 2015. Pneumococcal capsules and their types: past, present, and future. *Clin Microbiol Rev* 28:871–899. <https://doi.org/10.1128/CMR.00024-15>.
- Kjeldsen C, Slott S, Elverdal PL, Sheppard CL, Kapatai G, Fry NK, Skovsted IC, Duus J. 2018. Discovery and description of a new serogroup 7 *Streptococcus pneumoniae* serotype, 7D, and structural analysis of 7C and 7D. *Carbohydr Res* 463:24–31. <https://doi.org/10.1016/j.carres.2018.04.011>.
- Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabinowitz E, Collins M, Donohoe K, Harris D, Murphy L, Quail MA, Samuel G, Skovsted IC, Kalltoft MS, Barrell B, Reeves PR, Parkhill J, Spratt BG. 2006. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genet* 2:e31. <https://doi.org/10.1371/journal.pgen.0020031>.
- Grabenstein JD, Musey LK. 2014. Differences in serious clinical outcomes of infection caused by specific pneumococcal serotypes among adults. *Vaccine* 32:2399–2405. <https://doi.org/10.1016/j.vaccine.2014.02.096>.
- Harboe ZB, Thomsen RW, Riis A, Valentiner-Branth P, Christensen JJ, Lambertsen L, Krogfelt KA, Konradsen HB, Benfield TL. 2009. Pneumococcal serotypes and mortality following invasive pneumococcal disease: a population-based cohort study. *PLoS Med* 6:e1000081. <https://doi.org/10.1371/journal.pmed.1000081>.
- Jansen AGSC, Rodenburg GD, van der Ende A, van Alphen L, Veenhoven RH, Spanjaard L, Sanders EAM, Hak E. 2009. Invasive pneumococcal disease among adults: associations among serotypes, disease characteristics, and outcome. *Clin Infect Dis* 49:e23–e29. <https://doi.org/10.1086/600045>.
- Horácio AN, Diamantino-Miranda J, Aguiar SI, Ramirez M, Melo-Cristino J. 2013. The majority of adult pneumococcal invasive infections in Portugal are still potentially vaccine preventable in spite of significant declines of serotypes 1 and 5. *PLoS One* 8:e73704-9. <https://doi.org/10.1371/journal.pone.0073704>.
- Luján M, Gallego M, Belmonte Y, Fontanals D, Vallès J, Lisboa T, Rello J. 2010. Influence of pneumococcal serotype group on outcome in adults with bacteraemic pneumonia. *Eur Respir J* 36:1073–1079. <https://doi.org/10.1183/09031936.00176309>.
- Marsh RL, Smith-Vaughan H, Beissbarth J, Hare K, Kennedy M, Wigger C, Mellon G, Stubbs E, Gadil JR, Pettit A, Mackenzie G, Tipakalippa P, Morris PS, Leach AJ. 2007. Molecular characterisation of pneumococcal serotype 16F: established predominant carriage and otitis media serotype in the 7vPCV era. *Vaccine* 25:2434–2436. <https://doi.org/10.1016/j.vaccine.2006.09.016>.
- Duus J, Gotfredsen CH, Bock K. 2000. Carbohydrate structural determination by NMR spectroscopy: modern methods and limitations. *Chem Rev* 100:4589–4614. <https://doi.org/10.1021/cr990302n>.
- Enthart A, Freudenberger JC, Furrer J, Kessler H, Luy B. 2008. The CLIP/CLAP-HSQC: pure absorptive spectra for the measurement of one-bond couplings. *J Magn Reson* 192:314–322. <https://doi.org/10.1016/j.jmr.2008.03.009>.
- Bock K, Pedersen C. 1974. A study of ^{13}C coupling constants in hexopyranoses. *J Chem Soc Perkin Trans 2* 2:293–297. <https://doi.org/10.1039/p29740000293>.
- Skovsted IC, Kern MB, Sonne-Hansen J, Sauer LE, Nielsen AK, Konradsen HB, Petersen BO, Nyberg NT, Duus J. 2007. Purification and structure characterization of the active component in the pneumococcal 22F polysaccharide capsule used for adsorption in pneumococcal enzyme-linked immunosorbent assays. *Vaccine* 25:6490–6500. <https://doi.org/10.1016/j.vaccine.2007.06.034>.
- Agrawal PK. 1992. Nmr spectroscopy in the structural elucidation oligosaccharides and glycosides. *Phytochemistry* 31:3307–3330. [https://doi.org/10.1016/0031-9422\(92\)83678-R](https://doi.org/10.1016/0031-9422(92)83678-R).
- Abeygunawardana C, Bush CA, Cisar JO. 1991. Complete structure of the cell surface polysaccharide of *Streptococcus oralis* C104: a 600-MHz NMR study. *Biochemistry* 30:8568–8577. <https://doi.org/10.1021/bi00099a012>.
- Shashkov AS, Lipkind GM, Knirel YA, Kochetkov NK. 1988. Stereochemical factors determining the effects of glycosylation on the ^{13}C chemical shifts in carbohydrates. *Magn Reson Chem* 26:735–747. <https://doi.org/10.1002/mrc.1260260904>.
- Lundborg M, Widmalm G. 2011. Structural analysis of glycans by NMR chemical shift prediction. *Anal Chem* 83:1514–1517. <https://doi.org/10.1021/acs.1032534>.
- Mavroidi A, Aanensen DM, Godoy D, Skovsted IC, Kalltoft MS, Reeves PR, Bentley SD, Spratt BG. 2007. Genetic relatedness of the *Streptococcus pneumoniae* capsular biosynthetic loci. *J Bacteriol* 189:7841–7855. <https://doi.org/10.1128/JB.00836-07>.
- Aanensen DM, Mavroidi A, Bentley SD, Reeves PR, Spratt BG. 2007. Predicted functions and linkage specificities of the products of the *Streptococcus pneumoniae* capsular biosynthetic loci. *J Bacteriol* 189:7856–7876. <https://doi.org/10.1128/JB.00837-07>.
- Pelosi L, Boumedienne M, Saksouk N, Geiselmann J, Geremia RA. 2005. The glucosyl-1-phosphate transferase WchA (Cap8E) primes the capsular polysaccharide repeat unit biosynthesis of *Streptococcus pneumoniae* serotype 8. *Biochem Biophys Res Commun* 327:857–865. <https://doi.org/10.1016/j.bbrc.2004.12.082>.
- Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. *J Comput Biol* 7:203–214. <https://doi.org/10.1089/10665270050081478>.
- James DBA, Yother J. 2012. Genetic and biochemical characterizations of enzymes involved in *Streptococcus pneumoniae* serotype 2 capsule synthesis demonstrate that Cps2T (WchF) catalyzes the committed step by addition of β -1-4 rhamnose, the second sugar residue in the repeat unit. *J Bacteriol* 194:6479–6489. <https://doi.org/10.1128/JB.01135-12>.
- Mostowy RJ, Croucher NJ, De Maio N, Chewapreecha C, Salter SJ, Turner P, Aanensen DM, Bentley SD, Didelot X, Fraser C. 2017. Pneumococcal capsule synthesis locus cps as evolutionary hotspot with potential to generate novel serotypes by recombination. *Mol Biol Evol* 34:2537–2554. <https://doi.org/10.1093/molbev/msx173>.
- Kolkman MAB, Van Der Zeijst BAM, Nuijten P. 1997. Functional analysis

- of glycosyltransferases encoded by the capsular polysaccharide biosynthesis locus of streptococcus pneumoniae serotype 14. *J Biol Chem* 272:19502–19508. <https://doi.org/10.1074/jbc.272.31.19502>.
25. Calix JJ, Nahm MH, Zartler ER. 2011. Elucidation of structural and antigenic properties of pneumococcal serotype 11A, 11B, 11C, and 11F polysaccharide capsules. *J Bacteriol* 193:5271–5278. <https://doi.org/10.1128/JB.05034-11>.
 26. Lin FL, Vinogradov E, Deng C, Zeller S, Phelan L, Green BA, Jansen KU, Pavliak V. 2014. Structure elucidation of capsular polysaccharides from *Streptococcus pneumoniae* serotype 33C, 33D, and revised structure of serotype 33B. *Carbohydr Res* 383:97–104. <https://doi.org/10.1016/j.carres.2013.11.006>.
 27. Calix JJ, Porambo RJ, Brady AM, Larson TR, Yother J, Abeygunwardana C, Nahm MH. 2012. Biochemical, genetic, and serological characterization of two capsule subtypes among *Streptococcus pneumoniae* serotype 20 strains: discovery of a new pneumococcal serotype. *J Biol Chem* 287:27885–27894. <https://doi.org/10.1074/jbc.M112.380451>.
 28. Geno KA, Bush CA, Wang M, Jin C, Nahm MH, Yang J. 2017. WciG O-acetyltransferase functionality differentiates pneumococcal serotypes 35C and 42. *J Clin Microbiol* 55:2775–2784. <https://doi.org/10.1128/JCM.00822-17>.
 29. Geno KA, Saad JS, Nahm MH. 2017. Discovery of novel pneumococcal serotype 35D, a natural WciG-deficient variant of serotype 35B. *J Clin Microbiol* 55:1416–1425. <https://doi.org/10.1128/JCM.00054-17>.
 30. Calix JJ, Brady AM, Du VY, Saad JS, Nahm MH. 2014. Spectrum of pneumococcal serotype 11A variants results from incomplete loss of capsule O-acetylation. *J Clin Microbiol* 52:758–765. <https://doi.org/10.1128/JCM.02695-13>.
 31. Ravenscroft N, Omar A, Hlozek J, Edmonds-Smith C, Follador R, Serventi F, Lipowsky G, Kuttel MM, Cescutti P, Faridmoayer A. 2017. Genetic and structural elucidation of capsular polysaccharides from *Streptococcus pneumoniae* serotype 23A and 23B, and comparison to serotype 23F. *Carbohydr Res* 450:19–29. <https://doi.org/10.1016/j.carres.2017.08.006>.
 32. Richards JC, Perry MB. 1988. Structure of the specific capsular polysaccharide of *Streptococcus pneumoniae* type 23F (American type 23). *Biochem Cell Biol* 66:758–771. <https://doi.org/10.1139/o88-087>.
 33. Duda KA, Petersen S, Holst O. 2016. Structural characterization of the lipoteichoic acid isolated from *Staphylococcus sciuri* W620. *Carbohydr Res* 430:44–47. <https://doi.org/10.1016/j.carres.2016.04.026>.
 34. Ciucanu I, Kerek F. 1984. A simple and rapid method for the permethylation of carbohydrates. *Carbohydr Res* 131:209–217. [https://doi.org/10.1016/0008-6215\(84\)85242-8](https://doi.org/10.1016/0008-6215(84)85242-8).
 35. Lowry OH, Roberts NR, Ky L, M-L Wu, Al Farr. 1953. The quantitative histochemistry of brain. I. Chemical methods. *J Biol Chem* 207:1–17.
 36. Gerwig GJ, Kamerling JP, Vliegthart J. 1979. Determination of the absolute configuration of monosaccharides in complex carbohydrates by capillary g.l.c. *Carbohydr Res* 77:1–7. [https://doi.org/10.1016/S0008-6215\(00\)83788-X](https://doi.org/10.1016/S0008-6215(00)83788-X).
 37. Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J. 2005. ACT: the Artemis comparison tool. *Bioinformatics* 21:3422–3423. <https://doi.org/10.1093/bioinformatics/bti553>.

SUPPLEMENTAL MATERIAL

Table S1 Summary of J coupling constants*

| Sample | Unit | $^1J_{H1, C1}$ (Hz) | $^3J_{H1, H2}$ (Hz) | $^3J_{H2, H3}$ (Hz) | $^3J_{H3, H4}$ (Hz) | $^3J_{H4, H5}$ (Hz) |
|---------|----------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| 16F CPS | A (α -Rhap) | 170 | ~3.5 | ~4.0 | ~9.0 | ~9.5 |
| | B (β -Rhap) | 161 | ~4.0 | ~4.0 | ~10 | ~10 |
| | C (α -Glcp) | 173 | ~4.5 | ~9.5 | ~7.5 | ~11.5 |
| | D (β -Glcp) | 160 | ~8.0 | ~10.2 | - | ~9.5 |
| 16A CPS | A (α -Galp) | 178 | ~2.7 | ~2.4 | ~5.7 | ~3.5 |
| | B (α -Rhap) | 173 | ~3.5 | ~4.5 | ~10.0 | ~9.5 |
| | C (α -Galp) | 171 | ~3.5 | ~9.5 | ~4.5 | ~4.6 |
| | D (α -Rhap) | 172 | ~3.3 | ~4.5 | ~9.0 | ~9.3 |
| | E (β -Glcp) | 160 | ~8.2 | ~9.4 | ~7.1 | ~12.2 |
| | F (β -Galp) | 160 | ~7.8 | ~10.5 | ~3.7 | ~4.6 |

* $^1J_{H1, C1}$ coupling constants were determined by CLIP-HSQC and all $^3J_{H, H}$ coupling constants were determined using DQF-COSY with approximately ± 1 Hz error. $^3J_{H3, H4}$ coupling constants of **D** cannot be determined because of overlap.

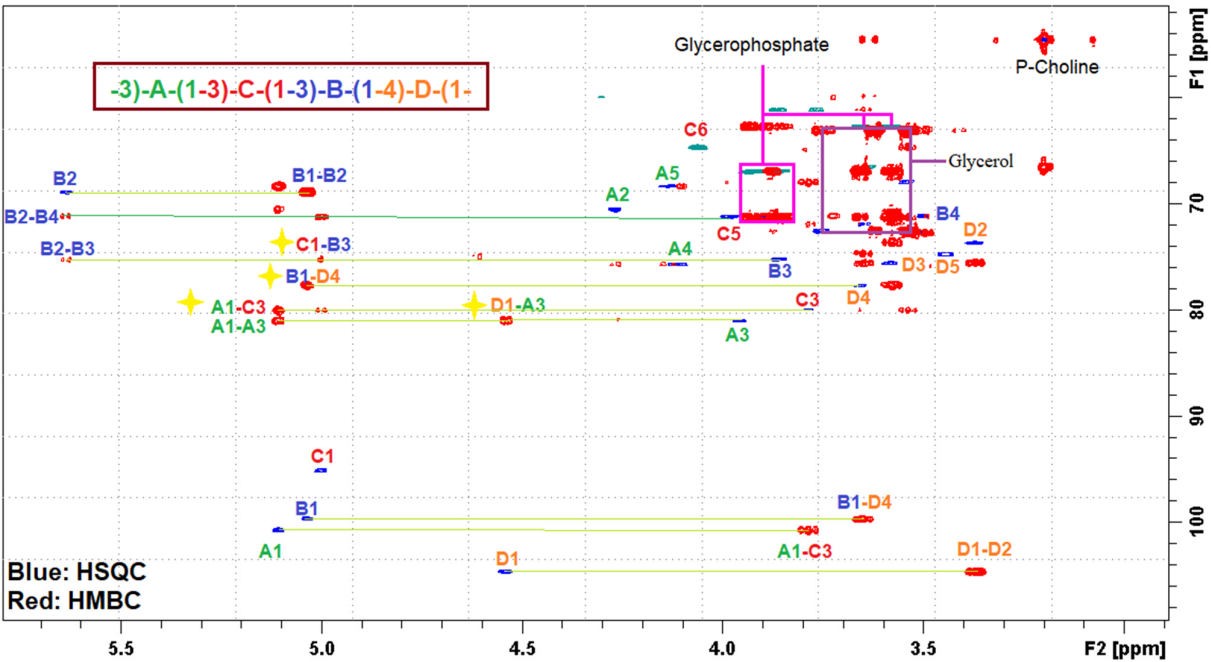


Fig. S1 Expansion and overlap of HSQC and HMBC spectra of ultra-sonicated 16F CPS

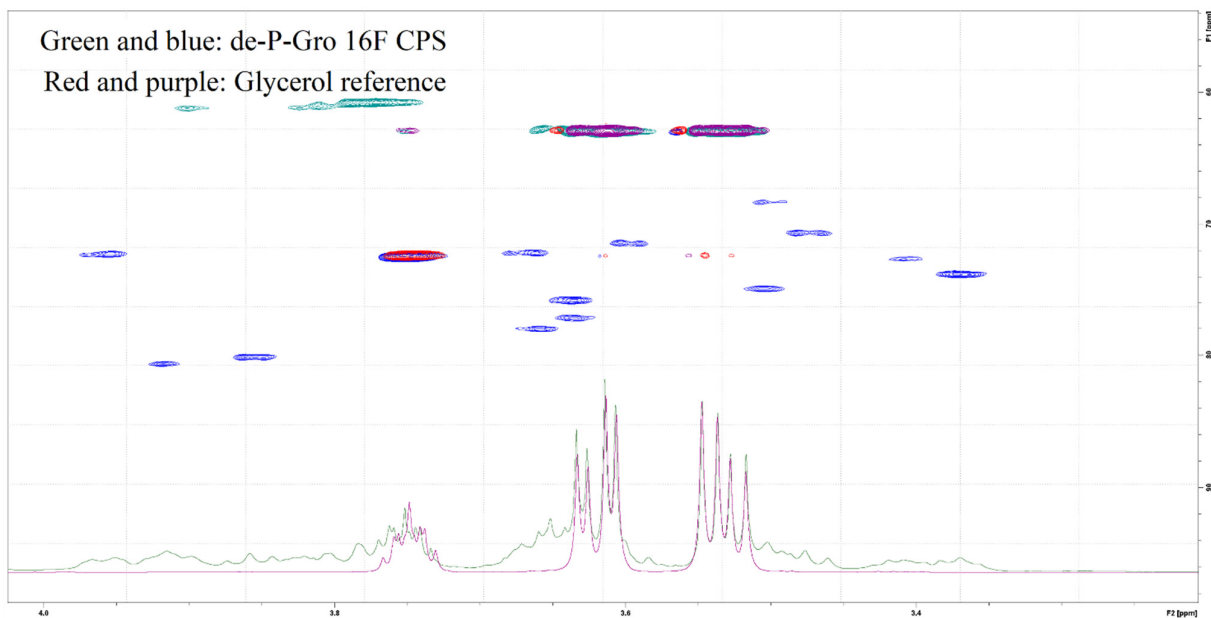


Fig. S2 Expansion and overlap ^1H and HSQC spectra of de-P-Gro16F CPS and glycerol reference

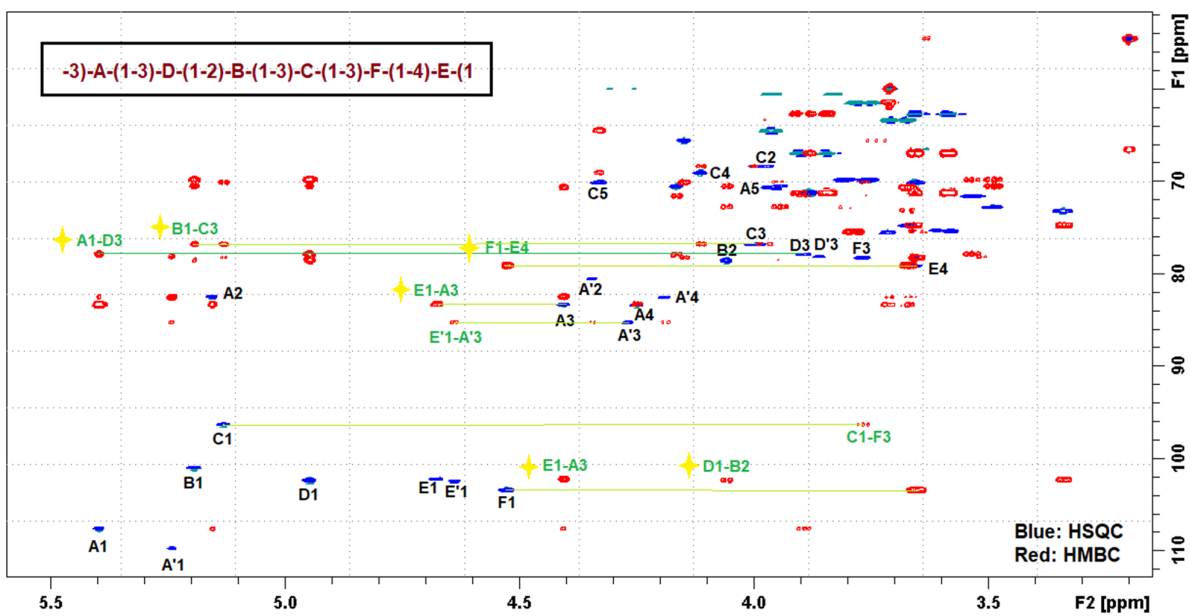
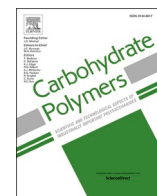


Fig. S3 Expansion HSQC and HMBC spectra of native 16A CPS

Paper 2

Li C, Duda KA, Elverdal PL, Skovsted IC, Kjeldsen C, Teze D, Duus JØ. 2021. Structural, biosynthetic and serological cross-reactive elucidation of capsular polysaccharides from *Streptococcus pneumoniae* serogroup 28. Carbohydr Polym 254:117323.



Structural, biosynthetic and serological cross-reactive elucidation of capsular polysaccharides from *Streptococcus pneumoniae* serogroup 28

Chengxin Li^a, Katarzyna A. Duda^b, Pernille L. Elverdal^c, Ian C. Skovsted^c, Christian Kjeldsen^a, David Teze^d, Jens Ø. Duus^{a,*}

^a Department of Chemistry, Technical University of Denmark, Kgs. Lyngby, Denmark

^b Junior Research Group of Allergobiology, Research Center Borstel, Leibniz Lungenzentrum, Airway Research Center North (ARCN), German Center for Lung Research (DZL), Borstel, Germany

^c SSI Diagnostica A/S, Hilleroed, Denmark

^d Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, Denmark

ARTICLE INFO

Keywords:

Streptococcus pneumoniae
Capsular polysaccharides; Serogroup 28
Structural characterization
Biosynthesis
Serological cross-reactivity

ABSTRACT

Capsular polysaccharides (CPS) are the key virulent factors in the pathogenesis of *Streptococcus pneumoniae*. The previously unknown CPS structures of the pneumococcal serotype 28F and 28A were thoroughly characterized by NMR spectroscopy, chemical analysis and AF4-MALS-dRI. The following repeat unit structures were determined: -4)[α -1-Rhap-[4-P-2-Gro]]-(1-3)- α -D-Sug-[6-P-Cho]-(1-3)- β -1-Rhap-[2-OAc]-(1-4)- β -D-Glcp-(1-; 28F: Sug = Glcp, Mw: 540.5 kDa; 28A: Sug = GlcpNAc, Mw: 421.9 kDa; The correlation of CPS structures with biosynthesis showed that glycosyltransferase WciU in serotypes 28F and 28A had different sugar donor specificity toward α -D-Glcp and α -D-GlcNAc, respectively. Furthermore, latex agglutination tests of de-OAc and de-PO₄ CPS were conducted to understand cross-reactions between serogroup 28 with factor antiserum 23d. Interestingly, the de-OAc 28F and 28A CPS can still weakly react with factor antiserum 23d, while de-PO₄ CPS did not react with factor antiserum 23d. This indicated that OAc group could affect the affinity and P-2-Gro was crucial for cross-reacting with factor antiserum 23d.

1. Introduction

Streptococcus pneumoniae is an encapsulated Gram-positive human pathogen, which can cause invasive pneumococcal disease (IPD), such as pneumonia, meningitis and sepsis (Kadioglu, Weiser, Paton, & Andrew, 2008). Pneumococcal disease with a significant mortality is a health burden worldwide. The thick polysaccharide capsule, protecting the bacterium from phagocytosis, plays a crucial role on virulence of this pathogen. Therefore, the capsular polysaccharides (CPS) are used as the main target of current vaccines. Pneumococcal vaccines target prevalent serotypes, which can elicit serotype-specific protections. However, the polysaccharide diversity makes it difficult to eliminate the disease. There are around 100 different serotypes reported to date and each serotype has a unique serological response due to having a unique CPS structure (Ganaie et al., 2020; Geno et al., 2015). The wide

implementation of pneumococcal vaccines is driving changes in prevalent serotypes and invasive serotypes in the population (Ladhani et al., 2018; Mehr & Wood, 2012; Tin Tin Htar et al., 2015). The currently rare or non-vaccine serotypes might become prevalent in the vaccinated population in the future. Thus, there is an increased need to accurately monitor serotype distribution in IPD and to understand non-vaccine serotypes to formulate the future vaccines.

Pneumococcal serogroup 28 contains serotype 28F and serotype 28A, which are both non-vaccine serotypes. Serotype 28F and 28A clustered with serotype 16F in genetic analysis because of their highly similar *cps* loci, where encoded genes responsible for capsule synthesized. In recent years, serotype 16F have become more prevalent and have increased fatality risks (Dube et al., 2018; Grabenstein & Musey, 2014; Horácio, Diamantino-Miranda, Aguiar, Ramirez, & Melo-Cristino, 2013). Serogroup 28 also have the potential to become prevalent in the

Abbreviations: AF4-MALS-dRI, Asymmetric Flow Field Flow Fractionation-Multi-Angle Light Scattering-differential Refractive Index; CPS, Capsular Polysaccharides; de-OAc, de-O-acetylation; de-PO₄, de-phosphorylation; GT, Glycosyl Transferase; IPD, Invasive Pneumococcal Disease; NMR, Nuclear Magnetic Resonance; OAc, O-Acetyl; P-2-Gro, glycerol-2-phosphate; P-Cho, phosphocholine.

* Corresponding author.

E-mail address: jduus@kemi.dtu.dk (J.Ø. Duus).

<https://doi.org/10.1016/j.carbpol.2020.117323>

Received 14 August 2020; Received in revised form 21 October 2020; Accepted 23 October 2020

Available online 4 November 2020

0144-8617/© 2020 Elsevier Ltd. All rights reserved.

future. Some studies reported that serotype 28F became prevalent and was found in both vaccinated and unvaccinated children in Africa (Ramdani-Bouguesse et al., 2015; Walekhwa, Muturi, Gunturu, Kenya, & Kabera, 2018). Thus, there is a need to know more about serogroup 28 and 16F. The structural information is essential when designing a novel vaccine and for accurately serotyping. Since genotyping gains popularity, it is significant to understand the genetic, structural and serological bases of pneumococcal capsules to differentiate serotypes accurately.

Recently, we determined the serotype 16F CPS structure and reported cross-reaction between serotype 16F with factor antiserum 28b, which was used for serotyping 28F (Li et al., 2019). Considering highly similar *cps* genes and serological reactivity between 16F and serogroup 28, our hypothesis is that the CPS structure of serogroup 28 should be similar to 16F. Therefore, the determination of chemical structure of serogroup 28 is crucial to verify presumable functions of common glycosyltransferase genes in serogroup 28 and 16F. Here we determined CPS structures of 28F and 28A using chemical analyses and NMR spectroscopy. A further correlation analysis of structure, biosynthesis and serological activity was conducted to achieve better understanding of the relationship between pneumococcal genotype and phenotype, as well as CPS structure and immunological property.

2. Materials and methods

2.1. Sample preparation

Purified pneumococcal CPS from serotypes 28F and 28A, as well as rabbit diagnostic antisera, were produced by SSI Diagnostica, Hillerød, Denmark. The polysaccharide samples were ultra-sonicated using a LABSONIC® P ultrasonic homogenizer with a 3 mm Probe for 2 hours under an ice bath, then desalted and concentrated using centrifugal filters (Amicon® Ultra-4, 3 kDa). De-OAc and de-PO₄ were performed using methods described by Richards and Perry (Richards & Perry, 1987). De-OAc CPS samples were obtained by treating approximately 5–10 mg of each polysaccharide with 1 mL of a 0.2 M NaOH in D₂O solution at room temperature for 2 h, followed by neutralizing with 2 M HCl and desalting (Amicon® Ultra-4, 3 kDa). Finally, these samples were transferred to 5 or 3 mm NMR tubes. De-PO₄ 28F and 28A CPS were generated by treating approximately 10 mg of each CPS in 2 mL of 1.0 M NaOH, containing a trace amount of sodium borohydride, at 100 °C for 4 h in screw-capped glass tubes filled with nitrogen. Then the solutions were cooled in an ice bath and neutralized with acetic acid, followed by desalting and removal of residuals using centrifugal filters (Amicon® Ultra-4, 3 kDa).

2.2. Molecular weight measurement by AF4-MALS-dRI

Molecular weight determination of serotype 28F, 28A and 16F CPS was performed on an Asymmetric Flow Field Flow Fractionation-Multi-Angle Light Scattering-differential Refractive Index (AF4MALS-dRI) system. The system consisted of a pump (Agilent 1260 Infinity II) with an online degasser. The separation was performed on an AF4 system (ECLIPSE; Wyatt Technology) using a polyethersulfone membrane with 10 kDa cut-off in a 17.5 cm separation channel with a 350 µm spacer. A constant flow of 2.0 mL/min was applied for 3 min before injection. 2 mg/mL CPS sample in 0.9 % NaCl filtered by Durapore 0.1 µm PVDF membrane filters (Merck Millipore) before injection. 0.9 % NaCl was used as carrier liquid. The samples were injected with a flow of 0.2 mL/min. Then, the samples were eluted at 2.0 mL/min for 5 min. A constant detector flow of 0.5 mL/min was used during the separation process, which included 3 stages: (1) linear ramped flow from 2 to 0.7 mL/min in 22 min; (2) exponential ramped flow from 0.7 to 0.05 mL/min in 25 min; and (3) linear ramped flow from 0.05 to 0.0 mL/min in 3 min and 0.0 mL/min flow, hold constant for 5 min. A TREOS II MALS detector (Wyatt Technology) and a differential refractive index detector, Optilab T-REX

(Wyatt Technology), were connected to the system. Data collection and processing were performed using the ASTRA software, V 7.1.4. For data analysis, the dn/dc of 0.130 mL/g for CPS concentration determination (Bednar & Hennessey, 1993) and a second-order Berry formalism for fitting light scattering data were employed.

2.3. NMR spectroscopy

The NMR experiments were carried out on an a Bruker Avance III (799.90 MHz for ¹H and 201.14 MHz for ¹³C) equipped with a 5 mm TCI ¹H/(¹³C, ¹⁵N) cryoprobe or a Bruker AVANCE 600 MHz instrument with 5 mm SmartProbe BB (F)-H-D (¹⁵N-³¹P, ¹H, ¹⁹F). All spectra were recorded at 50 °C. The assignments of 28F and 28A CPS repeat units were performed using the following experiments: 1D ¹H and ³¹P and the 2D NMR spectra, including double quantum filter correlated spectroscopy (DQF-COSY), total correlation spectroscopy (TOCSY) with 80 ms mixing time, nuclear Overhauser effect spectroscopy (NOESY) with 150–400 ms mixing time, multiplicity edited ¹H-¹³C heteronuclear single quantum coherence (HSQC), ¹H-¹³C HMBC optimized for 8 Hz long range coupling constants, ¹H-¹³C HSQC-TOCSY with 80 ms mixing time, Clean in-phase (CLIP)-HSQC and ¹H-³¹P HMBC. Experiments were conducted with standard gradient enhanced Bruker pulse sequences, except CLIP-HSQC (Enthart, Freudenberger, Furrer, Kessler, & Luy, 2008).

NMR spectra were processed with Topspin 4.0.7 (Bruker) using extensive zero filling in all dimensions. The one-dimensional ¹H spectra were processed with an exponential window function with a line broadening of 0.3 Hz for ¹H spectra. All 2D spectra were processed with shifted sine bell window functions in both dimensions. All spectra were referenced to residual cell wall polysaccharide phosphocholine signals (¹H 3.2 ppm and ¹³C 54.5 ppm and the shielded ³¹P signal at 1.30 ppm) (Skovsted et al., 2007).

2.4. General chemical analysis

The monosaccharide composition and linkage types in serotype 28F and 28A CPS repeat unit were determined by GC and GC-MS after dephosphorylating by HF using methods described previously (Li et al., 2019). The content of total phosphate was determined by colorimetric analysis according to Lowry et al. (Lowry, Roberts, Leiner, Wu, & Farr, 1954). The absolute configuration of monosaccharides was determined by GLC (Gerwig, Kamerling, & Vliegenthart, 1979), by comparison with authentic standards of the acetylated O-[(R)-2-Octyl] glycoside after methanolysis (0.5 M HCl/MeOH, 85 °C, 3.5 h), butanolysis (2 M HCl/(R)-2-Octanol, 65 °C, 4 h) and peracetylation (85 °C, 10 min).

2.5. Genetic analysis of serogroup 28

The published *cps* loci sequences have been downloaded from GenBank (<https://www.ncbi.nlm.nih.gov/nuccore>, accession number of serotype 16F, 28F, 28A: CR931668.1, CR931693.2, CR931692.1, respectively) (Bentley et al., 2006). Pairwise protein sequence identity was assessed using BLASTp. Comparison of similarities of the gene products (TBLASTX) of the *cps* biosynthetic loci was performed by the Artemis comparison tool (ACT) (Carver et al., 2005). Protein models of WciU were built by SWISS-MODEL (Waterhouse et al., 2018), and superimposed over the structure of *S. gordonii* (5E9U.pdb) using Pymol 2.3.3 (Schrodinger LLC). WciU models showing the smallest rmsd were then energy minimized through the YASARA server (Krieger et al., 2009). Minimized WciU models and experimental structures of GT4 N-acetylglucosaminyltransferases from *S. aureus* (6N1X.pdb), *B. subtilis* (5D00.pdb), *S. pneumoniae* (4PQG.pdb) were then superimposed over the structure of *S. gordonii* (5E9U.pdb), obtaining rmsd in the 1.1–2.2 Å² range.

2.6. Latex agglutination test

Latex Agglutination of native and degraded CPS with diagnostic antisera was performed at SSI Diagnostica A/S using ImmuLex™ Pneumotest Kit. Briefly, a drop (approximately 10 µL) of latex particles coated pneumococcal antiserum suspension was applied onto the reaction card and add 10 µL of approximately 2 mg/mL degraded CPS solution next to the latex suspension. Then, mix the two drops by a stick for maximum 10 seconds and observe for antigen agglutinations.

3. Results and discussion

3.1. Molecular weight distribution measured by AF4-MALS-dRI

Serotype 28F, 28A and 16F CPS NMR spectra all gave broad lines due to short transverse relaxation times (T2) of high molecular weight polysaccharides or the heterogeneity of CPS samples. In order to determine the molecular weight distribution and size, AF4-MALS-dRI was used to measure the molecular weight of 28F and 28A CPS samples (Table 1). The weight average molecular weight (Mw) 28F and 28A were around 540.5 kDa and 421.9 kDa, respectively. The polydispersity (Mw/Mn) of 28F and 28A were similar, around 1.2. The Mw of 16F CPS was 653.7 kDa, which had relatively higher molecular weight distribution than 28F and 28A CPS (Fig. 1). In Fig. 1, at the same retention time, 16F CPS had a higher molar mass than 28F and 28A CPS. Moreover, 16F CPS had a smaller RMS radius compared to 28F CPS, while it had a higher Mw than 28F. This indicated a relatively compact conformation of 16F CPS compared to 28F and 28A CPS. The dRI signal of all three CPS

Table 1

The molecular weight, polydispersity and root mean square radius (RMS radius) of serotype 28F, 28A and 16F CPS.

| Sample name | Mn (kDa) | Mw (kDa) | Polydispersity (Mw/Mn) | rn (nm) | rw (nm) |
|-------------|--------------|--------------|------------------------|------------|------------|
| 28F CPS | 476.9 ± 11.5 | 540.5 ± 9.1 | 1.13 ± 0.03 | 51.3 ± 3.9 | 55.5 ± 2.9 |
| 28A CPS | 363.6 ± 3.4 | 421.9 ± 3.3 | 1.16 ± 0.01 | 41.7 ± 0.9 | 45.1 ± 0.8 |
| 16F CPS | 563.1 ± 10.8 | 653.7 ± 11.8 | 1.16 ± 0.03 | 40.1 ± 4.2 | 46.4 ± 3.4 |

Mn: Number-average molecular weight; Mw: Weight-average molecular weight; rn: number-average mean square radius; rw: number-average mean square radius.

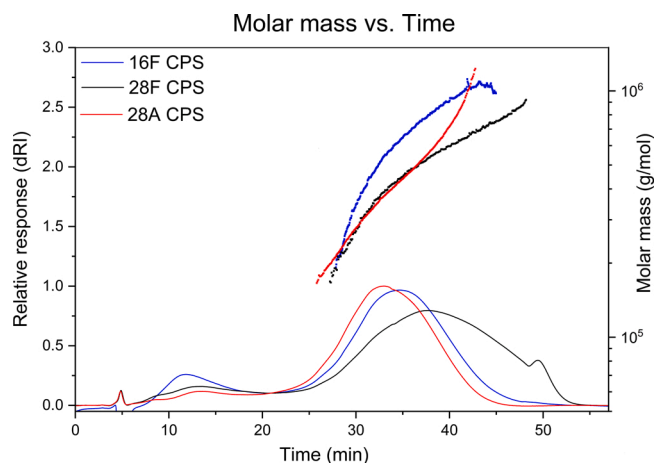


Fig. 1. Molar mass distribution (dots) and differential refractive index response (solid lines) of serotype 28F (black), 28A (red) and 16F (blue) CPS. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

showed small peaks at around 12 min that were cell wall polysaccharides impurities. Besides, 28F CPS had a small peak at 50 min, which was caused by aggregation.

3.2. Structural characterization of CPS from serogroup 28

The CPS repeat unit structures of serotype 28F and 28A were determined by NMR spectroscopy and general chemical analysis. The results of monosaccharides composition, phosphate content and methylation analysis were summarized in Table 2.

In order to obtain better NMR spectra for assignment of CPS, applying a higher temperature (50 °C) and slightly depolymerizing CPS by sonication can increase molecular tumbling rate to improve NMR line shapes. NMR structural analysis of serogroup 28 native and de-OAc CPS repeat units were conducted in two steps: identify individual spin systems (monosaccharides) and connecting the identified structural motifs to yield the repeat unit structure. Assignment of individual spin systems were based on 1D ¹H NMR and 2D DQF-COSY, TOCSY, HSQC, HSQC-TOCSY and H2BC spectra. Then glycosidic bonds were identified using ¹H-¹³C HMBC and NOESY spectra, and similarly were O-acetylated positions. Phosphodiester linkages were determined by ¹H-³¹P HMBC. CLIP-HSQC, providing ¹J_{C-H} coupling constants, was used to determine the anomeric configuration of pyranosides, especially for rhamnopyranosides. In addition, NMR analysis of de-OAc CPS was conducted to simplify spectra and confirm the position of substitutions. 1D ¹H spectra of sonicated and de-OAc 28F CPS and sonicated and de-OAc 28A CPS were shown in supplements (Fig. S1). There were some signals from cell wall polysaccharides in both 28F and 28A CPS NMR spectra (Skovsted et al., 2007), marked with * in Fig. 2.

3.2.1. NMR assignments for serotype 28F CPS repeat unit

According to 1D ¹H and HSQC spectra of 28F CPS, there were four different anomeric signals corresponding to four sugar residues, labelled as A, B, C, and D in order of descending ¹H chemical shift (Table 3). Combined with monosaccharides composition (Table 2), residues A and B were both identified as rhamnopyranosides as they had methyl groups as their 6-position. Based on ¹J_{H1, C1} measured from CLIP-HSQC, A, with ¹J_{H1, C1} of about 174 Hz, was identified as α-L-Rhap, while B ¹J_{H1, C1} of about 162 Hz, as β-L-Rhap (Duus, Gotfredsen, & Bock, 2000). Residue C was determined to be an α-D-glucopyranose according to ¹J_{H1, C1} of 170 Hz and ³J_{H, H} coupling pattern. D was determined to be a β-D-glucopyranose, as it showed only large ³J_{H, H} coupling constants and ¹J_{H1, C1} of 162 Hz. The monosaccharide composition analysis showed that 28F CPS were composed of rhamnose and glucose in a 1:1 ratio (Table 2), which was consistent with the results of the NMR analysis.

After the monosaccharide units were identified, HMBC was used to find the linkage of these monosaccharides (Fig. 2). The anomeric position of A had HMBC correlations with C3, while the anomeric position of C had correlations with B3. Similarly, there were HMBC cross peaks between B1 and D4 and between D1 and C4. In addition, NOESY also showed same correlations (Fig. S2). Therefore, these four sugar residues make up the backbone of the repeat unit in the order: 4)-[A-(1-3)]-C-(1-3)-B-(1-4)-D-(1-, which was further confirmed in methylation analysis. Furthermore, according to ³¹P and ¹H-³¹P HMBC spectra, there were two

Table 2

Results of monosaccharide composition, phosphate content and methylation analysis.

| Serotype | L-Rha | D-Glc | D-GlcN | PO ₄ | Linkage type |
|-------------------|-------|-------|--------|-----------------|---------------------------------|
| 28F CPS (nmol/mg) | 1605 | 1521 | – | 1686 | t-Rha, 3-Rha, 4-Glc, 3,4-Glc |
| Ratio | 1.1 | 1 | – | 1.1 | |
| 28A CPS (nmol/mg) | 1011 | 610 | +++ | 1505 | t-Rha, 3-Rha, 4-Glc, 3,4-GlcNAc |
| Ratio | 1.7 | 1 | 1 | 2.5 | |

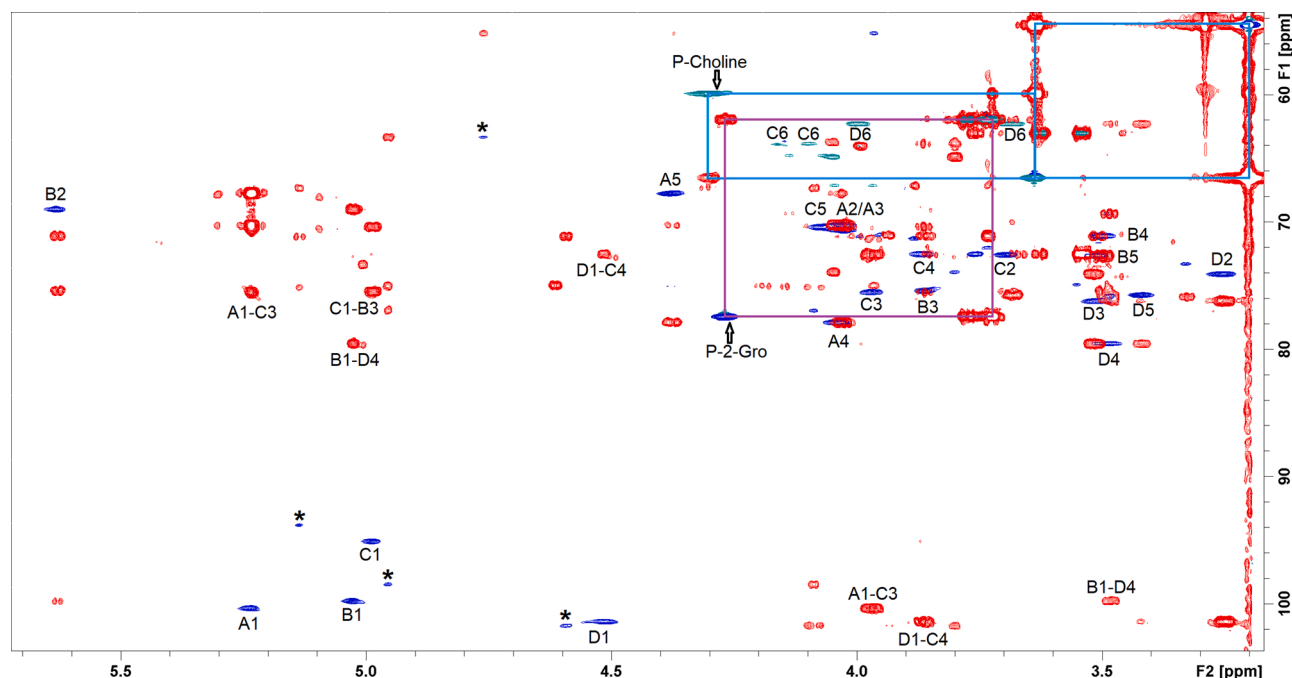


Fig. 2. Expansion and overlap of HSQC (blue) and HMBC (red) spectra of sonicated 28F CPS, * marked HSQC peaks from cell wall polysaccharides. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

phosphate groups attached to the repeat unit. One was assigned to be glycerol-2-phosphate substituted at the 4-position of the α -L-Rha A, based on ^1H - ^{31}P HMBC cross peaks of the ^{31}P with 2-position of glycerol and 4-position of α -L-Rha.

Another phosphate was determined to be phosphocholine, based on the ^1H - ^{31}P HMBC correlations between ^{31}P and choline. This phosphocholine was assigned to be attached to the 6-position of the α -Glc C for the relatively downfield proton chemical shifts of C6 at 4.141/4.101 ppm due to the deshielding effect of phosphate substitution. Moreover, the HMBC cross peak between carbonyl carbon of the OAc group and B2 proton suggested that the β -L-Rhap B was 2-OAc substituted, which would explain the extreme downfield chemical shift of B2 at 5.620/69.04 ppm. In order to prove the OAc substituted, the 28F CPS was subjected to de-O-acetylation. In the NMR spectra of de-OAc 28F CPS, the signals from the acetyl group disappeared and the B2 signal moved upfield to 4.254/68.22 ppm. Based on the above analysis, the repeat unit of serotype 28F CPS was determined to be: -4)-[α -L-Rhap-[4-P-2-Gro]]-(1-3)- α -D-Glcp-[6-P-Cho]-(1-3)- β -L-Rhap-[2-OAc]-(1-4)- β -D-Glcp-(1- (Fig. 3). Full assignments of the sonicated 28F CPS and de-OAc 28F CPS repeat unit can be found in Table 3.

3.2.2. NMR assignments for serotype 28A CPS repeat unit

NMR analysis of serotype 28A CPS was conducted in the same way as for the 28F CPS. Again, the repeat unit structure of 28A CPS consisted of four sugar residues, which was labelled as A, B, C, and D in order of descending ^1H anomeric chemical shift (Table 3). As for the 28F CPS, residues A and B were both identified as rhamnopyranosides, due to the presence of methyl groups as their 6-position and based on monosaccharides composition (Table 2). The anomeric configuration of the two rhamnopyranosides was determined using CLIP-HSQC, and A was determined as having β -configuration, while B had α -configuration, with $^1J_{\text{H1,C1}}$ coupling constants of 163 and 172 Hz, respectively. Residue C was determined to be a 2-acetamido-2-deoxy- α -glucopyranose according to $^1J_{\text{H1,C1}}$ of 172 Hz, $^3J_{\text{H,H}}$ coupling pattern and relative upfield carbon chemical shift of C2 at 53.61 ppm. D was determined to be a β -D-glucopyranose, as it only showed large $^3J_{\text{H,H}}$ coupling constants and a $^1J_{\text{H1,C1}}$ of 161 Hz. Monosaccharide composition analysis supported these findings, as it showed the 28A CPS to be a 2:1:1 ratio between Rha, Glc and GlcNAc (Table 2).

Following this assignment of the individual monosaccharide units, the linkages between them were determined using HMBC and NOESY (Fig. S3 and S4). The anomeric position of the β -L-Rha A had HMBC correlations with D4, while the anomeric position of the α -D-GlcNAc C had correlations with A3. Similarly, there were HMBC cross peaks between B1 and C3 and between D1 and C4. Therefore, these four sugar residues make up the repeat unit in the order: 4)-[B-(1-3)]-C-(1-3)-A-(1-4)-D-(1-, which were consistent with the linkage types of monosaccharides in methylation analysis. In ^{31}P and ^1H - ^{31}P HMBC two phosphate groups attached to the CPS could be observed. Similar to the 28F CPS, one was determined to be a glycerol-2-phosphate attached to the 4-position of the α -L-Rha side chain B, and the other was a phosphocholine connected to the 6-position of the α -D-GlcNAc C. The HMBC cross peak between carbonyl carbon of the OAc group and A2 proton suggested that the β -L-Rha A was 2-OAc substituted, which is supported by the downfield proton chemical shift of A2 at 5.540 ppm. This was further supported by deacetylation, as the corresponding proton signal in de-OAc 28A CPS had moved upfield to 4.199 ppm. Based on this analysis, the repeat unit of serotype 28A CPS was determined to be: -4)-[α -L-Rhap-[4-P-2-Gro]]-(1-3)- α -D-GlcpNAc-[6-P-Cho]-(1-3)- β -L-Rhap-[2-OAc]-(1-4)- β -D-Glcp-(1- (Fig. 3). Full assignments of the sonicated 28A CPS and de-OAc 28A CPS repeat unit is given in Table 3.

3.3. CPS biosynthesis of serogroup 28

The capsular biosynthetic gene clusters from at that time known 90 pneumococcal serotypes were sequenced and analyzed by Bentley (Bentley et al., 2006). The serotype 28F and 28A are genetically close to serotype 16F, which were classified into a same cluster (Mavroidi et al., 2007) (Mostowy et al., 2017). The distinct differences between serogroup 28 and serotype 16F were the phosphate synthetic genes and *wzy* polymerase genes. The phosphate synthetic gene in serotype 16F was glycerol-1-phosphate synthetic gene (*gct*). While phosphate synthetic genes in serogroup 28 were putative glycerol-2-phosphate synthetic genes (*gct1*, *gct2*, *gct3*), which was in agreement with determined 28F and 28A CPS structures. *Wzy* polymerase of 28F had 99.5 % identity with 28A and had only 26.5 % identity with 16F in BLASTp alignment (Nguyen, Ho, Harikrishna, Wong, & Abdul Rahim, 2007). According to

Table 3¹H and ¹³C NMR chemical shifts (δ, ppm) assignment of sonicated and de-OAc 28F CPS and 28A CPS.

| Sample | Residue | 1 | 2 | 3 | 4 | 5 | 6 | HMBC |
|-------------------|-------------------------|-------------------------------------|--------|-------|-------|-------|-------------|--------------------|
| Sonicated 28F CPS | α-L-Rhap-(1-A | 5.229 | 4.028 | 4.024 | 4.027 | 4.362 | 1.286 | A1-C3 |
| | 3)-β-L-Rhap-(1-B | 100.22 | 70.21 | 70.56 | 77.83 | 67.64 | 17.11 | |
| | | 5.024 | 5.620 | 3.851 | 3.496 | 3.498 | 1.358 | B1-D4 |
| | | 99.64 | 69.04 | 75.49 | 71.03 | 72.56 | 17.34 | |
| | | OAc: 2.155 / 20.74 C = O: 174.07 | | | | | | B2 |
| | 3,4)-α-D-Glcp-(1-C | 4.984 | 3.685 | 3.962 | 3.860 | 4.060 | 4.141/4.101 | C1-B3 |
| | | 95.06 | 72.56 | 75.49 | 72.56 | 70.33 | 63.87 | |
| | 4)-β-D-Glcp-(1-D | 4.518 | 3.251 | 3.512 | 3.482 | 3.417 | 3.993/3.675 | |
| | | 101.28 | 74.08 | 76.19 | 79.47 | 75.60 | 62.24 | D1-C4 |
| | P ₁ -2-Gro | 3.760 | | 3.76 | | | | |
| | | 3.734 | 4.262 | 3.734 | | | | |
| | | 61.89 | 77.36 | 61.89 | | | | |
| | P ₁ | 0.308 | | | | | | A4 |
| | P ₂ -Cho | 4.294 | 3.629 | 3.200 | | | | |
| De-OAc 28F CPS | | 59.89 | 66.47 | 54.50 | | | | |
| | P ₂ | -0.472 | | | | | | |
| | α-L-Rhap-(1-A | 5.251 | 4.053 | 4.059 | 4.045 | 4.400 | 1.305 | A1-C3 |
| | | 100.36 | 70.57 | 70.22 | 77.84 | 67.75 | 17.08 | |
| | 3)-β-L-Rhap-(1-B | 4.830 | 4.254 | 3.649 | 3.464 | 3.421 | 1.338 | B1-D4 |
| | | 101.06 | 68.22 | 78.78 | 70.80 | 72.56 | 17.44 | |
| | 3,4)-α-D-Glcp-(1-C | 5.050 | 3.745 | 4.051 | 3.905 | 4.190 | 4.180/4.083 | C1-B3 |
| | | 96.37 | 72.91 | 75.38 | 72.44 | 70.45 | 63.88 | |
| | 4)-β-D-Glcp-(1-D | 4.570 | 3.284 | 3.590 | 3.495 | 3.493 | 4.075 | |
| | | 101.41 | 74.09 | 76.31 | 79.25 | 75.85 | 62.36 | D1-C4 |
| | P ₁ -2-Gro | 3.771 | | 3.771 | | | | |
| | | 3.739 | 4.270 | 3.739 | | | | |
| | | 62.00 | 77.37 | 62.00 | | | | |
| | P ₁ | 0.308 | | | | | | A4 |
| Sonicated 28A CPS | P ₂ -Cho | 4.308 | 3.641 | 3.200 | | | | |
| | | 59.90 | 66.57 | 54.50 | | | | |
| | P ₂ | -0.472 | | | | | | |
| | 3)-β-L-Rhap-(1-A | 5.009 | 5.540 | 3.823 | 3.486 | 3.488 | 1.358 | A1-D4 |
| | | 99.47 | 69.56 | 75.78 | 70.97 | 72.61 | 17.26 | |
| | | OAc: 2.209 / 20.90 C = O: 174.07 | | | | | | A2 |
| | α-L-Rhap-(1-B | 4.958 | 3.809 | 3.928 | 4.014 | 4.325 | 1.275 | B1-C3 |
| | | 100.17 | 70.85 | 70.15 | 77.54 | 67.92 | 17.14 | B4-P ₁ |
| | 3,4)-α-D-GlcpNAc-(1-C | 4.951 | 4.183 | 3.975 | 3.972 | 4.119 | 4.175/4.108 | C1-A3 |
| | | 94.42 | 53.61 | 74.49 | 73.20 | 70.50 | 63.82 | C3-B1, C4-D1 |
| | 4)-β-D-Glcp-(1-D | 4.537 | 3.239 | 3.515 | 3.476 | 3.427 | 3.969/3.686 | D1-C4 |
| | | 101.34 | 74.02 | 76.25 | 79.06 | 75.66 | 62.18 | D4-A1 |
| | P ₁ -2-Gro | 3.755/3.725 | 4.256 | | | | | |
| | | -61.94 | 77.30 | | | | | |
| de-OAc 28A CPS | P ₁ | 0.5151 | | | | | | P ₁ -B4 |
| | P ₂ -Choline | 4.305 | 3.632 | 3.200 | | | | |
| | | -59.83 | -66.51 | 54.5 | | | | |
| | P ₂ | -0.343 | | | | | | |
| | 3)-β-L-Rhap-(1-A | 4.783 | 4.199 | 3.627 | 3.440 | 3.398 | 1.333 | A1-D4 |
| | | 101.18 | 68.11 | 78.43 | 70.70 | 72.45 | 17.44 | |
| | α-L-Rhap-(1-B | 4.978 | 3.836 | 3.965 | 4.037 | 4.389 | 1.286 | B1-C3 |
| | | 100.24 | 70.92 | 70.10 | 77.96 | 67.87 | 17.20 | |
| | 3,4)-α-D-GlcpNAc-(1-C | 4.963 | 4.197 | 4.061 | 3.994 | 4.203 | 4.207/4.099 | C1-A3 |
| | | 95.04 | 53.91 | 74.56 | 73.15 | 70.80 | 63.88 | |
| | 4)-β-D-Glcp-(1-D | 4.594 | 3.260 | 3.602 | 3.463 | 3.505 | 4.069/3.708 | D1-C4 |
| | | 101.41 | 74.09 | 76.20 | 79.25 | 75.84 | 62.35 | |
| | P ₁ -2-Gro | 3.732 | 4.279 | | | | | |
| | | 61.90 | 77.72 | | | | | |
| | P ₁ | 0.175 | | | | | | P ₁ -B4 |
| | P ₂ -Choline | 4.314 | 3.639 | 3.200 | | | | |
| | | 60.01 | 66.48 | 54.50 | | | | |
| | P ₂ | -0.420 | | | | | | |

the determined CPS structures, Wzy polymerases of 28F and 28A were assigned to polymerize at β-D-Glcp-(1-3)-α-D-Glcp and β-D-Glcp-(1-3)-α-D-GlcpNAc, respectively. While, 16F CPS repeat unit was polymerized at the β-D-Glcp-(1,3)-α-L-Rhap linkage with a higher molecular weight than 28F and 28A (Fig. 1).

The serotypes 28F, 28A and 16F have common glycosyltransferase (GT) genes (*wchF*, *wciU*, *wcnN*) (Fig. 4). The GT WchF was reported to be the β-1-4 rhamnosyltransferase (James & Yother, 2012), which was consistent with CPS structures from 27 serotypes that had the *wchF* gene,

including 28F and 28A. The GT WciU of serotype 28F had 94.1 % identity with serotype 28A in BLASTp (Fig. 4). The phylogenetic analysis indicated that the split between serotype 28F and 28A was due to the recombination at the *wciU* gene (Mostowy et al., 2017). The *wciU* genes are present in *cps* loci of seven serotypes (16F, 18F, 18A 18B, 18C, 28F and 28A) among all the pneumococcal serotypes (Bentley et al., 2006; Li et al., 2019; Mavroidi et al., 2007). According to determined CPS structures, WciU in 28F, 16F, 18F, 18B and 18C were assigned specific for adding α-D-Glcp to 3-β-L-Rhap (Table 4), while in 28A and 18A WciU

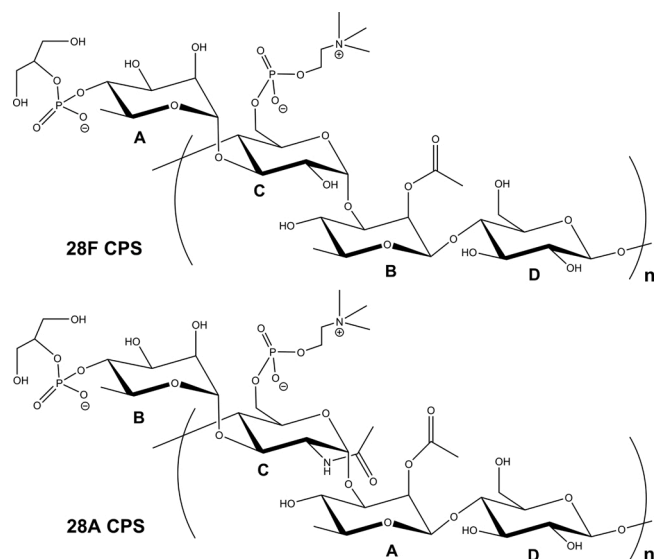


Fig. 3. Repeat unit structures of 28F and 28A CPS.

were specific for transferring α -D-GlcNAc to 3- β -L-Rhap. WciU belongs to the retaining GT4 family enzyme in Carbohydrate-Active Enzymes (CAZy) data bank with a GT-B fold (Aanensen, Mavroidi, Bentley, Reeves, & Spratt, 2007). In order to understand what resulted in the change of substrate specificity of WciU, multiple sequence alignment of WciU from the seven serotypes was shown in Fig. 4. Among the 423 amino acids of the WciU sequence, there are 18 amino acids, highlighted with red frames in Fig. 4 (c), which are identical in WciU of 28A and

18A, but different from those in corresponding sites of the other five serotypes. To identify the key residue positions that could be responsible for sugar donor specificity of WciU, molecular models of WciU from 28F and 28A were built (Fig. 5). By comparison with the sequences and the structures of distant GT4 N-acetylglucosaminyltransferases from *S. aureus* (6N1X.pdb), *S. gordonii* (5E9U.pdb), *S. pneumoniae* (4PQG.pdb) and *B. subtilis* (5D00.pdb), the residue in position 346 (Trp in the Glc-specific WciU, Cys in the GlcNAc one) seems to be the only one in the close vicinity (<5 Å) of the 2-acetamido group among the non-identical residues between the two isoforms of WciU (Fig. 5). The second closest residue (342) is located nearly 10 Å away from the 2-acetamido group, and represent a small change (Ile versus Val). While the orientation of side-chain residues or the exact location of a residue from *in silico* WciU models are not to be trusted, the overall position of the residues is likely correct, and thus the residue 346 is deemed most likely to affect the

Table 4

A summary of pneumococcal serotypes with WciU, their protein sequence identity and catalytic donor, linkage and acceptor.

| Serotype with WciU | Percentage identity (refer to WciU of 28F) | Donor | Linkage | Acceptor |
|--------------------|--|----------|----------------|-----------------------|
| 28F | 100.0 % | | | |
| 18F | 99.1 % | D-Glcp | α (1-3) | β -L-Rhap-2-OAc |
| 16F | 98.6 % | | | |
| 18B | 92.6 % | | | |
| 18C | 92.6 % | D-Glcp | α (1-3) | β -L-Rhap |
| 18A | 89.0 % | D-GlcNAc | α (1-3) | β -L-Rhap |
| 28A | 94.1 % | D-GlcNAc | α (1-3) | β -L-Rhap-2-OAc |

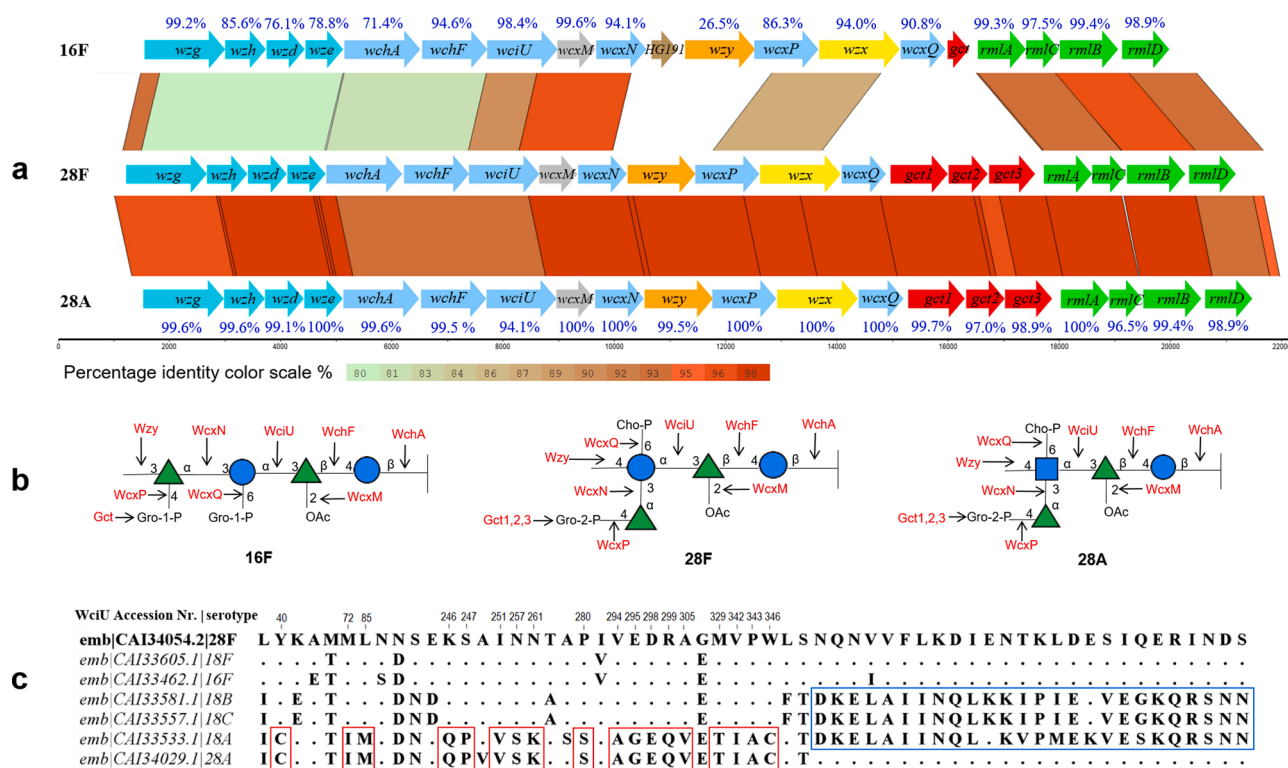


Fig. 4. (a) Comparison of serotype 28F, 28A and 16F cps loci. The results of a pairwise tBLASTx comparison are indicated with different red color scale (percent identity cutoff: 80 %, score cutoff: 1000). The percentage identity of a pairwise BLASTp comparison of each gene product in 16F and 28A compared to a counterpart in 28F are shown in blue characters. (b) Correlation of CPS biosynthetic enzymes (red letters above black arrows) and repeat unit structures of serotype 16F, 28F and 28A; The ●, ▲ and ■ refer to pyranoside Glc, Rha and GlcNAc, respectively. (c) Protein sequence alignment of WciU from seven serotypes (identical parts were hidden); Red frames highlighted identical amino acids from WciU of 28A and 18A that were different from those in the other five serotypes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

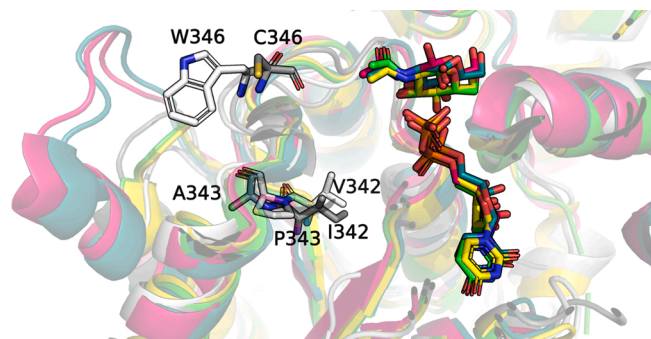


Fig. 5. Superimposed protein models of WciU of 28F (white) and 28A (grey) and protein structures of *N*-acetylglucosaminyltransferases from *S. aureus* (6N1X .pdb, yellow), *S. pneumoniae* (4PQG.pdb, blue), *S. gordonii* (5E9U.pdb, purple) and *B. subtilis* (5D00.pdb, green). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

specificity. Moreover, the mutation of a bulky residue (Arg) into a Cys (R194C) in the GT4 WciY from *E. coli* O107 resulted in a similar donor specificity shift from Glc towards GlcNAc (Kocov et al., 2020). The superimposed models of WciY from *E. coli* O107 with WciU in 28A showed that the C194 that was responsible for the specificity shift was in the close vicinity of C346 (<2 Å). However, experimental validations are required to confirm the predicted key residue. In addition, interestingly, the residues from position 384–423 of WciU in 28F, 28A, 18F and 16F are highly similar but quite different from those in 18A, 18B and 18C (highlighted with the blue frame in Fig. 4 (c)), which correspond to the presence of 2-OAc in the acceptor β -L-Rhap. However, if the *O*-acetylation of β -L-Rhap happened after donor residues were added, the acceptors of WciU in the seven serotypes in Table 4 could be the same. WcxN protein sequence in serotype 28F and 28A are 100 % identical and share 94.1 % identity with 16F. The CPS structure of 28F support the function of glycosyltransferase WcxN proposed in serotype 16F, which transfers α -L-Rhap to 3- α -D-Glcp (Fig. 4) (Li et al., 2019). However, the acceptor of WcxN in 28A is 3- α -D-GlcNAcp, which means WcxN in 28A or 28F could transfer α -L-Rhap to 3- α -D-Glcp and 3- α -D-GlcNAcp. Besides, there are two phosphotransferases in serogroup 28, WcxP (the putative glycerol phosphotransferase) and WcxQ (the putative LicD-family phosphotransferase). The WcxQ in serogroup 28 is most likely the phosphorylcholine transferase LicD1, which transfers P-Chol to the 6-position of α -D-Glcp or α -D-GlcpNAc. While the WcxP in serogroup 28 should be responsible for addition of P-2-Gro to the 4-position of α -L-Rhap. Accordingly, the WcxQ and WcxP in serotype 16F could be responsible for adding P-1-Gro to the 6-position of α -D-Glcp and the 4-position of α -L-Rhap, respectively. However, further experiments are required to validate the putative functions of these two phosphotransferases. Above discussion did not take into account the possibility that there might be other gene clusters involved in the actual CPS biosynthesis.

3.4. Serological cross-reactive analysis

The factor antiserum 28b and 28c were designed for serotyping 28F and 28A, respectively. Based on determined CPS structures of 28F and 28A, α -D-Glcp in 28F CPS is the key epitope for factor antiserum 28b binding, while α -D-GlcpNAc in 28A CPS is the key epitope for factor antiserum 28c binding. Furthermore, the latex agglutination of de-OAc and de-PO₄ 28F and 28A CPS with factor antiserum 28b and 28c were all negative, which indicated that the OAc and the P-2-Gro groups could also affect factor antiserum 28b and 28c binding. In addition, we reported that serotype 28F and 28A reacted with factor antiserum 23d, which was used for serotyping 23B within serogroup 23 by Quellung reaction (Li et al., 2019). The cross-reaction between serogroup 28 and serotype 23B indicated that there were structural similarities in their

Table 5

Cross-reactions between serogroup 28 and serogroup 23 using latex agglutination.

| Samples | Factor antisera Latex solution | | | | |
|----------------------------|--------------------------------|-----|-----|-----|-----|
| | 28b | 28c | 23b | 23c | 23d |
| 28F native CPS | +++ | – | – | – | +++ |
| 28F de-OAc CPS | – | – | – | – | + |
| 28F de-PO ₄ CPS | – | – | – | – | – |
| 28A native CPS | – | +++ | – | – | +++ |
| 28A de-OAc CPS | – | – | – | – | + |
| 28A de-PO ₄ CPS | – | – | – | – | – |

“+++”: positive agglutination reaction; “+”: weakly positive agglutination reaction; “–”: negative agglutination reaction; “/”: not tested.

CPS antigens. The CPS repeat unit structures of serogroup 28 and serotype 23B all consist of a trisaccharide backbone with P-2-Gro on the external surface (Ravenscroft et al., 2017). Therefore, the surface P-2-Gro groups could be important for cross-reaction between serogroup 28 CPS with antiserum 23d. In order to prove the assumption, we tested de-OAc and de-PO₄ 28F and 28A CPS with factor antiserum 23d using latex agglutination. Interestingly, de-OAc 28F and 28A CPS both still can react with antiserum 23d weakly, while de-PO₄ 28F and 28A cannot react with 23d factor antiserum (Table 5). Thus, it suggests that the OAc groups can affect the affinity of antibody binding to the CPS. The P-2-Gro groups are necessary for antibodies in factor antiserum 23d binding onto the CPS.

4. Conclusion

The CPS structures of serotype 28F and 28A were characterized by NMR spectroscopy and chemical analysis, which filled the gap of structure and biosynthesis within pneumococcal serogroup 28/16. The comparison of CPS biosynthesis between serogroup 28 and 16F was consistent with previously proposed functions of glycosyltransferase WcxN. However WcxN in serogroup 28 appeared to show broad acceptor specificity, which can catalyze α -L-Rhap to 3- α -D-Glcp and 3- α -D-GlcNAcp. The GT WciU in serotypes 28F and 28A appeared to show different sugar donor specificity toward α -D-Glcp and α -D-GlcNAcp, respectively, which resulted in the CPS structural difference. Furthermore, homology modeling of WciU and comparison with published GT structure revealed that the residue in position 346 could be most likely to affect the donor specificity of WciU. In addition, serological test of diagnostic antisera with partially degraded CPS revealed the importance of *O*-Acetyl and phosphate substitutions. As CPS structures are determined, biosynthesis and serological activity can be better understood, which will be useful for developing future serotyping and diagnostic methods as well as vaccines.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgements

The 800 MHz NMR spectra were recorded on the NMR Center-DTU, supported by the Villum Foundation. C.L. appreciate the financial support of China Scholarship Council (PhD scholarship No. 201708310119) and Oticon Fonden for financial support during her external stay in Research Center Borstel. D.T. acknowledges the support of the NovoNordisk Foundation (NNF10CC1016517). We thank Kasper Enemark-Rasmussen for excellent technical assistance with NMR instrument setup, Mariusz Kubus and Tao Jiang for the assist of AF4-MALS measurements, Katharina Jakob (Research Center Borstel) for compositional chemical analyses and Katrine Helander Pedersen for serological test.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.carbpol.2020.117323>.

References

- Aanensen, D. M., Mavroidi, A., Bentley, S. D., Reeves, P. R., & Spratt, B. G. (2007). Predicted functions and linkage specificities of the products of the *Streptococcus pneumoniae* capsular biosynthetic loci. *Journal of Bacteriology*, 189(21), 7856–7876. <https://doi.org/10.1128/JB.00837-07>
- Bednar, B., & Hennessey, J. P. (1993). Molecular size analysis of capsular polysaccharide preparations from *Streptococcus pneumoniae*. *Carbohydrate Research*, 243(1), 115–130. [https://doi.org/10.1016/0008-6215\(93\)84085-K](https://doi.org/10.1016/0008-6215(93)84085-K)
- Bentley, S. D., Aanensen, D. M., Mavroidi, A., Saunders, D., Rabinowitsch, E., Collins, M., et al. (2006). Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genetics*, 2(3), e31. <https://doi.org/10.1371/journal.pgen.0020031>
- Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G., & Parkhill, J. (2005). ACT: The Artemis comparison tool. *Bioinformatics*, 21(16), 3422–3423. <https://doi.org/10.1093/bioinformatics/bti553>
- Dube, F. S., Ramjith, J., Gardner-Lubbe, S., Nduru, P., Robberts, F. J. L., Wolter, N., et al. (2018). Longitudinal characterization of nasopharyngeal colonization with *Streptococcus pneumoniae* in a South African birth cohort post 13-valent pneumococcal conjugate vaccine implementation. *Scientific Reports*, 8(1), 1–9. <https://doi.org/10.1038/s41598-018-30345-5>
- Duus, J., Gotfredsen, C. H., & Bock, K. (2000). Carbohydrate structural determination by NMR spectroscopy: Modern methods and limitations. *Chemical Reviews*, 100(12), 4589–4614. <https://doi.org/10.1021/cr990302n>
- Enthart, A., Freudenberger, J. C., Furrer, J., Kessler, H., & Luy, B. (2008). The CLIP/CLAP-HSQC: Pure absorptive spectra for the measurement of one-bond couplings. *Journal of Magnetic Resonance*, 192(2), 314–322. <https://doi.org/10.1016/j.jmr.2008.03.009>
- Ganaie, F., Saad, J. S., McGee, L., van Tonder, A. J., Bentley, S. D., Lo, S. W., et al. (2020). A new pneumococcal capsule type, 10D, is the 100th serotype and has a large cps fragment from an oral streptococcus. *MBio*, 11(3), 1–15. <https://doi.org/10.1128/mBio.00937-20>
- Geno, K. A., Gilbert, G. L., Song, J. Y., Skovsted, I. C., Klugman, K. P., Jones, C., et al. (2015). Pneumococcal capsules and their types: Past, present, and future. *Clinical Microbiology Reviews*, 28(3), 871–899. <https://doi.org/10.1128/CMR.00024-15>
- Gerwig, G. J., Kamerling, J. P., & Vliegthart, J. F. G. (1979). Determination of the absolute configuration of monosaccharides in complex carbohydrates by capillary g. l.c. *Carbohydrate Research*, 77(1), 1–7. [https://doi.org/10.1016/S0008-6215\(00\)83788-X](https://doi.org/10.1016/S0008-6215(00)83788-X)
- Grabenstein, J. D., & Musey, L. K. (2014). Differences in serious clinical outcomes of infection caused by specific pneumococcal serotypes among adults. *Vaccine*, 32(21), 2399–2405. <https://doi.org/10.1016/j.vaccine.2014.02.096>
- Horácio, A. N., Diamantino-Miranda, J., Aguiar, S. I., Ramirez, M., & Melo-Cristino, J. (2013). The majority of adult pneumococcal invasive infections in Portugal are still potentially vaccine preventable in spite of significant declines of serotypes 1 and 5. *PloS One*, 8(9), 1–9. <https://doi.org/10.1371/journal.pone.0073704>
- James, D. B. A., & Yother, J. (2012). Genetic and biochemical characterizations of enzymes involved in *Streptococcus pneumoniae* serotype 2 capsule synthesis demonstrate that Cps2T (WchF) catalyzes the committed step by addition of β 1-4 rhamnose, the second sugar residue in the repeat unit. *Journal of Bacteriology*, 194(23), 6479–6489. <https://doi.org/10.1128/jb.01135-12>
- Kadioglu, A., Weiser, J. N., Paton, J. C., & Andrew, P. W. (2008). The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. *Nature Reviews Microbiology*, 6(4), 288–301. <https://doi.org/10.1038/nrmicro1871>
- Kocev, A., Melamed, J., Torgov, V., Danilov, L., Veselovsky, V., & Brockhausen, I. (2020). The wclY gene of *Escherichia coli* serotype O117 encodes an α 1,4-glucosyltransferase with strict acceptor specificity but broad donor specificity. *Glycobiology*, 00(00), 1–12. <https://doi.org/10.1093/glycob/cwaa045>
- Krieger, E., Joo, K., Lee, J., Lee, J., Raman, S., Thompson, J., et al. (2009). Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Protein Structure Function and Bioinformatics*, 77(S9), 114–122. <https://doi.org/10.1002/prot.22570>
- Ladhani, S. N., Collins, S., Djennad, A., Sheppard, C. L., Borrow, R., Fry, N. K., et al. (2018). Rapid increase in non-vaccine serotypes causing invasive pneumococcal disease in England and Wales, 2000–17: A prospective national observational cohort study. *The Lancet Infectious Diseases*, 18(4), 441–451. [https://doi.org/10.1016/S1473-3099\(18\)30052-5](https://doi.org/10.1016/S1473-3099(18)30052-5)
- Li, C., Duda, K. A., Elverdal, P. L., Skovsted, I. C., Kjeldsen, C., & Duus, J. Ø. (2019). Structural, biosynthetic, and serological cross-reactive elucidation of capsular polysaccharides from *Streptococcus pneumoniae* serogroup 16. *Journal of Bacteriology*, 201(20), 1–13. <https://doi.org/10.1128/JB.00453-19>
- Lowry, O. H., Roberts, N. R., Leiner, K. Y., Wu, M. L., & Farr, A. L. (1954). The quantitative histochemistry of brain. I. Chemical methods. *The Journal of Biological Chemistry*, 207(1), 1–17.
- Mavroidi, A., Aanensen, D. M., Godoy, D., Skovsted, I. C., Kalfott, M. S., Reeves, P. R., et al. (2007). Genetic relatedness of the *Streptococcus pneumoniae* capsular biosynthetic loci. *Journal of Bacteriology*, 189(21), 7841–7855. <https://doi.org/10.1128/JB.00836-07>
- Mehr, S., & Wood, N. (2012). *Streptococcus pneumoniae* - a review of carriage, infection, serotype replacement and vaccination. *Paediatric Respiratory Reviews*, 13(4), 258–264. <https://doi.org/10.1016/j.prrv.2011.12.001>
- Mostowy, R. J., Croucher, N. J., De Maio, N., Chewapreecha, C., Salter, S. J., Turner, P., et al. (2017). Pneumococcal capsule synthesis locus cps as evolutionary hotspot with potential to generate novel serotypes by recombination. *Molecular Biology and Evolution*, 34(10), 2537–2554. <https://doi.org/10.1093/molbev/msx173>
- Nguyen, P. D., Ho, C.-L., Harikrishna, J. A., Wong, M. C. V.-L., & Abdul Rahim, R. (2007). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Trees (Berlin, Germany: West)*, 21(5), 3389–3402. <https://doi.org/10.5511/plantbiotechnology.19.145>
- Ramdani-Bougues, N., Ziane, H., Bekhoucha, S., Guechi, Z., Azzam, A., Touati, D., et al. (2015). Evolution of antimicrobial resistance and serotype distribution of *Streptococcus pneumoniae* isolated from children with invasive and noninvasive pneumococcal diseases in Algeria from 2005 to 2012. *New Microbes and New Infections*, 6, 42–48. <https://doi.org/10.1016/j.nmni.2015.02.008>
- Ravenscroft, N., Omar, A., Hlozek, J., Edmonds-Smith, C., Follador, R., Serventi, F., et al. (2017). Genetic and structural elucidation of capsular polysaccharides from *Streptococcus pneumoniae* serotype 23A and 23B, and comparison to serotype 23F. *Carbohydrate Research*, 450, 19–29. <https://doi.org/10.1016/j.carres.2017.08.006>
- Richards, J. C., & Perry, M. B. (1987). Structure of the specific capsular polysaccharide of *Streptococcus pneumoniae* type 23F (American type 23). *Biochemistry and Cell Biology*, 66(29219), 758–771. <https://doi.org/10.1139/o88-087>
- Skovsted, I. C., Kerrn, M. B., Sonne-Hansen, J., Sauer, L. E., Nielsen, A. K., Konradsen, H. B., et al. (2007). 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(35), 6490–6500. <https://doi.org/10.1016/j.vaccine.2007.06.034>
- Tin Tin Htar, M., Christopoulou, D., & Schmitt, H. J. (2015). Pneumococcal serotype evolution in Western Europe. *BMC Infectious Diseases*, 15(1), 1–10. <https://doi.org/10.1186/s12879-015-1147-x>
- Walekhwa, M., Muturi, M., Gunturu, R., Kenya, E., & Kabera, B. (2018). *Streptococcus pneumoniae* serotype epidemiology among PCV-10 vaccinated and unvaccinated children at Gertrude's Children's Hospital, Nairobi County: A cross-sectional study. *F1000Research*, 7, 879. <https://doi.org/10.12688/f1000research.14387.2>
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gky427>

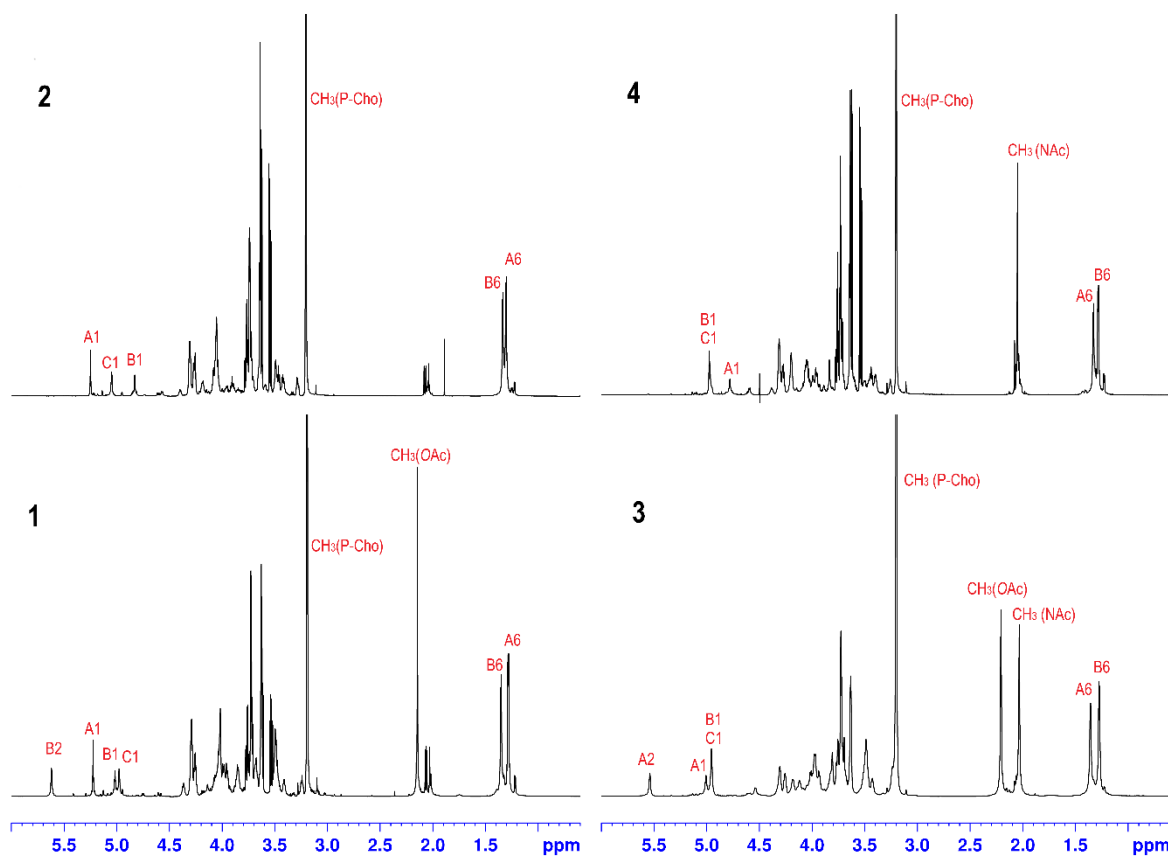


Fig. S1 1D ^1H spectra of sonicated 28F CPS(1), de-OAc 28F CPS(2), sonicated 28A CPS(3) and de-OAc 28A CPS(4)

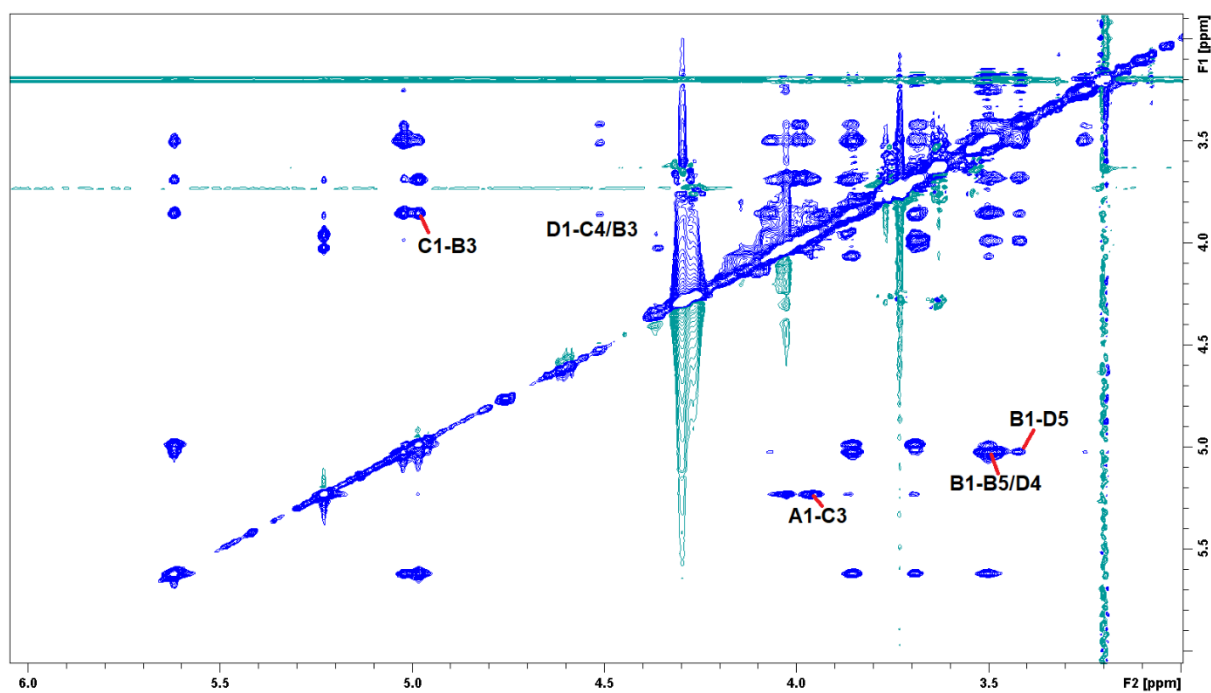


Fig. S2 Expansion of NOESY spectra of sonicated 28F CPS

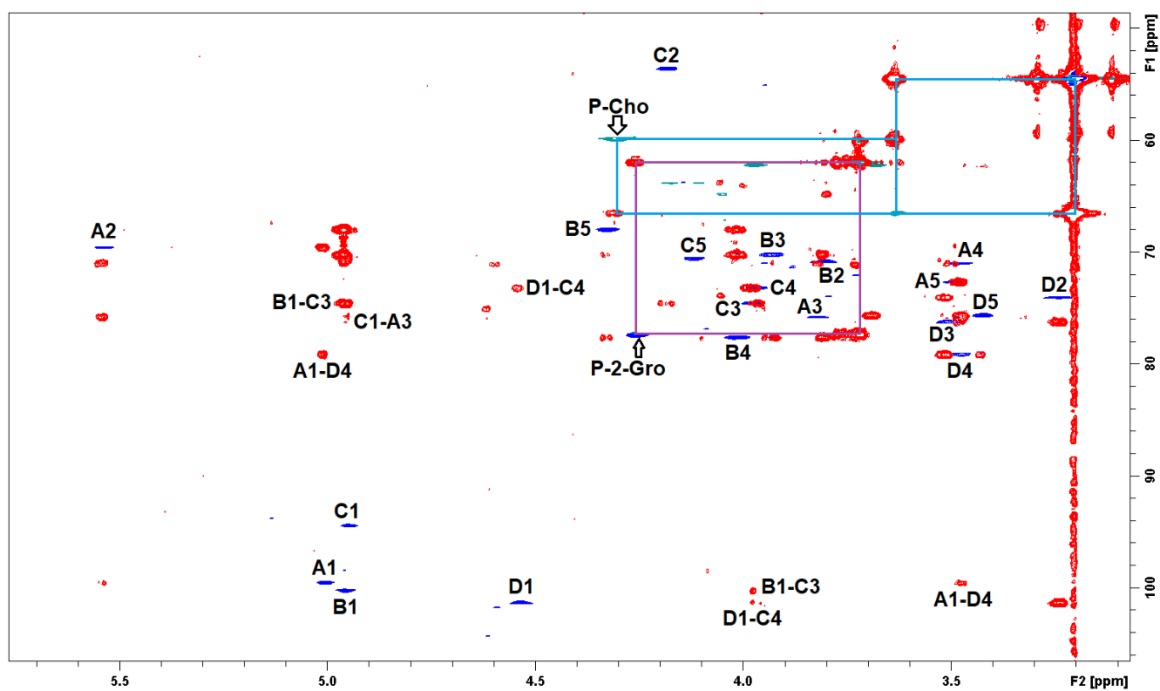


Fig. S3 Expansion and overlap of HSQC (blue) and HMBC (red) spectra of sonicated 28A CPS

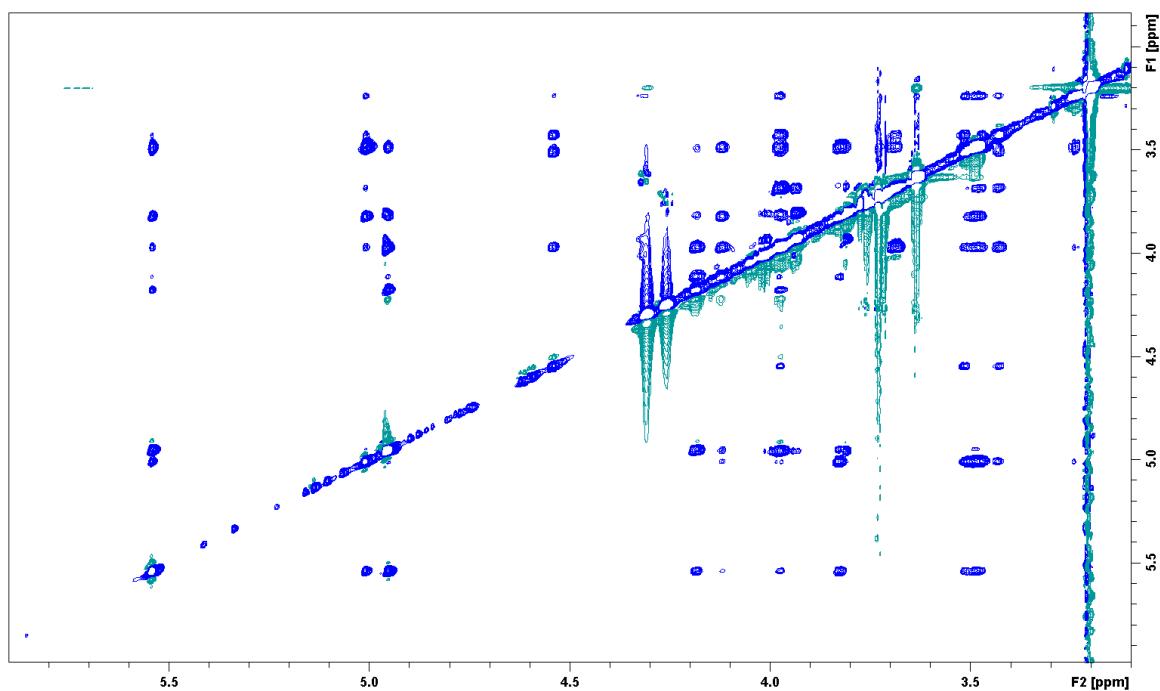


Fig. S4 Expansion of NOESY spectra of sonicated 28A CPS