Accepted Manuscript

Genetic and structural elucidation of capsular polysaccharides from *Streptococcus pneumoniae* serotype 23A and 23B, and comparison to serotype 23F

Neil Ravenscroft, Aneesa Omar, Jason Hlozek, Cesarina Edmonds-Smith, Rainer Follador, Fabio Serventi, Gerd Lipowsky, Michelle M. Kuttel, Paola Cescutti, Amirreza Faridmoayer

PII: S0008-6215(17)30519-0

DOI: 10.1016/j.carres.2017.08.006

Reference: CAR 7431

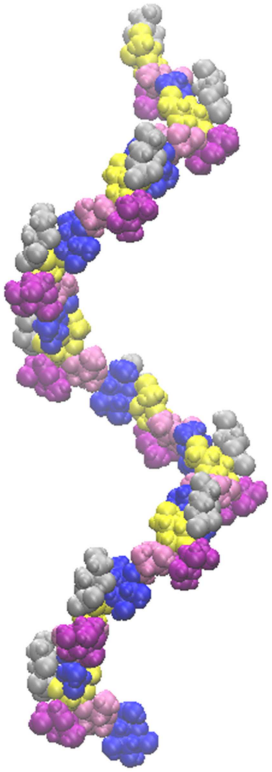
To appear in: *Carbohydrate Research*

Received Date: 25 July 2017

Revised Date: 11 August 2017

Accepted Date: 14 August 2017

Please cite this article as: N. Ravenscroft, A. Omar, J. Hlozek, C. Edmonds-Smith, R. Follador, F. Serventi, G. Lipowsky, M.M. Kuttel, P. Cescutti, A. Faridmoayer, Genetic and structural elucidation of capsular polysaccharides from *Streptococcus pneumoniae* serotype 23A and 23B, and comparison to serotype 23F, *Carbohydrate Research* (2017), doi: 10.1016/j.carres.2017.08.006.

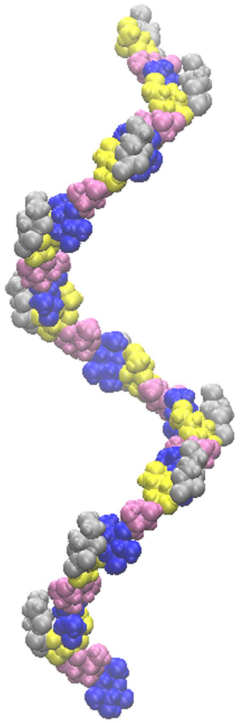


23F

α-L-Rha*p*-(1
↓
2
→4)-β-D-Glc*p*-(1→4)-β-D-Gal*p*-(1→4)-β-L-Rha*p*-(1→
3
↑
Gro-(2→P

23A

α-L-Rha*p*-(1
↓
2
Gro-(2→P→3)-β-D-Gal*p*-(1
↓
4
→4)-β-D-Glc*p*-(1→3)-β-L-Rha*p*-(1→



23B

→4)-β-D-Glc*p*-(1→4)-β-D-Gal*p*-(1→4)-β-L-Rha*p*-(1→
3
↑
Gro-(2→P

Genetic and structural elucidation of capsular polysaccharides from *Streptococcus pneumoniae* serotype 23A and 23B, and comparison to serotype 23F

Neil Ravenscroft[a]*, Aneesa Omar[a], Jason Hlozek[a], Cesarina Edmonds-Smith[a], Rainer Follador[b], Fabio Serventi[b], Gerd Lipowsky[b], Michelle M. Kuttel[c], Paola Cescutti[d], Amirreza Faridmoayer[b]

[a] Department of Chemistry, University of Cape Town, Rondebosch 7701, South Africa

[b] LimmaTech Biologics AG, Grabenstrasse 3, Schlieren, Switzerland

[c] Department of Computer Science, University of Cape Town, Rondebosch 7701, South Africa

[d] Department of Life Sciences, Blg. C11, Università di Trieste, via L. Giorgieri 1, 34127 Trieste, Italy

* Corresponding author: Tel: +27 21 650 4354; E-mail address: neil.ravenscroft@uct.ac.za (N. Ravenscroft)

Abstract

Streptococcus pneumoniae is a globally important encapsulated human pathogen with approximately 100 different serotypes recognized. Serogroup 23 consists of serotype 23F, present in licensed vaccines, and emerging serotypes 23A and 23B. Here, we report the previously unknown structures of the pneumococcal capsular polysaccharides serotype 23A and 23B determined using genetic analysis, NMR spectroscopy, composition and linkage analysis and Smith degradation (of polysaccharide 23A). The structure of the serotype 23A capsular polysaccharide is: →4)-β-D-Glc*p*-(1→3)-[[α-L-Rha*p*-(1→2)]-[Gro-(2→P→3)]-β-D-Gal*p*-(1→4)]-β-L-Rha*p*-(1→. This structure differs from polysaccharide 23F as it features a disaccharide backbone and the di-substituted β-Gal is linked to β-Rha as a side chain. This is due to the different polymerization position catalysed by the unusually divergent repeat unit polymerase Wzy in the 23A *cps* biosynthesis locus. Steric crowding in 23A, confirmed by molecular models, causes the NMR signal for H-1 of the di-substituted 2,3-β-Gal to resonate in the α- anomeric region. The structure of the serotype 23B capsular polysaccharide is the same as 23F, but without the terminal α-Rha: →4)-β-D-Glc*p*-(1→4)-[Gro-(2→P→3)]-β-D-Gal*p*-(1→4)-β-L-Rha*p*-(1→. The immunodominant terminal α-Rha of 23F is more sterically crowded in 23A and absent in 23B. This may explain the reported typing cross reactions for serotype 23F: slight with 23A and none with 23B.

Keywords: Streptococcus pneumoniae, Capsular polysaccharide, Serotype 23A, Serotype 23B, NMR spectroscopy.

1. Introduction

Streptococcus pneumoniae is a globally important encapsulated human pathogen and approximately 100 different serotypes have been recognized [1]. Despite extensive genetic and serological studies, the polysaccharide structures of several serotypes have yet to be determined. Serogroup 23 consists of serotypes 23F, 23A and 23 B, of which only the structure of polysaccharide 23F has been published [2]. Based on the epidemiology at the time, serotype 23F was chosen for inclusion in the 23-valent polysaccharide vaccine [3] and 23F is currently present in all licensed conjugate vaccines. Epitope specificity studies on synthetic conjugates and killed *S. pneumoniae* 23F in animals showed that the terminal α-Rha is immunodominant [4]; this has also been observed in human sera from subjects immunized with PPV23 [5]. Early studies showed that typing antiserum prepared in rabbits with type 23F bacteria reacts only slightly with serotype 23A and hardly at all with serotype 23B [3] and therefore cross-protection from the 23F polysaccharide and conjugate vaccines is not expected. It is therefore unsurprising that serotype 23A and 23B have been identified as emerging pathogens due to a combination of serotype replacement and antimicrobial resistance [6-11].

The genes required for pneumococcal capsular polysaccharide (CPS) synthesis are generally encoded on the *cps* locus [12]. The locus contains three types of enzymes: those responsible for (i) biosynthesis of nucleotide-activated sugars, (ii) polysaccharide repeat-unit synthesis and (iii) assembly of the repeat units and transport across the membrane. In 23F, WchA initiates the repeat-unit synthesis by catalysing the production of an undecaprenyl pyrophosphate (UndPP)-D-Glucose. In successive steps, WchF adds a L-rhamnose to the UndPP-Glucose, followed by WchV linking a D-galactose. This D-galactose is extended by WchW and WchX adding a L-rhamnose and a glycerol-2-phosphate, respectively. Wzx flips

a single repeat unit into the periplasm, where the Wzy polymerase links the D-glucose at reducing end of the growing chain to the position 4 of the single repeat units' D-galactose, resulting in the mature polysaccharide [13]. Wzg, Wzh, Wzd, and Wze are involved in the modulation of capsule synthesis [14]. Biosynthesis of dTDP-L-rhamnose is achieved via the RmlACBD genes, while CDP-2-glycerol biosynthesis requires Gtp123 [15]. The serogroup 23 *cps* loci sequences share the same 18 genes with a varying degree of similarity.

The published structure for the 23F polysaccharide was elucidated by use of chemical and spectroscopic analysis performed on the native and de-phosphorylated polysaccharide and fragments generated by partial hydrolysis and periodate treatment [2]. ^{1}H NMR assignments were presented for the native and de-phosphorylated polysaccharide, however, the ^{13}C NMR spectrum was not assigned. Here we describe detailed 1D and 2D ^{1}H, ^{13}C and ^{31}P NMR experiments performed on the 23F polysaccharide in order to make full NMR assignments that were used to facilitate the structural elucidation of the structurally related serotype 23A and 23B polysaccharides.

2. Results and discussion

2.1. Genetic analysis and predicted structures for serotype 23A and 23B capsular polysaccharide repeating units

The presence of the same glycosyltransferase (GT) genes (wchA, wchF, wchV, wchW and wchX) that are present in the *cps* locus of serotype 23F, indicates the possibility that serotype 23A and 23B contain the same monosaccharide composition as serotype 23F (Fig. 1).

The *cps* cluster of serotype 23A is similar to 23F, with the notable exception of the oligosaccharide polymerase Wzy. A comparison of Wzy sequences from all *S. pneumoniae* serotype *cps* clusters (see Suppl. Fig. S1) reveals that even though 23A Wzy is most closely related to Wzy of 23F and 23B, the divergence is as high as what can be observed between polymerases belonging to different serogroups. The specificity of Wzy protein sequence to the polysaccharide subunit has been postulated and exploited for serotyping purposes [16]. Thus it can be hypothesized that the repeating unit of the 23A CPS is identical to the repeating unit of 23F CPS, but that the serotype difference occurs due to a different polymerization linkage of the single repeating units.

In the *cps* cluster of serotype 23B, the highest sequence diversity towards other GTs of the 23 group is observed for WchA and WchW. The function of WchA is biochemically demonstrated as undecaprenyl-phosphate glucosyl-1-phosphate transferase [17] which is a highly conserved function irrelevant to sequence diversity within *S. pneumoniae* [12]. WchW is a putative α-1,2-L-rhamnosyltransferase [13]. A search for WchW homologues revealed no hits outside of serogroup 23 *cps* loci within *S. pneumoniae;* however a homologue to WchW was found in the *cps* locus of *Klebsiella pneumoniae* serotype K56, designated as WcpL (accession CZQ24634 [18]). The *K. pneumoniae* K56 CPS repeating unit also contains a rhamnose which is α-1,2-linked to galactose [19]. The fact that WchW of 23B has the lowest sequence identity to WchW of 23A/F suggests that there may be a different bond between the L-rhamnose and the D-galactose, possibly explaining the serotyping differences. The close homology between the 23B and 23F repeat unit polymerases (Wzy), suggests a similar polymerization for both serotypes.

2.2. NMR assignments for serotype 23F capsular polysaccharide repeating unit

Chemical analysis of the 23F polysaccharide gave the expected results. GC analysis of the alditol acetate derivatives confirmed the presence of Rha, Gal and Glc in the molar ratio 1.8 : 0.9 : 1.0, whereas GC analysis of the chiral glycosides showed that the hexoses were in the D absolute configuration and Rha in the L absolute configuration. The linkage positions for the constituent sugars were determined by GC and GC-MS analysis of the partially-methylated alditol acetate (PMAA) derivatives (Table 1, columns I and II). The 23F polysaccharide contains terminal Rha (t-Rha), 4-linked Rha (4-Rha), 4-linked Glc (4-Glc) and 2,3,4-linked Gal (2,3,4-Gal). The presence of 2,4-linked Gal (2,4-Gal) is due to loss of the 3-linked phosphoglycerol substituent during the longer base treatment required to achieve higher levels of methylation. The low amount of terminal Rha detected was attributed to decomposition during the strong acid hydrolysis conditions employed and to the volatile nature of trimethylated deoxy sugars.

The ^{1}H NMR spectrum (Fig. 2A) shows the expected signals for the 23F tetrasaccharide repeating unit (RU): four H-1, ring signals, sharp peaks from glycerol and two methyl signals from α- and β-Rha, together with small signals from residual cell wall polysaccharide (CWPS). The diagnostic anomeric and methyl proton signals were used as starting points for the ^{1}H-^{1}H correlation experiments (COSY and TOCSY) which elucidated H-1 to H-6 for β-Glc, α- and β-Rha and H-1 to H-4 for β-Gal. H-5 of β-Gal was assigned from the H-1/H-5 crosspeak in the NOESY experiment and H-6 from the H-4/C-6 crosspeak in the HSQC-NOESY experiment. All of the HSQC crosspeaks (Fig. 3) could be assigned from the proton assignments already established aided by overlays with 1D TOCSY (200 ms), HSQC-TOCSY, HSQC-NOESY and HMBC experiments. The ^{1}H and ^{13}C NMR data are collected in Table 2. The deshielded carbons and glycosylation shifts compared to the corresponding

monosaccharide [21] confirmed the linkage positions: C-2 (+2.73 ppm), C-3 (+4.45 ppm) and C-4 (+4.65 ppm) of β-Gal, C-4 (+4.96 ppm) of β-Rha and C-4 (+6.90 ppm) of β-Glc. The relatively small glycosylation shift for C-2 of Gal has been observed for other 2,3-β-Gal residues in serotypes 15B and 33F and was attributed to the strong steric hindrance imposed by vicinal 2,3-disubstitution [22]. The sequence of sugar residues indicated by glycosylation shifts followed from the HMBC interresidue correlations (Suppl. Fig. S2A) and transglycosidic correlations in the NOESY experiment. The ^{1}H-^{31}P HMBC experiment showed major crosspeaks from the phosphodiester signal at -0.09 ppm to H-3 of β-Gal at 4.33 ppm and H-2 (and H-1/H-3) of Gro confirming the presence of the Gro-(2→P→3)-β-D-Gal*p*-linkage. An expansion of the fully assigned ^{13}C NMR spectrum is shown in Fig. 4A; the splitting of C-2 of glycerol (6 Hz) is from ^{31}P coupling. Lastly the proton-coupled ^{13}C spectrum gave $J_{H1,C1}$ for the anomeric carbons confirming the α- configuration of the terminal Rha (174 Hz) and β- for the remaining residues (162-168 Hz). Thus NMR analysis confirmed the structure of the tetrasaccharide repeating unit of serotype 23F polysaccharide as →4)-β-D-Glc*p*-(1→4)-[α-L-Rha*p*-(1→2)]-[Gro-(2→P→3)]-β-D-Gal*p*-(1→4)-β-L-Rha*p*-(1→.

2.3. Structure of serotype 23A capsular polysaccharide repeating unit

Composition analysis of the 23A polysaccharide gave similar results to those obtained for 23F. GC analysis of the alditol acetate derivatives showed the presence of Rha, Gal and Glc in the molar ratio 2.3 : 0.7 : 1.0. This was confirmed by GC-MS analysis of the TMS methyl glycosides (Suppl. Fig. S3A) which also gave poor release of Gal (relative peak areas of 0.45 : 0.20 : 1.00). A small amount of glycerol was also detected by MS. GC analysis of the chiral glycosides showed that the hexoses were in the D absolute configuration and Rha in the L absolute configuration as for 23F. The linkage positions for the constituent sugars were

determined by GC and GC-MS analysis of the PMAA derivatives (Table 1, columns III and IV). In contrast to the 23F polysaccharide, the 23A polysaccharide contains terminal Rha (t-Rha), 4-linked Glc (4-Glc) and 2,3-linked Gal (2,3-Gal) instead of 2,3,4-Gal. The low amounts of 2-linked Gal (2-Gal) are due to some loss of the 3-linked phosphoglycerol substituent during the potassium dimsyl base treatment. Base treatment for 2 h resulted in higher levels of methylation (Table 1, column IV) and showed the presence of 3,4-linked Rha (3,4-Rha) not detectable in the first methylation analysis. The linkage analysis showing the presence of 2,3-Gal and the doubly-branched Rha was confirmed by NMR analysis.

The ^{1}H NMR spectrum (Fig. 2B) shows signals for the 23A tetrasaccharide RU: four H-1, ring signals, sharp peaks from glycerol and two methyl signals from α- and β-Rha, together with small signals from residual CWPS. Similar chemical shifts and coupling constants were observed for α- and β-Rha and β-Glc compared to the spectrum of 23F (Fig. 2A). The major difference is the presence of a new H-1 signal at 5.17 ppm attributed to Gal. This chemical shift is in the α- anomeric region, however, the large coupling with H-2 (7.8 Hz) is characteristic of β-Gal [23]. A full set of ^{1}H, ^{13}C and ^{31}P 1D and 2D NMR experiments were performed, as described for the 23F polysaccharide. As for 23F, the COSY and TOCSY experiments elucidated H-1 to H-4 for Gal and H-5 was assigned from the H-1/H-5 crosspeak in the NOESY experiment; this constitutes further proof of the β- configuration for Gal. As for 23F, all of the HSQC crosspeaks (Fig. 5) could be assigned from the proton assignments aided by appropriate overlays with hybrid and HMBC experiments. The ^{1}H and ^{13}C NMR data are collected in Table 3. The deshielded carbons and glycosylation shifts established the linkage positions: C-2 (+2.83 ppm) and C-3 (+4.85 ppm) of β-Gal, C-3 (+9.47 ppm) and C-4 (+1.10 ppm) of β-Rha and C-4 (+6.50 ppm) of β-Glc. The small glycosylation shift for C-4 of β-Rha has been observed for 3,4-β-Rha in serotype 17F [24]; this can be attributed to the

strong steric hindrance imposed by vicinal disubstitution. The sequence of sugar residues indicated by glycosylation shifts followed from the HMBC interresidue correlations (Suppl. Fig. S2B) and transglycosidic correlations in the NOESY experiment. The ^{1}H-^{31}P HMBC experiment showed major crosspeaks from the phosphodiester signal at -0.68 ppm to H-3 of β-Gal at 4.25 ppm and H-2 (and H-1/H-3) of Gro confirming the presence of the Gro-(2→P→3)-β-D-Gal*p*-linkage. An expansion of the fully assigned ^{13}C NMR spectrum is shown in Fig. 4B; the splitting of C-2 of glycerol (6 Hz) is from ^{31}P coupling. Lastly the proton-coupled ^{13}C spectrum gave $J_{H1,C1}$ for the anomeric carbons confirming the α-configuration of the terminal Rha (173 Hz) and β- for the remaining residues (162-168 Hz) including the Gal (168 Hz). Thus NMR analysis established the structure of the tetrasaccharide repeating unit of serotype 23A polysaccharide as →4)-β-D-Glc*p*-(1→3)-[[α-L-Rha*p*-(1→2)]-[Gro-(2→P→3)]-β-D-Gal*p*-(1→4)]-β-L-Rha*p*-(1→. The repeating unit structure and anomeric configuration of Gal was confirmed by Smith degradation studies which yielded a major oligosaccharide product **23ASD**.

Periodate oxidation of the proposed repeating unit structure would be expected to depolymerize the polysaccharide by oxidation of the 4-linked β-Glc in the sugar backbone and to oxidize the terminal α-Rha on the 2,3-linked Gal side chain to yield an oligosaccharide product that would be amenable to analysis. ^{1}H NMR analysis of the major Smith degradation product, **23ASD**, showed the presence of β-Rha (H-1 at 4.77 and H-6 at 1.37 ppm) and H-1 of β-Gal at 4.64 ppm, the expected chemical shift region for a β-linked Gal (Figure 6).

Full NMR characterization of **23ASD** elucidated the oligosaccharide as β-D-Gal*p*-(1→4)-β-L-Rha*p*-(1→2)-threitol; the labelled HSQC spectrum and chemical shift data are presented in

Suppl. Fig. S4 and Table 3 (lower panel), respectively. The threitol is derived from oxidation followed by borohydride reduction of the 4-linked Glc and β-Gal is terminal due to oxidation and cleavage of the α-Rha linked to C-2 and loss of the 3-linked phosphoglycerol substituent during $NaBH_4$ treatment. The disaccharide chemical shift data are in good agreement with those predicted by CASPER [25]. These results unambiguously prove that the 2,3-linked Gal residue in serotype 23A has the β- configuration. H-1 of 2,3,4-β-Gal linked to C-4 of Rha in serotype 23F resonates at 4.95 ppm, however, it is strongly deshielded to 5.17 ppm in the 23A polysaccharide which has the 2,3-β-Gal linked to C-4 of the 3,4-disubstituted Rha.

2.4. Structure of serotype 23B capsular polysaccharide repeating unit

Composition analysis of the 23B polysaccharide gave similar results to those obtained for 23F and 23A. GC-MS analysis of the TMS methyl glycosides (Fig. S2B) showed the presence of Rha, Gal and Glc (relative peak areas of 0.22 : 0.53 : 1.00) together with trace amounts of glycerol. The ^{1}H NMR spectrum (Fig. 2C) shows signals for the 23B trisaccharide RU: three H-1, ring signals, sharp signals from glycerol and one methyl signal from β-Rha, together with signals from residual CWPS. The 1D spectrum indicated that 23B has the same structure as 23F, but without the terminal α-Rha. This was confirmed by the full set of ^{1}H, ^{13}C and ^{31}P 1D and 2D NMR experiments performed. All of the HSQC crosspeaks (Fig. 7) could be assigned from the proton assignments aided by appropriate overlays with hybrid and HMBC experiments. The ^{1}H and ^{13}C NMR data are collected in Table 4. The deshielded carbons and glycosylation shifts established the linkage positions: C-3 (+4.15 ppm) and C-4 (+4.38 ppm) of β-Gal, C-4 (+8.96 ppm) of β-Rha and C-4 (+7.10 ppm) of β-Glc. The similarity of the glycosylation shifts to those obtained for 23F provide proof that the bonds have the same configuration factors i.e. that the hexoses have the D absolute configuration and Rha the L absolute configuration. The sequence of sugar residues indicated

by glycosylation shifts followed from the HMBC interresidue correlations (Suppl. Fig. S2C) and transglycosidic correlations in the NOESY experiment. The ^{1}H-^{31}P HMBC experiment showed major crosspeaks from the phosphodiester signal at -0.26 ppm to H-3 of β-Gal at 4.26 ppm and H-2 (and H-1/H-3) of Gro confirming the presence of the Gro-(2→P→3)-β-D-Gal*p*-linkage. An expansion of the fully assigned ^{13}C NMR spectrum is shown in Fig. 4C; the splitting of C-2 of glycerol (5.8 Hz) is from ^{31}P coupling. Lastly the proton-coupled ^{13}C spectrum gave $J_{H1,C1}$ for the anomeric carbons confirming the β-configuration for all the residues (162-165 Hz). Thus NMR analysis established the tetrasaccharide repeating unit of serotype 23B polysaccharide as →4)-β-D-Glc*p*-(1→4)-[Gro-(2→P→3)]-β-D-Gal*p*-(1→4)-β-L-Rha*p*-(1→.

2.5 Molecular models

Molecular models of 10RU of the three polysaccharides were built with CarbBuilder [26] and subsequently minimized. The models for 23F and 23B show the same loose helical conformation (Fig. 8), with a helical pitch of approximately 40 Å and 15 residues per helical turn. However, the immunodominant terminal α-Rha in 23F (absent in 23B) is clearly exposed on the edge of the helix (purple residues in Fig. 8), and would present a markedly different surface for antibody binding than 23B. In contrast to the conformations of 23F and 23B, the model for 23A is not a helix at all, but a slightly twisted flat ribbon, with clear steric crowding at the β-L-Rha branch point: the β-Glc is in close proximity to β-Gal (< 3 Å). This model thus explains the strong deshielding of H-1 of 2,3-β-Gal observed in the NMR spectrum of polysaccharide 23A. Further, the presentation of the terminal α-Rha in 23A is quite different to 23F: the α-Rha forms a long, almost straight line along the chain in 23A, as opposed to its orientation in the 23F helix. The different conformations depicted in these

preliminary models suggest little likelihood of cross-protection between either 23F or 23B with 23A.

3. Conclusions

Structural predictions of the 23A and 23B polysaccharide based on the genetic analyses are in agreement with the experimentally-obtained structures. The biological repeat units of the two polysaccharides can be identified with confidence, and the glycosyltransferases responsible for each elongation step can be assigned by comparison with the 23F *cps* locus (Fig. 9).

The low similarity between the polymerase Wzy of the 23A and 23F *cps* locus is reflected in the different polymerization which results in a significantly divergent polysaccharide structure, where the backbone is constituted by the repetition of the →3)-β-L-Rha*p*-(1→4)-β-D-Glc*p*-(1→ disaccharide. This is the first report describing a polymerization position on the second sugar from the reducing end of the repeat unit in *Streptococcus pneumoniae*.

WchW activity in 23B is not only different to its 23F counterpart as predicted, but it is even absent. The lack of the α-1-2-linked-rhamnose constitutes the only difference between 23F and 23B capsular polysaccharides. The reasons for WchW's missing activity are not clearly deducible from a comparison of the known homologues; it could be due to either several inactivating mutations in the protein, comparing to 23F, or a halt at the transcriptional level. As Geno at al. [27] have recently shown, a subtle difference of only two amino acid substitutions can be responsible for the inactivation of the acetyltransferase in an isolate of serotype 35B, resulting in a novel, non-acetylated serotype.

Molecular modelling shows similar helical structures for 23F and 23B, but a markedly different sterically-crowded ribbon-like structure for 23A. The repeating unit structures for 23A and 23B may explain why the typing antiserum prepared in rabbits with type 23F bacteria reacts only slightly with serotype 23A and hardly at all with serotype 23B [3]. In 23A, the immunodominant terminal α-Rha [5] is no longer a pendant group at C-2 of the main backbone 2,3,4-Gal as in 23F, but on C-2 of the sterically constrained 2,3-Gal, now present as a side chain (Fig. 8). This means that the terminal α-Rha of 23A will be less accessible to 23F antibody directed against this dominant epitope. The terminal α-Rha is absent in 23B, which results in little or no cross reaction with 23F antisera as reported.

4. Experimental

Purified pneumococcal polysaccharide serotypes 23A and 23B were purchased from Statens Serum Institut (SSI). A second sample of polysaccharide 23A and the comparator polysaccharide 23F were obtained from GlaxoSmithKline Biologicals (Rixensart, Belgium).

4.1. Genetic analysis of serogroup 23 *cps* locus sequence

The published *cps* locus sequences (serotype 23a: accession CR931683; 23b: CR931684; 23f: CR931685) and Wzy sequences [12] have been downloaded from GenBank (https://www.ncbi.nlm.nih.gov/nuccore). Pairwise protein sequence identity has been assessed using BLASTp [28]. Multiple sequence alignments have been performed using T-Coffee and standard parameters (v11.00) [29]. Wzy phylogeny was inferred from multiple sequence alignments by running RAxML using a gamma distribution to model site-specific rate variation and 100 bootstrap replicates [30].

4.2. Monosaccharide composition analysis by GC and GC-MS

Hydrolysis of polysaccharide 23F and 23A samples (0.5 mg) was performed with 2M TFA for 2 h at 125 °C and alditol acetates prepared as previously described [31]. GC analysis was performed on a Perkin–Elmer Autosystem XL gas chromatograph equipped with a flame ionisation detector and SP2330 column (30 m); temperature program: 200 °C for 1 min, 200 – 245 °C at 4 °C/min, and 245 °C for 16 min. A mixture of standard monosaccharides (with inositol as an internal standard) was used to determine the retention times and response factors for each sugar.

Methanolysis (3 M HCl) of polysaccharide 23F, 23A and 23B samples (0.5 – 1 mg) was performed in a CEM Discover SP-d Microwave reactor at 120 W and 121 °C for 5 minutes and the tri-methyl silyl ether (TMS) derivatives prepared as described by Kim et al. [32]. GC-MS analysis was performed on an Agilent 8720A Gas Chromatograph equipped with a Agilent 5975 mass spectrometer and a DB-1MS column (30 m); temperature program: 50 °C for 2 min, 50 - 150 °C at 30 °C/min, 150 - 220 °C at 3 °C/min, 220 - 300 °C at 30 °C/min and 300 °C for 10 min. The inlet temperature was set at 250 °C and the MS transfer line at 300 °C. The MS acquisition parameters were set to scan at m/z 50-550 in electron impact (EI) mode. GC-MS data was processed using Agilent Chemstation software. A mixture of standard monosaccharides was used to determine the retention times and corresponding mass spectra for each sugar derivative.

4.3. Monosaccharide absolute configuration analysis by GC and GC-MS

Determination of the absolute configuration of the monosaccharide residues in 23F and 23A was performed according to Gerwig et al. [33]. Poor recovery using the standard method was addressed by additional steps of sample preparation. The samples were sonicated using a Branson sonicator equipped with a microtip at 2.8 Å (3X for 60 sec at power 4 in ice, at 1 min intervals). Prior hydrolysis of sonicated polysaccharide 23F and 23A samples (0.5 mg)

was performed (2M TFA for 2 h at 125 °C) was followed by butanolysis (1 M HCl) in S-(+)-2-butanol for 16 h at 80 °C and TMS derivatization. GC analysis was performed on an Agilent Technologies 6850 gas chromatograph equipped with a flame ionisation detector and an HP-1 column (30 m); temperature program: 50 °C for 1 min, 50 – 130 °C at 45 °C/min, 130 °C for 1 min, 130 – 200 °C at 1°C/min, and 200 °C for 10 min. GC–MS (e.i.) analyses were carried out on an Agilent Technologies 7890A gas chromatograph coupled to an Agilent Technologies 5975C VL MSD, using an HP-1 column (30 m) and the same temperature program. TMS derivatives of monosaccharide standards (all with the D configuration, except L-Rha) were prepared using butanolysis (1 M HCl) in S-(+)-2-butanol or R-(-)-2-butanol. Attribution to the D- or L- absolute configuration was achieved by comparing the elution time of the samples with those of the monosaccharide standards. GC-MS was used to confirm the data obtained with GC and to identify all peaks present in the chromatograms.

4.4. Linkage analysis by methylation and GC-MS

Permethylation of polysaccharide 23F and 23A samples (0.5 mg), hydrolysis and derivatization to partially methylated alditol acetates (PMAA) was achieved following the methods described by Harris et al. [34] and Albersheim et al. [31], respectively. Poor recovery using these standard methods was addressed by additional steps: prior sonication of the polysaccharides as described in section 2.2, initial addition of a small amount of potassium dimsyl and CH_3I, in order to achieve some methylation of hydroxyl functions which aids solubilization, and by repeating the methylation step with potassium dimsyl and CH_3I for 30 min instead of 10 min. A second set of methylation experiments were performed using an even longer incubation time of 2 h. PMAA derivatives were analyzed by GC and GC-MS. Identification of the sugar type followed from retention times and the ring size and the linkage positions of the glycosidic bonds from the corresponding mass spectra.

Quantification of each sugar derivative was achieved by correcting the corresponding area of the gas chromatogram by an effective carbon response factor according to Sweet et al. [20]. GC analysis was performed on a Perkin–Elmer Autosystem XL gas chromatograph equipped with a flame ionisation detector and an HP-1 column (30 m); temperature program: 125 °C for 1 min, 125 – 240 °C at 4 °C/min, and 240 °C for 2 min. GC–MS (e.i.) analyses were carried out on an Agilent Technologies 7890A gas chromatograph coupled to an Agilent Technologies 5975C VL MSD, using an HP-1 column (30 m) and the same temperature program.

4.5. Smith degradation of polysaccharide 23A

Polysaccharide 23A (23 mg) was subjected to complete oxidation with 0.18 mmol of $NaIO_4$ at 10 °C for 6 days in the dark [35,36]. The reaction was stopped by the addition of glycerol and the products were reduced with $NaBH_4$. Addition of 50% CH_3COOH after 16 h destroyed the excess of reducing reagent, the sample was dialysed and the product recovered by lyophilisation. Mild hydrolysis (0.5 M TFA) was conducted at room temperature for 6 days. The solution was taken to dryness under reduced pressure, dissolved in water, its pH adjusted to neutrality, and the product recovered under reduced pressure. It was then separated on a Bio Gel P2 column (1.6 cm i.d. x 90 cm) equilibrated in 50 mM $NaNO_3$ which was also used as eluent. The flow rate was 6 mL/h and fractions were collected at 15 min intervals. Elution was monitored using a refractive index detector (WGE Dr. Bures, LabService Analitica) which was connected to a paper recorder and interfaced with a computer via picolog software. One major oligosaccharide, named **23ASD**, was obtained from the chromatographic separation and purified by dialysis (Float-A-Lyzer, MWCO 100-500 Da) and treatment with MTO-Dowex marathon (H^+, OH^-) resin to remove residual salt. The **23ASD** oligosaccharide was fully characterized by NMR spectroscopy.

4.6. NMR spectroscopy

Polysaccharide samples (~10 mg) were lyophilized and exchanged twice with 99.9 % deuterium oxide (Sigma Aldrich), then dissolved in 600 µL of D_2O and introduced into a 5 mm NMR tube for data acquisition. Preliminary NMR studies yielded broad lines and poor 2D crosspeaks for polysaccharide 23A and 23F, the spectral resolution was improved by placing the NMR sample in a Branson 1200 Sonicator water bath for 1-2 days. 1D ^{1}H, ^{13}C and ^{31}P and 2D, COSY, TOCSY, NOESY, HSQC, HMBC and hybrid H2BC, HSQC-TOCSY and HSQC-NOESY NMR spectra were obtained using a Bruker Advance III 600 MHz NMR spectrometer equipped with a BBO Prodigy cryoprobe and processed using standard Bruker software (Topspin 3.2). The probe temperature was set at 313 or 323 K. 2D TOCSY experiments were performed using mixing times of 120 or 180 ms and the 1D variants using mixing times up to 200 ms. The HSQC experiment was optimized for J = 145 Hz (for directly attached ^{1}H-^{13}C correlations), and the HMBC experiment optimized for a coupling constant of 6 Hz (for long-range ^{1}H-^{13}C correlations). HSQC-TOCSY and HSQC-NOESY NMR spectra were recorded using mixing times of 120 and 250 ms respectively. Polysaccharide spectra were referenced to residual cell wall polysaccharide signals (phosphocholine ^{1}H signal at 3.23 ppm and ^{13}C signal at 54.5 ppm and the shielded ^{31}P signal at 1.30 ppm) [37]. Spectra recorded for oligosaccharide **23ASD** were referenced relative to H6/C6 of β-Rha: ^{1}H at 1.37 ppm, ^{13}C at 17.5 ppm.

4.7. Molecular modeling

Optimal dihedral angle conformations for the glycosidic linkages were taken from the corresponding disaccharide potential of mean force free energy surfaces calculated with the metadynamics routine incorporated into NAMD [38], with the ϕ,ψ glycosidic linkage torsion

angles used as collective variables. The optimal conformations are listed in Table 5. Molecular models of 10 repeat units of 23F, 23A and 23B were built with CarbBuilder version 2.1.17 [26] using the dihedral angles listed in Table 5. We added bond, angle and dihedral parameters to the CHARMM36 additive force field for carbohydrates [39,40] to represent the 2-phosphate substitution on glycerol, as well as the glycosidic phosphodiester (2->3) linkage. These parameters were adapted from the ribitol phosphodiester parameters previously added to the force field [41]. These initial oligosaccharide structures were optimized through 20000 steps of standard NAMD (version 2.9) minimization in vacuum [42].

Acknowledgements

This project was funded by LimmaTech Biologics AG. Computations were performed using facilities provided by the University of Cape Town's ICTS High Performance Computing team: http://hpc.uct.ac.za. The authors thank the South African National Research Foundation for NMR equipment funding (NRF Grant 86038).

Table 1

Determination of the glycosidic linkages in pneumococcal polysaccharides 23F and 23A by GC-MS of PMAA derivatives.

Linkage[a]		Relative molar ratio[c]			
	RRT[b]	I[d]	II[e]	III[f]	IV[g]
t-Rha	0.60	0.48	0.14	0.20	0.15
4-Rha	0.75	0.61	0.64		
3,4-Rha	0.87			n.d.[h]	0.94
4-Glc	1.00	1.00	1.00	1.00	1.00
2-Gal	1.01			0.13	0.11
2,3-Gal	1.13			0.50	0.59
2,4-Gal	1.13	0.21	0.22		
2,3,4-Gal	1.21	0.29	0.41		

[a] the numbers indicate the position of the glycosidic linkages, e.g. t-Rha= terminal non-reducing rhamnose; [b] Relative retention time; [c] Peak areas were corrected by the effective carbon response factor (Sweet et al., 1975,[20]) and the molar ratio are expressed relative to 4-Glc (set as 1.00); [d] I = Pn23F polysaccharide methylated for 30 min; [e] II = Pn23F polysaccharide methylated for 2 h; [f] III = Pn23A polysaccharide methylated for 30 min; [g] IV = Pn23A polysaccharide methylated for 2 h; [h] n.d. = not detected.

Table 2

^{1}H and ^{13}C NMR chemical shifts (δ, ppm) for the serotype 23F polysaccharide repeating unit

Residue	H-1 C-1	H-2 C-2	H-3 C-3	H-4 C-4	H-5 C-5	H-6 C-6
α-L-Rha*p*-(1→	5.10	4.15	3.82	3.47	4.10	1.27
α-R	101.66	70.30	70.51	72.37	69.20	16.94
→2,3,4)-β-D-Gal*p*-(1→	4.95	3.82	4.33	4.42	3.81	3.94
GA	101.22	75.69	78.23	74.34	74.45	61.08
→ 4)-β-L-Rha*p*-(1→	4.86	4.04	3.80	3.70	3.44	1.36
β-R	101.05	71.55	73.91	77.79	71.41	17.59
→4)-β-D-Glc*p*-(1→	4.83	3.36	3.68	3.64	3.53	3.94, 3.83
G	102.76	73.81	75.88	77.61	74.78	61.42

Phosphoglycerol at O-3 of Gal: ^{1}H, ^{13}C and ^{31}P assignments; δ CH: 4.29, 77.49; δ CH_2: 3.77, 62.03 and ^{31}P at -0.09 ppm.

Table 3

^{1}H and ^{13}C NMR chemical shifts (δ, ppm) for the serotype 23A polysaccharide repeating unit (upper panel) and for the Smith degradation product **23ASD** (lower panel)

Residue	**H-1** **C-1**	**H-2** **C-2**	**H-3** **C-3**	**H-4** **C-4**	**H-5** **C-5**	**H-6** **C-6**
α-L-Rha*p*-(1→	5.06	4.13	3.82	3.49	4.07	1.27
α-R	101.50	70.28	70.85	72.36	69.06	16.93
→2,3)-β-D-Gal*p*-(1→	5.17	3.67	4.25	4.18	3.66	~3.81
GA	99.53	75.79	78.63	68.51	74.89	61.31
→3,4)-β-L-Rha*p*-(1→	4.91	4.32	3.95	3.94	3.45	1.37
β-R	100.83	71.37	83.23	73.93	71.45	17.51
→4)-β-D-Glc*p*-(1→	4.67	3.40	3.66	3.65	3.54	3.91, 3.85
G	104.00	73.69	76.03	77.21	74.95	61.15
β-D-Gal*p*-(1→	4.64	3.53	3.65	3.91	3.67	~3.77
GA	104.4	72.4	73.5	69.3	75.8	61.5
→4)-β-L-Rha*p*-(1→	4.77	4.04	3.82	3.63	3.48	1.37
β-R	99.8	71.5	73.4	81.6	71.4	17.5

Residue	H-1 C-1	H-2 C-2	H-3 C-3	H-4 C-4	H-5 C-5	H-6 C-6
→2)-Threitol	3.72, 3.82	3.83	3.83	3.68, 3.73		
T	61.2	80.6	71.6	63.3		

Phosphoglycerol at O-3 of Gal: ^{1}H, ^{13}C and ^{31}P assignments; δ CH: 4.27, 77.51; δ CH_2: 3.77, 61.96 and ^{31}P at -0.68 ppm.

Table 4

^{1}H and ^{13}C NMR chemical shifts (δ, ppm) for the serotype 23B polysaccharide repeating unit

Residue	H-1 C-1	H-2 C-2	H-3 C-3	H-4 C-4	H-5 C-5	H-6 C-6
→3,4)-β-D-Gal*p*-(1→	4.77	3.80	4.26	4.39	3.73	3.82, 3.76
GA	104.10	71.34	77.93	74.07	74.77	60.92
→ 4)-β-L-Rha*p*-(1→	4.86	4.09	3.82	3.63	3.49	1.37
β-R	101.02	71.02	73.27	81.79	71.12	17.35
→4)-β-D-Glc*p*-(1→	4.81	3.34	3.68	3.64	3.53	3.94, 3.82
G	102.76	73.87	75.91	77.69	74.79	61.43

Phosphoglycerol at O-3 of Gal: ^{1}H, ^{13}C and ^{31}P assignments; δ CH: 4.29, 77.69; δ CH_2: 3.76, 61.9; ^{31}P at -0.26 ppm.

Table 5

Optimal values for the ϕ,ψ glycosidic linkage torsion angles determined from vacuum metadyamics.

Disaccharide	ϕ, ψ
α-L-Rha*p*-(1→2)-β-D-Gal*p*	39, 21
β-D-Glc*p*-(1→3)-β-L-Rha*p*	46, 11; **59 -13**[a]
β-D-Gal*p*-(1→4)-β-L-Rha*p*	26, 26
β-D-Glc*p*-(1→4)-β-D-Gal*p*	44,16
β-L-Rha*p*-(1→4)-β-D-Glc*p*	-51, -8

[a] Value used for 23A, to avoid atomic collisions. This value is still within the vacuum global minimum energy well

List of Figures

repeating unit proton/carbon crosspeaks have been labeled according to the carbon atom of the corresponding residue β-R = β-Rha, G = Glc, GA = Gal and Gro = glycerol).

Figure 8: Minimized molecular models for 10RU of 23F (left) 23A (middle) and 23B (right), shown in space-filling representation and colored according to residue type. The models for 23F and 23B show a similar loose helical conformation, the model for 23A is a slightly twisted flat ribbon, with clear steric crowding at the β-L-Rha branch point.

Figure 9: Proposed glycosyltransferase and polymerase activity in serogroup 23 polysaccharides. Glycosyltransferases responsible for each elongation step are listed above the respective glycosidic linkage in italics. The polymerization site is marked by an arrow.

References

[1] A.J. van Tonder, J.E. Bray, S.J. Quirk, G. Haraldsson, K.A. Jolley, M.C. Maiden, S. Hoffmann, S.D. Bentley, Á. Haraldsson, H. Erlendsdóttir, K.G. Kristinsson, Microb. Genom. 2 (2016) 1-17.

[2] J.C. Richards, M.B. Perry, Biochem. Cell Biol. 66 (1988) 758-771.

[3] J.B. Robbins, R. Austrian. C.J. Lee, S.C. Rastogi, G. Schiffman, J. Henrichsen, P.H. Mäkelä, C.V. Broome, R.R. Facklam, R.H. Tiesjema, J.C. Parke, J. Infect. Dis. 148 (1983) 1136-1159.

[4] E.A. De Velasco, A.F. Verheul, A.M. Van Steijn, H.A. Dekker, R.G. Feldman, I.M. Fernandez, J.P. Kamerling, J.F. Vliegenthart, J. Verhoef, H. Snippe, Infect. Immun. 62 (1994) 799-808.

[5] S. Park, M.H. Nahm, PLoS One. 8 (2013) e83810.

[6] R.E. Gertz, Z. Li, F.C. Pimenta, D. Jackson, B.A. Juni, R. Lynfield, J.H. Jorgensen, M. da Gloria Carvalho, B.W. Beall, Infect. Dis. 201 (2010) 770-775.

[7] S.S. Richter, D.J. Diekema, K.P. Heilmann, C.L. Dohrn, F. Riahi, G.V. Doern, Antimicrob. Agents Chemother. 58 (2014) 6484-6489

[8] R.A. Gladstone, J.M. Jefferies, A.S. Tocheva, K.R. Beard, D. Garley, W.W. Chong, S.D. Bentley, S.N. Faust, S.C. Clarke, Vaccine. 33 (2015) 2015-2021.

[9] M. van der Linden, S. Perniciaro, M. Imöhl, BMC Infect. Dis. 15 (2015) 207.

[10] C. Hays, Q. Vermee, A. Agathine, A. Dupuis, E. Varon, C. Poyart, M.C. Ploy, J. Raymond, J. Eur. Clin. Microbiol. Infect. Dis. (2016) 1-8.

[11] V.T. Devine, D.W. Cleary, J.M. Jefferies, R. Anderson, D.E. Morris, A.C. Tuck, R.A. Gladstone, G. O'Doherty, P. Kuruparan, S.D. Bentley, S.N. Faust, Vaccine. 35 (2017) 1293-1298.

[12] S.D. Bentley, D.M. Aanensen, A. Mavroidi, D. Saunders, E. Rabbinowitsch, M. Collins, K. Donohoe, D. Harris, L. Murphy, M.A. Quail, G. Samuel, PLoS Genet. 2 (2006) e31.

[13] J.K. Morona, D.C. Miller, T.J. Coffey, C.J. Vindurampulle, B.G. Spratt, R. Morona, J.C. Paton, Microbiology. 145 (1999) 781-789.

[14] J. Yother, Annu. Rev. Microbiol. 65 (2011) 563-581.

[15] Q. Wang, Y. Xu, A.V. Perepelov, W. Xiong, D. Wei, A.S. Shashkov, Y.A. Knirel, L. Feng, L. Wang, J. Bacteriol. 192 (2010) 5506-5514.

[16] F. Kong, W. Wang, J. Tao, L. Wang, Q. Wang, A. Sabananthan, G.L. Gilbert, J. Med. Microbiol. 54 (2005) 351-356.

[17] L. Pelosi, M. Boumedienne, N. Saksouk, J. Geiselmann, R.A. Geremia, Biochem. Biophys. Res. Commun. 327 (2005) 857-865.

[18] R. Follador, E. Heinz, K.L. Wyres, M.J. Ellington, M. Kowarik, K.E. Holt, N.R. Thomson, Microb. Genom. 2 (2016) e000073.

[19] Y.M. Choy, G.G. Dutton, Can. J. Biochem. 51 (1973) 3021-3026.

[20] D.P. Sweet, R.H. Shapiro, P. Albersheim, Carbohydr. Res. 40 (1975) 217-225.

[21] P.E. Jansson, L. Kenne, G. Widmalm, Carbohydr. Res. 188 (1989) 169-191.

[22] P.E. Jansson, L. Kenne, T. Wehler, Carbohydr. Res. 179 (1988) 359-368.

[23] J.Ø. Duus, C.H. Gotfredsen, K. Bock, Chem. Rev. 100 (2000) 4589-4614.

[24] C. Jones, C. Whitley, X. Lemercinier, Carbohydr. Res. 325 (2000) 192-201.

[25] M. Lundborg, C. Fontant, G. Widmalm, Biomacromolecules. 12 (2011) 3851-3855.

[26] M.M. Kuttel, J. Ståhle, G. Widmalm, J. Comput. Chem. 37 (2016) 2098-2105.

[27] K.A. Geno, J.S. Saad, M.H. Nahm, J. Clin. Microbiol. (2017) 1416-1425.

[28] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T.L. Madden, BMC Bioinf. 10 (2009) 421.

[29] C. Notredame, D.G. Higgins, J. Heringa, J. Mol. Biol. 302 (2000) 205-217.

[30] A. Stamatakis, Bioinformatics. 22 (2006) 2688-2690.

[31] P. Albersheim, D.J. Nevins, P.D. English, A. Karr, Carbohydr. Res. 5 (1967) 340-345.

[32] J.S. Kim, E.R. Laskowich, R.G. Arumugham, R.E. Kaiser, G.J. MacMichael, Anal. Biochem. 347 (2005) 262-274.

[33] G.J. Gerwig, J.P. Kamerling, J.F. Vliegenthart, Carbohydr. Res. 77 (1979) 1-7.

[34] P.J. Harris, R.J. Henry, A.B. Blakeney, B.A. Stone, Carbohydr. Res. 127 (1984) 59-73.

[35] G.W. Hay, B.A. Lewis, F. Smith, Methods Carbohydr. Chem. 5 (1965) 357-361.

[36] I.J. Goldstein, G.W. Hay, B.A. Lewis, F. Smith, Methods Carbohydr. Chem. 5 (1965) 361-370.

[37] S. Vialle, P. Sepulcri, J. Dubayle, P. Talaga, Carbohydr. Res. 340 (2005) 91-96.

[38] A. Laio, M. Parrinello, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 12562-12565.

[39] O. Guvench, E. Hatcher, R.M. Venable, R.W. Pastor, A.D. MacKerell Jr, J. Chem. Theory Comput. 5 (2009) 2353-2370.

[40] S.S. Mallajosyula, O. Guvench ,E. Hatcher, A.D. MacKerell Jr, J. Chem. Theory Comput. 8 (2012) 759-776.

[41] M.M. Kuttel, G.E. Jackson, M. Mfata, N. Ravenscroft, Carbohydr. Res. 406 (2015) 27-33.

[42] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, K. Schulten, J. Comput. Chem. 26 (2005) 1781-1802.

Figure 1

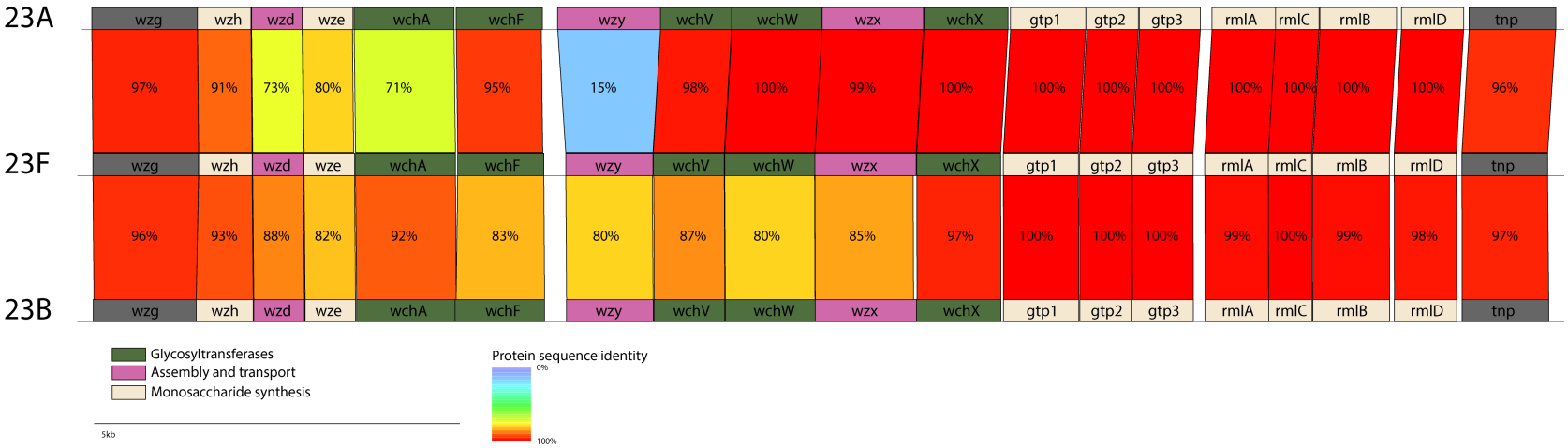


Figure 2

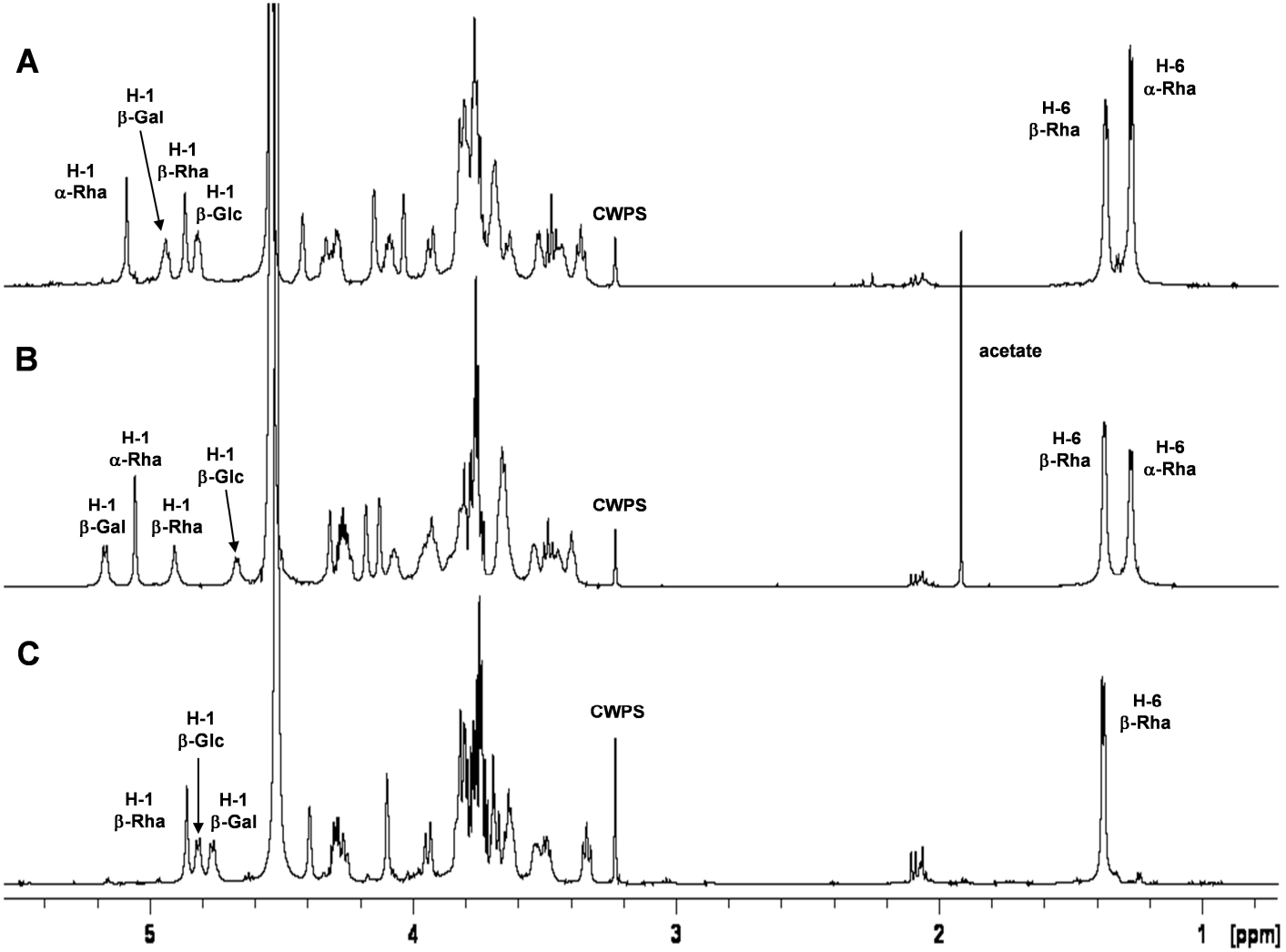

A
H-1
α-Rha
H-1
β-Gal
H-1
β-Rha
H-1
β-Glc
CWPS
H-6
β-Rha
H-6
α-Rha
B
acetate
H-1
α-Rha
H-1
β-Glc
H-1
β-Gal
H-1
β-Rha
CWPS
H-6
β-Rha
H-6
α-Rha
C
H-1
β-Glc
CWPS
H-6
β-Rha
H-1
β-Rha
H-1
β-Gal
5
4
3
2
1
[ppm]

Figure 3

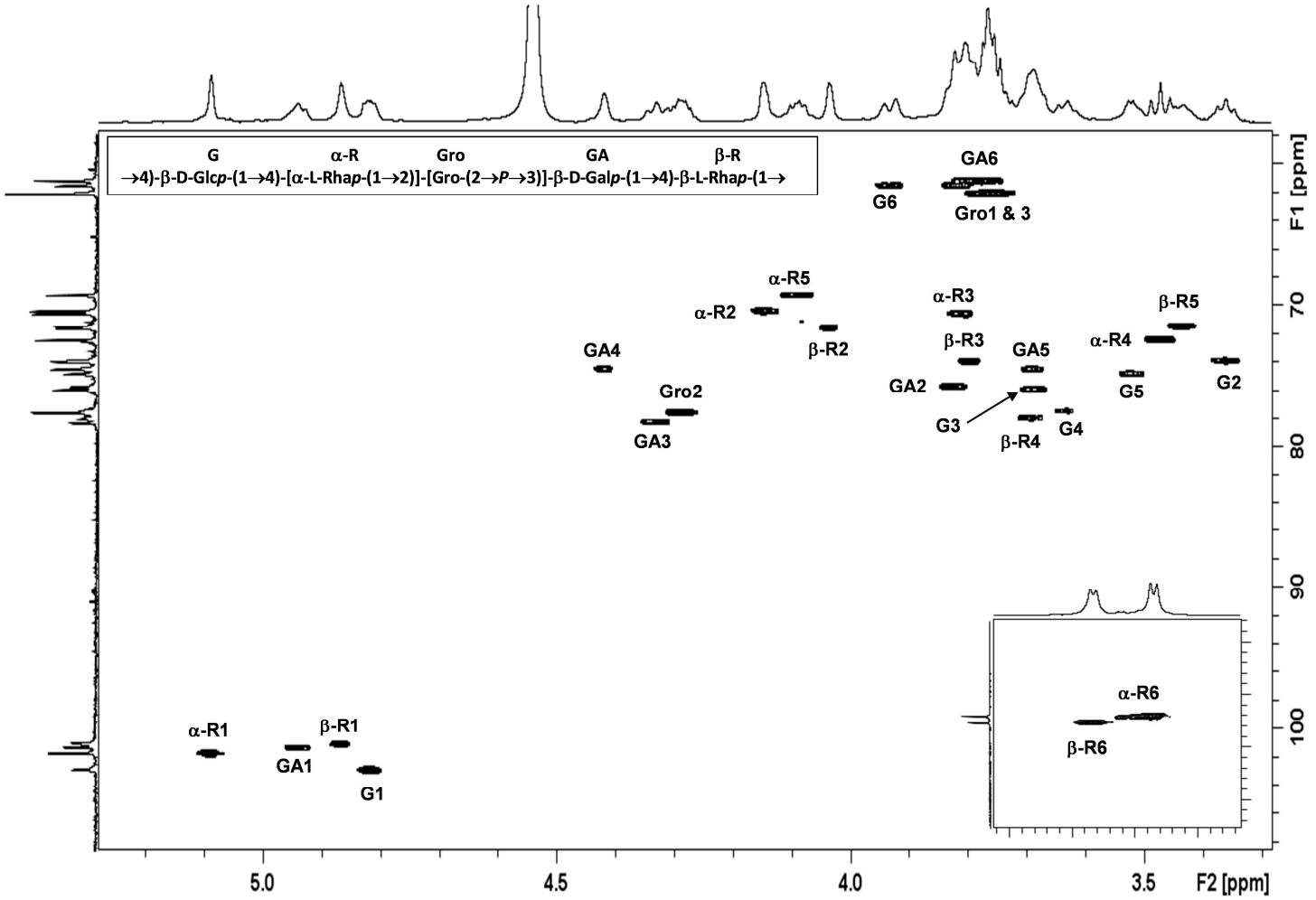

G α-R Gro GA β-R
→4)-β-D-Glcp-(1→4)-[α-L-Rhap-(1→2)]-[Gro-(2→P→3)]-β-D-Galp-(1→4)-β-L-Rhap-(1→
G6
GA6
Gro1 & 3
α-R5
α-R2
β-R2
α-R3
β-R3
GA4
Gro2
GA3
GA5
GA2
G3
β-R4
G4
β-R5
α-R4
G5
G2
α-R1
β-R1
GA1
G1
α-R6
β-R6
F1 [ppm]
70
80
90
100
5.0
4.5
4.0
3.5
F2 [ppm]

Figure 4

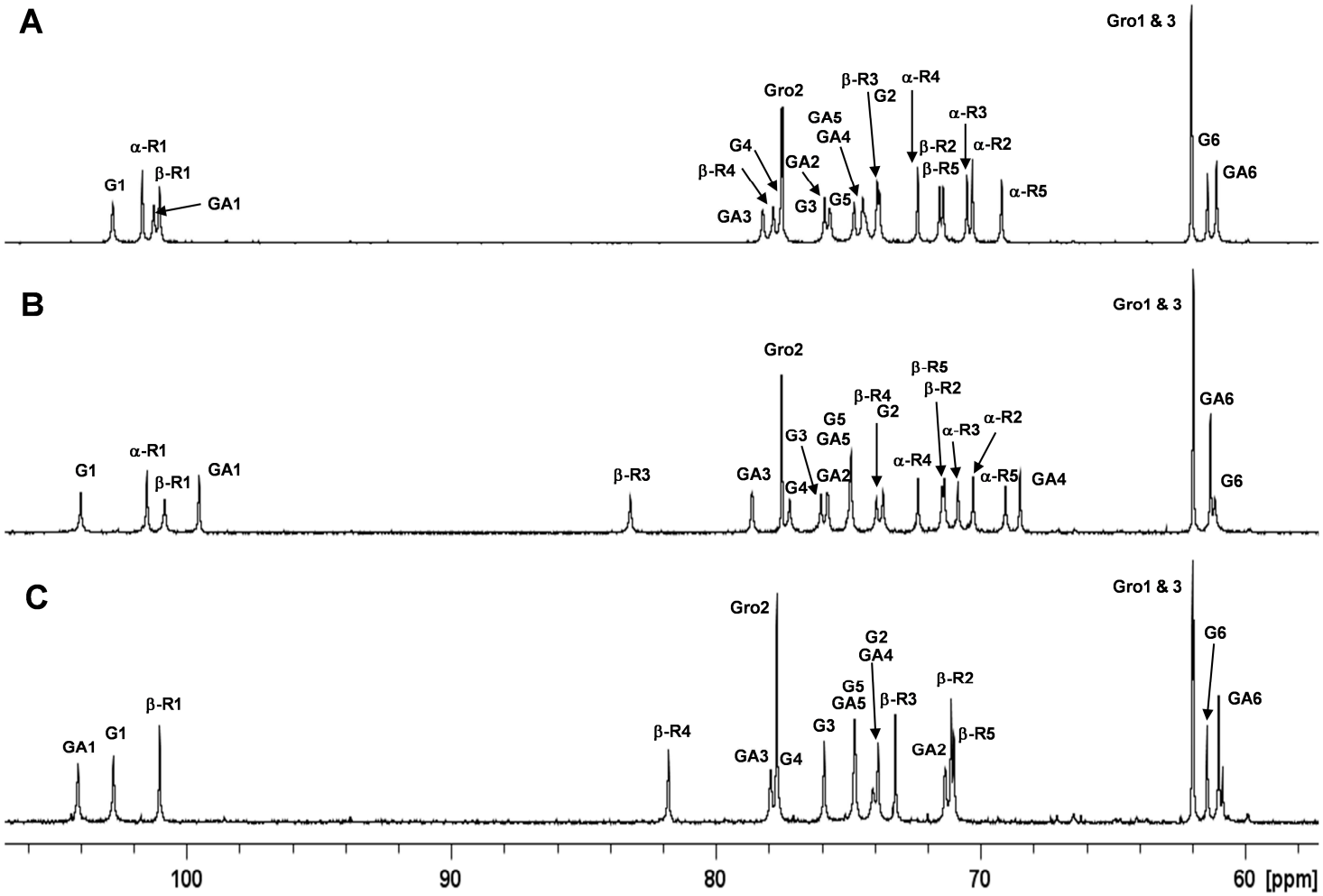


Figure 5

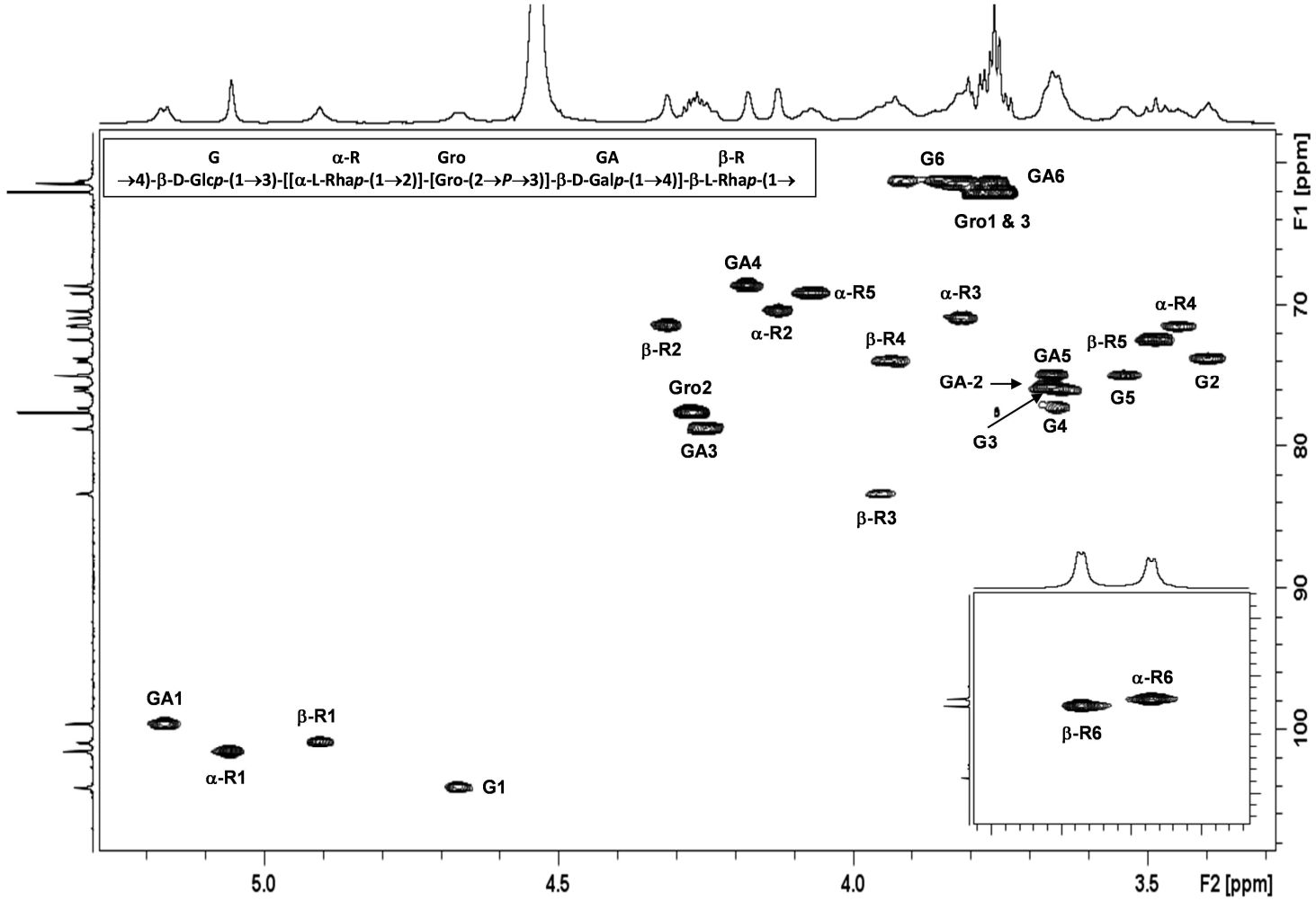

G α-R Gro GA β-R
→4)-β-D-Glcp-(1→3)-[[α-L-Rhap-(1→2)]-[Gro-(2→P→3)]-β-D-Galp-(1→4)]-β-L-Rhap-(1→
G6
GA6
Gro1 & 3
GA4
α-R5
α-R3
α-R4
α-R2
β-R2
β-R4
β-R5
GA5
GA-2
G5
G2
Gro2
G4
G3
GA3
β-R3
α-R6
β-R6
GA1
β-R1
α-R1
G1
F1 [ppm]
70
80
90
100
5.0
4.5
4.0
3.5
F2 [ppm]

Figure 6

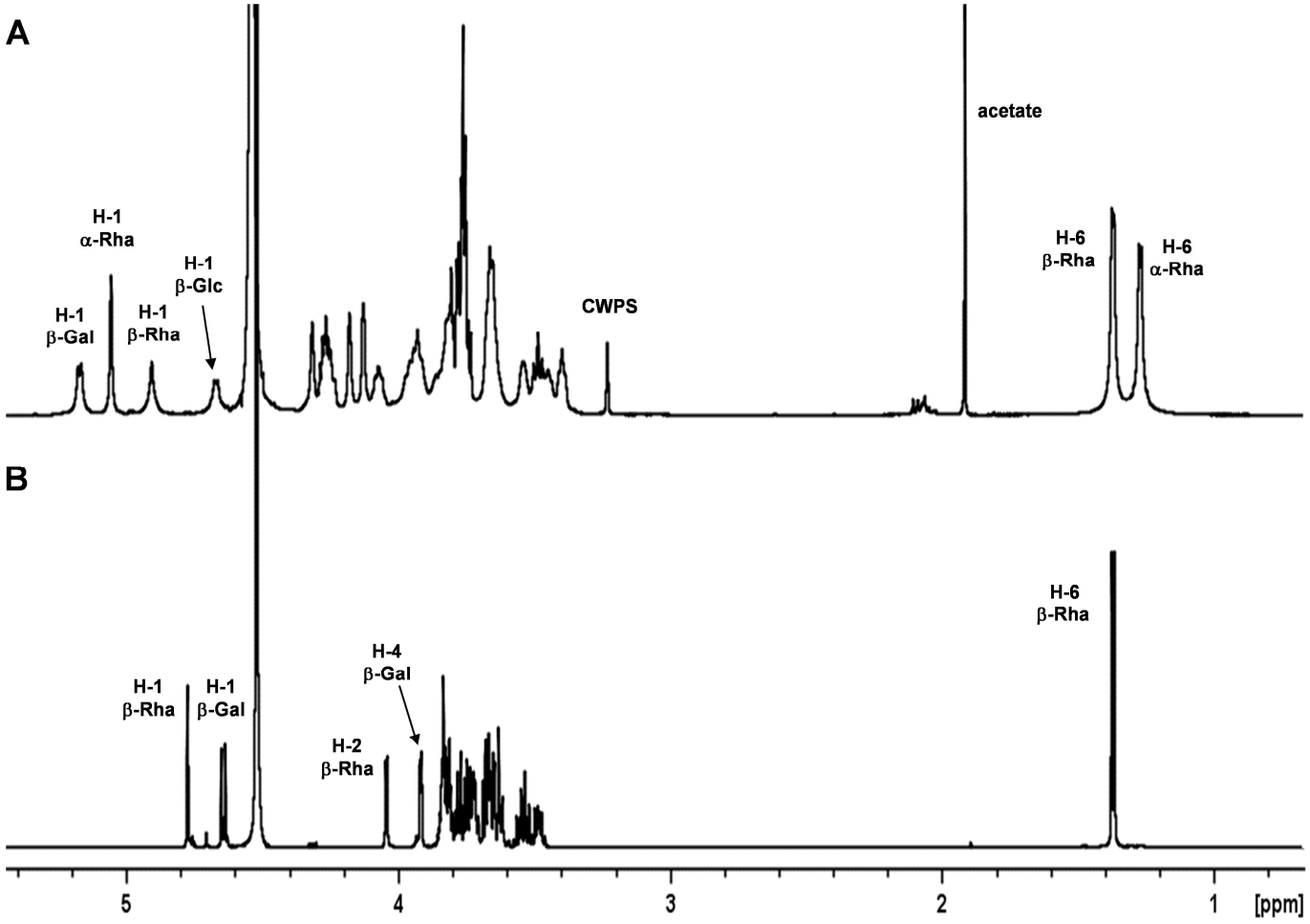

A
H-1
α-Rha
H-1
β-Gal
H-1
β-Rha
H-1
β-Glc
CWPS
acetate
H-6
β-Rha
H-6
α-Rha
B
H-1
β-Rha
H-1
β-Gal
H-2
β-Rha
H-4
β-Gal
H-6
β-Rha
5
4
3
2
1
[ppm]

Figure 7

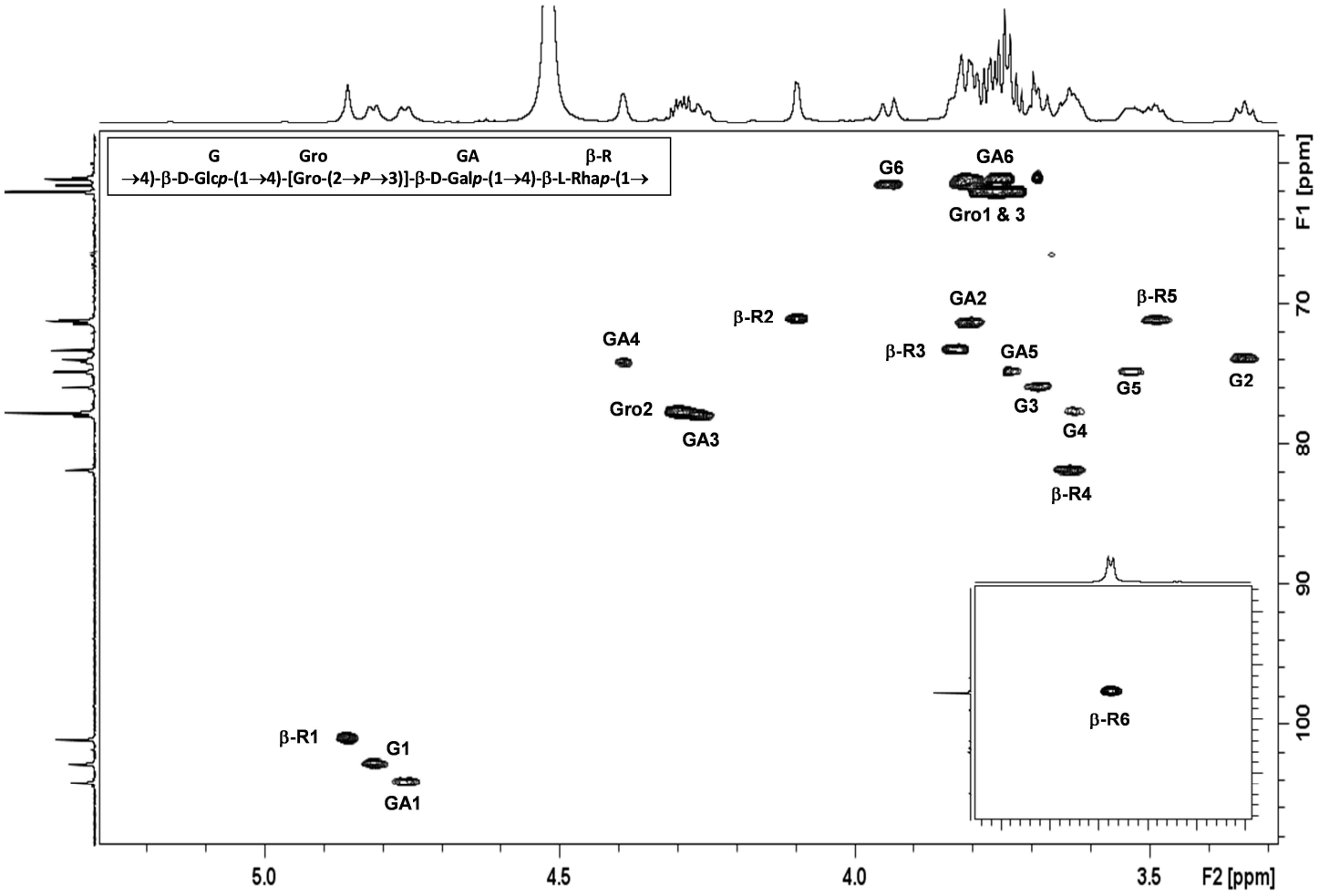

G Gro GA β-R
→4)-β-D-Glcp-(1→4)-[Gro-(2→P→3)]-β-D-Galp-(1→4)-β-L-Rhap-(1→
G6
GA6
Gro1 & 3
GA2
β-R5
β-R2
GA4
β-R3
GA5
G5
G2
G3
Gro2
G4
GA3
β-R4
β-R6
β-R1
G1
GA1
5.0
4.5
4.0
3.5
F2 [ppm]
F1 [ppm]
70
80
90
100

Figure 8

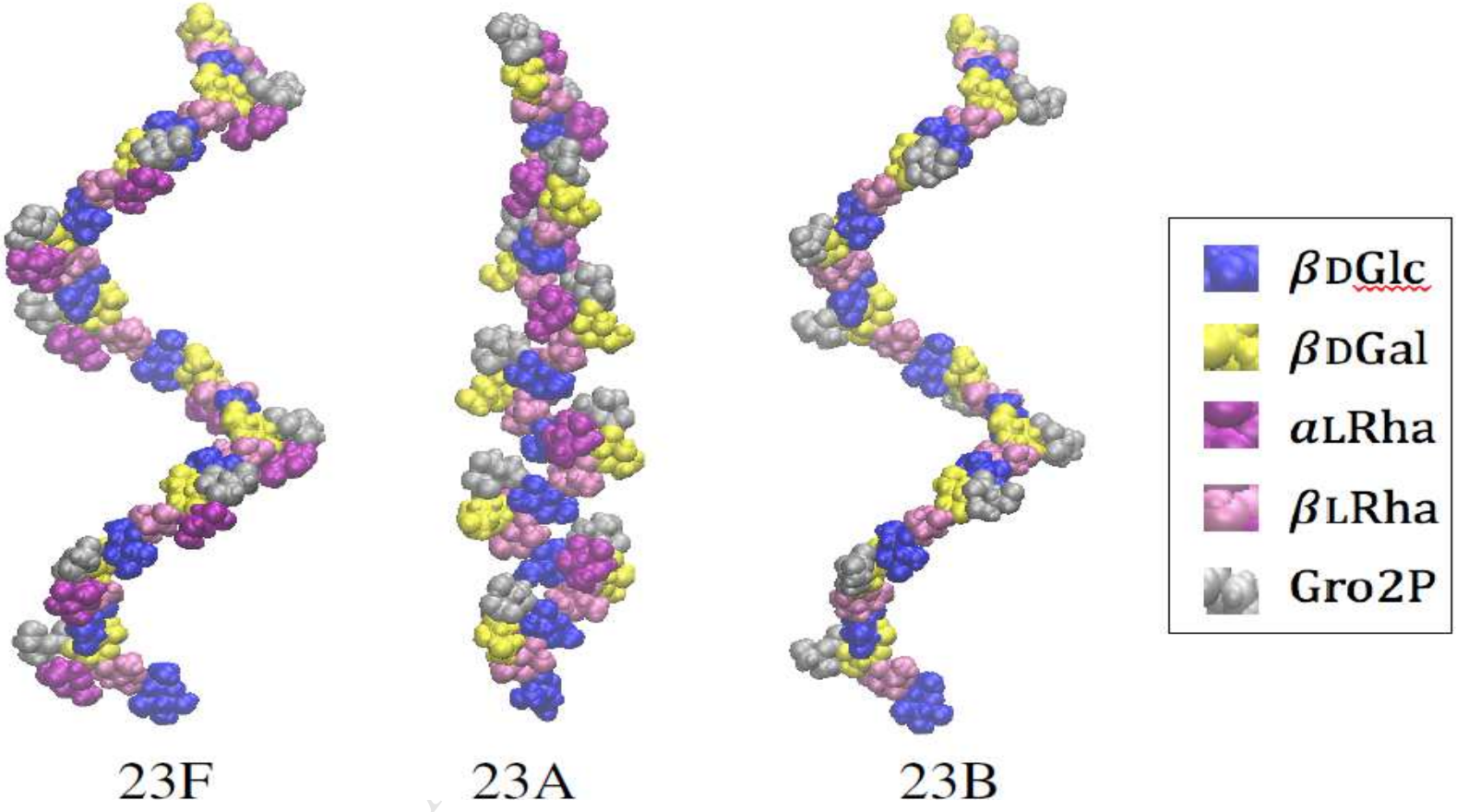
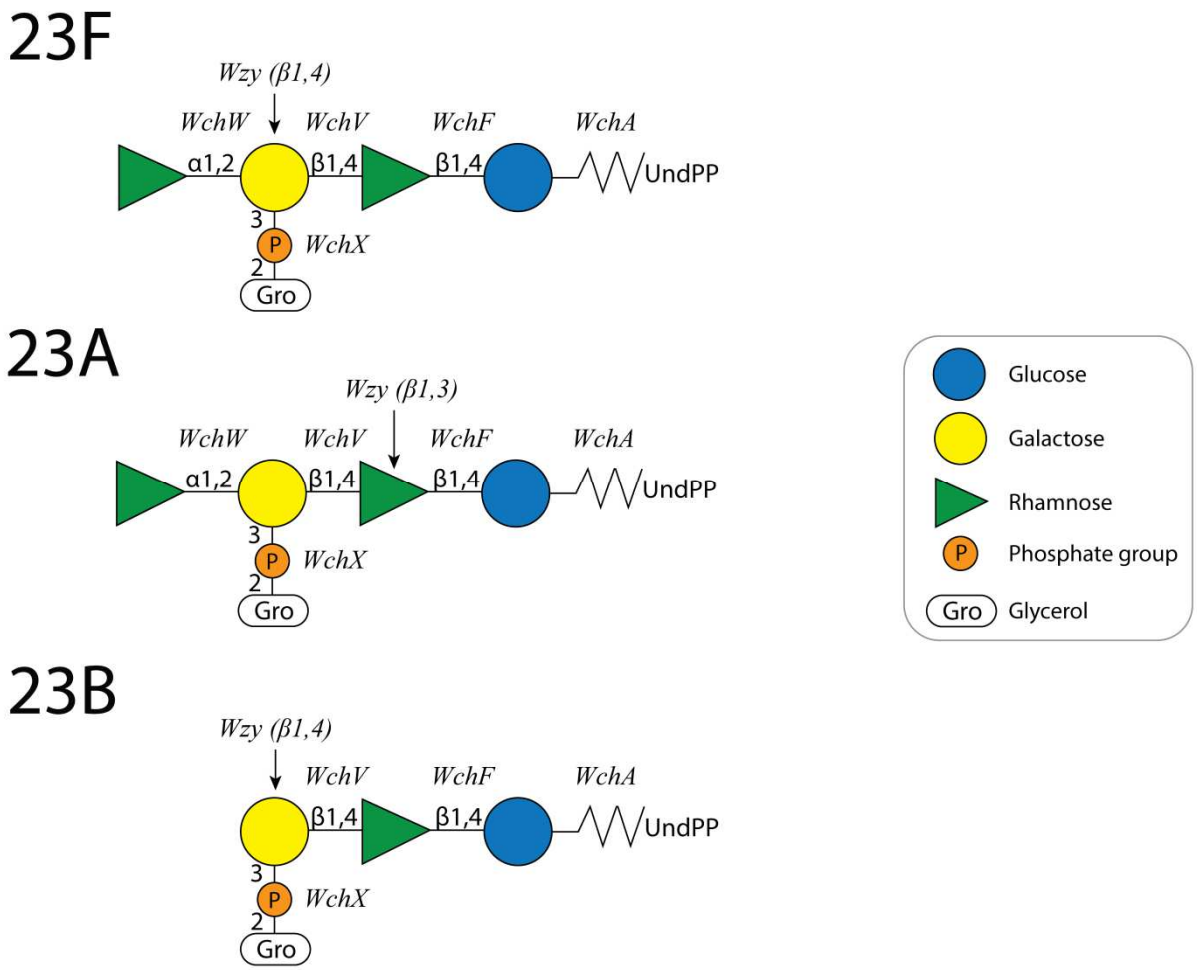


Figure 9


23F
Wzy (β1,4)
WchW
WchV
WchF
WchA
α1,2
β1,4
β1,4
UndPP
3
P
WchX
2
Gro
23A
Wzy (β1,3)
WchW
WchV
WchF
WchA
α1,2
β1,4
β1,4
UndPP
3
P
WchX
2
Gro
23B
Wzy (β1,4)
WchV
WchF
WchA
β1,4
β1,4
UndPP
3
P
WchX
2
Gro
Glucose
Galactose
Rhamnose
P
Phosphate group
Gro
Glycerol

Highlights

- Serotype 23A and 23B subunit structures were predicted from genome sequence analysis
- The structures of the 23A and 23B polysaccharides were elucidated
- The 23A polysaccharide has a disaccharide backbone and disubstituted 2,3-β-Gal as a side chain
- The 23B polysaccharide is the same as 23F but without the terminal α-Rha