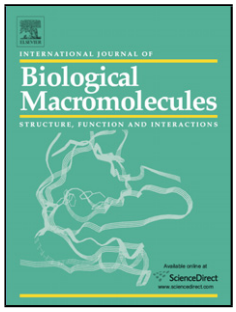
Accepted Manuscript



Title: *Acinetobacter baumannii* K11 and K83 capsular polysaccharides have the same 6-deoxy-L-talose-containing pentasaccharide K units but different linkages between the K units

Authors: Johanna J. Kenyon, Alexander S. Shashkov, Sof'ya N. Senchenkova, Mikhail M. Shneider, Bin Liu, Anastasiya V. Popova, Nikolay P. Arbatsky, Konstantin A. Miroshnikov, Lei Wang, Yuriy A. Knirel, Ruth M. Hall

PII: S0141-8130(17)30573-1
DOI: http://dx.doi.org/doi:10.1016/j.ijbiomac.2017.05.082
Reference: BIOMAC 7571

To appear in: *International Journal of Biological Macromolecules*

Received date: 17-2-2017
Revised date: 18-4-2017
Accepted date: 15-5-2017

Please cite this article as: Johanna J.Kenyon, Alexander S.Shashkov, Sof'ya N.Senchenkova, Mikhail M.Shneider, Bin Liu, Anastasiya V.Popova, Nikolay P.Arbatsky, Konstantin A.Miroshnikov, Lei Wang, Yuriy A.Knirel, Ruth M.Hall, Acinetobacter baumannii K11 and K83 capsular polysaccharides have the same 6-deoxy-L-talose-containing pentasaccharide K units but different linkages between the K units, International Journal of Biological Macromoleculeshttp://dx.doi.org/10.1016/j.ijbiomac.2017.05.082

Acinetobacter baumannii K11 and K83 capsular polysaccharides have the same 6-deoxy-L-talose-containing pentasaccharide K units but different linkages between the K units

Johanna J. Kenyon,[a,b,†,*] Alexander S. Shashkov,[c,†] Sof'ya N. Senchenkova,[c,†] Mikhail M. Shneider,[d] Bin Liu,[e] Anastasiya V. Popova,[f,g] Nikolay P. Arbatsky,[c] Konstantin A. Miroshnikov,[d] Lei Wang,[e] Yuriy A. Knirel,[c] Ruth M. Hall[b,h]

[a] *Institute of Health and Biomedical Innovation, School of Biomedical Sciences, Queensland University of Technology, Brisbane, Australia*

[b] *School of Molecular Bioscience, The University of Sydney, Sydney, Australia*

[c] *N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Moscow, Russia*

[d] *M. M. Shemyakin & Y. A. Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia*

[e] *TEDA Institute of Biological Sciences and Biotechnology, Nankai University, TEDA, Tianjin, PR China*

[f] *Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region, Russia*

[g] *State Research Center for Applied Microbiology and Biotechnology, Obolensk, Moscow Region, Russia*

[h] *School of Life and Environmental Sciences, The University of Sydney, Sydney, Australia*

[*] Corresponding author. Tel.: +617 3138 2552

E-mail address: johanna.kenyon@qut.edu.au (Johanna J. Kenyon).

[†]Contributed equally

Abstract

Acinetobacter baumannii produces a variety of capsular polysaccharides (CPS) via genes located at the chromosomal K locus and some KL gene clusters include genes for the synthesis of specific sugars. The structures of K11 and K83 CPS produced by isolates LUH5545 and LUH5538, which carry related KL11a and KL83 gene clusters, respectively, were established by sugar analysis and one- and two-dimensional ^{1}H and ^{13}C NMR spectroscopy. Both CPS contain L-rhamnose (L-Rha) and 6-deoxy-L-talose (L-6dTal), and both KL gene clusters include genes for dTDP-L-Rha*p* synthesis and a *tle* (talose epimerase) gene encoding an epimerase that converts dTDP-L-Rha*p* to dTDP-L-6dTal*p*. The K11 and K83 repeat units are the same pentasaccharide, consisting of D-glucose, L-Rha, L-6dTal, and *N*-acetyl-D-glucosamine, except that L-6dTal is 2-O-acetylated in K83. However, the K units are linked differently, with L-Rha in the main chain in K11, but as a side-branch in K83. KL11 and KL83 encode unrelated Wzy polymerases that link the K units together and different acetyltransferases, though only Atr8 from KL83 is active. The substrate specificity of each Wzy polymerase was assigned, and the functions of all glycosyltransferases were predicted. The CPS structures produced by three closely related K loci, KL29, KL105 and KL106, were also predicted.

Abbreviations

6dTal, 6-deoxytalose; Atr, acetyltransferase; COSY, correlation spectroscopy; CPS, capsular polysaccharide; dTDP, deoxythymidine diphosphate; GLC, gas-liquid chromatography; Gtr, glycosyltransferase; HMBC, heteronuclear multiplebond correlation; HSQC, heteronuclear single-quantum coherence; KL, K locus; MLST, multilocus sequence typing; NMR, nuclear magnetic resonance; PSgc, polysaccharide gene cluster; Rha, rhamnose; ROESY, rotating frame Overhauser effect spectroscopy; TOCSY, total correlation spectroscopy; UDP, uridine diphosphate; WGS, Whole Genome Shotgun.

Keywords: Acinetobacter baumannii; Capsular polysaccharide; 6-Deoxy-L-talose.

1. Introduction

Acinetobacter baumannii is an important human opportunistic nosocomial pathogen [1]. Different *A. baumannii* isolates produce structurally distinct capsular polysaccharides (CPS) on the cell surface, a number of which contain unusual monosaccharides. In the *A. baumannii* genome, the K locus (KL), which is located between *fkpA* and *lldP*, harbours the many different clusters of genes for the CPS synthesis [2,3]. Genes for the synthesis of specific sugars are often observed in the variable central region of the KL gene cluster that is flanked by sugar synthesis and CPS export genes that are shared by all *A. baumannii* KL gene clusters [2]. Bioinformatics methodologies and use of experimental information from other organisms have allowed predictions to be made about the sugars produced by the products of these genes, and these predictions can then be confirmed if these sugars are present in the CPS structure from the same isolate. To date, the role of genes predicted to be involved in the synthesis of specific sugars has been confirmed for seventeen gene modules, and in some cases the sugar composition and structures of CPS produced by multiple KL gene clusters carrying the same sugar synthesis gene module are available (for example, [4,5]). This approach allows genes to be unambiguously assigned to sugar biosynthesis pathways, broadening the capacity for accurate annotation of new KL.

There are now several cases where the sequence of the *A. baumannii* CPS gene cluster and the CPS structure produced by the same strain are available (as referenced in [6]). However, there are many further *A. baumannii* CPS gene cluster sequences that lack corresponding structural data [2,3,7]. For example, Hu et al. [7] reported the CPS gene cluster arrangements for 24 *A. baumannii* and 3 *A. nosocomialis* strains, but structural data was available for only 12 of them. In that study, the correlation of these CPS structures with their gene clusters revealed that at least two strain mixups had occurred, indicating the need to reinvestigate all of the CPS structures and gene clusters from this set of isolates. For three isolates (LUH5537, LUH5533, and LUH5550) the structures have since been

re-examined [8-10], and were found to be different from the structures previously reported for these strains but consistent with the available KL (PSgc) sequence.

Several *A. baumannii* gene cluster pairs, which differ only in a short segment containing one or two genes, have also been found [2,6]. For instance, related gene clusters that differ only in the sequence for the gene encoding the Wzy polymerase that is responsible for linking K units together to form the CPS have been found [5,11]. Determination of the structures of CPS from such pairs of isolates has revealed that they are comprised of the same K unit linked differently, and established the Wzy linkage specificities. This also informs the specificity of the Itr initiating transferase, which can be otherwise difficult to establish.

Here, we have determined the CPS sugar composition and structure of two further isolates from the collection studied by Hu et al., which carry related gene sets at KL [7]. This study establishes the role of further sugar synthesis genes and Wzy polymerases.

2. Materials and Methods

2.1. Bacterial strains and cultivation

A. baumannii LUH5538 and LUH5545 (laboratory stock numbers G4770 and G4777) [7] was originally from the W. H. Traub collection at the Institut für Medizinische Mikrobiologie und Hygiene, Universität des Saarlandes (Saarland, Germany). Bacteria were cultivated in 2TY media overnight; cells were harvested by centrifugation (10,000×*g*, 20 min), washed with phosphate-buffered saline, suspended in aqueous 70 % acetone, precipitated, and dried.

2.2. Isolation and O-deacetylation of CPSs

CPSs were isolated by phenol-water extraction [12] of bacterial cells, and the extract was dialyzed without layer separation and freed from insoluble contaminations by centrifugation. The resultant solution was treated with cold (4 °C) aqueous 50 % CCl_3CO_2H; after centrifugation the supernatant was dialyzed against distilled water and freeze-dried to give CPS preparations.

A sample of the LUH5538 CPS was treated with 12.5% aqueous ammonia at 37°C for 2 h, and an O-deacetylated CPS was isolated by gel-permeation chromatography on a column (80 × 1.6 cm) of TSK HW-40 (S) in 1 % AcOH monitored using a differential refractometer (Knauer, Germany).

2.3. Monosaccharide analyses

A CPS sample (1 mg) was hydrolyzed with 2 M CF_3CO_2H (120 °C, 2 h). Monosaccharides were analyzed by GLC of the alditol acetates on a Maestro (Agilent 7820) chromatograph (Interlab, Russia) equipped with an HP-5 column (0.32 mm × 30 m) using a temperature program of 160 °C (1 min) to 290 °C at 7 °C min^{-1}.

2.4. Smith degradation

A CPS sample of LUH5545 was oxidized with 0.1 M $NaIO_4$ in the dark for 72 h at 20°C, reduced with an excess of $NaBH_4$ and desalted by gel-permeation chromatography on TSK HW-40 (S) as described above. The polymer obtained was hydrolyzed with aqueous 2% HOAc for 2 h at 100°C, the products were reduced with $NaBH_4$ and fractionated by chromatography on the same gel to give oligosaccharide **1** (5.2 mg).

2.5. NMR spectroscopy

Samples were deuterium-exchanged by freeze-drying from 99.9 % D_2O and then examined as solutions in 99.95 % D_2O. NMR spectra were recorded on a Bruker Avance II 600 MHz spectrometer (Germany) at 20°C (for oligosaccharide **1**) or 70°C (for CPS). Sodium 3-trimethylsilylpropanoate-2,2,3,3-d_4 (δ_H 0, δ_C −1.6) was used as the internal reference for calibration. Two-dimensional NMR spectra were obtained using standard Bruker software, and Bruker TopSpin 2.1 program was used to acquire and process the NMR data. 60-ms MLEV-17 spin-lock time and 150-ms mixing time were used in the TOCSY and ROESY experiments, respectively. A 60-ms delay was used for evolution of long-range couplings to optimize $^1H,^{13}C$ HMBC experiments for coupling constant $J_{H,C}$ 8 Hz.

2.6. Bioinformatics

The KL11 gene cluster from the multiply antibiotic resistant Australian isolate, 99-015-2025 (also known as J9; equivalent to WM97a in [13]), was extracted from a draft genome sequence determined as described elsewhere [3], annotated and deposited into GenBank under accession number KF002790. The CPS gene cluster sequences from *A. baumannii* LUH5545 and LUH5538 were obtained from GenBank accession numbers KC526904.1 and KC526898.1, respectively. Both sequences were re-annotated using the nomenclature system established by Kenyon and Hall [2]. Functions of the gene products were assigned as described previously [6]. Additional KL sequences were retrieved from the WGS and non-redundant nucleotide databases, and annotated only if the complete gene cluster was present in no more than two contigs. MLST was performed by submitting draft genome sequences to https://pubmlst.org/abaumannii/.

3. Results

3.1. The KL11 and KL29 capsule biosynthesis gene clusters

The gene cluster for the synthesis of the K11 capsule was initially identified in the genome of the *A. baumannii* isolate, AB900 [14]. It shares 99% identity with the gene cluster (GenBank accession number KF002790) from strain 99-015-2025 also known as J9, a multiply antibiotic resistant isolate recovered in 1999 at Westmead hospital in Sydney, Australia. Annotation of the KL11 gene cluster in 99-015-2025 revealed four glycosyltransferase genes (*gtr26-29*), a gene coding for an ItrA3 initiating transferase, an acetyltransferase gene (*atr6*), and a *wzy* polymerase gene for the polymerisation of K units to form the CPS (Fig. 1). KL11 does not contain a *gne1* gene for the conversion of UDP-D-GlcNAc to UDP-D-GalNAc in the usual location between *gpi* and *pgm,* and it was predicted that the K11 structure would not include GalNAc.

KL11 also contains *rmlBDAC* genes for the synthesis of dTDP-L-Rha*p*. The predicted product of an additional gene located between the *atr6* and *gtr29* genes, matches mostly putative epimerases without a specific predicted function in a BLASTp similarity search. However, there are clear hits to a sequence annotated as a CDP-paratose 2-epimerase (71% identical to RfbE from *Faecalibacterium*

prausnitzii; GenPept accession number CUP38397.1) for production of CDP-tyvelose, and also a dTDP-L-Rha*p* 4-epimerase (69% identical to WbiB from *Mucilaginibacter gotjawali*; GenPept accession number BAU55752.1) for the conversion of dTDP-L-Rha*p* to dTDP-L-6dTal*p*. Given the presence of genes for synthesis of dTDP-L-Rha*p* in the KL11 gene cluster, dTDP-L-Tal*p* may be produced.

A closely related gene cluster, KL29, was found in three isolates, 2007-16-27-01, 2007-16-25-01-7, and TG27339 from the same study [15] and in isolate ABBL038 (GenBank accession number LLDS01000040.1). In KL29, a single *gtr* has been partially replaced, and an *atr8* acetyltransferase gene is in place of *atr6* (Fig. 1). The N terminus (125 amino acids (aa)) of the $Gtr26_{K11}$ (300 aa) and $Gtr60_{K29}$ (302 aa) proteins is conserved, and it is possible that the two proteins perform the same function. Atr6 from KL11 (GenPept accession number AOX98974.1) and Atr8 from KL29 (GenPept accession number KRJ36102.1) belong to different hexapeptide transferase protein families (PF14602 and PF00132, respectively). Though both families include acetyltransferases, Atr6 and Atr8 are not significantly related, suggesting that they may perform different functions.

3.2. Synthesis of dTDP-L-6dTal in *A. baumannii*

The synthesis of dTDP-L-6dTal*p* has been experimentally confirmed in some species, and is known to occur via one of two pathways (Fig. 2). In the first, RmlA, RmlB, RmlC and RmlD form dTDP-L-Rha*p* that can then be reversibly converted to dTDP-L-6dTal*p* by a 4-epimerase, designated $WbiB_{BTh}$ in *Burkholderia thailandensis* [16]. In the other pathway, the intermediate produced by RmlA, RmlB and RmlC is converted to dTDP-L-6dTal*p* by a 4-reductase, known as Tal in *Kitasatospora kifunensis* [17] or Tll in *Actinobacillus actinomycetemcomitans* [18]. KL11 carries genes for RmlA, RmlB, RmlC and RmlD, and the putative epimerase from KL11 (GenPept accession number AOX98975.1) shares 37% identity with the *B. thailandensis* $WbiB_{BTh}$ epimerase (GenPept accession number ABC37367.1) and belongs to the PF01370 epimerase protein family (Pfam). Thus it was likely that the *A. baumannii* epimerase reversibly converts dTDP-L-Rha*p* to dTDP-L-6dTal*p*

(Fig. 2). In keeping with the *A. baumannii* nomenclature system [2], the epimerase gene was annotated as *tle* for dTDP-6-deoxy-L-talose 4-epimerase (Fig. 1).

3.3. Relationship of PSgc8 and PSgc17 to the KL11 capsule biosynthesis gene cluster

Following the discovery of KL11 in AB900 and 99-015-2025, and KL29 in 2007-16-27-01, 2007-16-25-01-7 and TG27339, two capsule gene clusters, PSgc8 and PSgc17 (GenBank accession numbers KC526898 and KC526904) in the genomes of *A. baumannii* isolates LUH5545 and LUH5538, respectively [7], were reported. They were found to contain the same *rmlBDAC* module and also the same epimerase gene (99-100% identical) as in KL11. The putative epimerase gene had been annotated as *gnaB* for an epimerase that would convert UDP-D-Glc*p*NAcA to UDP-D-Gal*p*NAcA [7]. The PSgc8 and PSgc17 gene clusters were re-annotated in this study using the more widely adopted nomenclature system for *A. baumannii* [2]. The PSgc17 gene cluster from LUH5545 has the same genetic content as KL11, but includes an additional ISAba13 insertion sequence (IS) between *gtr29* and *itrA3* (Fig. 1). Thus, PSgc17 was re-named here KL11a to differentiate it from the uninterrupted form in isolates AB900 and 99-015-2025. Given that the IS does not appear to interrupt any genes, LUH5545 is expected to produce the complete K11 CPS.

The PSgc8 gene cluster from LUH5538, here renamed KL83, shares the *gtr27* and *gtr29* glycosyltransferase genes with KL11 (Fig. 1). However, as in KL29, KL83 contains *gtr60* and *atr8* in place of *gtr28* and *atr6*, and in addition *gtr26* has been replaced by *gtr154.* Though the *wzx* flippase gene is shared between the three gene clusters, KL83 contains a *wzy* gene encoding a Wzy polymerase (Wzy_{K83}) unrelated to $Wzy_{K11/K29.}$

3.4. K11 and K83 contain L-Rha and L-6dTal

K11 and K83 CPSs were isolated by phenol-water extraction from strains LUH5545 and LUH5538, respectively. Full acid hydrolysis of the K11 CPS followed by GLC analysis of the derived alditol acetates revealed Rha, 6dTal, Glc, and GlcNAc (Fig. S1). The D configuration of Glc and GlcNAc and the L configuration of Rha and 6dTal were determined using known regularities in glycosylation

effects on ^{13}C NMR chemical shifts [19]. GLC analysis of the acetylated alditols showed that the K83 CPS had the same composition as the K11 CPS. Therefore, the K11 and K83 units both contain L-6dTal*p* confirming the role of the Tle epimerase in the conversion of dTDP-L-Rha*p* to dTDP-L-6dTal*p* as predicted.

3.5. Structure of the K11 CPS from *A. baumannii* LUH5545

The ^{1}H NMR and ^{13}C NMR (Fig. 3A) spectra showed that the K11 CPS has a regular structure. Studies using two-dimensional NMR spectroscopy, including ^{1}H,^{1}H COSY, ^{1}H,^{1}H TOCSY, ^{1}H,^{1}H ROESY, ^{1}H,^{13}C HSQC, and ^{1}H,^{13}C HMBC experiments, demonstrated spin systems for five monosaccharide residues, including two residues of Glc*p*NAc (α-linked **D** and β-linked **A**) and one residue each of β-Glc*p* (**C**), α-Rha*p* (**E**), and α-6dTalp (**B**) (Table 1). Downfield displacements of the signals for the linkage carbons, C-2 of unit **C** and C-3 of the other units, as compared with their positions in the corresponding non-substituted sugars [20,21], revealed the substitution pattern of the monosaccharides. Their sequence in the K11 unit was determined by the ^{1}H,^{13}C HMBC (Fig. S2) and ROESY (Fig. S3) experiments, which showed correlations between the anomeric protons and linkage carbons and *vice versa* or between anomeric protons and protons at the linkage carbons, respectively (Supplementary materials, Table S1).

The structure of the K11 CPS was confirmed by Smith degradation, which resulted in oligosaccharide **1**, whose structure shown in Fig. 4 was established by two-dimensional NMR spectroscopy as described above for the CPS (for assigned ^{1}H and ^{13}C NMR chemical shifts of **1** see Table 1). Based on the data obtained, it was concluded that the K11 CPS of *A. baumannii* LUH5545 has the structure **2** shown in Fig. 4.

3.6. Structure of the K83 CPS from *A. baumannii* LUH5538

The ^{1}H NMR and ^{13}C NMR (Fig. 3B) spectra of the K83 CPS of LUH5538 were similar to those of K11 CPS but showed signals for an O-acetyl group at δ_H 2.14 and δ_C 21.9. Assignment of the ^{1}H NMR and ^{13}C NMR (Fig. 3B) spectra of the O-deactylated K83 CPS (Table 1) revealed more differences from the

K11 CPS: i) the signal for C-4 of GlcNAc **D** shifted significantly downfield from δ 70.0 to δ 74.4 and the signal for C-3 shifted upfield from δ 81.1 to δ 76.3, and ii) the signal for C-3 of Rha shifted upfield from δ 81.7 to δ 71.4. These displacements resulted evidently from a conversion of the 3-substituted GlcNAc **D** into the 3,4-disubstituted residue and the 3-substituted Rha residue into the terminal side-chain residue.

The attachment of GlcNAc **A** to O-4 of GlcNAc **D** was confirmed by two-dimensional ROESY (Fig. S4) and $^{1}H,^{13}C$ HMBC (Fig. S5) experiments (Supplementary materials, Table S1), which, particularly, showed correlations of GlcNAc **A** H-1 with GlcNAc **D** H-4 and C-4 at δ 4.59/3.91 and δ 4.59/74.4, respectively. These experiments also confirmed that the other linkages in the K11 and K83 CPSs are the same.

A comparison of the ^{1}H NMR and ^{13}C NMR (Fig.s 3B and 3C) spectra demonstrated a significant downfield shift of the signal for H-2 of 6dTal from δ 3.78 in the initial CPS to δ 5.00 in the O-deacetylated K83 CPS, which was evidently due to a deshielding effect of O-acetylation. Accordingly, the signals for C-1 and C-3 of this residue shifted upfield from δ 103.1 and 76.3 to δ 100.6 and 75.5, respectively (β-effects of O-acetylation). These displacements indicated that the 6dTal residue is O-acetylated at position 2.

Therefore, the K83 CPS of *A. baumannii* LUH5538 has the structure **3** shown in Fig. 4. It differs from the K11 CPS in i) one linkage, namely β-D-Glc*p*NAc-(1→4)-α-D-Glc*p*NAc in the former versus β-D-Glc*p*NAc-(1→3)-α-L-Rha*p* in the latter, and ii) 2-O-acetylation of the 6dTal residue.

3.7. Initiation of K-unit synthesis and Wzy specificity

Both the KL11 and KL83 gene clusters include an *itrA3* gene for initiating transferases that are 99% identical to each other. These proteins share ~96% identity with ItrA3 proteins from *A. baumannii* KL43, KL45, KL47, KL48, and KL88, which is consistent with GlcNAc as the first sugar of the respective K units [22,23]. However, there are two GlcNAc sugars in both the K11 and K83 structures complicating

identification of the first sugar. However, one of the GlcNAc residues in both the K11 and K83 units is α-(1→2) linked to a Glc residue, and this is the only linkage in either structure that requires a retaining glycosyltransferase. A shared gene, *gtr27,* encodes a retaining glycosyltransferase that belongs to the GT4 family of glycosyltransferases in the Carbohydrate Active enZymes (CAZy) database [24]. In addition, $Gtr27_{K11}$ (GenPept accession number AOX98972.1) and $Gtr27_{K83}$ (GenPept accession number AHB32310.1; WafY), which are identical, share 43% identity with WfaB from *Escherichia coli* O66 (GenPept accession number AAZ20762.1), and there is a similar α-D-Glc*p*NAc-(1→2)-D-Glc*p*3Ac linkage in the O66 polysaccharide [25]. The assignment of Gtr27 to this linkage indicates that the other GlcNAc residue in both structures is the first sugar, as shown in Fig. 4, and that Wzy_{K11} (GenPept accession numbers AHB32462.1) links K units via the β-D-Glc*p*NAc-(1→3)-α-L-Rha*p* linkage, and Wzy_{K83} (GenPept accession number AHB32311.1) catalyses formation of the β-D-Glc*p*NAc-(1→4)-α-D-Glc*p*NAc linkage between K units (Fig. 4).

3.8. Assignment of functions to the remaining glycosyltransferases

$Gtr26_{K11}$ (GenPept accession number AOX98970.1 and AHB32461.1; WafX) shares 80% identity with $Gtr154_{K83}$ (GenPept accession number AHB32312.1; also WafX), with differences found only in last 175 of 302 amino acids (aa). BLASTp similarity searches for both protein sequences revealed significant similarity to several predicted α-(1,3)-rhamnosyltransferases, including WsaD (GenPept accession number AAR99614.1) from *Geobacillus stearothermophilus*, which shares 31% identity (51% similarity) with $Gtr26_{K11}$ and 33% identity (55% similarity) with $Gtr154_{K83}$. WsaD has been shown to transfer L-Rha*p* from dTDP-β-L-Rha*p* to D-Gal*p* making an α-(1→3) linkage [26]. As there is a similar α-L-Rha*p*-(1→3)-D-Glc*p*NAc linkage in both K11 and K83 (Fig. 4), $Gtr26_{K11}$ and $Gtr154_{K83}$ are likely to form this linkage.

$Gtr28_{K11}$ (GenPept accession number AHB32464.1) and $Gtr60_{K83}$ (GenPept accession number AHB32309.1) are only 46% identical to each other, though both were previously assigned the same

name, WafZ [7]. In comparison, Gtr29 is encoded by both KL11 and KL83 gene clusters (GenPept accession numbers AHB32306.1 and AHB32467.1; WagB), and the sequences are 97% identical to each other. $Gtr28_{K11}$ and $Gtr60_{K83}$ are respectively 39% and 36% identical to WfaC (GenPept accession number AAZ20763.1) from *Escherichia coli* O66, and Gtr29 shares ~35% identity with WfaE from the same strain (GenPept accession number AAZ20765.1). Though the functions of these glycosyltransferases have not been experimentally confirmed, the *E. coli* O66 polysaccharide contains a β-D-Glc*p*3Ac-(1→3)-α-L-6dTal*p*-(1→3)-α-D-Glc*p*NAc fragment in the main chain [25], and both K11 and K83 contain a similar fragment, β-D-Glc*p*-(1→3)-α-L-6dTal*p*-(1→3)-β-D-Glc*p*NAc (Fig. 4), suggesting a role for Gtr29 and $Gtr28_{K11}$/$Gtr60_{K83}$ in the joining of these sugars.

In *A. baumannii* so far, the first sugar is always 3-substituted, and the gene that encodes the glycosyltransferase that links the second sugar via this substitution is always the last glycosyltransferase gene found in the gene cluster, usually just upstream of the *itr* gene. Based on this pattern, we predicted that Gtr29 forms the α-L-6dTal*p*-(1→3)-β-D-Glc*p*NAc linkage, while $Gtr28_{K11}$ and $Gtr60_{K83}$ catalyse formation of the β-D-Glc*p*-(1→3)-α-L-6dTal*p* linkage. This is consistent with the location of the L-6dTal*p*-transferase gene, *gtr29*, adjacent to the *tle* epimerase gene required for the synthesis of that sugar. In addition, the *gtr29*, *gtr60* and *gtr27* genes (annotated as *wagB, wafZ* and *wafY*, respectively) have each been knocked out in *A. nosocomialis* strain M2 [27], which carries a KL sequence (WGS accession number AWOW01000001.1) that is 95% identical to KL106 (see Fig. 1B). The length of the single K unit decorating the pilin was shown to decrease stepwise indicating that the first linkage is formed by Gtr29, and Gtr60 then Gtr27 function next in the pathway.

3.9. Other gene clusters with *rml* and *tle*

The Whole Genome Shotgun (WGS) database was searched for additional *A. baumannii* isolates with KL that carry *rmlBDAC* and *tle.* A further seven draft genomes with KL11 from isolates recovered in various countries were found (Table 2). Most of these isolates belonged to sequence type (ST) 49 and

ST128 in the Pasteur and Oxford MLST schemes, respectively. Only one further isolate, ABOB15, recovered from the USA carried KL83, but there were no additional isolates with KL29.

In addition to the KL already identified above, five further KL have the same combination of *rmlBDAC* and *tle* genes. Two of these, KL105 and KL106, are close relatives of KL11/KL29/KL83 (Fig. 1A). The structures of K29 and K105 CPS can be predicted to be identical to K11 and K83, respectively, except that in K29 the L-6dTal would be acetylated and in K105 it would not be acetylated. KL106 lacks *gtr154* and, as a result, K106 likely lacks the L-Rha residue.

Three further *tle*-containing gene clusters, KL26, KL36, and KL87a (Fig. 1B), are found in the genomes of single isolates (WGS/GenBank accession numbers ALOH01000183.1, AMFI01000021.1, and KC526918.1 respectively), and it is likely that these isolates produce capsules containing L-6dTal or both L-6dTal and L-Rha. In addition, KL26, KL36, and KL87a include *manC, mnaA,* and *ugd* genes for synthesis of UDP-D-mannose, UDP-*N*-acetyl-D-mannosamine, and UDP-D-glucuronic acid, respectively, suggesting that the corresponding CPSs may also contain these sugars.

4. Discussion

This paper reports CPS structures for two more strains from the set sequenced by Hu et al. [7]. It is also the first to report L-6dTal in *A. baumannii*, adding another sugar to the list of sugars that are incorporated into the CPS of strains from this species. L-6dTal is relatively uncommon in nature and has so far only been found in a few bacterial species (Bacterial Carbohydrate Structure Database at http://csdb.glycoscience.ru/bacterial/). The CPS structures determined here confirm the role of *rmlBDAC* in the synthesis of L-Rha*,* and Tle as a dTDP-L-Rha 4-epimerase. The L-6dTal residue is O-acetylated in K83 but not in K11 (Fig. 4), and Atr8 encoded by KL83 would carry out this O-acetylation. However, a role could not be assigned to *atr6* gene that is present in KL11. The related loci shown in Fig. 1 also include either *atr6* or *atr8* gene, and it is likely that the CPS produced by those that carry *atr8* will also be acetylated.

Although both K11 and K83 have the same K units, the different *wzy* genes result in a change in the preference of the Wzy polymerase for its acceptor substrate. Wzy_{K83} recognises an internal GlcNAc residue in the K-unit forcing the L-Rha*p* residue to become a side-branch, whereas Wzy_{K11} recognises the terminal L-Rha*p* sugar residue and forms a linear polysaccharide. In this regard, KL11 and KL83 gene clusters resemble KL27 and KL44, which also encode different Wzy polymerases leading to different final CPS structures made up of the same K unit [5]. In addition, the presence of the wzy_{K83} gene in KL106 may allow the K unit to be polymerised despite the lack of a glycosyltransferase (either Gtr26 or Gtr154) to add L-Rha to the K unit.

The fact that the first 125 of the 302 aa $Gtr26_{K11}$ are conserved in $Gtr154_{K83}$, suggests that a hybrid may have arisen due to a recombination event that would have replaced the adjacent *wzy* gene in one of the gene clusters. The same is also true for $Gtr28_{K11}$ and $Gtr60_{K83}$, which may be linked to the replacement of the adjacent *atr* gene. However, it was not possible to predict which gene cluster was the original.

These studies further highlight the potential and importance of comparing structural and genetic data to soundly assign roles to genes found in novel KL gene clusters. The predictions for the Gtrs were further strengthened by the experimental data for three of the *gtr* genes that are also found in *A. nosocomialis* strain M2 [27], as this confirms our observation of a correlation between the order of *gtr* genes in the K locus and the order of linkages in the K unit [28]. Here, it was also possible to predict the structure of CPS produced by strains with KL closely related to those for which the CPS structure was determined.

Acknowledgements

This work was supported by the Russian Foundation for Basic Research (project No.17-04-01254), Australian National Health and Medical Research Council (NHMRC) project grant 1026189, and National Natural Science Foundation of China (NSFC) Program (31371259).

References

1. D. Wong, T.B. Nielsen, R.A. Bonomo, P. Pantapalangkoor, B. Luna, B. Spellberg, Clin. Microbiol. Rev. 30 (2017) 409-447.
2. J.J. Kenyon, R.M. Hall, PLoS One. 8 (2013) e62160.
3. K.E. Holt, J.J. Kenyon, M. Hamidian, M.B. Schultz, D.J. Pickard, G. Dougan, R.M. Hall, MGen. 2 (2016) 1-16.
4. J.J. Kenyon, A.M. Marzaioli, C. De Castro, R.M. Hall, Carbohydr. Res. 409 (2015) 30-35.
5. A.S. Shashkov, J.J. Kenyon, S.N. Senchenkova, M.M. Shneider, A.V. Popova, N.P. Arbatsky, K.A. Miroshnikov, N.V. Volozhantsev, R.M. Hall, Y.A. Knirel, Glycobiology. 26 (2016) 501-508.
6. J.J. Kenyon, M.M. Shneider, S.N. Senchenkova, A.S. Shashkov, M.N. Siniagina, S.Y. Malanin, A.V. Popova, K.A. Miroshnikov, R.M. Hall, Y.A. Knirel, Microbiology. 162 (2016) 1479-1489.
7. D. Hu, B. Liu, L. Dijkshoorn, L. Wang, P.R. Reeves, PLoS One. 8 (2013) e70329.
8. N.P. Arbatsky, M.M. Shneider, J.J. Kenyon, A.S. Shashkov, A.V. Popova, K.A. Miroshnikov, N.V. Volozhantsev, Y.A. Knirel. Carbohydr. Res. 413 (2015) 12-15.
9. A.S. Shashkov, S.N. Senchenkova, A.V. Popova, Z. Mei, M.M. Shneider, B. Liu, K.A. Miroshnikov, N.V. Volozhantsev, Y.A. Knirel. Russ. Chem. Bull. 64 (2015) 1196–1199.
10. S.N. Senchenkova, A.V. Popova, A.S. Shashkov, M.M Shneider, Z. Mei, N.P. Arbatsky, B. Liu, K.A. Miroshinikov, N.V. Volozhantsev, Y.A. Knirel, Carbohydr. Res. 407 (2015) 154-157.
11. J.J. Kenyon, A.M. Marzaioli, C. De Castro, R.M. Hall, Glycobiology. 25 (2015) 644-654.

12. O. Westphal, K. Jann, Bacterial lipopolysaccharides: extraction with phenol-water and further applications of the procedure, in: R.L. Whistler (Ed). Methods in Carbohydrate Chemistry, Academic Press, New York, 1965, p.p. 83-91.

13. J.K. Valenzuela, L. Thomas, S.R. Partridge, T. van der Reijden, L. Dijkshoorn, J. Iredell, J. Clin. Microbiol. 45 (2007) 453-460.

14. M.D. Adams, K. Goglin, N. Molyneaux, K.M. Hujer, H. Lavender, J.J. Jamison, I.J. MacDonald, K.M. Martin, T.A. Russo, A.A. Campagnari, J. Bacteriol. 190 (2008) 8053-8064.

15. J. Sahl, J.D. Gillece, J.M. Schupp, V.G. Waddell, E.M. Driebe, D.M. Engelthaler, P. Keim, PLoS One. 8 (2013) e54287.

16. H.G. Yoo, S.Y. Kwon, S. Karki, H.J. Kwon, Bioorg. Med. Chem. Lett. 21 (2011) 3914-3917.

17. S. Karki, H.G. Yoo, S.Y. Kwon, J.W. Suh, H.J. Kwon, Carbohydr. Res. 345 (2010) 1958-1962.

18. Y. Nakano, N. Suzuki, Y. Yoshida, T. Nezu, Y. Yamashita, T. Koga. J. Biol. Chem. 275 (2000) 6806-6812.

19. A.S. Shashkov, N.A. Paramonov, S.P. Veremeychenko, H. Grosskurth, G.M. Zdorovenko, Y.A. Knirel, N.K. Kochetkov. Carbohydr. Res. 306 (1998) 297-303.

20. G.M. Lipkind, A.S. Shashkov, S.S. Mamian, N.K. Kochetkov, Carbohydr. Res. 181 (1988) 1-12.

21. Y.A. Knirel, N.A. Paramonov, A.S. Shashkov, N.K. Kochetkov, R.G. Yarullin, S.M. Farber, V.I. Efremenko, Carbohydr. Res. 233 (1992) 185-193.

22. A.S. Shashkov, J.J. Kenyon, N.P. Arbatsky, M.M. Shneider, A.V. Popova, K.A. Miroshnikov, N.V. Volozhantsev, Y.A. Knirel, Carbohydr. Res. 417 (2015) 81-88.

342 23. A.S. Shashkov, J.J. Kenyon, N.P. Arbatsky, M.M. Shneider, A.V. Popova, K.A.
343 Miroshnikov, R.M. Hall, Y.A. Knirel, Carbohydr. Res. 435 (2016) 173-179.
344 24. V. Lombard, H. Golaconda Ramulu, E. Drula, P.M. Coutinho, B. Henrissat, Nucleic
345 Acids Res. 42 (2014) D490-D495.
346 25. B. Jann, A. Shashkov, V. Torgov, H. Kochanowski, G. Seltmann, K. Jann, Carbohydr.
347 Res. 278 (1995) 155-165.
348 26. K. Steiner, R. Novotny, D.B. Werz, K. Zarschler, P.H. Seeberger, A. Hofinger, P.
349 Kosma, C. Schäffer, P. Messner, J. Biol. Chem. 283 (2008) 21120-21133.
350 27. C.M. Harding, M.A. Nasr, R.L. Kinsella, N.E. Scott, L.J. Foster, B.S. Weber, S.E.
351 Fiester, L.A. Actis, E.N. Tracy, R.S. Munson Jr, M.F. Feldman, Mol. Microbiol. 96
352 (2015) 1023-1041.
353 28. J.J. Kenyon, A.A. Kasimova, M.M. Shneider, A.S. Shashkov, N.P Arbatsky, A.V.
354 Popova, K.A. Miroshnikov, R.M. Hall, Y.A. Knirel, Microbiology. 2017. DOI:
355 10.1099/mic.0.000430.

LEGENDS TO FIGURES

Fig. 1. Comparison of the *A. baumannii* CPS biosynthesis gene clusters containing the *tle* gene. The black vertical arrows indicate the position of insertion sequences (IS) in KL11a and KL87a (GenBank accession numbers KC526904.1 and KC526918.1, respectively). Light grey genes encode glycosyltransferases, and dark grey are for acetyltransferases. Pathways for products of gene modules are shown above. Dark grey shading between gene clusters shows regions of >97% identity, whereas light shading is regions of 75-97% identity. Figure is drawn to scale from GenBank accession numbers KF002790 (KL11), AMHJ01000024.1 (KL29), KC526898.1 (KL83), JEXD01000015.1 (KL105), JVPN01000008.1 (KL106), ALOH01000183.1 (KL26), AMFI01000021.1 (KL36), and KC526918.1 (KL87a), and scale bar is shown below.

Fig. 2. Synthesis pathways of dTDP-L-6dTal. The Tal/Tll-dependent pathway was identified in *K. kifunensis* [17] and *A. actinomycetemcomitans* [18], and the $WbiB_{BTh}$/Tle-dependent pathway was identified in *B. thailandensis* [16] and predicted in *A. baumannii* (this work).

Fig. 3. ^{13}C NMR spectra of the K11 CPS from *A. baumannii* LUH5545 (A), K83 CPS from LUH5538 (B), and O-deacetylated CPS from *A. baumannii* LUH5538 (C). Numbers refer to carbons in sugar residues denoted by letters as indicated in Table 1 and Fig. 4.

Fig. 4. Structures of oligosaccharide **1** derived by Smith degradation of CPS of *A. baumannii* LUH5545, K11 CPS of *A. baumannii* LUH5545 (**2**), and K83 CPS of LUH5538 (**3**). Enzymes are shown in bold next to the linkage each is predicted to be responsible for.

Table S1. Correlations for H-1 and C-1 in the two-dimensional $^{1}H,^{1}H$ ROESY and $^{1}H,^{13}C$ HMBC spectra of the CPSs.

Fig. S1. GLC profile of the alditol acetates derived from the K11 CPS of strain LUH5545.

Fig. S2. Part of a $^{1}H,^{1}H$ ROESY spectrum of the K11 CPS of strain LUH5545. The corresponding parts of the ^{1}H NMR spectrum are shown along the axes.

Fig. S3. Parts of a $^{1}H,^{13}C$ HMBC spectrum of the K11 CPS of strain LUH5545. The corresponding parts of the ^{1}H and ^{13}C NMR spectra are shown along the horizontal and vertical axes, respectively.

Fig. S4. Part of a $^{1}H,^{1}H$ ROESY spectrum of the K83 CPS of strain LUH5538. The corresponding parts of the ^{1}H NMR spectrum are shown along the axes.

Fig. S5. Parts of a $^{1}H,^{13}C$ HMBC spectrum of the K83 CPS of strain LUH5538. The corresponding parts of the ^{1}H and ^{13}C NMR spectra are shown along the horizontal and vertical axes, respectively.

Chemical shifts for NAc are δ_C 23.4-23.7 (CH_3), 174.5-175.8 (CO), δ_H 2.03-2.08.

Table 2. *A. baumannii* genomes carrying KL with *rmlBDAC* and *tle* genes

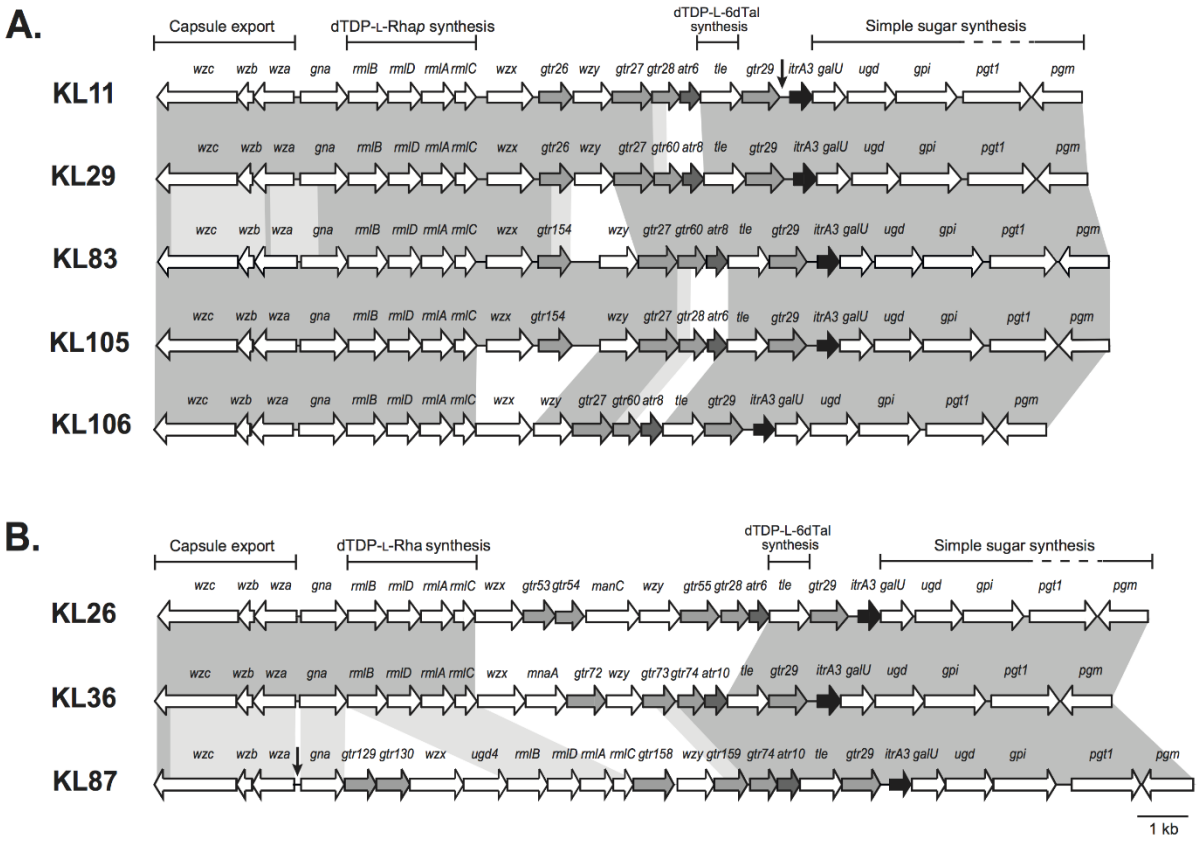
KL	Strain	ST_{Ps}	ST_{Ox}	Country	Year	GenBank/WGS accession no.
KL11	AB900	49	128	USA	2003	ABXK01000027.1
	99-015-2025	49	128	Australia	1999	KF002790.1
	LUH5545	-[1]	-[1]	Germany	-	KC526904.1
	3207	422	1321	Mexico	2008	CP015364.1
	OIFC111	49	128	USA	2003	AMFY01000028.1
	ABBL071	49	128	USA	2009	LLGB01000093.1
	1461963	49	128	USA	-	JEWQ01000019.1
	1293320	49	128	USA	-	JFEE01000018.1
	3.5D	388	836	Germany	2013	MABZ01000025.1
	Ab18	-[2]	-[2]	Brazil	2004	LMBN01000003.1
KL29	2007-16-27-01	241	613	-	2007	AMHJ01000024.1
	2007-16-25-01-7	241	613	-	2007	AMHI01000016.1
	TG27339	241	613	-	2005	AMIR01000046.1
	ABBL038	-[2]	-[2]	USA	2007	LLDS01000040.1
KL83	LUH5538	-[1]	-[1]	Germany	-	KC526898.1
	ABOB15	New	-[2]	USA	-	LLJE01000049.1
KL105	625974	509	1056	USA	-	JEXD01000015.1
KL106	219_ABAU	333	-[2]	USA	-	JVPN01000008.1
	TG22198	164	234	-	-	ASFT01000013.1
KL26	BZICU-2	218	184	China	2010	ALOH01000183.1
KL36	Naval-72	405	1025	USA	2006	AMFI01000021.1
KL87a	LUH5547	-[1]	-[1]	Germany	-	KC526918.1

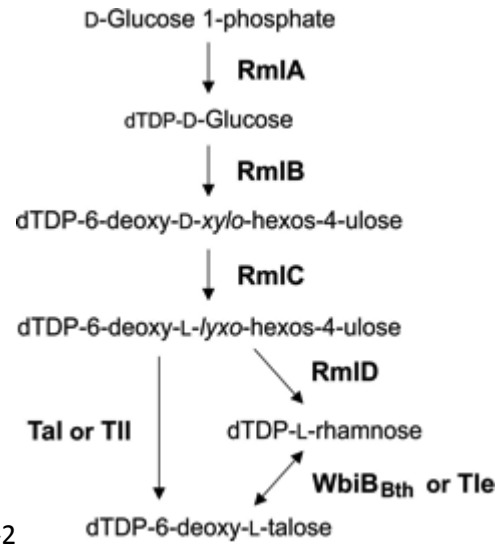
[1] Draft genome sequence not available.

[2] Missing alleles. Cannot be typed.

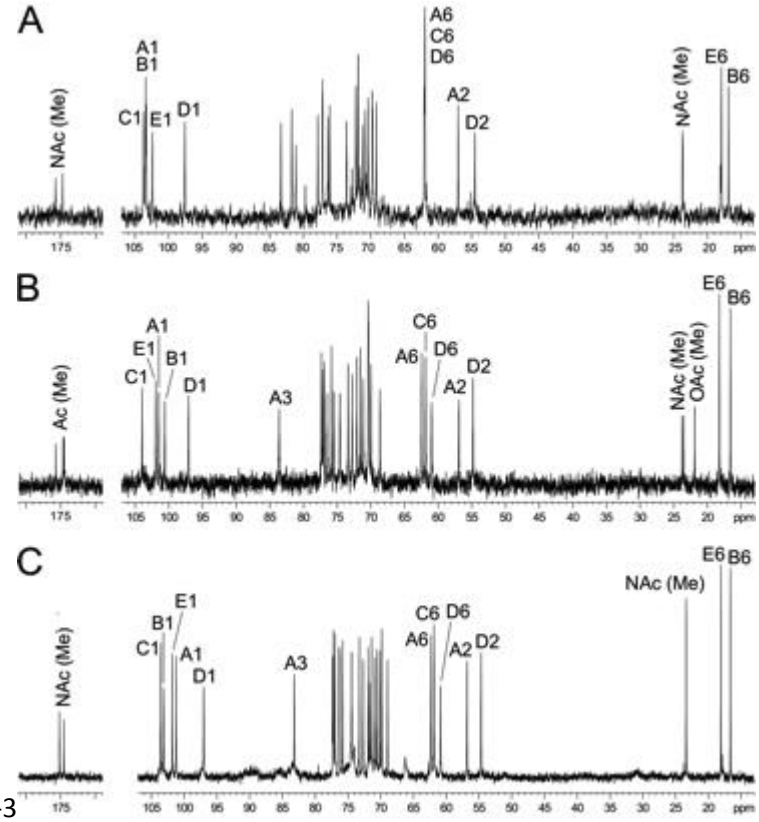
Supplementary materials

Figr-1





Figr-2



Figr-3

Figr-4

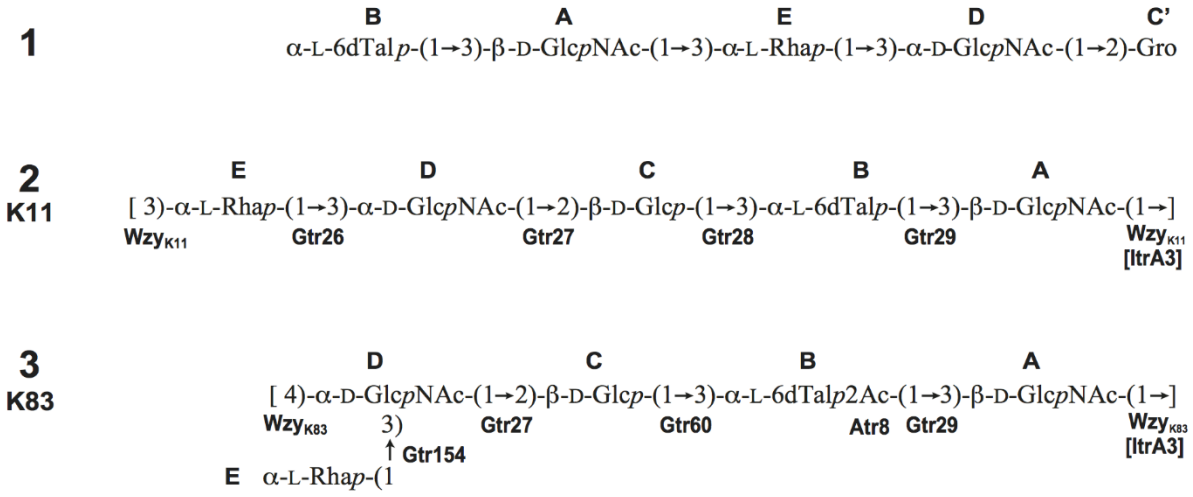


Table 1. ^{1}H and ^{13}C NMR chemical shifts (δ, ppm)

Residue	C-1	C-2	C-3	C-4	C-5	C-6
	H-1 (a, b)	*H-2*	*H-3 (a, b)*	*H-4*	*H-5*	*H-6 (a, b)*
Oligosaccharide **1**						
→2)-Gro	62.2	78.7	62.4			
C'	*3.85, 3.79*	*3.83*	*3.73, 3.63*			
→3)-α-Glc*p*NAc-(1→	97.6	54.5	80.9	69.6	73.8	61.8
D	*5.18*	*4.07*	*3.82*	*3.57*	*3.89*	*3.92, 3.81*
→3)-α-L-Rha*p*-(1→	102.2	71.6	81.6	72.0	70.2	17.7
E	*4.87*	*3.98*	*3.77*	*3.49*	*4.00*	*1.23*
→3)-β-Glc*p*NAc-(1→	103.2	56.8	82.5	69.6	77.0	61.8
A	*4.71*	*3.84*	*3.68*	*3.54*	*3.47*	*3.87, 3.79*
α-L-6dTal*p*-(1→	103.1	71.5	66.8	73.4	68.9	16.7

B	*5.00*	*3.70*	*3.87*	*3.74*	*4.28*	*1.22*
CPS of *A. baumannii* LUH5545						
→3)-β-Glc*p*NAc-(1→	103.4	57.0	83.3	69.8	77.1	62.0
A	*4.69*	*3.82*	*3.63*	*3.50*	*3.45*	*3.80, 3.91*
→3)-α-L-6dTal*p*-(1→	103.3	71.2	76.3	71.8	69.1	16.8
B	*4.97*	*3.89*	*3.91*	*3.63*	*4.23*	*1.20*
→2)-β-Glc*p*-(1→	103.7	77.9	76.1	70.8	77.1	61.9
C	*4.76*	*3.51*	*3.58*	*3.47*	*3.45*	*3.76, 3.87*
→3)-α-Glc*p*NAc-(1→	97.6	54.5	81.1	69.8	73.6	61.9
D	*5.42*	*4.07*	*3.79*	*3.57*	*4.08*	*3.84, 3.89*
→3)-α-L-Rha*p*-(1→	102.4	71.8	81.7	72.2	70.4	17.9
E	*4.83*	*3.92*	*3.76*	*3.47*	*3.98*	*1.21*
O-Deacetylated CPS of *A. baumannii* LUH5538						
→3)-β-Glc*p*NAc-(1→	101.2	56.8	83.2	70.1	77.2	62.3
A	*4.59*	*3.86*	*3.66*	*3.44*	*3.44*	*3.85, 3.93*
→3)-α-L-6dTal*p*-(1→	103.1	70.9	76.3	71.7	69.0	16.6
B	*5.00*	*3.91*	*3.89*	*3.64*	*4.23*	*1.23*
→2)-β-Glc*p*-(1→	103.5	77.3	75.8	70.6	77.0	61.9
C	*4.79*	*3.52*	*3.60*	*3.48*	*3.47*	*3.78, 3.90*

→3,4)-α-Glc*p*NAc-(1→	97.0	54.7	76.3	74.4	72.7	60.9
D	*5.42*	*4.17*	*3.98*	*3.91*	*4.11*	*3.90, 3.78*
α-L-Rha*p*-(1→	101.8	71.9	71.4	73.2	69.8	18.1
E	*4.95*	*3.78*	*3.79*	*3.44*	*4.22*	*1.31*