K17 capsular polysaccharide produced by *Acinetobacter baumannii* isolate G7 contains an amide of 2-acetamido-2-deoxy-D-galacturonic acid with D-alanine

Johanna J. Kenyon,[a,†,1] Sof'ya N. Senchenkova,[b,†] Alexander S. Shashkov,[b] Mikhail M. Shneider,[c,d] Anastasia V. Popova,[d,e,f] Yuriy A. Knirel,[b,*] Ruth M. Hall[g,*]

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

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

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

[d] *Institute of Antimicrobial Chemotherapy, Smolensk State Medical University, Smolensk, Russia*

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

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

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

[1] Corresponding author. Tel.: +617 3138 2552; E-mail: johanna.kenyon@qut.edu.au

[†] and [*] Contributed equally

Keywords: *Acinetobacter baumannii*; capsular polysaccharide; KL17; K locus; D-alanine

Abbreviations

Alt, alanine transferase; COSY, correlation spectroscopy; CPS, capsular polysaccharide; GC, global clone; Gtr, glycosyltransferase; HMBC, heteronuclear multiplebond correlation; HSQC, heteronuclear single-quantum coherence; KL, K locus; ROESY, rotating frame Overhauser effect spectroscopy; TOCSY, total correlation spectroscopy; UDP, uridine diphosphate; UndP, undecaprenyl phosphate; WGS, Whole Genome Shotgun.

Abstract- The K17 capsular polysaccharide (CPS) produced by *Acinetobacter baumannii* G7, which carries the KL17 configuration at the capsule biosynthesis locus, was isolated and studied by chemical methods along with one- and two-dimensional ^{1}H and ^{13}C NMR spectroscopy. Selective cleavage of the glycosidic linkage of a 2,4-diacetamido-2,4,6-trideoxy-D-glucose (D-QuiNAc4NAc) residue by i) trifluoroacetic acid solvolysis or ii) alkaline β-elimination (NaOH-$NaBH_4$) of the 4-linked D-alanine amide of a 2-acetamido-2-deoxy-D-galacturonic acid residue (D-GalNAcA6DAla) yielded trisaccharides that were isolated by Fractogel TSK HW-40 gel-permeation chromatography and identified by using NMR spectroscopy and high-resolution electrospray ionization mass spectrometry. The following structure was established for the trisaccharide repeat (K unit) of the CPS: →4)-α-D-Gal*p*NAcA6DAla-(1→4)-α-D-Gal*p*NAcA-(1→3)-β-D-Qui*p*NAc4NAc-(1→ . The presence of the *itrA1* gene coding for the initial glycosylphosphotransferase in the KL17 gene cluster established the first sugar of the K unit as D-Qui*p*NAc4NAc. KL17 includes genes for three transferases that had been annotated previously as glycosyltransferases (Gtrs). As only two Gtrs are required for the K17 structure and one D-Gal*p*NAcA residue is modified by a D-alanine amide, these assignments were re-assessed. One transferase was found to belong to the ATPgrasp_TupA protein family that includes D-alanine-D-alanine ligases, and thus was renamed Alt1 (alanine transferase). Alt1 represents a novel family that amidate the carboxyl group of D-Gal*p*NAcA or D-Gal*p*A.

1. Introduction

The nosocomial pathogen, *Acinetobacter baumannii*, is recognised among the most significant bacterial species posing a threat to global health due to increasing levels of resistance to all therapeutically suitable antibiotics. The prevalence of extensively antibiotic resistant *A. baumannii* infections can be attributed to the evolution and expansion of successful clonal lineages [1-3], including the two globally distributed clonal groups, Global Clone 1 (GC1) and Global Clone 2 (GC2). However, the K locus, which includes the majority of genes for synthesis of the capsular polysaccharide (CPS) [4], often differs between isolates of the same clonal lineage [1, 4-7]. As the CPS is a major virulence determinant for the species, providing an external barrier surrounding the cells, and is a key cell surface target for novel vaccine and phage therapies [8-10], knowledge of CPS variation within a clone is important.

Each of the various CPS structures produced by *A. baumannii* are made up of many copies of identical oligosaccharide 'K units' linked together to form polysaccharide chains. These K units consist of two or more different sugars joined via specific linkages, and the linkage between K units in the polymer is also specific. Originally, eight distinct CPS biosynthesis gene clusters (KL1, KL4, KL12, KL15, KL17, KL20, KL25 and KL40) were identified in genomes of isolates belonging to *A. baumannii* GC1 [1]. A ninth configuration, the KL107 gene cluster, was recently reported in GC1 [11]. The chemical structures of six of the corresponding CPSs have been elucidated (K1 [10]; K4 [12]; K12 [13]; K15 [14]; K20 [15]; and K25 [16]), and each correlates with the genetic content of the corresponding gene cluster.

Of the remaining KL types in GC1, the KL17 gene cluster was originally identified in the draft genome sequence of *A. baumannii* isolate G7 [1], a sporadic GC1 isolate recovered in 2002 from a patient at the Alfred Hospital in Melbourne, Australia [17]. The sequence had been

annotated and deposited into GenBank under accession number KC118541.2. In this study, we determined the structure of the K17 CPS produced by this isolate, correlated the genes found at the genomic K locus to the functions required for CPS production, and revised the gene annotations accordingly.

2. Materials and methods

2.1. Bacterial strain and cultivation

Acinetobacter baumannii isolate G7 was isolated in 2002 at the Alfred Hospital, Melbourne, Australia and belongs to global clone 1 [17]. 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 of CPS

Bacterial cells were extracted with aqueous phenol [18], and the extract was dialysed without layer separation and freed from insoluble contaminations by centrifugation. The supernatant was treated with cold (4 °C) aqueous 50 % CCl_3CO_2H (Reakhim, Moscow, Russia); after centrifugation the supernatant was dialysed against distilled water and freeze-dried to give a CPS sample.

2.3. Analysis of alanine

A CPS sample (1 mg) was hydrolysed with 2 M HCl (0.5 mL) at 100 °C for 2 h, acid was removed with a stream of air, the residue was dissolved in saturated aqueous $NaHCO_3$ and N-acetylated with acetanhydride (3 × 20 μL, each portion being added in 15 min) at 0 °C. At 30 min after adding the last portion, the reaction mixture was diluted with water and desalted with KU-2 cation-exchange resin (H^+-form). The resin was removed by filtration, the solution was

evaporated, the residue was treated with (+)-(*S*)-2-octanol (0.1 mL) and CF_3CO_2H (15 μL) at 100 °C for 16 h. The reagents were removed with a stream of air, the residue was acetylated with a 1:1 acetanhydride-pyridine mixture (0.4 mL) at 20 °C for 16 h, and the reagents were removed as above. The residue was dissolved in ethyl acetate and analysed by GLC on a Maestro (Agilent 7820) chromatograph (Interlab, Russia) equipped with an HP-5 column (0.32 mm × 30 m) using a temperature program of 100 °C (1 min) to 145 °C at 2 °C min^{-1}.

2.4. Selective cleavage of CPS

A CPS sample (12 mg) was treated with CF_3CO_2H (0.5 mL) at 50 °C for 3 h under anhydrous conditions. After evaporation, the reaction products were dissolved in water and fractionated by gel-permeation chromatography on a column (85 × 1.6 cm) of Fractogel TSK HW-40 in aqueous 0.1% HOAc using a differential refractometer (Knauer, Germany) for monitoring elution to give fractions 1, 2, and 3 (OS1) in yields 3.7, 4.3, and 3.7 mg, respectively. The combined fractions 1 и 2 were treated as described above for the CPS to afford an additional amount of OS1 (2.2 mg).

A CPS sample (28 mg) was treated with 0.1 M NaOH (5 mL) in the presence of $NaBH_4$ (95 mg) at 37 °C for 24 h, the solution was neutralised with concentrated HOAc to pH 5, and the products were fractionated by gel-permeation chromatography on Fractogel TSK HW-40 as above to give fractions 1-3 in yields 6.2, 6.2, and 5.5 mg, respectively. Fraction 3 was rechromatographed on the same column to yield OS2 (2.6 mg). Fractions 1 and 2 were combined and treated as described above for the CPS to afford an additional amount of OS2 (1.5 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 (Deutero GmbH, Kastellaun, Germany). ^{1}H and ^{13}C NMR spectra were recorded at 50 °C for CPS or 30 °C for OS1 and OS2 on a Bruker Avance II 600 MHz spectrometer (Germany) equipped with an inverse broadband 5-mm probehead. Sodium 3-trimethylsilylpropanoate-2,2,3,3-d_4 (Sigma-Aldrich) (δ_H 0, δ_C −1.6) was used as internal reference for calibration. Bruker TopSpin 2.1 program was used to acquire and process the NMR data. Two-dimensional NMR experiments were performed using standard Bruker software. A spin-lock time 60 ms and a mixing time 150 ms were used in the TOCSY and ROESY experiments, respectively. The HMBC experiment was optimised for the coupling constant of 8 Hz.

2.6. Mass spectrometry

High-resolution electrospray ionization mass spectrometry was performed in the negative ion mode using a microTOF instrument (Bruker Daltonics). Capillary entrance voltage was set to 3200 V; the drying nitrogen temperature was 180 °C. A sample (~50 ng μL^{-1}) was dissolved in a 1:1 (v/v) H_2O/MeCN mixture and sprayed at a flow rate of 3 μL min^{-1}. Acquisition range was *m/z* 50–3000, internal calibration was done with ESI Tuning Mix (Agilent).

2.7. Bioinformatics analysis

The draft genome sequence of *A. baumannii* isolate G7 was reported previously [1], and annotations for the KL17 gene cluster were deposited in GenBank accession number KC118541.2. Functions of the proteins encoded by KL17, KL18 and the *A. haemolyticus* gene clusters were examined using a range of bioinformatic approaches as described previously [4].

3. Results

3.1. The KL17 gene cluster

According to published annotations [1], the KL17 gene cluster (Fig. 1A) includes a module of genes (*gna/gne2*) for the synthesis of UDP-2-acetamido-2-deoxy-D-galacturonic acid (UDP-D-Gal*p*NAcA) [12, 19] and *qhbCB/gdr* genes for the synthesis of UDP-2,4-diacetamido-2,4,6-trideoxy-D-glucose (UDP-D-Qui*p*NAc4NAc) [4, 20]. The cluster also includes an *itrA1* gene encoding the ItrA1 initiating transferase known to initiate K-unit synthesis by transferring D-Qui*p*NAc4NAc-1-*P* from UDP-D-Qui*p*NAc4NAc to the undecaprenyl phosphate (UndP) lipid carrier in the inner membrane [4, 20]. Hence, it is expected that the K17 unit will begin with D-Qui*p*NAc4NAc and contain at least one D-Gal*p*NAcA residue. In addition, the KL17 gene cluster includes three genes (*gtr38, gtr39,* and *gtr40*) that were predicted to encode glycosyltransferases [1], suggesting the K unit would contain four monosaccharides. Typical of *A. baumannii* K loci, unique forms of the *wzx* and *wzy* genes for K-unit translocase and polymerase, respectively, were present.

3.2. K17 CPS structure elucidation

The ^{13}C NMR spectrum (Fig. 2) of the K17 CPS isolated from *A. baumannii* G7 by phenol-water extraction showed signals for three anomeric atoms at δ_H 4.65-5.09 and δ_C 98.6-102.5, two methyl groups (H-6/C-6 of a 6-deoxy sugar and H-3/C-3 of alanine) at δ_H 1.07 and 1.37, δ_C 17.9 and 19.0, five nitrogen-bearing carbons of amino sugars and alanine at δ 50.5-58.4, two free carboxyl groups at δ 174.4 (a hexuronic acid) and 178.4 (alanine), one carboxamide group (a hexuronic acid amide) at δ 170.5, four N-acetyl groups at δ_H 1.93-2.07, δ_C 23.4-23.8 (Me) and 174.9-176.1 (CO), other protons at δ_H 3.34-4.70 and other carbons at δ_C 67.9-79.7 (assignment of the CO groups was performed using the $^{1}H,^{13}C$ HMBC spectrum of the CPS; Supplementary material, Fig. S1).

Assignment of the ^{1}H and ^{13}C NMR spectra of the CPS was performed using two-dimensional $^{1}H,^{1}H$ COSY, $^{1}H,^{1}H$ TOCSY, $^{1}H,^{1}H$ ROESY, and $^{1}H,^{13}C$ HSQC experiments

(Table 1), and spin systems for alanine and three monosaccharides, including β-Qui*p*NAc4NAc (residue **A**) and two α-Gal*p*NAcA residues (**B** and **C**), were identified. The assignment was based on the following correlations: H-1/H-2,3,4 and H-6/H-5,4,3 for QuiNAc4NAc in the TOCSY spectrum, H-1/H-2,3,4 and H-4/H-5 for GalNAcA in the TOCSY and ROESY spectrum, respectively.

The acidic components were confirmed by the $^{1}H,^{13}C$ HMBC spectrum, which showed correlations between C-1 (CO) and H-2,3 for Ala at δ 178.4/4.33 and 178.4/1.37, and between C-6 (CO) and H-5 of GalNAcA residues **B** and **C** at δ 174.4/4.19 and 170.5/4.70, respectively. The attachment of Ala to the carboxyl group of residue **C** was inferred from a correlation between H-2 of Ala and C-6 of GalNAcA **C** at δ 4.33/170.5 (Supplementary material, Fig. S1). The D configuration of Ala was established by GLC of the N-acetylated (*S*)-2-octyl ester. The D configuration of the monosaccharides was not confirmed chemically but was inferred from genetic data (see below).

The following correlations between anomeric protons and protons at the linkage carbons were observed in the ROESY spectrum of the CPS: **A** H-1/**C** H-4, **C** H-1/**B** H-4, and **B** H-1/**A** H-3 at δ 4.65/4.32, 5.05/4.38, and 5.09/3.85, respectively (Supplementary material, Fig. S2). These data defined the positions of substitution and the sequence of the monosaccharide residues in the K unit.

The CPS structure established by NMR spectroscopy (Fig. 3) was confirmed by selective chemical cleavages (Fig. 4). Solvolysis of the CPS with CF_3CO_2H [21] split off selectively cleaved the glycosidic linkage of residue **A** to give a trisaccharide (OS1) with a QuiNAc4NAc residue at the reducing end. β-Elimination in the hexuronamide residue **C** caused by treatment of the CPS with alkali in the presence of $NaBH_4$ [22] afforded a trisaccharide (OS2) with a β-L-*threo*-hex-4-enuronic acid (HexenNAcA) residue **C'** at the one end and a QuiNAcNAc-derived alditol (QuiNAcNAc-ol) **A'** at the other end.

The oligosaccharides OS1 and OS2 were isolated by gel-permeation chromatography on Fractogel TSK HW-40 and their structures shown in Fig. 4 were established by two-dimensional NMR spectroscopy as described above for the CPS (the assigned ^{1}H and ^{13}C NMR chemical shifts of OS1 and OS2 are shown in Table 1). Structures of OS1 and OS2 were confirmed by negative ion high-resolution electrosptay ionizaton mass spectra, which showed $[M–H]^-$ ion peaks at *m/z* 750.2682 and 734.2732 (calculated values are *m/z* 750.2687 and 734.2738 for compounds with the molecular formulae $C_{29}H_{45}N_5O_{18}$ and $C_{29}H_{45}N_5O_{17}$, respectively).

Hence, the K17 CPS is composed of repeats of the →4)-α-D-Gal*p*NAcA6DAla-(1→4)-α-D-Gal*p*NAcA-(1→3)-β-D-Qui*p*NAc4NAc-(1→ trisaccharide.

3.3. Correlation of the K17 structure with the KL17 capsule biosynthesis gene cluster

The K17 CPS includes both D-Gal*p*NAcA and D-Qui*p*NAc4NAc consistent with the presence of *gna/gne2* and *qhbCB/gdr* gene modules in the KL17 gene cluster. As the presence of *itrA1* indicates that D-Qui*p*NAc4NAc is the first residue in the K unit, Wzy_{K17} polymerase (GenPept accession number AIT75781.1) would catalyse formation of the β-D-Qui*p*NAc4NAc-(1→4)-D-Gal*p*NAcA6DAla linkage between the K units in the K17 CPS (Fig. 3). Though three glycosyltransferase genes (*gtr38, gtr39,* and *gtr40*) were initially identified in KL17 [1], the construction of the K unit requires only two glycosyltransferases for the internal linkages. However, a D-alanine transferase is needed for the attachment of D-alanine to the terminal D-Gal*p*NAcA residue. The possibility that that one of the predicted *gtr* genes may encode the D-alanine transferase was examined as follows.

3.4. Assignment of glycosyltransferase and D-alanine transferase genes

The $Gtr40_{K17}$ glycosyltransferase (GenPept accession number AIT75782.1) is 99.5%

identical to Gtr40$_{K35}$ (GenPept accession number AHB32264.1) encoded by the *A. baumannii* KL35 gene cluster. The K35 structure has been elucidated and Gtr40$_{K35}$ was predicted to form an α-D-Gal*p*NAcA-(1→3)-D-Qui*p*NAc4NAc linkage [14]. The same linkage is present in K17, and Gtr40$_{K17}$ was assigned to this linkage (Fig. 3). Gtr39$_{K17}$ (GenPept accession number AIT75780.1) is 34% identical to Gtr11$_{K4}$ (GenPept accession number AGK44884.1) from *A. baumannii* KL4, and Gtr11$_{K4}$ was predicted previously to catalyse formation of the α-D-Gal*p*NAc-(1→4)-D-Gal*p*NAcA linkage in the K4 CPS [12]. A similar linkage between two D-Gal*p*NAcA residues is found in the K17 structure, and Gtr39$_{K17}$ was assigned to this linkage.

The remaining gene, previously annotated as *gtr38*, encodes a protein (GenPept accession number AIT75779.1) that returns multiple hits to sequences annotated as either 'carboxylate–amine ligase' (45-87% identity), 'hypothetical protein'/'uncharacterised protein' (38-87% identity), 'tupA-like ATPgrasp family protein' (38-60% identity), or 'glycosyltransferase' (38-87% amino acid identity) in a BLASTp search against NCBI databases. The hits to 'glycosyltransferase' enzymes explain the previous annotation of Gtr38. However, this protein belongs to the ATPgrasp_TupA family of proteins (PF14305), which comprises a variety of enzymes including D-alanine-D-alanine ligases [23]. In light of the K17 CPS structure determined here, this protein was assigned to the addition of D-alanine to the carboxyl group of one of the Gal*p*NAcA residues. The *gtr38* gene was therefore reannotated as *alt1* for alanine transferase to be consistent with transparent gene designations in the established nomenclature system for *A. baumannii* CPS biosynthesis genes [4], and the revised annotations are shown in Fig. 1B.

3.5. D-Alanine transferases in *Acinetobacter* species

In *A. baumannii*, the *alt1* gene is only otherwise found in the KL18 gene cluster (Fig. 1B), which was first identified in the draft genome sequences of *A. baumannii* isolates OIFC065 and IS-116 (WGS accession numbers AMFV01000002.1 and AMGF01000004.1, respectively). The arrangement of KL18 is nearly identical to the KL17 gene cluster, though it differs from KL17 in the presence of an additional *gne1* gene in the module of genes for the synthesis of simple sugars (Fig. 1B).

The structure of a polysaccharide (referred to as O-antigen) produced by *Acinetobacter haemolyticus* type strain ATCC 17606 [24] is identical to the structure of the K17 CPS produced by *A. baumannii* G7. An *alt1* gene was found in the recently released draft genome of ATCC 17606 (also known as CIP 64.3 or NCTC 10305; WGS accession number APQQ01000000) and in draft genome sequences of several other *A. haemolyticus* isolates. The Alt1 gene product from *A. haemolyticus* shares 87% identity with Alt1 from *A. baumannii*, consistent with the two proteins having the same function.

The draft genome sequence of ATCC 17606 was found to carry a gene cluster closely related to KL17, with the same genetic content (85% nucleotide sequence identity) except that an additional *mnaA* gene is inserted between *wza* and *gna* (Fig. 1C). MnaA is an UDP-*N*-acetyl-glucosamine 2-epimerase that converts UDP-D-Glc*p*NAc to UDP-D-Man*p*NAc, and as D-Man*p*NAc is not present in the structure, the *mnaA* gene is not involved in the synthesis of the polysaccharide. The presence of CPS export genes in the *A. haemolyticus* gene cluster supports the conclusion that the polysaccharide studied by Haseley et al. was CPS.

4. Discussion

In this study, we assigned the formation of linkages between sugar residues found in the K17 CPS structure determined here to two glycosyltransferases and the Wzy polymerase encoded by the gene cluster. Preliminary annotations of the KL17 gene cluster were re-examined in

light of the structural data, and a transferase responsible for the addition of D-alanine to one of the two D-Gal*p*NAcA residues was identified. Accordingly, the earlier annotation of *gtr38* was corrected to *alt1* encoding an alanine transferase, Alt1. Granted that only one of the two D-Gal*p*NAcA residues in K17 is modified by the addition of D-Ala, it seems likely that this residue is added to the completed K unit rather than the UDP-D-Gal*p*NAcA precursor prior to incorporation of the terminal D-Gal*p*NAcA residue into the growing UndPP-linked trisaccharide.

K17 is the first D-alanine containing CPS structure to be reported for *A. baumannii,* and to date the *alt1* gene is only found in the KL17 and KL18 gene clusters in *A. baumannii*. These KL can be considered a gene cluster pair as they differ only in the presence of *gne1* in KL18 (Fig. 1B). As Gne1 converts D-Glc*p*NAc to D-Gal*p*NAc and the K17 unit does not include D-Glc*p*NAc, it is likely that the presence of Gne1 would not alter the CPS structure. Gne1 is not required for synthesis of the D-Gal*p*NAcA sugars and these are synthesised by Gna in combination with another 4-epimerase, Gne2. Therefore, the KL18 gene cluster is expected to produce K17 CPS.

The polysaccharide produced by *A. haemolyticus* ATCC 17606 is identical to K17 and it is likely that the ATCC 17606 polysaccharide is a CPS rather than an O-antigen as described [24]. The gene cluster found in the recently reported draft genome of ATCC 17606 includes all KL17 genes. However, the sequences of KL17 and the *A. haemolyticus* locus have diverged substantially sharing only 85% identity on average. Hence, it cannot be argued that the KL17 gene cluster in isolate G7 was acquired recently from that species. The ATCC 17606 locus also contains an additional gene, *mnaA,* that appears to be redundant and may be a remnant of an evolutionary event.

The closest relatives of the *Acinetobacter* Alt1 proteins found by searching the NCBI non-redundant database include the product of *wfdG* from *Escherichia coli* O167 (GenPept

accession number ACD37044.1) and *Shigella boydii* B3 (GenPept accession number ACD37035.1) O-antigen gene clusters and WfdR from *E. coli* O143 (GenPept accession number ACA24759.1) and *S. boydii* B8 (GenPept accession number ACA24749.1) [25, 26]. Like Alt1, both WfdG and WfdR belong to the ATPgrasp_TupA protein family and they share between 33-36% identity with Alt1. However, WfdG is likely to be responsible for the addition of L-alanine to a D-Gal*p*A residue of the O167/B3 O-antigens [26, 27]. The O143/B8 O-antigens do not include an alanyl residue but these structures both include a 2-amino-1,3-propanediol linked to a D-Gal*p*A residue [25, 28], and it is likely that WfdR catalyses formation of this linkage. As neither transfers a D-alanine residue, the homology shared with Alt1 can be explained by the similarity of the site of amidation at the carboxyl group of Gal*p*A in *E. coli/S. boydii* and Gal*p*NAcA in *Acinetobacter*. Hence, these transferases represent a new family that amidate the carboxyl group of Gal*p*NAcA and Gal*p*A with small molecules. However, further experiments would be needed to confirm the specificity and function of these predicted transferases.

Acknowledgements

Authors thank Dr. A.O. Chizhov for help with mass spectrometry. This work was financially supported by the Russian Science Foundation (Grant No. 19-14-00273) to YAK, an Australian Research Council (ARC) DECRA Fellowship DE180101563 to JJK, and a National Health and Medical Research Council (NHMRC) grant GNT1079616 to RMH.

References

[1] K.E. Holt, J.J. Kenyon, M. Hamidian, M.B. Schultz, D.J. Pickard, G. Dougan, R.M. Hall, Five decades of genome evolution in the globally distributed, extensively antibiotic resistant *Acinetobacter baumannii* global clone 1, MGen. 2 (2016) e000052.

[2] L. Diancourt, V. Passet, A. Nemec, L. Dijkshoorn, S. Brisse, The population structure of *Acinetobacter baumannii*: Expanding multiresistant clones from an ancestral susceptible genetic pool, PLoS One. 5 (2010) e10034.

[3] J. Sahl, M. Del Franco, S. Pournaras, R. Colman, N. Karah, L. Dijkshoorn, R. Zarrilli, Phylogenetic and genomic diversity in isolates from the globally distributed *Acinetobacter baumannii* ST25 lineage. Sci. Rep. 5 (2015) 15188.

[4] J.J. Kenyon, R.M. Hall, Variation in the complex carbohydrate biosynthesis loci of *Acinetobacter baumannii* genomes. PLoS One. 8 (2013) e62160.

[5] M.S. Wright, D. Haft, D. Harkins, F. Perez, K. Hujer, S. Bajaksouzian, M. Bernard, M. Jacobs, R. Bonomo, M. Adams, New insights into dissemination and variation of the health care-associated pathogen *Acinetobacter baumannii* from genomic analysis, mBio. 5 (2014) e00963-13

[6] M.B. Schultz, D.P. Thanh, N. Hoan, R. Wick, D.J. Ingle, J. Hawkey, D. Edwards, J.J. Kenyon, N. Lan, J. Campbell, G. Thwaites, N. Nhu, R.M. Hall, A. Fournier-Level, S. Baker, K.E. Holt, Repeated local emergence of carbapenem-resistant *Acinetobacter baumannii* in a single hospital ward. MGen 2 (2016) e000050.

[7] D. Hu, B. Liu, L. Dijkshoorn, L. Wang, P.R. Reeves, Diversity in the major polysaccharide antigen of *Acinetobacter baumannii* assessed by DNA sequencing, and development of a molecular serotyping scheme, PLoS One 8 (2013) e70329.

[8] H. Oliveira, A. Costa, N. Konstantinides, A. Ferreira, E. Akturk, S. Sillankorva, A. Nemec, M.M. Shneider, A. Dötsch, J. Azeredo, Ability of phages to infect

Acinetobacter calcoaceticus-Acinetobacter baumannii complex species through acquisition of different pectate lyase depolymerase domains, Environ. Microbiol. 19 (2017) 5060-5077.

[9] H. Oliveira, A. Costa, A. Ferreira, N. Konstantinides, S. Santos, M. Boon, J. Noben, R. Lavigne, J. Azeredo, Functional analysis and antivirulence properties of a new depolymerase from a Myovirus that infects *Acinetobacter baumannii* capsule K45, J. Virol. 93 (2019) e01163-18.

[10] T.A. Russo, J. Beanan, R. Olson, U. MacDonald, A. Cox, F. St. Michael, E. Vinogradov, B. Spellberg, N. Luke-Marshall, A.A. Campagnari, The K1 capsular polysaccharide from *Acinetobacter baumannii* is a potential therapeutic target via passive immunization, Infect. Immun. 81 (2013) 915-922.

[11] M. Hamidian, J. Hawkey, R. Wick, K.E. Holt, R.M. Hall, Evolution of a clade of *Acinetobacter baumannii* global clone 1, lineage 1 via acquisition of carbapenem- and aminoglycoside-resistance genes and dispersion of ISAba1. MGen. 6 (2019) e000242.

[12] J.J. Kenyon, I. Speciale, R.M. Hall, C. De Castro, Structure of repeating unit of the capsular polysaccharide from *Acinetobacter baumannii* D78 and assignment of the K4 gene cluster, Carbohydr. Res. 434 (2016) 12-17.

[13] J.J. Kenyon, A.M. Marzaioli, R.M. Hall, C. De Castro, Structure of the K12 capsule containing 5,7-di-N-acetylacinetaminic acid from *Acinetobacter baumannii* isolate D36. Glycobiology. 25 (2015) 881-887.

[14] A.S. Shashkov, B. Liu, J.J. Kenyon, A.V. Popova, M.M. Shneider, S.N. Senchenkova, N.P. Arbatsky, K.A. Miroshnikov, L. Wang, Y.A. Knirel, Structures of the K35 and K15 capsular polysaccharides of *Acinetobacter baumannii* LUH5535 and LUH5554 containing amino and diamino uronic acids. Carbohydr. Res. 448 (2017) 28-34.

[15] A.A. Kasimova, J.J. Kenyon, N.P. Arbatsky, A.S. Shashkov, A.V. Popova, M.M. Shneider, Y.A. Knirel, R.M. Hall, *Acinetobacter baumannii* K20 and K21 capsular polysaccharide structures establish roles for UDP-glucose dehydrogenase Ugd2, pyruvyl transferase Ptr2 and two glycosyltransferases, Glycobiology. 28 (2018) 876-884.

[16] S.N. Senchenkova, A.S. Shashkov, A.V. Popova, M.M. Shneider, N.P. Arbatsky, K.A. Miroshnikov, N. Volozhantsev, Y.A. Knirel, Structure elucidation of the capsular polysaccharide of *Acinetobacter baumannii* AB5075 having the KL25 capsule biosynthesis locus, Carbohydr. Res. 408 (2015) 8-11.

[17] M.M. Hamidian, R.M. Hall, AbaR4 replaces AbaR3 in a carbapenem-resistant *Acinetobacter baumannii* isolate belonging to global clone 1 from an Australian hospital, J. Antimicrob. Chemother. 66 (2011) 2484-2491.

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

[19] J.J. Kenyon, M.M. Shneider, S.N. Senchenkova, A.S. Shashkov, M. Siniagina, S. Malanin, A.V. Popova, K.A. Miroshnikov, R.M. Hall, Y.A. Knirel, K19 capsular polysaccharide of *Acinetobacter baumannii* is produced via a Wzy polymerase encoded in a small genomic island rather than the KL19 capsule gene cluster. Microbiology. 162 (2016) 1479-1489.

[20] M.J. Morrison, B. Imperiali. Biosynthesis of UDP-N,N'-diacetylbacillosamine in *Acinetobacter baumannii*: Biochemical characterization and correlation to existing pathways. Arch. Biochem. Biophys. 536 (2013) 72-80.

[21] V. L'vov, A. Filatov, A. Perepelov, A. Shpirt, A.S. Shashkov, A. Chizhov, Y.A. Knirel, Solvolysis with trifluoroacetic acid: an efficient method for selective cleavage of polysaccharides, Mendeleev Commun. 26 (2016) 279-281.

[22] Y.A. Knirel, O. Naumenko, S.N. Senchenkova, A. Perepelov. Chemical methods for selective cleavage of glycosidic linkages in structural analysis of bacterial polysaccharides. Russ. Chem. Rev. 88 (2019) 406-424.

[23] M.Y. Galperin, E.V. Koonin, A diverse superfamily of enzymes with ATP-dependent carboxylate-amine/thiol ligase activity, Protein Sci. 6 (1997) 2639-2643.

[24] S. Haseley, O. Holst, H. Brade, Structural and serological characterisation of the O-antigenic polysaccharide of the lipopolysaccharide from *Acinetobacter haemolyticus* strain ATCC 17906, Eur. J. Biochem. 244 (1997) 761-766.

[25] C. Landersjo, A. Weintraub, G. Widmalm. Structure determination of the O-antigen polysaccharide from the enteroinvasive *Escherichia coli* (EIEC) O143 by component analysis and NMR spectroscopy. Carbohydr. Res. 291 (1996) 209-216.

[26] B. Liu, Y.A. Knirel, L. Feng, A. Perepelov, S.N. Senchenkova, Q. Wang, P.R. Reeves, L. Wang, Structure and genetics of *Shigella* O antigens. FEMS Microbiol. Rev. 32 (2008) 627-653.

[27] M. Linnerborg, R. Wollin, G. Widmalm, Structural studies of the O-antigenic polysaccharide from *Escherichia coli* O167, Eur. J. Biochem. 246 (1997) 565-573.

[28] V. L'vov, N. Tokhtamysheva, A.S. Shashkov, B. Dmitriev, N. Kochetkov, Structure and ^{13}C NMR spectrum of the polysaccharide chain of *Shigella boydii* type 8 lipopolysaccharide. Bioorg. Khim. 9 (1983) 60-73.

Figures

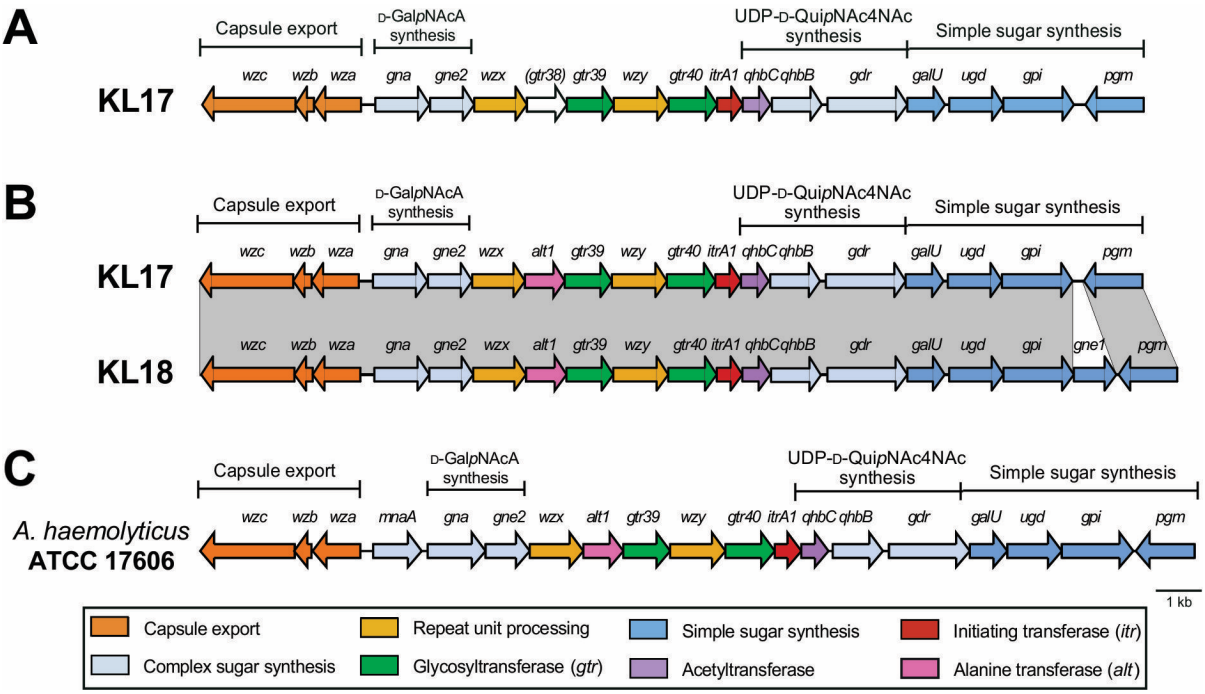

Figure 1. A. The original annotations of the KL17 capsule biosynthesis gene cluster from *A. baumannii* isolate G7 (Holt et al. 2016). Annotation of the white gene has been revised in this study. Figure is drawn to scale from GenBank accession number KC118541.2. **B.** Relationship of the *A. baumannii* KL17 and KL18 capsule biosynthesis gene clusters annotated in this work. Dark grey shading between the gene clusters indicates >90% nucleotide sequence identity. Figures are drawn to scale from GenBank accession numbers KC118541.2 (KL17) and AMFV01000002.1 (KL18). **C.** *A. haemolyticus* polysaccharide gene cluster drawn to scale from WGS accession number APQQ01000000. Colour scheme for genes shown below indicates the predicted functions of the encoded products.

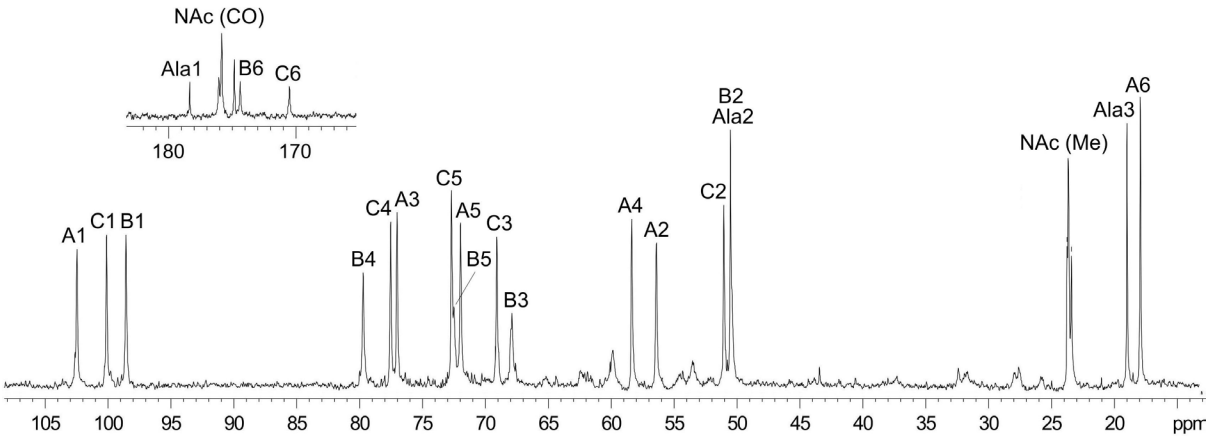

Fig. 2. ^{13}C NMR spectrum of the K17 capsular polysaccharide of *A. baumannii* G7. Numbers refer to carbons in alanine and sugar residues denoted by letters as indicated in Table 1.

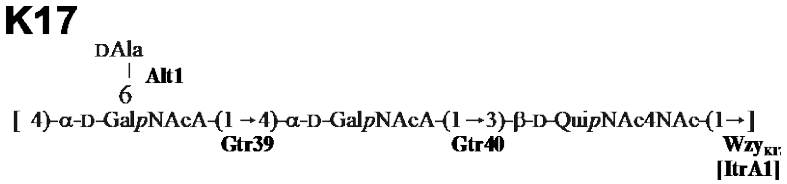

Fig. 3. Structure of the K17 CPS from *A. baumannii* G7. The glycosyltransferases, alanine transferase, initiating transferase and Wzy polymerase are shown in bold next to the linkage each are assigned to.

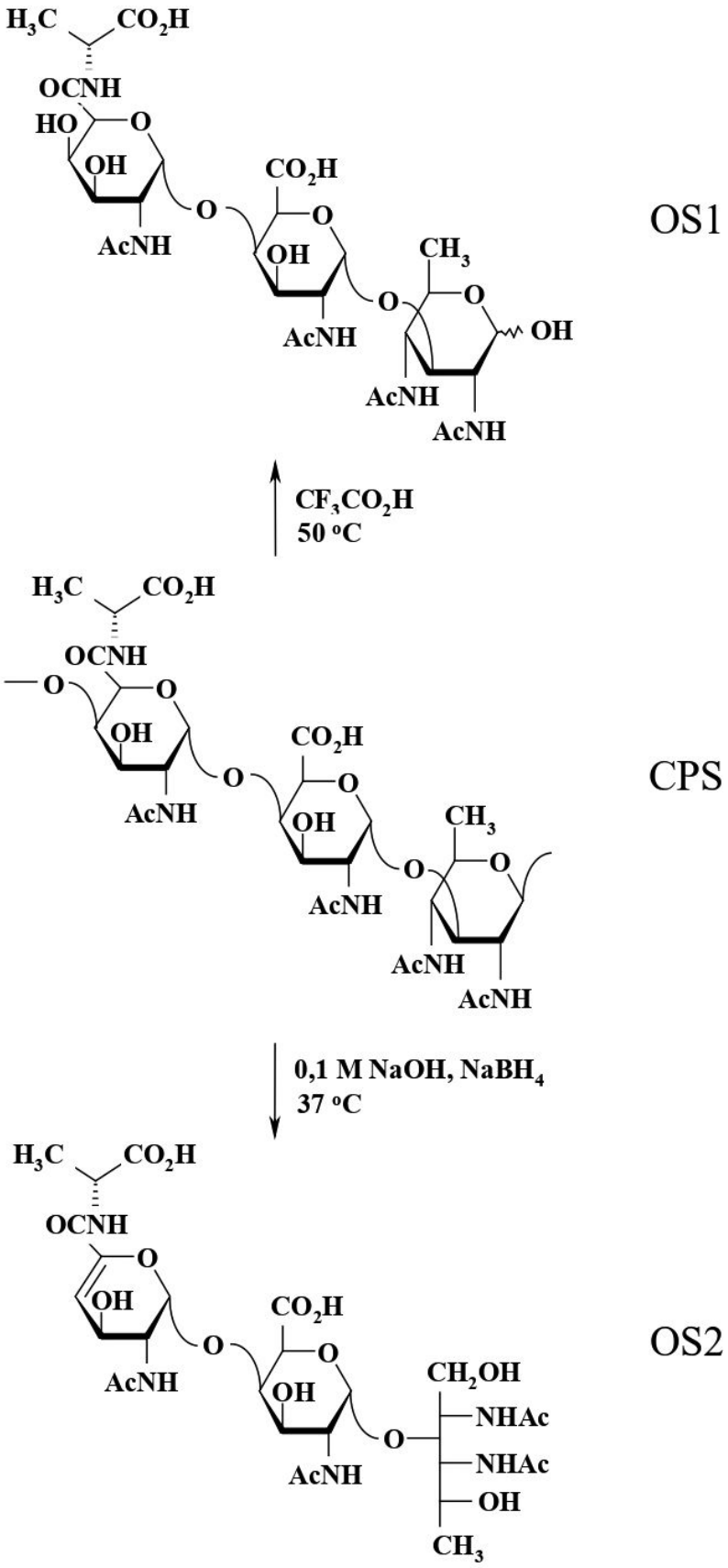

Figure 4. Chemical selective cleavages of the K17 capsular polysaccharide of *A. baumannii* G7.

Supplementary Material

Fig. S1. Parts of a $^{1}H,^{13}C$ HMBC spectrum of the K17 capsular polysaccharide of *A. baumannii* G7 showing correlations for CO groups. Numbers refer to protons and carbons in N-acetyl groups, alanine and sugar residues denoted by letters as indicated in Table 1.

Fig. S2. Part of a $^{1}H,^{1}H$ ROESY spectrum of the K17 capsular polysaccharide of *A. baumannii* G7 showing correlations for anomeric protons. Numbers refer to protons in sugar residues denoted by letters as indicated in Table 1.

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

Residue	*H-1* *(H-1a,1b)* C-1	*H-2* C-2	*H-3* C-3	*H-4* C-4	*H-5* C-5	*H-6* *(H-6a,6b)* C-6
CPS						
→3)-β-D-Qui*p*NAc4NAc-(1→	*4.65*	*3.68*	*3.85*	*3.66*	*3.34*	*1.07*
A	102.5	56.4	77.0	58.4	72.0	17.9
→4)-α-D-Gal*p*NAcA-(1→	*5.09*	*4.16*	*3.88*	*4.38*	*4.19*	
B	98.6	50.5	67.9	79.7	72.5	174.4
→4)-α-D-Gal*p*NAcA-(1→	*5.05*	*4.13*	*4.05*	*4.32*	*4.70*	
C	100.1	51.1	69.1	77.5	72.7	170.5
D-Ala		*4.33*	*1.37*			
	178.4	50.5	19.0			
OS1						
→3)-α-D-Qui*p*NAc4NAc	*5.10*	*4.01*	*3.99*	*3.77*	*3.95*	*1.13*
Aα	92.2	54.4	74.9	58.3[a]	68.0	18.0
→3)-β-D-Qui*p*NAc4NAc	*4.70*	*3.72*	*3.81*	*3.77*	*3.52*	*1.16*
Aβ	95.9	57.4	77.0	58.5[a]	72.6	18.0
→4)-α-D-Gal*p*NAcA-(1→	*5.19*	*4.18*	*3.91*	*4.42*	*4.18*	
B	98.6, 98.9	50.4	67.9	79.7, 79.8	72.5	174.4
α-D-Gal*p*NAcA-(1→	*5.08*	*4.22*	*4.01*	*4.30*	*4.77*	
C	100.2	50.8	68.8	70.3	73.0	171.8
D-Ala		4.37	1.39			
	178.4	50.4	18.2			
OS2						
→3)-D-QuiNAc4NAc-ol	*3.48, 3.62*	*4.00*	*4.28*	*3.86*	*3.64*	*1.16*
A'	61.8	55.1	77.0	57.1	67.1	20.6
→4)-α-D-Gal*p*NAcA-(1→	*5.20*	*4.10*	*4.09*	*4.58*	*4.40*	
B	100.0	50.8	67.0	79.3	72.3	175.0
Hexen*p*NAcA-(1→	*5.28*	*4.08*	*4.52*	*5.94*		
C'	100.3	53.3	64.7	110.7	143.3	169.6

D-Ala		*4.39*	*1.44*
	177.6	50.4	17.5

[1]H NMR chemical shifts are italicised.

Chemical shifts for the N-acetyl groups are δ_H 1.93-2.08, δ_C 23.2-23.8 (Me) and 175.1-176.1 (CO).

[a]Assignment could be interchanged.