A novel ItrA4 D-galactosyl 1-phosphate transferase is predicted to initiate synthesis of an amino sugar-lacking K92 capsular polysaccharide of *Acinetobacter baumannii* B8300

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

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

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

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

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

[e] *Department of Clinical Microbiology, Christian Medical College, Vellore, 632004 Tamil Nadu, India*

[f] *Department of Microbiology, Molecular Genetics and Immunology, University of Kansas Medical Center, Kansas City, KS 66160, USA*

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

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

[†] and [*] Contributed equally.

Abstract

The K92 capsular polysaccharide (CPS) from *Acinetobacter baumannii* B8300 was studied by sugar analysis, Smith degradation, and one- and two-dimensional ^{1}H and ^{13}C NMR spectroscopy. The elucidated CPS includes a branched pentasaccharide repeat unit containing one D-Gal*p* and four L-Rha*p* residues; an atypical composition given that all *A. baumannii* CPS structures determined to date contain at least one amino sugar. Accordingly, biosynthesis of *A. baumannii* CPS types are initiated by initiating transferases (Itrs) that transfer 1-phosphate of either a 2-acetamido-2-deoxy-D-hexose, a 2-acetamido-2,6-dideoxy-D-hexose or a 2-acetamido-4-acylamino-2,4,6-trideoxy-D-hexose to an undecaprenyl phosphate (UndP) carrier. However, the KL92 capsule biosynthesis gene cluster in the B8300 genome sequence includes a gene for a novel Itr type, ItrA4, which is predicted to begin synthesis of the K92 CPS by transferring D-Gal*p* 1-phosphate to the UndP lipid carrier. The *itrA4* gene was found in a module transcribed in the opposite direction to the majority of the K locus. This module also includes an unknown open reading frame (*orf*$_{KL92}$), a *gtr166* glycosyltransferase gene, and a *wzi* gene predicted to be involved in the attachment of CPS to the cell surface. Investigation into the origins of *orf*$_{KL92}$*-gtr166-itrA4-wzi*$_{KL92}$ revealed it might have originated from *Acinetobacter junii*.

Keywords: *Acinetobacter baumannii*; capsular polysaccharide; capsule biosynthesis; ItrA4; initiating transferase; galactosyl 1-phosphate transferase

1. Introduction

Most bacterial pathogens, including the important nosocomial species, *Acinetobacter baumannii,* are able to produce capsular polysaccharide (CPS) that plays an important role in virulence, providing a protective barrier between the cell and the external environment. This outer polysaccharide layer includes chains of repeating oligosaccharides (K units) and may be incredibly diverse between different isolates in both sugar content and the linkages between sugars. For *A. baumannii*, more than 130 different genetic arrangements have been found at the chromosomal K locus (KL) that directs CPS biosynthesis [1]. Due to this diversity, the construction of the CPS can involve multiple sugar synthesis and/or assembly pathways [2-4].

The construction of K units always occurs in the cytoplasm, beginning with the transfer of a nucleotide-linked sugar to an undecaprenol phosphate (UndP) lipid carrier in the inner membrane [5]. This process is carried out by an enzyme known as the initiating transferase (Itr). Six different Itr proteins have been described for *A. baumannii*, though only five of them have been shown to be functional, each being able to transfer a different nucleotide-linked sugar to initiate K-unit synthesis. The Itr types fall into one of two major families, named ItrA and ItrB, which are differentiated by the length of the proteins [4], being 200-210 or 335-338 amino acids (aa), respectively. Each family includes three different sequence types (determined by an 80% aa identity cutoff), which can be further distinguished by the substrates they preferentially use. Among ItrAs, ItrA1 is specific for UDP-D-Qui*p*NAc4NAc [6-8], while ItrA2 and ItrA3 utilize UDP-D-Gal*p*NAc [8-16] and UDP-D-Glc*p*NAc [16,17], respectively. For ItrBs, ItrB1 and ItrB3 recognise UDP-D-Qui*p*NAc [8] and UDP-D-Fuc*p*NAc [18,19], respectively.

Correlation of a CPS structure with the genetic content of the K locus of the same isolate has been used to analyse the functional roles of CPS biosynthesis proteins *in silico*, including ItrA and ItrB types. In this study, we used the same approach to describe a novel

ItrA type presumably involved in the synthesis of the K92 CPS produced by *A. baumannii* B8300, a twitching-positive clinical isolate from India that is susceptible to all commonly used antibiotics [20].

2. Materials and methods

2.1 Cultivation of the bacterium and isolation of CPS

A. baumannii B8300 was isolated from the bloodstream infection of a 29-year-old male patient from the Christian Medical College, Vellore, India in 2014 [20]. It was cultivated in 2× TY medium for 24 h. Cells were collected by centrifugation at 12,000*g*, washed with a 7 : 3 acetone–water mixture, and dried.

CPS was isolated by phenol–water extraction [21] of bacterial mass (0.85 g), the crude extract was dialyzed without layer separation and freed from insoluble contaminations by centrifugation. As *A. baumannii* produces lipooligosaccharide (LOS) devoid of an O-antigen polysaccharide chain [1,4], only pure CPS sample was recovered from the extract**.** Proteins and nucleic acids were removed by precipitation with aqueous 50% CCl_3CO_2H at 0 °C; after centrifugation, the supernatant was dialyzed against distilled water and lyophilized. A purified CPS sample (29 mg) was obtained by heating the crude CPS with 2% HOAc (100 °C, 5 h) followed by centrifugation and lyophilization.

2.2 Composition analysis

A CPS sample was hydrolysed with 2 M CF_3CO_2H (120°C, 2 h). Monosaccharides were analysed by GLC of the alditol acetates on a Maestro (Agilent 7820) chromatograph (Interlab, Russia) equipped with a HP-5 column (0.32 mm × 30 m) using a temperature program of 160 °C (1 min) to 290 °C at 7 °C/min. The absolute configurations of monosaccharides were determined by GLC of the acetylated (*S*)-2-octyl glycosides [22].

2.3 Smith degradation

A CPS sample was oxidized with aqueous 1% $NaIO_4$ (1 mL) at 20 °C for 48 h in the dark, reduced with $NaBH_4$ (30 mg) for 3 h, the excess of $NaBH_4$ was destroyed by adding concentrated HOAc, the solution was evaporated, the residue was evaporated with methanol (three times) and applied to a column (80 × 1.6 cm) of Fractogel TSK HW-40S. Elution with 1% HOAc monitored using a differential refractometer (Knauer, Germany) afforded an oxidized polysaccharide, which was hydrolysed with 2% AcOH (100 ºC, 3 h) to give a modified polysaccharide (MPS) isolated as above. An MPS sample was subjected to Smith degradation and the products were fractionated by gel-permeation chromatography on Fractogel TSK HW-40S as above to give an oligosaccharide (OS).

2.4 NMR spectroscopy

Samples were deuterium-exchanged by lyophilization from 99.9% D_2O and then examined as solutions in 99.95% D_2O at 60 °C for the polysaccharides or 30 °C for the oligosaccharide using internal sodium 3-trimethylsilylpropanoate-2,2,3,3-d_4 (δ_H 0.0, δ_C −1.6) as a reference for calibration. The ^{13}C NMR spectra were recorded on a Bruker DRX-500 instrument, and two-dimensional NMR spectra were run on a Bruker Avance II 600 MHz spectrometer (Germany) using standard Bruker software. A spin-lock time of 60 ms and a mixing time of 150 ms were used in TOCSY and ROESY experiments, respectively. A 60-ms delay was applied for evolution of long-range coupling to optimize the $^{1}H,^{13}C$ HMBC experiment. Other NMR parameters were set as previously described [7]. The Bruker TopSpin 2.1 program was used to acquire and process the NMR data.

2.5 Bioinformatics

The sequence at the K locus in *A. baumannii* B8300 was obtained from NCBI GenBank accession number CP021347.1 (base positions 1420707 to 1449646). Genes were identified and annotated using the nominated nomenclature system [4], and protein functions were predicted as described previously [23]. Sequence and gene annotations for KL92 and KL99 were submitted to GenBank under accession numbers MN958100.1 and MN958101.1, respectively.

3. Results

3.1 CPS gene cluster in the chromosome of Acinetobacter baumannii *B8300*

The CPS gene cluster from *A. baumannii* B8300 (GenBank accession number MN958100.1) has a typical arrangement for gene clusters found at the K locus in *A. baumannii* [1,4]. It includes a central region specific to the CPS type that is flanked on one side by a module of genes for CPS export and on the other side by genes for synthesis of common sugars (Fig. 1). As the central region of the B8300 gene cluster includes a novel arrangement of genes for *A. baumannii*, the gene cluster was assigned the name KL92 (GenBank accession number MN958100.1). This specific portion of KL92 (Fig. 1) includes *rml* genes for the synthesis of dTDP-L-rhamnose (dTDP-L-Rha*p*), *wzx* and *wzy* processing genes, and three *gtr* genes (*gtr164, gtr165, gtr166*) that encode predicted glycosyltransferases responsible for linking the sugars together to form the K unit.

Interestingly, the *gtr166* gene is positioned within a 2.8 kb segment that is transcribed in the opposite direction to the majority of other genes at the locus (Fig. 1). This segment also includes an open reading frame (orf_{KL92}) encoding a product of unknown function, and a gene whose product (GenPept accession number KMV25998.1) shares 58% identity to the last 235

aa of WbaP from *Salmonella enterica* sv. Typhimurium Group B (GenPept accession number AAC44095.1). WbaP is an initiating transferase that transfers D-Gal*p*-1-phosphate from UDP-D-Gal*p* to UndP [24]. Remarkably, this protein does not share significant identity with any Itr previously reported for *A. baumannii.* Thus, granted the predicted protein is 232 aa in length, it was designated ItrA4 representing the fourth ItrA type in the species. Given that there is no alternate *itr* gene in KL92, ItrA4 would likely initiate the synthesis of K92, utilizing UDP-D-Gal*p* as a source of D-Gal*p*-1-phosphate.

The *itrA4-gtr166- orf*$_{KL92}$ segment is immediately adjacent to a gene for a putative Wzi protein, which shares 42.4% identity with Wzi from *Escherichia coli* O9a:K30 (GenPept accession number AAD21561.1) known to encode an outer membrane protein anchor required for organisation of the CPS on the cell surface [25,26]. In *A. baumannii,* a putative *wzi* gene has only ever been reported outside of the K locus, adjacent to a *lysS* lysyl-tRNA synthetase gene elsewhere in the *A. baumannii* chromosome [4]. However, its role in CPS synthesis or attachment has not been experimentally confirmed. The B8300 genome also contains this same chromosomal *wzi* gene (designated *wzi*$_C$), and the Wzi$_C$ product (GenPept accession number KMV25104.1) is 67% identical to Wzi$_{KL92}$ (GenPept accession number KMV25997.1). The two proteins belong to the same Caps_assemb_Wzi protein family (PF14052), though experimental validation would be needed to determine whether one or both proteins are functional.

Taken together, the analysis of the KL92 gene cluster enables the prediction that the K92 CPS structure will be made up of a pentasaccharide repeating unit composed of D-Gal*p* and L-Rha*p*, with a D-Gal*p* residue representing the first sugar.

3.2 Structure elucidation of the CPS

A CPS preparation was isolated from cells of *A. baumannii* B8300 by phenol–water

extraction and purified by precipitation of contaminating proteins and nucleic acids with trichloroacetic acid. Sugar analysis using GLC of the alditol acetates derived after full acid hydrolysis of the CPS revealed Rha and Gal in the ratio ~4 : 1 (detector response). GLC analysis of the acetylated (*S*)-2-octyl glycosides indicated that Gal has the D configuration and Rha has the L configuration.

The ^{1}H NMR spectrum of the CPS showed signals for five anomeric protons at δ 4.63-5.08, four CH_3-C groups (Rha H-6) at δ 1.29-1.36, and other sugar protons at δ 3.48-4.17 (Table 1). The ^{13}C NMR spectrum (Fig. 2) contained signals for five anomeric carbons between 103.1 and 104.2 ppm, four *C*H_3-C groups (Rha C-6) between 17.8-18.3 ppm, one C-*C*H_2O group (Gal C-6) at δ 62.1, and other sugar carbons at δ 69.3-81.8 (Table 1).

Assignment of the NMR spectra of the CPS was complicated by close positions of the anomeric signals of four rhamnose residues (units **B-E**) at δ_H 5.05-5.08. Therefore, the CPS was subjected to Smith degradation, and the resulting modified polysaccharide (MPS) was subjected to Smith degradation again to give an oligosaccharide (OS). Assignment of the ^{1}H NMR and ^{13}C NMR (Fig. 2) spectra of the OS and MPS (Table 1) was performed using ^{1}H,^{1}H COSY, ^{1}H,^{1}H TOCSY, ^{1}H,^{13}C HSQC ^{1}H,^{13}C HSQC-TOCSY and ^{1}H,^{13}C HMBC experiments, and, based on these data, the NMR spectra of the CPS were assigned (Table 1).

In accordance with sugar composition of the CPS, spin systems were identified for one β-galactopyranose residue (unit **A**) and four α-rhamnopyranose residues (units **B-E**). Low-field positions of the signals for C-3 of units **A**, **B**, and **C**, and C-3 and C-4 of unit **D** at δ 78.4-81.8, as compared with their positions in the corresponding non-substituted monosaccharides at δ 71.3-74.1 [27], showed that the CPS is branched and defined the glycosylation pattern in the repeating unit. The ^{13}C NMR chemical shifts of C-2−C-6 of unit **E** were similar to those of unsubstituted α-Rha*p* and thus indicated a terminal side-chain position of this monosaccharide [27]. Using the ^{13}C NMR chemical shift data and one-

dimensional NOE or two-dimensional ^{1}H,^{1}H ROESY and two-dimensional ^{1}H,^{13}C HMBC experiments, the sequence of the monosaccharide residues was established stepwise in the OS, MPS, and CPS (Supplementary Table 1). Based on the data obtained, it was concluded that these compounds have the structures shown in Fig. 3. From them, the linear MPS was derived from the branched CPS by elimination of the side-chain Rha*p* residue (unit **E**), and the OS from the MPS by destruction of the 4-substituted Rha*p* residue (unit **D**). A peculiar feature of the CPS of *A. baumannii* B8300 is the absence of any amino sugars, whereas all other *A. baumannii* CPSs with known structures include at least one 2-acetamido-2-deoxyhexose [6-19, 23, 28, 29].

3.3 Confirmation of ItrA4, a new A. baumannii initiating transferase

The absence of *N*-acetylated amino sugars in K92, combined with the finding of an *itrA4* gene for a possible D-Gal*p* 1-phosphate transferase, indicates that the D-Gal*p* residue represents the first sugar of the K unit. This finding and elucidation of the K92 CPS structure (Fig. 3) together demonstrate that the K92 units are joined by the β-D-Gal*p*-(1→4)-L-Rha*p* linkage. Formation of this linkage is catalysed by the Wzy$_{K92}$ polymerase (GenPept accession number KMV26001.1), which is responsible for elongating the CPS polymer prior to export to the cell surface.

3.4 Assignment of glycosyltransferases to glycosidic linkages in the K unit

The K92 unit contains four L-Rha*p* residues and a single D-Gal*p* residue (Fig. 3). This sugar composition correlates with the presence of *rml* and *gne1* genes in the KL92 gene cluster, respectively. The finding of a pentasaccharide unit suggests the requirement for four glycosyltransferases to form the four internal linkages. However, there are only three *gtr* genes present in the KL92 gene cluster (Fig. 1). Though it is possible that *orf$_{KL92}$* may encode a

fourth glycosyltransferase, the predicted product does not belong to a known protein family (Pfam), and similarity searches using BLASTp and ISFinder could not shed light on its potential function. Therefore, specific linkages were assigned to the three predicted Gtrs based on homology to other proteins of known or deduced functions.

Gtr166$_{KL92}$ (GenPept accession number KMV25999.1) is 66% identical to Gtr82 encoded by the *A. baumannii* KL55, KL74 and KL85 gene clusters (GenBank accession numbers MN148383.1, MN148381.1 and KC526897.2). Recently, Gtr82 has been inferred to form an α-L-Rha*p*-(1→3)-D-Glc*p*NAc linkage in the K55, K74 and K85 structures [28], which is similar to α-L-Rha*p*-(1→3)-D-Gal*p* in K92 (Fig. 3). Therefore, it is predicted that Gtr166$_{KL92}$ forms this linkage.

The remaining two rhamnosidic linkages in the K92 main chain, and the side-branch linkage to this main chain, are all α-L-Rha*p*-(1→3)-L-Rha*p*. Therefore, either Gtr164$_{K92}$ or Gtr165$_{K92}$ is capable of catalysing more than one of these L-Rha*p* linkages. Gtr164$_{KL92}$ (GenPept accession number KMV26004.1) shares 23% identity with Gtr79$_{K55}$ (GenPept accession number QHE90303.1) that was unambiguously assigned to a α-L-Rha*p*-(1→3)-L-Rha*p* linkage that terminates the side branch of the K55 and K74 CPS structures [28]. Thus, it is suggested that Gtr164$_{K92}$ is responsible for the formation of the same rhamnosidic linkage that terminates the K92 unit (Fig. 3). This is consistent with the observation that the *gtr* found nearest to the *wza-wzb-wzc* gene module, like *gtr164$_{K92}$* (Fig. 1), usually encodes the Gtr that forms the last (often the side-branch) linkage in the K unit [28].

Gtr165$_{K92}$ (GenPept accession number KMV26002.1) shares 36% identity to Gtr112 also encoded by the KL55, KL74 and KL85 gene clusters. The Gtr112 protein (GenPept accession number QHE90312.1) was recently predicted to be a multifunctional glycosyltransferase capable of forming two α-(1→2) linkages and one α-(1→3) linkage between L-Rha*p* residues in the K55, K74 and K85 structures [28]. Since the K92 main chain

contains multiple α-(1→3)-linked L-Rha*p* residues, it is probable that Gtr165$_{KL92}$ is also multifunctional, forming both α-L-Rha*p*-(1→3)-L-Rha*p* linkages in the K-unit main chain.

3.5 Origins of the itrA4 *sequence*

The *itrA4* gene was identified in a further three *A. baumannii* genome sequences available in the NCBI non-redundant and whole genome sequence (WGS) databases. However, all three carry novel gene clusters at the K locus, and hence were designated KL99, KL142, and KL143 (see Table 2 for strain details). Though the arrangements are different to KL92 and to each other, the *itrA4* gene is found in the same orientation flanked by *gtr166* and *wzi* in each gene cluster (Fig. 4A). All gene clusters also include *rmlBDAC* genes. However, the *orf$_{KL92}$* sequence appears to be unique to KL92. Though the common *gtr166-itrA4-wzi* segment shares >95% nucleotide sequence identity between the four gene clusters (Fig. 4A), the other shared sequence portions have various ranges of sequence identity in pairwise alignments suggesting a complex evolutionary history.

Outside *A. baumannii*, the same *orf$_{KL92}$-gtr166-itrA4-wzi* segment is only otherwise found in *Acinetobacter junii*. A gene cluster closely resembling KL92 (Fig. 4B) was identified in the *A. junii* SH205 draft genome sequence (WGS accession number ACPM01000120.1), and the gene cluster is ~80% identical across 22479 bp of the 28940 bp KL92 locus, suggesting that the sequences may have separated at a time closer to species divergence. The *A. junii* SH205 gene cluster is missing *gna* and *pgt1* genes that are present in KL92. However, as typical of other *A. baumanni* CPS gene clusters, these genes are not expected to play a role in CPS synthesis, and the *A. junii* SH205 CPS structure would likely be identical to the K92 structure resolved in this study.

4. Discussion

The K92 CPS whose structure was elucidated in this study represents the first *A. baumannii* CPS described to date that does not contain any *N*-acetylamino sugar, and includes D-Gal*p* as the first monosaccharide of the K unit. Here, we deduce that the transfer of this sugar to the UndP lipid carrier to initiate K-unit synthesis would be carried out by the ItrA4 protein, which is related to the WbaP initiating transferase in *S. enterica* [24]. ItrA4 therefore represents a new ItrA initiating transferase in the species, increasing the number of substrates that may be used to initiate CPS synthesis in *A. baumannii*.

The *itrA4* gene is located within a module that is unusually transcribed in the opposite orientation to the majority of the K locus, and is flanked by orf_{KL92} and *gtr166* on one side, and *wzi* on the other. The *gtr166, itrA4,* and *wzi* genes were only otherwise found in three other *A. baumannii* KL gene clusters (Fig. 4A), and in all cases, were found together as a module suggesting that the *gtr166-itrA4-wzi* genes may move together as a unit. Investigation into the origins of the orf_{KL92}-*gtr166-itrA4-*wzi_{KL92} module revealed that it might have originated from *A. junii*, though the sequences have remained separate for some time. It is also interesting to note the combination of genes present with *itrA4* in this module. Given that ItrA4 initiates CPS synthesis with transfer of D-Gal*p* 1-phosphate, and Gtr166 uses D-Gal*p* as an acceptor substrate, the synthesis of the CPS relies on the two genes coding for these proteins to be present together. Thus, it is also plausible to suggest that Wzi_{KL92} may be needed for the cell surface attachment of CPS polymers beginning with D-Gal*p*. Nonetheless, the four different KLs that carry *itrA4* are each found in only 1-2 isolates amongst >3400 *A. baumannii* sequences available in NCBI, indicating the rare occurrence of *gtr166-itrA4-wzi* in the species.

The K92 structure represents another example for which the number of internal linkages in a K unit does not match the number of Gtrs encoded by the *A. baumannii* K locus. Recently, a multifunctional glycosyltransferase (Gtr112) capable of forming α-(1→2) and α-(1→3) linkages between L-Rha*p* residues was described for the K55, K74 and K85 CPSs [28].

Gtr165 shares similarity with this glycosyltransferase, therefore we suggest that Gtr165 is also able to catalyse formation of more than one α-L-Rha*p*-(1→3)-L-Rha*p* linkage in the K92 CPS. However, biochemical studies are needed to confirm the multifunctional role of this protein and the specific linkages formed by Gtr164 and Gtr166 in the construction of the K92 unit.

Conflicts of interest

None declared.

Acknowledgements

This work was supported by the Russian Science Foundation (project No. 19-14-00273 to YAK), and an Australian Research Council (ARC) DECRA Fellowship DE180101563 to JJK.

References

[1] Wyres KL, Cahill SM, Holt KE, Hall RM, Kenyon JJ. Identification of *Acinetobacter baumannii* loci for capsular polysaccharide (KL) and lipooligosaccharide outer core (OCL) synthesis in genome assemblies using curated reference databases compatible with Kaptive. Microb Genom 2020;6:e000339.

[2] Willis LM, Whitfield C. Structure, biosynthesis, and function of bacterial capsular polysaccharides synthesized by ABC transporter-dependent pathways. Carbohydr Res 2013;378:35-44.

[3] Islam S, Lam J. Synthesis of bacterial polysaccharides via the Wzx/Wzy-dependent pathway. Can J Microbiol 2014;60:697-716.

[4] Kenyon JJ, Hall RM. Variation in the complex carbohydrate biosynthesis loci of *Acinetobacter baumannii* genomes. PLoS One 2013;8:e62160.

[5] Whitfield C. Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli.* Annu Rev Biochem 2006;75:39-68.

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

[7] Senchenkova SN, Shashkov AS, Popova AV, Shneider MM, Arbatsky NP, Miroshnikov KA, et al. Structure elucidation of the capsular polysaccharide of *Acinetobacter baumannii* AB5075 having the KL25 capsule biosynthesis locus. Carbohydr Res 2015;408:8-11.

[8] Kenyon JJ, Speciale I, Hall RM, De Castro C. Structure of repeating unit of the capsular polysaccharide from *Acinetobacter baumannii* D78 and assignment of the K4 gene cluster. Carbohydr Res 2016;434:12-7.

[9] Kenyon JJ, Marzaioli AM, Hall RM, De Castro C. Structure of the K2 capsule associated with the KL2 gene cluster of *Acinetobacter baumannii*. Glycobiology 2014;24:554-63.

[10] Kenyon JJ, Marzaioli AM, Hall RM, De Castro C. Structure of the K6 capsular polysaccharide from *Acinetobacter baumannii* isolate RBH4. Carbohydr Res 2015;409:30-5.

[11] Shashkov AS, Kenyon JJ, Senchenkova SN, Shneider MM, Popova AV, Arbatsky NP, et al. *Acinetobacter baumannii* K27 and K44 capsular polysaccharides have the same K unit but different structures due to the presence of distinct *wzy* genes in otherwise closely related K gene clusters. Glycobiology 2016;26:501-8.

[12] Arbatsky NP, Shneider MM, Kenyon JJ, Shashkov AS, Popova AV, Miroshnikov KA, et al. Structure of the neutral capsular polysaccharide of *Acinetobacter baumannii* NIPH146 that carries the KL37 capsule gene cluster. Carbohydr Res 2015;413:12-5.

[13] Shashkov AS, Senchenkova SN, Popova AV, Mei Z, Shneider MM, Liu B, et al. Revised structure of the capsular polysaccharide of *Acinetobacter baumannii* LUH5533 (serogroup O1) containing di-*N*-acetyllegionaminic acid. Russ Chem Bull 2015;64:1196–9.

[14] Senchenkova SN, Popova AV, Shashkov AS, Shneider MM, Mei Z, Arbatsky NP, et al. Structure of a new pseudaminic acid-containing capsular polysaccharide of *Acinetobacter baumannii* LUH5550 having the KL42 capsule biosynthesis locus. Carbohydr Res 2015;407:154-7.

[15] Lees-Miller RG, Iwashkiw JA, Scott NE, Seper A, Vinogradov E, Schild S, et al. A common pathway for O-linked protein-glycosylation and synthesis of capsule in *Acinetobacter baumannii.* Mol Microbiol 2013;89:816-30.

[16] Harding C, Haurat F, Vinogradov E, Feldman M. Distinct amino acid residues confer one of three UDP-sugar substrate specificities in *Acinetobacter baumannii* PglC phosphoglycosyltransferases. Glycobiology 2018;28:522–33.

[17] Shashkov AS, Kenyon JJ, Arbatsky NP, Shneider MM, Popova AV, Miroshnikov KA, et al. Structures of three different neutral polysaccharide of *Acinetobacter baumannii*, NIPH190, NIPH201, and NIPH615, assigned to K30, K45, and K48 capsule types, respectively, based on capsule biosynthesis gene clusters. Carbohydr Res 2015;417:81-8.

[18] Kenyon JJ, Marzaioli AM, Hall RM, De Castro C. Structure of the K12 capsule containing 5,7-di-*N*-acetylacinetaminic acid from *Acinetobacter baumannii* isolate D36. Glycobiology 2015;25:881-7.

[19] Shashkov AS, Shneider MM, Senchenkova SN, Popova AV, Nikitina A, Babenko V, et al. Structure of the capsular polysaccharide of *Acinetobacter baumannii* 1053 having the KL91 capsule biosynthesis gene locus. Carbohydr Res 2015;404:79-82.

[20] Vijaykumar S, Balaji V, Biswas I. Complete genome sequence of *Acinetobacter baumannii* strain B8300, which displays high twitching motility. Genome Announc 2015;3:e00956-15.

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

[22] Leontein K, Lönngren J. Determination of the absolute configuration of sugars by gas–liquid chromatography of their acetylated 2-octyl glycosides. Methods in Carbohydrate Chemistry 1993;9:87-9.

[23] Kenyon JJ, Shneider MM, Senchenkova SN, Shashkov AS, Siniagina M, Malanin S, et al. 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 2016;162:1479-89.

[24] Reeves PR, Cunneen MM, Liu B, Wang L. Genetics and evolution of the *Salmonella* galactose-initiated group of O antigens. PLoS One. 2013;8:e69306.

[25] Rahn A, Kostantinos B, Naismith J, Whitfield C. A novel outer membrane protein, Wzi, is involved in surface assembly of the *Escherichia coli* K30 Group 1 capsule. J Bacteriol 2003;185(19):5882.

[26] Bushell SR, Mainprize IL, Wear MA, Lou H, Whitfield C, Naismith JH. Wzi Is an outer membrane lectin that underpins Group 1 capsule assembly in *Escherichia coli.* Structure. 2013;21(5):844-53.

[27] Lipkind GM, Shashkov AS, Knirel YA, Vinogradov EV, Kochetkov NK. A computer-assisted structural analysis of regular polysaccharides on the basis of ^{13}C-n.m.r. data. Carbohydr Res 1988;175:59–75.

[28] Kenyon JJ, Arbatsky NP, Sweeney EL, Zhang Y, Senchenkova SN, Popova AV, Shneider MM et al. Involvement of a multifunctional rhamnosyltransferase in the synthesis of three related *Acinetobacter baumannii* capsular polysaccharides, K55, K74 and K85. Int J Biol Macromol 2020; https://doi.org/10.1016/j.ijbiomac.2020.11.005.

[29] Knirel YA, Van Calsteren MR. Bacterial Exopolysaccharides in Chemistry, Molecular Sciences and Chemical Engineering. Elsevier 2020.

Legends to Figures

Fig. 1. The KL92 gene cluster from *A. baumannii* B8300. Gene shading represents the functional class of encoded proteins (scheme shown below). Figure is drawn to scale using GenBank accession number MN958100.1, and the scale bar is shown below.

Fig. 2. ^{13}C NMR spectra of the OS (top), MPS (middle), and CPS (bottom) from *A. baumannii* B8300. Numbers refer to carbons in sugar residues denoted by letters as indicated in Table 1 and Fig. 3.

Fig. 3. Structures of the OS, MPS, and CPS from *A. baumannii* B8300. 1dThr-ol indicates 1-deoxythreitol derived from 4-substituted Rha*p* (unit **D**). Glycosyltransferases (Gtr) are shown in bold at the linkage they are predicted to form.

Fig. 4. *Acinetobacter* KL gene clusters carrying the *gtr166-itrA4-wzi* module. **A.** Relationship between *A. baumannii* KL92, KL99, KL142 and KL143 gene clusters, drawn to scale from GenBank/WGS accession numbers listed in Table 2. Scheme shown below denotes grey scale between gene clusters representing nucleotide sequence identity. **B.** Gene cluster representative from *A. junii* carrying *orf$_{KL92}$-gtr166-itrA4-wzi*, drawn to scale from WGS accession number ACPM01000120.1 (base range 1-27000).

Tables

Table 1. ^{1}H and ^{13}C NMR chemical shifts of the OS, MPS, and CPS from *A. baumannii* B8300

Residue	C-1	C-2	C-3	C-4	C-5	C-6
	H-1	*H-2*	*H-3*	*H-4 (4a,4b)*	*H-5*	*H-6 (6a,6b)*
OS						
→3)-1dThr-ol[a]	18.5	68.6	85.4	62.2		
D'	*1.21*	*4.03*	*3.84*	*3.72, 3.77*		
→3)-β-D-Gal*p*-(1→	104.0	71.9	81.6	69.7	76.4	61.9
A	*4.59*	*3.70*	*3.72*	*4.01*	*3.71*	*3.75, 3.80*
→3)-α-L-Rha*p*-(1→	103.5	71.2	79.6	72.6	70.6	18.0
B	*5.03*	*4.17*	*3.92*	*3.56*	*3.87*	*1.29*
α-L-Rha*p*-(1→	103.7	71.5	71.4	73.3	70.4	17.9
C	*5.06*	*4.08*	*3.85*	*3.47*	*3.85*	*1.30*
MPS						
→3)-β-D-Gal*p*-(1→	104.9	72.5	81.7	69.7	76.4	62.1
A	*4.69*	*3.69*	*3.71*	*4.01*	*3.68*	*3.74, 3.79*
→3)-α-L-Rha*p*-(1→	103.3	71.3	79.5	72.8	70.6	18.0
B	*5.03*	*4.15*	*3.90*	*3.57*	*3.87*	*1.30*
→3)-α-L-Rha*p*-(1→	103.3	71.4	79.7	72.8	70.7	18.0
C	*5.03*	*4.14*	*3.90*	*3.56*	*3.84*	*1.28*
→4)-α-L-Rha*p*-(1→	103.3	71.6	71.6	82.6	69.0	18.2
D	*5.05*	*4.08*	*4.06*	*3.72*	*3.91*	*1.36*
CPS						

→3)-β-D-Gal*p*-(1→	104.2	71.7	81.8	69.7	76.1	62.1
A	*4.63*	*3.65*	*3.71*	*4.04*	*3.67*	*3.76, 3.81*
→3)-α-L-Rha*p*-(1→	103.3	71.2	79.7	72.6	70.5[b]	17.9[c]
B	*5.08*	*4.17*	*3.92*	*3.59*	*3.88*	*1.29*
→3)-α-L-Rha*p*-(1→	103.2	71.3[c]	79.4	72.6	70.6[b]	17.9[c]
C	*5.06*	*4.14*	*3.92*	*3.59*	*3.88*	*1.30*
→3,4)-α-L-Rha*p*-(1→	103.1	71.5[c]	80.5	78.4	69.3	18.3
D	*5.05*	*4.14*	*4.10*	*3.86*	*3.96*	*1.36*
α-L-Rha*p*-(1→	103.6	71.2	71.6	73.4	70.4	17.8[c]
E	*5.08*	*4.17*	*3.85*	*3.48*	*3.86*	*1.29*

[a]1dThr-ol indicates 1-deoxythreitol derived from 4-substituted Rha*p* (unit **D**).

[b,c] Assignment could be interchanged.

Table 2. Details of *A. baumannii* strains carrying *gtr166-itrA4-wzi* at the K locus

Strain	**KL**	**GenBank/WGS accession number**	**Base range**	**Country**	**Date**
A. baumannii 8300	KL92	MN958100.1	1.. 25980	India	2014
A. baumannii XH694	KL99	LYIX01000435.1	31582..59274	China: Hangzhou	2014
A. baumannii ABNIH28	KL142	CP026125.1	2955672..2982270	USA	2016
A. baumannii AB-HN-S73	KL143	PRHL01000062.1	32075..60630	China: Hangzhou	2014

Figures

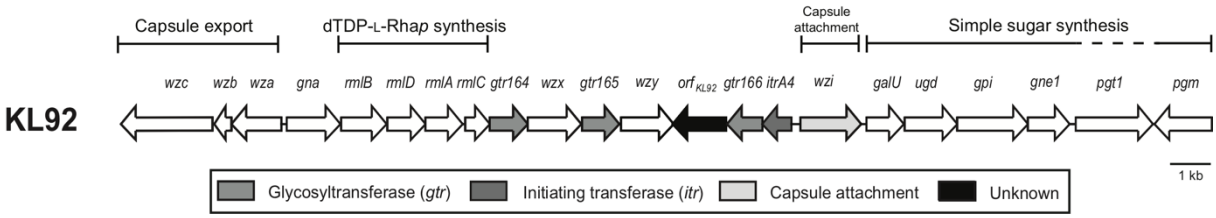

Fig. 1. The KL92 gene cluster from *A. baumannii* B8300. Gene shading represents the functional class of encoded proteins (scheme shown below). Figure is drawn to scale using GenBank accession number MN958100.1, and the scale bar is shown below.

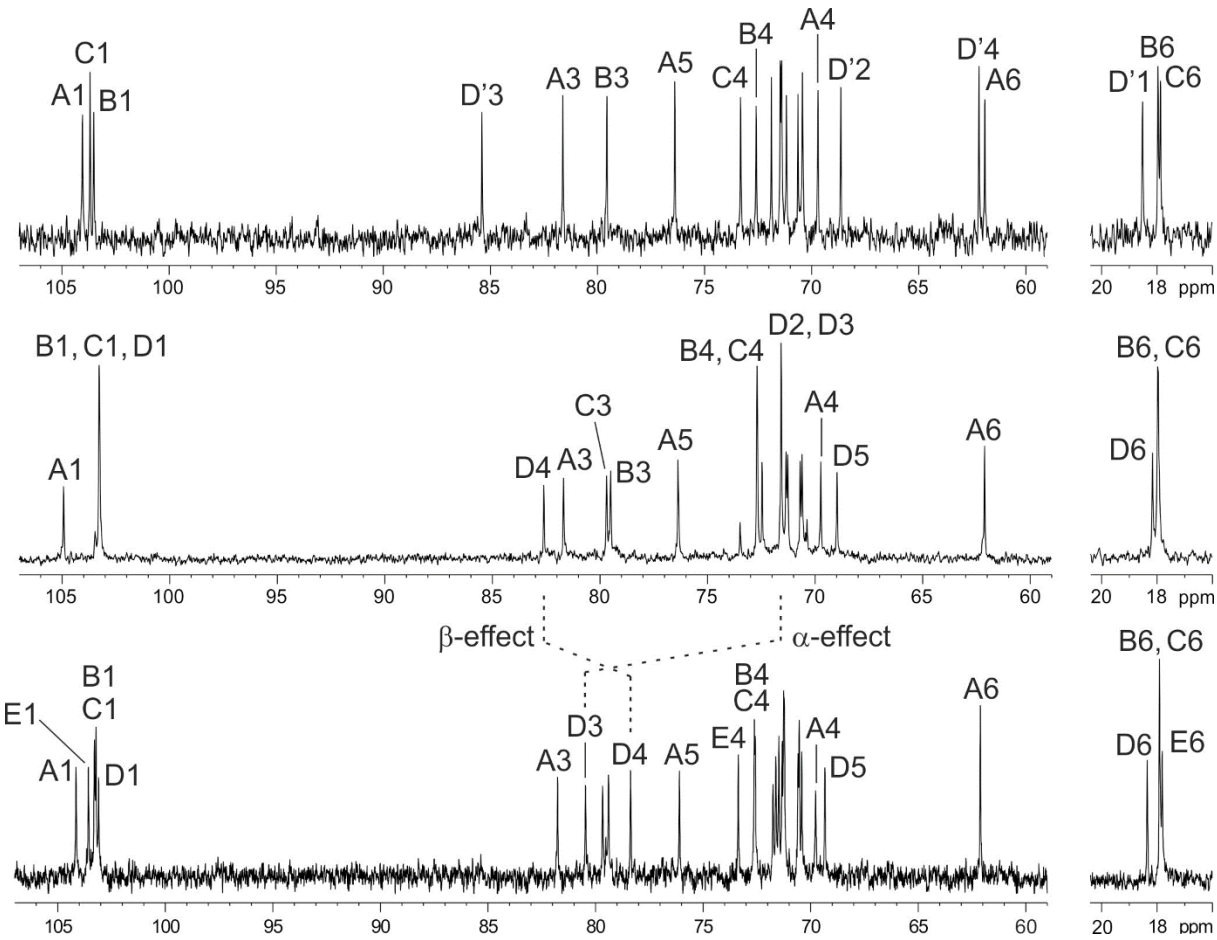

Fig. 2. ^{13}C NMR spectra of the OS (top), MPS (middle), and CPS (bottom) from *A. baumannii* B8300. Numbers refer to carbons in sugar residues denoted by letters as indicated in Table 1 and Fig. 3.

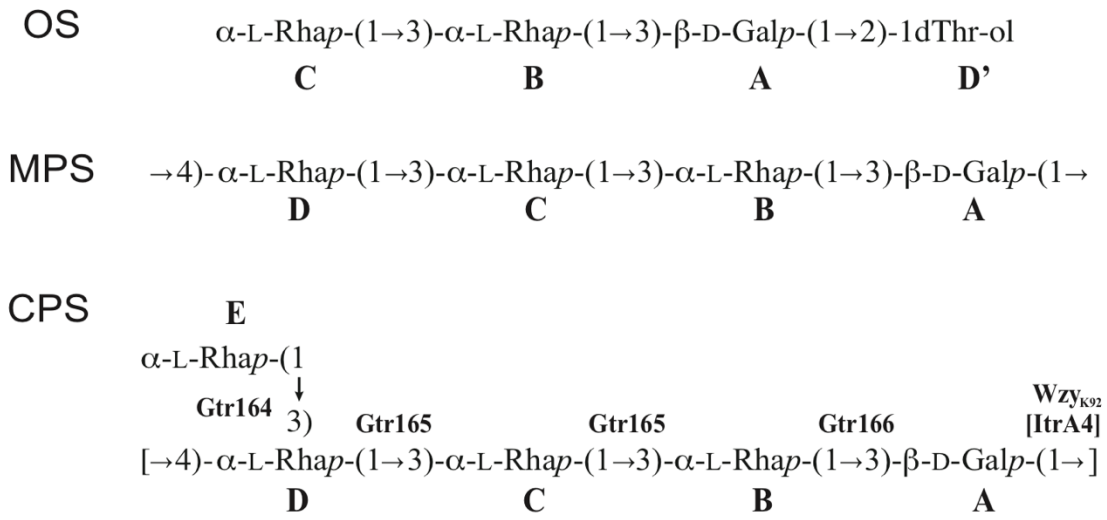

Fig. 3. Structures of the OS, MPS, and CPS from *A. baumannii* B8300. 1dThr-ol indicates 1-deoxythreitol derived from 4-substituted Rha*p* (unit **D**). Glycosyltransferases (Gtr) are shown in bold at the linkage they are predicted to form.

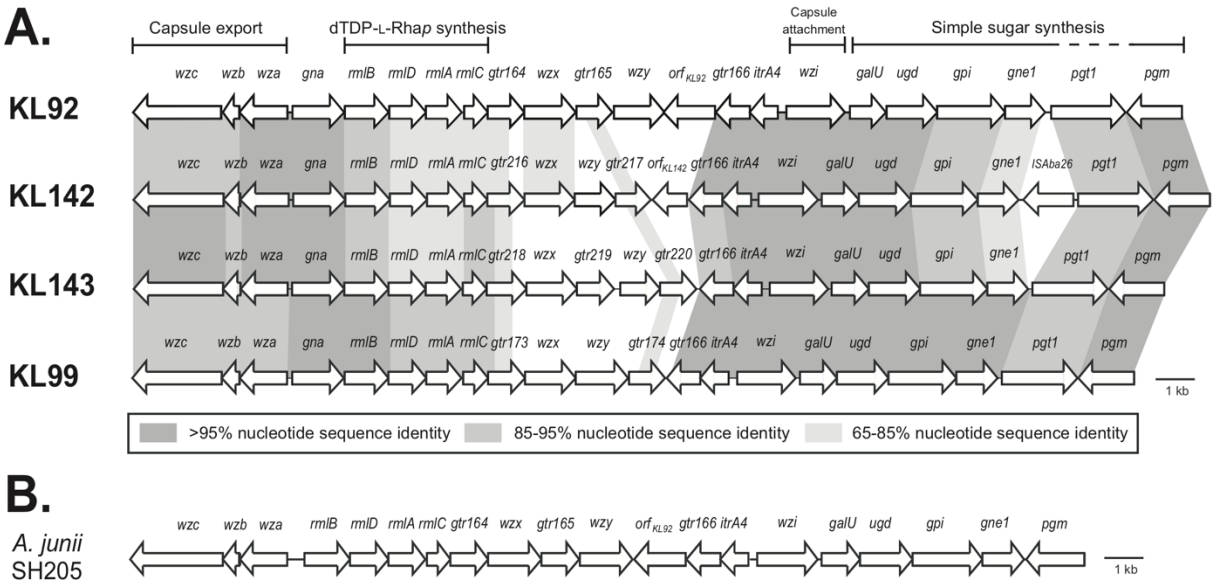

Fig. 4. *Acinetobacter* KL gene clusters carrying the *gtr166-itrA4-wzi* module. **A.** Relationship between *A. baumannii* KL92, KL99, KL142 and KL143 gene clusters, drawn to scale from GenBank/WGS accession numbers listed in Table 2. Scheme shown below denotes grey scale between gene clusters representing nucleotide sequence identity. **B.** Gene cluster representative from *A. junii* carrying *orf*$_{KL92}$*-gtr166-itrA4-wzi*, drawn to scale from WGS accession number ACPM01000120.1 (base range 1-27000).

Supplementary Table 1. Correlations for H-1 and C-1 in the two-dimensional $^{1}H,^{1}H$ ROESY (for MPS) or one-dimensional NOE (for OS) and $^{1}H,^{13}C$ HMBC spectra of the MPS and CPS from *A. baumannii* B8300.

Anomeric atom in sugar residue (δ)	Correlations to atom in sugar residue (δ)	
	Two-dimensional ROESY or one-dimensional NOE	HMBC
OS		
A H-1 (4.59)	**D'** H-3 (3.84), **D'** H-4 (3.72, 3.77, both w), **A** H-3 (3.72), **A** H-5 (3.71),	
A C-1 (104.0)		
B H-1 (5.03)	**A** H-3 (3.72), **B** H-2 (4.17), **B** H-3 (3.92, w)	
B C-1 (103.5)		
C H-1 (5.06)	**B** H-3 (3.92), **C** H-2 (4.08)	
C C-1 (103.7)		
MPS		
A H-1 (4.69)	**D** H-4 (3.72), **A** H-3 (3.71), **A** H-5 (3.68)	**D** C-4 (82.6)
A C-1 (104.9)		**D** H-4 (3.72), **A**H-2 (3.69)
B H-1 (5.03)	**A** H-3 (3.71), **B** H-2 (4.15)	**A** C-3 (81.7), **B** C-3 (79.5), **B** C-5 (70.6)
B C-1 (103.3)		**A** H-3 (3.71)
C H-1 (5.03)	**B** H-3 (3.90), **C** H-2 (4.14)	**B** C-3 (79.5), **C** C-3 (79.7), **C** C-5 (70.7)
C C-1 (103.3)		**B** H-3 (3.90)
D H-1 (5.05)	**C** H-3 (3.90), **D** H-2 (4.08)	**C** C-3 (79.7), **D** C-3 (71.6), **D** C-5 (69.0)

D C-1 (103.3)	**C** H-3 (3.90)
CPS	
A H-1 (4.63)	**D** C-4 (78.4)
A C-1 (104.2)	**D** H-4 (3.86), **A** H-2 (3.65), **A** H-5 (3.67)
B H-1 (5.08)	**A** C-3 (81.8), **B** C-3 (79.7), **B** C-5 (70.5)
B C-1 (103.3)	**A** H-3 (3.71)
C H-1 (5.06)	**B** C-3 (79.7), **C** C-3 (79.4), **C** C-5 (70.6)
C C-1 (103.2)	**B** H-3 (3.92)
D H-1 (5.05)	**C** C-3 (79.4), **D** C-3 (80.5), **D** C-5 (69.3)
D C-1 (103.1)	**C** H-3 (3.92)
E H-1 (5.08)	**D** C-3 (80.5), **E** C-3 (71.6), **E** C-5 (70.4)
E C-1 (103.6)	**D** H-3 (4.10)