



**WROCŁAWSKA SZKOŁA DOKTORSKA
INSTYTUTÓW POLSKIEJ AKADEMII NAUK**

**Badania strukturalne antygenów O i identyfikacja *loci* O
nietypowalnych izolatów klinicznych *Klebsiella pneumoniae***

Daria Artyszuk

**Promotor:
prof. dr hab. inż. Jolanta Łukasiewicz**

Praca doktorska
wykonana w Laboratorium Immunochemii Drobnoustrojów i Szczepionek
Instytutu Immunologii i Terapii Doświadczalnej im. Ludwika Hirszfelda
Polskiej Akademii Nauk



Wrocław 2023



WROCŁAW DOCTORAL SCHOOL
OF INSTITUTES OF POLISH ACADEMY OF SCIENCES

**O antigen structural studies and O *loci* identification of
nontypeable clinical isolates of *Klebsiella pneumoniae***

Daria Artyszuk

**Supervisor:
prof. dr hab. inż. Jolanta Łukasiewicz**

This doctoral dissertation is based on experimental work
performed in the Laboratory of Microbial Immunochemistry and Vaccines,
Hirsfeld Institute of Immunology and Experimental Therapy,
Polish Academy of Sciences



Wrocław 2023

Badania prowadzono w ramach projektu badawczego
finansowanego przez Narodowe Centrum Nauki,
Opus 16 2018/31/B/NZ7/04002,
“Nietypowalne antygeny O *Klebsiella pneumoniae* – struktura
i seroepidemiologia”

This study is part of a research project funded
by National Science Center of Poland,
Opus 16 2018/31/B/NZ7/04002,
“Nontypeable O antigens of *Klebsiella pneumoniae* – structure
and epidemiology”

Podziękowania

Pani Profesor Dr hab. Jolancie Łukasiewicz składam ogromne podziękowania za opiekę merytoryczną, zaufanie oraz wsparcie i sugestie – nie tylko w kwestiach niniejszej pracy. Przede wszystkim dziękuję za daną mi szansę i otwartość na moje pomysły – mniej lub bardziej immunochemiczne.

Serdecznie dziękuję Profesorowi Dr hab. Markowi Gniadkowskiemu oraz Dr hab. Radosławowi Izdebskiemu z Zakładu Mikrobiologii Molekularnej Narodowego Instytutu Leków w Warszawie za pomoc w przygotowaniu genetycznej części tej pracy oraz za niezwykle dokładność edytorską wspólnych publikacji.

Składam podziękowania Profesorowi Dr hab. Łukaszowi Łaczańskiemu oraz Dr Dariuszowi Martynowskiemu z Laboratorium Genomiki i Bioinformatyki za pomoc w sekwencjonowaniu oraz analizach bioinformatycznych.

Dziękuję również Oli Herud za cierpliwe szkolenie z analiz genomowych oraz pomoc w składaniu sekwencji.

Dr hab. Wojciechowi Jachymkowi serdecznie dziękuję za możliwość niekończących się konsultacji, wiarę w moje umiejętności i wsparcie podczas wielu badań i analiz.

Niezmiennie dziękuję Tomkowi Wołkowiczowi z Zakładu Bakteriologii i Zwalczania Skażeń Biologicznych Narodowego Instytutu Zdrowia Publicznego za ciągłą zachętę do stawiania sobie coraz większych wyzwań – nawet na odległość.

Koleżankom i Kolegom z Laboratorium Immunochemii Drobnoustrojów i Szczepionek dziękuję za okazaną pomoc i życzliwość, już od pierwszego dnia pracy.

Szczególne podziękowania składam Moim Najbliższym,

Rodzicom za niezachwianą wiarę w moje możliwości i spełnianie wszystkich marzeń, ogromne wsparcie oraz ciężką pracę, dzięki którym zawsze mogłam, mogę i będę mogła osiągać założone przez siebie cele.

Szymonowi za zaangażowanie, wszystkie dyskusje skutkujące nowymi pomysłami, sporą dawkę naukowego humoru oraz zdolność do poświęceń dla mojego ciągłego rozwoju – już od naszego pierwszego spotkania.

Results of this thesis were included in the following publications:

Artyszuk D, Izdebski R, Maciejewska A, Kaszowska M, Herud A, Szijártó V, Gniadkowski M, Lukaszewicz J. The Impact of Insertion Sequences on O-Serotype Phenotype and Its O-Locus-Based Prediction in *Klebsiella pneumoniae* O2 and O1. Int J Mol Sci. 2020, 21(18):6572. doi: 10.3390/ijms21186572.

Artyszuk D, Jachymek W, Izdebski R, Gniadkowski M, Lukaszewicz J. The OL101 O antigen locus specifies a novel *Klebsiella pneumoniae* serotype O13 structure. Carbohydr. Polym. 2024, 326, 121581. doi: 10.1016/j.carbpol.2023.121581.

Table of contents

1. Streszczenie	13
2. Abstract	15
3. Abbreviations	17
4. Introduction	21
4.1 Characteristics of <i>Klebsiella pneumoniae</i> species	21
4.2 <i>Klebsiella pneumoniae</i> major virulence factors	24
4.2.1 Fimbriae	25
4.2.2 Outer membrane proteins (OMP).....	26
4.2.3 Efflux pumps	27
4.2.4 Siderophores	28
4.2.5 Capsular polysaccharide (CPS).....	29
4.2.6 Lipopolysaccharide (LPS)	31
4.2.6.1 Lipid A.....	32
4.2.6.2 Core oligosaccharide.....	34
4.2.6.3 O-specific polysaccharide (OPS, O antigen).....	35
4.2.6.3.1 Structures of <i>K. pneumoniae</i> O antigens.....	36
4.2.6.3.2 Gene clusters required for O antigen biosynthesis.....	37
4.3 Therapeutic strategies against <i>K. pneumoniae</i> infections	39
5. Aims of the study	42
6. Materials and Methods	44
6.1 Media and buffers	44
6.2 Bacteria and growth conditions	44
6.3 Lipopolysaccharide preparation	46
6.4 O-specific polysaccharide preparation	46
6.5 Compositional analysis	47
6.5.1 Sugar analysis	47
6.5.2 Methylation analysis.....	48
6.5.3 Gas chromatography coupled to mass spectrometry (GC-MS)	48
6.6 NMR spectroscopy of LPS and OPS	48
6.7 Absolute configuration of sugar components analysis by NMR spectroscopy	49
6.8 Matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS)	50
6.9 Electrophoresis under denaturing conditions (SDS-PAGE)	50
6.10 Silver staining	51
6.11 Genomics analyses	51

6.11.1 DNA isolation	51
6.11.2 Whole-genome sequencing.....	51
6.11.3 Genomic sequence analysis.....	51
6.11.4 <i>In silico</i> studies for O <i>loci</i> prevalence among publicly available <i>K. pneumoniae</i> genomes.....	52
6.11.5 Data availability	53
6.12 Software used in this study	53
7. Results.....	54
7.1 Isolation and preliminary analysis of lipopolysaccharides.....	54
7.2 O antigen and O <i>loci</i> analysis of <i>K. pneumoniae</i> BIDMC 7B and ABC152.....	57
7.2.1 Structural analysis of BIDMC 7B and ABC152 O antigens	57
7.2.2 Bioinformatics analysis of BIDMC 7B and ABC152 O <i>loci</i>	59
7.2.3 Insertion sequences occurrence in <i>K. pneumoniae</i> O2v2 and O1v2 <i>loci</i> — <i>in silico</i> study of publicly available genomes.....	62
7.3 O antigen and O <i>loci</i> analysis of <i>K. pneumoniae</i> ABC122, BC738, BC13-986 and 3936/19.....	66
7.3.1 Structural analysis of <i>K. pneumoniae</i> ABC122, BC738, BC13-986 and 3936/19 O antigens.....	66
7.3.2 Clonality of the isolates and genetic analysis of the OL101 <i>loci</i>	79
7.3.3 OL101 prevalence among the <i>K. pneumoniae</i> isolates with publicly available genomes.....	84
7.4 O antigen and O <i>loci</i> analysis of <i>K. pneumoniae</i> Kp164, Kp165 and Kp166.....	85
7.4.1 Structural analysis	85
7.4.2 Genetic analysis of the <i>rfb</i> regions	87
7.4.3 MALDI-TOF Biotyper analysis of <i>K. pneumoniae</i> PCM15 reference strain.....	90
7.5. O antigen and O <i>loci</i> analysis of Kp174 and Kp177.....	90
7.5.1 Structural analysis	90
7.5.2 Genetic analysis of the <i>rfb</i> regions	91
8. Discussion.....	93
9. Conclusions.....	104
10. References	105
11. Appendix	116
12. List of figures.....	117
13. List of tables	119
14. Scientific achievements	120

1. Streszczenie

Klebsiella pneumoniae jest patogenem występującym w szpitalach i jednym z patogenów priorytetowych, wskazanych przez Światową Organizację Zdrowia (WHO) jako krytyczny ze względu na bardzo ograniczone możliwości leczenia powodowanych przez tę bakterię zakażeń. Lipopolisacharyd (LPS, antygen O) i polisacharyd otoczkowy (CPS, antygen K) to główne czynniki zjadliwości i antygeny powierzchniowe *K. pneumoniae* określające serotypy O i K, kodowane przez klastry genów nazywane *locus* O lub K. W przeciwieństwie do dużej zmienności strukturalnej w obrębie antygenów K, gatunek ten charakteryzuje się ograniczoną różnorodnością antygenów O (do tej pory zidentyfikowano 11 serotypów O). Cecha ta sprawia, że LPS jest atrakcyjnym celem dla terapii opartych na przeciwciałach (szczepionki oraz immunizacja bierna) jako alternatywy dla antybiotyków. Aby taka immunoterapia była skuteczna, ważne jest poszerzanie wiedzy na temat struktur antygenów O oraz dystrybucji serotypów wśród izolatów klinicznych *K. pneumoniae*.

Obecnie analiza strukturalna antygenów O jest wspierana przez bioinformatykę. Genotypowanie oparte na analizie porównawczej *loci* O (regionów *rfb*) i K za pomocą reakcji łańcuchowej polimerazy (PCR) lub analizie wyników sekwencjonowania całego genomu (WGS) – w tym na analizach wykorzystujących ogólnodostępne narzędzie Kaptive (<https://kaptive-web.erc.monash.edu/>), zostało zaproponowane jako narzędzie diagnostyczne, a wyniki genotypowania wskazują na większą niż opisana dotychczas różnorodność antygenu O wśród izolatów *K. pneumoniae*.

Praca doktorska obejmowała grupę jedenastu LPS wyizolowanych z izolatów klinicznych *K. pneumoniae* (BIDMC 7B, ABC152, ABC122, BC738, BC13-986, 3936/19, Kp164, Kp165, Kp166, Kp174, Kp177). Szczepy te oceniono wstępnie jako nietypowalne, biorąc pod uwagę brak ich reaktywności z przeciwciałami skierowanymi przeciwko znanym serotypom O lub różnice w sekwencjach regionów *rfb*. Niniejsza praca doktorska realizowana była w ramach projektu OPUS16 finansowanego przez Narodowe Centrum Nauki (2019-2025) pt. „Nietypowalne antygeny O *Klebsiella pneumoniae* – struktura i seroepidemiologia”.

Wyizolowane lipopolisacharydy analizowano za pomocą spektroskopii ^1H , ^{13}C HR-MAS NMR (*high-resolution-magic angle spinning nuclear magnetic resonance*). W przypadku wybranych LPS, badania uzupełniono o izolację polisacharydów O-swoistych i analizę składu cukrowego metodami chemicznymi oraz

za pomocą spektroskopii NMR. Dodatkowo, analizę strukturalną wzbogacono identyfikacją *locus* O za pomocą narzędzia Kaptive, analizując odpowiednio regiony *rfb* kodujące poszczególne serotypy O. Ze wszystkich 11 nietypowalnych izolatów *K. pneumoniae* wyizolowano DNA, poddano sekwencjonowaniu, a otrzymane sekwencje analizowano za pomocą oprogramowania Kaptive. Ponadto dla każdego izolatu przeprowadzono analizę MLST (ang. *multi-locus sequence typing*).

W wyniku przeprowadzonych analiz, badane lipopolisacharydy przydzielono do czterech grup: i) grupy 1 (szczepy BIDMC 7B, ABC122) reprezentującej fenotyp O2 wariant 1 (O2v1) z sekwencją insercyjną w regionie *rfb* (genotyp O2v2); ii) grupy 2 (szczepy ABC152, BC738, BC13-986, 3936/19) charakteryzującej się identycznym i nowym serotypem O13; iii) grupy 3 (szczepy Kp164, Kp165, Kp166) reprezentującej znany serotyp O4 oraz iv) grupy 4 (szczepy Kp174, Kp177) charakteryzującej się obecnością szorstkiego LPS, pozbawionego polisacharydu O-swoistego.

W wyniku przeprowadzonych badań opublikowane zostały dwa oryginalne osiągnięcia. W przypadku grupy 1 zidentyfikowano przyczynę rozbieżności pomiędzy serotypem przewidzianym przez Kaptive w oparciu o sekwencję *locus* O (O2v2; struktura podjednostki antygeny O: $\rightarrow 3$)- β -D-Galp-(1 \rightarrow 3)-[α -D-Galp-(1 \rightarrow 4)]- α -D-Galp-(1 \rightarrow) a zidentyfikowaną strukturą chemiczną (O2v1; struktura podjednostki antygeny O; $\rightarrow 3$)- β -D-Galp-(1 \rightarrow 3)- α -D-Galp-(1 \rightarrow). W regionach *rfb* szczepów BIDMC 7B i ABC122 zidentyfikowano sekwencje insercyjne (IS) w genie *gmlB*, odpowiedzialnym za biosyntezę terminalnej reszty α -D-Galp w przypadku wariantu O2v2. Zjawisko poddano szerszej analizie w oparciu o analizę *in silico* 8130 genomów *K. pneumoniae* dostępnych w publicznych bazach danych (dn. 3 grudnia 2019). Dla ~10% serotypów O1v2 i ~28% O2v2 wykazano występowanie różnych typów sekwencji IS (np. ISR1, IS903B, ISKpn14 lub ISKpn26) w genach regionu *rfb* o potencjalnym znaczeniu dla fenotypu antygeny O (Artyszuk *et al.* 2020).

Drugim osiągnięciem było powiązanie jednego z nowych *loci* O *K. pneumoniae* (OL101) zidentyfikowanego u szczepów ABC152, BC738, BC13-986, 3936/19, dla którego dotychczas nie opisano kodowanej przez ten region struktury antygeny O, z nowo opisaną strukturą antygeny O, β -Kdop- $\rightarrow 3$)- α -L-Rhap-(1 \rightarrow 4)- α -D-Glcp]_n. Izolaty te reprezentują nowy serotyp O13, stanowiący około 6,55% wszystkich 71377 analizowanych genomów *K. pneumoniae* dostępnych w bazach danych na dzień 27 lipca 2023 (Artyszuk *et al.* 2024).

2. Abstract

Klebsiella pneumoniae is a nosocomial pathogen and one of the priority species, pointed out by the World Health Organisation (WHO) as critical regarding highly limited options of treatment of infections caused by the species. Lipopolysaccharide (LPS) and capsular polysaccharide (CPS, K antigen) are its major virulence factors and surface antigens, determining O and K serotypes and encoded by O or K *loci*, respectively. Contrary to CPS, the species has been perceived as of limited variety of O antigens (11 O serotypes identified to date) and that trait makes LPS an attractive target for antibody-based therapies (vaccines and passive immunization) as an alternative to antibiotics. To make such immunotherapy effective, knowledge about O antigen structures, drift, and distribution among clinical isolates is important. At present, the structural analysis of O antigens is efficiently supported by bioinformatics. O and K *loci*-based genotyping by polymerase chain reaction (PCR) or whole-genome sequencing WGS has been proposed as a diagnostic tool, including the Kaptive (<https://kaptive-web.erc.monash.edu/>), and obtained results have indicated indicating higher diversity of the O antigen *loci*.

The research included the group of eleven LPSs isolated from clinical isolates of *K. pneumoniae* (BIDMC 7B, ABC152, ABC122, BC738, BC13-986, 3936/19, Kp164, Kp165, Kp166, Kp174, Kp177). Strains were selected as nontypeable strains considering their reactivity with antibodies against known O serotypes or differences in O *loci* sequences. This PhD project was part of the OPUS16 grant founded by National Science Center (2019-2025) and entitled “Nontypeable O antigens of *Klebsiella pneumoniae* – structures and seroepidemiology”.

Lipopolysaccharides were isolated, purified and analysed as a native molecules by ¹H, ¹³C HR-MAS NMR (*high-resolution-magic angle spinning nuclear magnetic resonance*). For selected strains, O-specific polysaccharides were isolated and analysed by chemical methods and NMR spectroscopy. Additionally, structural analyses were supported by molecular biology and O *locus* prediction by Kaptive. DNA was isolated from all 11 nontypeable *K. pneumoniae* isolates, sequenced and obtained contigs analysed by Kaptive. Moreover, multi-*locus* sequence typing (MLST) analyses were performed for each isolate.

The spectra of selected LPSs were compared with the spectra of *K. pneumoniae* LPSs of known O serotypes, which allowed the LPSs to be assigned to four groups: i) group 1 (strains BIDMC 7B, ABC122) representing O2 variant 1 (O2v1) phenotype with insertion sequence in *rfb* regions predicted by Kaptive as O2v2 genotype; ii) group 2 (strains ABC152, BC738, BC13-986, 3936/19) characterised by identical and novel O13 serotype; iii) group 3 (strains Kp164, Kp165, Kp166) represented known O4 serotype, and iv) group 4 (strains Kp174, Kp177) expressing rough LPS devoid of O-specific polysaccharide.

Two original achievements have been made and published. For BIDMC 7B and ABC122 strains the reasons for discrepancies in O2 serotyping between Kaptive-based predictions (indicating the following O antigen repeating unit $\rightarrow 3$)- β -D-Galp-(1 \rightarrow 3)-[α -D-Galp-(1 \rightarrow 4)]- α -D-Galp-(1 \rightarrow (O2v2 serotype)) and the phenotype $\rightarrow 3$)- β -D-Galf-(1 \rightarrow 3)- α -D-Galp-(1 \rightarrow (O2v1)) have been identified. As a reason, insertion sequences (ISs) were identified within *rfb* regions in the gene *gmlB* responsible for biosynthesis of the terminal α -D-Galp residue. Discovered phenomenon were reported by *in silico* analysis of 8130 genomes available in public databases (December 3, 2019) for ~10% and ~28% of O1v2 and O2v2 genomes, respectively indicating a broader distribution of ISs (e.g. ISR1, IS903B, ISKpn14, or ISKpn26) in *rfb* regions that may influence the O antigen chemical structure (Artyszuk *et al.* 2020).

As a second accomplishment, one of the novel *K. pneumoniae* O loci, for which the antigen structure has not been elucidated so far, OL101 locus, have been characterised by identification of encoded O antigen structure. In this study, four clinical isolates predicted by Kaptive as OL101 (ABC152, BC738, BC13-986, 3936/19) were characterized and found to have the O antigen structure composed of β -Kdop- $\rightarrow 3$)- α -L-Rhap-(1 \rightarrow 4)- α -D-Glcp]_n, representing a novel serotype O13 and occurring in ~6,55% of isolates in the dataset of *K. pneumoniae* 71377 genomes (July 27, 2023) screened by *in silico* analysis (Artyszuk *et al.* 2024).

3. Abbreviations

Abbreviation	Abbreviation explanation
ABC	ATP-binding Cassette
Hep	Heptose
Fuc	Fucose
TRIS	2-Amino-2-hydroxymethyl-propane-1,3-diol
NIAID	National Institute of Allergy and Infectious Diseases
NIH	National Institutes of Health
USA	United States of America
Ac	Acetyl group
Ara	Arabinose
AraN	Aminoarabinose
BHI	Brain Heart Infusion
BSA	Bovine Serum Albumin
BTS	Bacterial Test Standard
CDS	Coding Sequence
CGE	Centre for Genomic Epidemiology
<i>cKp</i>	Classical <i>Klebsiella pneumoniae</i>
COSY	Correlation Spectroscopy
CPKP	Carbapenemase-Producing <i>K. pneumoniae</i>
CPS	Capsular Polysaccharide; K antigen
CRE	Carbapenem Resistant
CRKP	Carbapenem Resistant <i>Klebsiella pneumoniae</i>
CRM197	Cross-Reactive Material 197
DEPT	Distortionless Enhancement by Polarization Transfer
DMSO	Dimethylsulfoxide
DTT	Dithiothreitol
ESBL	Extended-Spectrum β -Lactamases
ESKAPE	<i>Enterococcus faecium</i> , <i>Staphylococcus aureus</i> , <i>Klebsiella pneumoniae</i> , <i>Acinetobacter baumannii</i> , <i>Pseudomonas aeruginosa</i> and <i>Enterobacter</i> spp.
Gal	Galactose
GalA	Galacturonic Acid
GalN	Glucosamine
GC-MS	Gas chromatography coupled to Mass Spectrometry
Glc	Glucose
GlcA	Glucuronic Acid
HCCA	α -Cyano-4-hydroxycinnamic Acid
HMBC	Heteronuclear Multiple-Bond Correlation Spectroscopy
HR-MAS	High Resolution-Magic Angle Spinning
HSQC	Heteronuclear Single Quantum Correlation
hv <i>Kp</i>	Hipervirulent <i>Klebsiella pneumoniae</i>

ICU	Intensive Care Unit
IL	Interleukine
IM	Inner Membrane
IS	Insertion Sequence
Kdo	3-deoxy-D-manno-2-octulosonic acid
KPC	<i>Klebsiella pneumoniae</i> carbapenemase
LPS	Lipopolysaccharide
MALDI- TOF MS	Matrix-Assisted Laser Desorption Ionization-Time Of Flight Mass Spectrometry
Man	Mannose
MDR	Multidrug-Resistant
Me	Methyl group
MeP	Methyl Phosphate
MFS	Major Facilitator Superfamily
MLST	Multi- <i>Locus</i> Sequence Typing
mQ	Milli-Q Water
NBD	Nucleotide-Binding Domain
NDM-1	New Delhi metallo- β -lactamase 1
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
OM	Outer Membrane
OMP	Outer Membrane Protein
OPS	O-specific Polysaccharide
P	Phosphate group
PBS	Phosphate Buffered Saline
PCM	Polish Collection of Microorganisms
PCR	Polymerase Chain Reaction
PDR	Pan-Drug Resistant
PEtn	Phosphoethanolamine
Rha	Rhamnose
Rib	Ribose
R-LPS	Rough-Lipopolisaccharide
RND	Resistance-Nodulation-Division
RPM	Revolutions Per Minute
RS-LPS	Rough-Smooth-Lipopolysaccharide
RU	Repeating Unit
SDS-PAGE	Sodium Dodecyl Sulphate–Polyacrylamide Gel Electrophoresis
S-LPS	Smooth-Lipopolysaccharide
SMB	<i>O</i> -(<i>S</i>)-2-methyl butyrate
SNP	Single Nucleotide Polymorphism
ST	Sequence Type
TLR4/MD-2	Toll-like receptor 4 (TLR4)/Myeloid Differentiation factor 2 (MD-2)
TMD	Transmembrane Domain

TNF- α	Tumour Necrosis Factor α
TOC	Total Organic Carbon
TOCSY	Total Correlation Spectroscopy
UAE	United Arab Emirates
VIM-1	Verona Integron-encoded Metallo- β -lactamase 1
WGS	Whole-Genome Sequencing
WHO	World Health Organization
XDR	Extensively Drug-resistant
Xyl	Xylose

4. Introduction

4.1 Characteristics of *Klebsiella pneumoniae* species

Klebsiella pneumoniae was first isolated from the lungs of pneumonia patient by German bacteriologist Edwin Klebs in 1875. In 1882, Carl Friedländer characterized this genus (Köhler and Mochmann 1987). *Klebsiella pneumoniae* is a Gram-negative, non-motile, facultatively anaerobic bacterium that is part of the human microbiota. However, it is also a common cause of nosocomial and community-acquired infections in neonates, the elderly and immunocompromised patients. Recent studies have reported the prevalence of *Klebsiella* colonization ranges from 18,8 to 87,7% in Asia and 5 to 35% in Western countries (Lin *et al.* 2012; Russo and Marr 2019). The species belongs to the ESKAPE group of pathogens (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Enterobacter* spp) and is on the WHO list of priority "critical" pathogens (Tacconelli *et al.* 2018), which have been indicated as the primary target for the development of new therapeutic strategies against infections. *Klebsiella pneumoniae* can be broadly divided into two subtypes: classical and hypervirulent strains. Most infections are caused by strains referred to as classical *K. pneumoniae* (*cKp*). These strains persist in the hospital environment and cause infections in debilitated patients affected by systemic diseases, surgical operations and intensive care unit treatments. Classic strains are usually associated with pneumonia, urinary tract infection, nosocomial infections and sepsis in immunocompromised patients. *cKp* strains are characterized by a tendency to acquire resistance to antimicrobial agents (Vandhana, Vishwas Saralaya, *et al.* 2022). Risk factors for *K. pneumoniae* infections are treatment with corticosteroids, chemotherapy, transplantation and neutropenia. Moreover, risk factors for *K. pneumoniae* bacteraemia include: chronic liver disease, solid-organ transplantation, dialysis, malignancy and diabetes (Paczosa and Meccas 2016).

Hypervirulent variants of *K. pneumoniae* (*hvKp*) have been reported with increasing frequency worldwide. They were first recognized in Taiwan at the end of 20th century and have caused liver abscesses and meningitis (Liu and Guo 2019). The *hvKp* strains can cause serious infections in both young, immunocompetent, sick and healthy people. The infection often involves multiple organs (Liu and Guo 2019). The high virulence of *hvKp* is mainly related to the increased production of the capsular polysaccharide (CPS, K antigen) leading to the hypermucoviscous phenotype that is

contributed by a plasmid. Therefore not all hvKp strains have hypermucoviscous phenotype, but certain classic strains can also be hypermucoviscous (Nassif 1989; Russo and Marr 2019; Vandhana, Saralaya, *et al.* 2022). The increasingly frequent occurrence of multi-organ clinical symptoms associated with *K. pneumoniae* infections indicates a high incidence of hvKp strains, becoming a global threat requiring immediate action. Due to the high mortality rate, early diagnosis play a key role in reducing the severity of disease symptoms and death due to hvKp (Vandhana, Vishwas Saralaya, *et al.* 2022).

In recent years, *K. pneumoniae* has become clinically significant owing to its increased antibiotic resistance and ability to produce serious health consequences. Exposure of strains to sublethal concentrations of antibiotics may contribute to the development of resistance among pathogens. Moreover, the increased and uncontrolled use of antibiotics is an additional important factor. The global emergence of multidrug-resistant (MDR) strains, especially *K. pneumoniae* producing extended-spectrum β -lactamase (ESBL) and/or carbapenemase (carbapenem-resistant *K. pneumoniae* strains, CRKP strains), has become a major public health challenge. Multidrug-resistant bacteria cause serious community and hospital-acquired infections that are difficult to eradicate. *Klebsiella pneumoniae* has been described as the main source of nosocomial pneumonia, ranking second among Gram-negative bacteria and accounting for approximately 10% of all worldwide nosocomial infections (Li and Webster 2018; Chang *et al.* 2021; Vandhana *et al.* 2022).

New Delhi metallo- β -lactamases (NDM)-producing Enterobacterales, including *K. pneumoniae*, are one of the leading threats in infectious diseases. The term "*Klebsiella* New Delhi" is associated with a specific antibiotic-resistant *K. pneumoniae* strain known as New Delhi metallo- β -lactamase-1 (NDM-1)-producing. This strain has gained attention due to its high level of resistance to various antibiotics, including β -lactams, making it challenging to treat infections caused by the bacterium. The NDM-1 is an enzyme that confers resistance to a broad range of β -lactam antibiotics, including carbapenems. The term "New Delhi" is derived from the city where *K. pneumoniae* NDM-1 was first identified in a Swedish patient who had travelled to India in 2008. This strain often carries additional resistance genes, further complicating the selection of effective antimicrobial therapies (Yong *et al.* 2009; Biedrzycka *et al.* 2021).

NDM-1-producing *K. pneumoniae* has been identified in various parts of the world, contributing to the global dissemination of antibiotic resistance. It is frequently

associated with healthcare-associated infections and has been found in both clinical and environmental settings.

A wide variety of plasmids harbour *bla*NDM genes are found in a large taxonomic and clonal diversity of organisms, indicating efficient horizontal transfer of *bla*NDMs. However, clonal spread of NDM producers has been significant in some areas too (ECDC, 2012).

The *K. pneumoniae* population is dominated by a number of pandemic clones, some of which broadly disseminate resistance genes, including encoding carbapenemases. High-risk clones, such as ST11 or ST15, are genetically diverse, whereas the more recently emerged ones, such as ST258, ST307 and ST147, have been less differentiated so far (ECDC, 2012; Biedrzycka *et al.*, 2021).

The one of the largest genomic studies on the *K. pneumoniae* ST147 clone occurring in Poland, has showed that since late 2012 a nation-wide outbreak of NDM-1-producing *K. pneumoniae* ST11 has been observed in Poland, converting into endemicity in consecutive regions. With time, new imports of NDM producers occurred, including in March 2015, when Polish nationals who suffered a terrorist attack in Tunisia and were treated locally, were then transported to Warsaw. Since 2011, the number of registered NDM+ isolates and related infections are constantly increasing, reaching 126 cases in the end of 2019 (ECDC, 2012; Biedrzycka *et al.*, 2021).

Treatment options against CRPK are few and usually limited to last-line antibiotics such as colistin, fosfomycin, and tigecycline (Paczosa and Mecsas 2016; Grundmann *et al.* 2017). The emergence of extensively drug-resistant (XDR) and pan-drug-resistant (PDR) *K. pneumoniae* isolates is of great concern worldwide and increases the risk of antibiotic treatment failure in humans (Hafiz *et al.* 2023).

Vaccination is one of the most effective medical strategies protecting human health, and it has been shown to be the most effective intervention to control infectious diseases. Carbohydrates, covering the surface of bacteria, play a key role in the development of infection, acting as virulence factors and hiding the microorganisms from the immune system of the host or activating inflammatory responses. For these reasons, cell surface polysaccharides have been successfully targeted for the development of bacterial vaccines. For instance glycoconjugate vaccines, Quimi-Hib for *Haemophilus influenzae* and Synflorix or Prevnar for *Streptococcus pneumoniae* (Del Bino *et al.* 2022). Identification of highly conserved antigens of *K. pneumoniae* strains is essential to develop vaccines with extensive protection (Lundberg *et al.* 2013; Chang *et al.* 2021).

4.2 *Klebsiella pneumoniae* major virulence factors

To cause infection, *K. pneumoniae* must overcome mechanical and chemical barriers and escape the host's humoral and cellular innate defences. Upon accessing the host, invading organisms are recognized by immune cells through pattern recognition receptors and trigger the production of various immune mediators (Silhavy, Kahne, and Walker 2010). *Klebsiella pneumoniae* can produce several virulence factors that are essential for colonization, invasion and infection progression, many of which include the presence of LPS, CPS production, adherence factors and siderophore activity (Figure 1a).

As *K. pneumoniae* is a Gram-negative bacterium its cell envelope is a complex multilayered structure with three principal layers in the envelope; the outer membrane (OM), the periplasm with peptidoglycan layer, and the cytoplasmic or inner membrane (IM) (Figure 1b) (Silhavy, Kahne, and Walker 2010; Sun *et al.* 2021). Some of virulence factors are surface antigens anchored within the outer membrane and periplasm of bacterial cell wall (e.i. LPS, CPS, OMP, efflux pumps) and some are secreted outside the cell (e.i. siderophores) (Figure 1a).

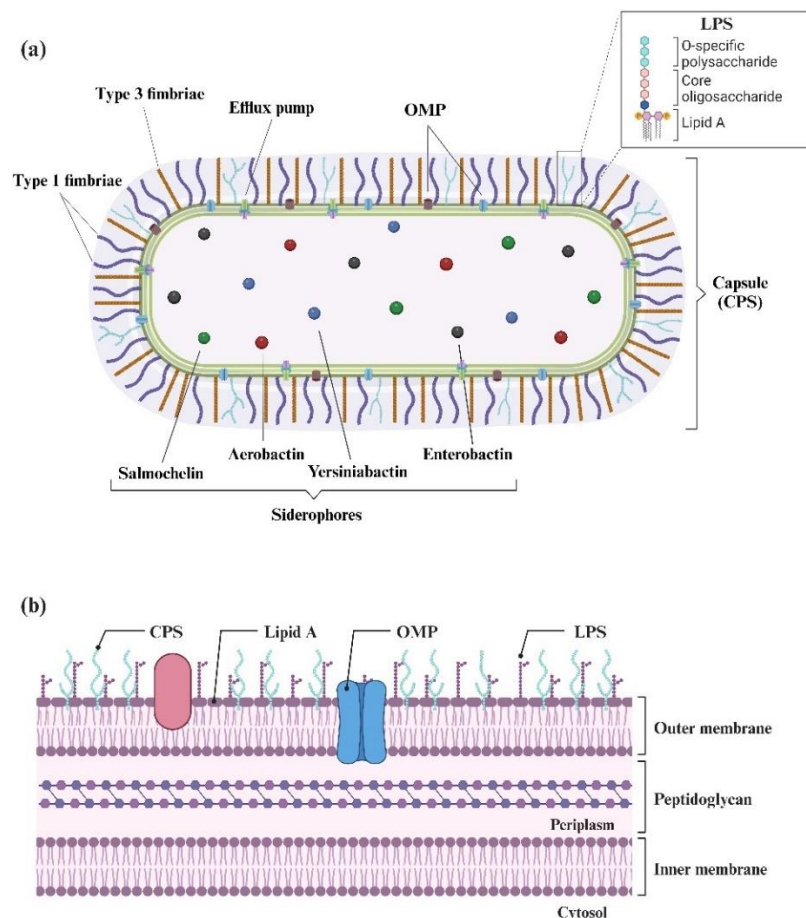


Figure 1. Major virulence factors of *K. pneumoniae* (a); the scheme of the cell envelope of Gram-negative bacteria (b).

4.2.1 Fimbriae

Central to *K. pneumoniae* pathogenic arsenal are fimbriae, hair-like appendages that extend from the bacterial surface and contribute significantly to colonization, host adherence, and evasion of the host immune response. Fimbriae are key players in the establishment of infections, and among the various fimbrial types identified in *K. pneumoniae*, type 1 and type 3 fimbriae have emerged as particularly crucial (Struve, Bojer, and Krogfelt 2008; Paczosa and Meccas 2016).

Fimbriae Type 1 are long, filamentous structures composed of repeating subunits of a major structural protein FimA with the FimH subunit on the tip. The *fim* gene cluster, responsible for the expression of fimbriae, typically includes genes such as *fimA*, *fimF*, *fimH*, and *fimI*, which collectively encode the various components of the fimbrial structure (Struve, Bojer, and Krogfelt 2008).

The primary function of these fimbriae is their ability to mediate adhesion to host tissues. The adhesin protein, FimH, located at the tip of the fimbrial structure, interacts with specific receptors on host cells, initiating the process of colonization. This adherence is crucial for establishing infection, as it allows the bacterium to resist mechanical clearance and immune responses, ultimately promoting the formation of bacterial communities.

Type 1 fimbriae also play a key role in biofilm formation. Biofilms are structured communities of bacteria embedded in a self-produced extracellular matrix. The adherence mediated by fimbriae contributes to the initial attachment of bacteria to surfaces, facilitating subsequent biofilm development (Schroll *et al.* 2010; Alcantar-Curiel *et al.*, 2012).

The expression of type 1 fimbriae is tightly regulated, responding to environmental cues such as temperature, nutrient availability, and host signals. The *fim* gene cluster is subject to phase variation, a reversible ON/OFF switching mechanism, allowing *K. pneumoniae* to adapt to changing host environments (Struve, Bojer, and Krogfelt 2008).

Type 3 fimbriae are helix-like, membrane-bound, adhesive structures on the surface of *K. pneumoniae*. These subunits assemble to form hair-like appendages extending from the bacterial surface. It is composed primarily of MrkA subunits, with the MrkD subunit on the tip. MrkD has specifically been found to bind extracellular

matrix, such as that exposed on damaged tissues. MrkA binds abiotic surfaces, such as medical devices (Paczosa and Meccas 2016).

The primary function of type 3 fimbriae is adherence to host tissues, a critical step in the pathogenesis of *K. pneumoniae*. By promoting bacterial attachment to host cells, fimbriae facilitates the initiation of infections in various anatomical sites, including the respiratory and urinary tracts (Schurtz Sebghati *et al.* 1998).

Moreover, type 3 fimbriae play a key role in the formation of biofilms and binding to the medical devices. In addition to its adhesive functions, fimbriae also modulate the host immune response. By interacting with immune cells and components, these fimbriae contribute to the evasion of host defences, allowing *K. pneumoniae* strains to establish persistent infections (Alcántar-Curiel *et al.* 2013; Paczosa and Meccas 2016).

4.2.2 Outer membrane proteins (OMP)

Klebsiella pneumoniae possesses outer membrane proteins (OMP) known as porins, which play a crucial role in regulating the passage of ions and small molecules across the bacterial outer membrane. Porins contribute to the permeability of the outer membrane and are essential for the bacteria survival and interaction with its environment, including host interactions.

Klebsiella pneumoniae porins typically form trimeric structures, with each monomer contributing to the formation of a channel in the outer membrane. These bacteria commonly express two major porins - OmpK35 and OmpK36, with distinct molecular characteristics and roles in permeability (Paczosa and Meccas 2016; Sugawara, Kojima, and Nikaido 2016).

OmpK35 has a molecular weight of approximately 35 kDa. It exhibits selective permeability to small hydrophilic molecules, such as ions and nutrients. Moreover, OmpK35 is involved in the diffusion of essential nutrients and ions across the outer membrane. OmpK36 has a molecular weight of around 36 kDa. It plays a role in the permeability of antibiotics, influencing the susceptibility of *K. pneumoniae* (Tsai *et al.* 2011).

Expression of porins in *K. pneumoniae* is influenced by environmental factors, including nutrient availability and exposure to antibiotics. Various regulatory proteins control the expression of porins, modulating the permeability of the outer membrane in response to environmental cues. Changes in the expression of porins, especially

OmpK36, have been linked to antibiotic resistance in *K. pneumoniae*. Moreover, porins contribute to the *K. pneumoniae* virulence by influencing its ability to acquire nutrients and survive within host environments (Tsai *et al.* 2011; Sugawara, Kojima, and Nikaido 2016).

4.2.3 Efflux pumps

One of the mechanisms contributing to MDR phenotype of bacteria is the presence of efflux pumps, intricate molecular machines that actively expel a variety of antimicrobial compounds from the bacterial cell, rendering it less susceptible to antibiotics (Paczosa and Mecsas 2016).

Efflux pumps in *K. pneumoniae* exhibit a complex architecture comprising transmembrane domains (TMDs) and cytoplasmic domains. TMDs, embedded in the bacterial membrane, form a channel through which substrates are transported. Cytoplasmic domains, often nucleotide-binding domains (NBDs), energize the pump by hydrolysing ATP. In *K. pneumoniae*, efflux pumps are categorized into several families, with notable examples including the Resistance-Nodulation-Division (RND) family, Major Facilitator Superfamily (MFS), and the ATP-binding cassette (ABC) superfamily (Padilla *et al.* 2010; Ni *et al.* 2020).

For example, AcrAB-TolC system is a tripartite efflux pump and a member of the RND family in *K. pneumoniae*. The AcrAB complex spans the inner membrane, with AcrA protein linking the inner membrane pump AcrB protein to the outer membrane protein TolC. The AcrB acts as the major multidrug transporter, recognizing diverse substrates, for example antibiotics, dyes, bile salts and detergents (Padilla *et al.* 2010; Zwama and Yamaguchi 2018; Ni *et al.* 2020).

A member of the MFS family, KpnEF is implicated in tetracycline resistance in *K. pneumoniae*. This pump actively expels tetracycline from the bacterial cell, contributing to the bacterium's resistance profile (Srinivasan and Rajamohan 2013).

An ATP-binding cassette transporter, KexD is associated with resistance to aminoglycosides and fluoroquinolones. Its ATP-binding cassette hydrolyses ATP to provide energy for substrate transport (Ogawa *et al.* 2012).

The expression of efflux pumps is tightly regulated to ensure the bacterium optimally balances the benefits of efflux with the energy cost. Regulatory mechanisms

involve transcriptional regulators, such as MarA and SoxS, responding to environmental signals and modulating the expression of efflux pump genes.

Efflux pumps contribute significantly to antibiotic resistance, and targeting these pumps may provide novel therapeutic strategies to enhance the efficacy of existing antibiotics.

4.2.4 Siderophores

Iron is a compound necessary for both the growth and enzymatic reactions of bacterial cells. In living organisms, the concentration of free iron ions available to bacteria is low. In the human body, iron occurs in a bound form, therefore the acquisition of iron ions in a bound form is an extremely important property of bacterial cells (Paczosa and Mecsas 2016). Bacteria have developed mechanisms that enable them to assimilate iron from the external environment. One of such mechanisms enabling the growth and colonization of host tissues is iron uptake with the participation of siderophores. Siderophores are low molecular weight chelators with high affinity for metal ions. Siderophores are released into the environment, where they bind free Fe^{3+} ions. Then, this complex is recognized and bound by receptor proteins located on the surface of the bacterial membrane. The active uptake of the siderophore- Fe^{3+} ion complex into the bacterial cell is conditioned by the presence of ABC transporters. Among the siderophores secreted by *K. pneumoniae*, there are aeroactin, enterobactin, yersiniabactin and colibactin.

Almost all *K. pneumoniae* strains produce enterobactin and the genes that are required for enterobactin biosynthesis are carried on the chromosome in the *entABCDEF* gene cluster (Hsieh *et al.* 2008; Paczosa and Mecsas 2016).

Yersiniabactin was originally discovered in *Yersinia* as a highly pathogenic island member, but this siderophore has also been identified in other bacteria, such as *K. pneumoniae*. Yersiniabactin was observed in only 18% of classical but in 90% of clinical hvKp isolates. However, in association with enterobactin, it is overrepresented in respiratory *K. pneumoniae* isolates (Bach and Carniel 2000; Hsieh *et al.* 2008).

Salmochelins are the C-glucosylated form of enterobactin. This modification is carried out by genes located either on the chromosome or on a plasmid within the *iroA*, *iroBCDE* gene cluster. This siderophore occurs only in approximately 2-4% of

K2 strains are the most prevalent type of *K. pneumoniae* strain, followed by K1 strains. There are several possible reasons for the increased incidence of virulence of K1 and K2 strains. One is that strains of the K1 and K2 serotypes are more resistant to phagocytosis than other strains. Furthermore, recently studies have shown that K1/K2 strains may be more resistant to opsonophagocytic uptake by macrophages via mannose/lectin receptors than strains of other serotypes (Shu *et al.* 2009; Paczosa and Mecsas 2016; Opoku-Temeng, Kobayashi, and DeLeo 2019).

The mucus viscosity-associated gene (*magA*) is restricted to the K1 serotype, while the envelope-associated gene (*k2A*) is restricted to the K2 serotype. The *magA* gene is an important virulence gene in invasive *K. pneumoniae*. This gene encodes the Wzy enzyme, which plays the role of a polymerase in the synthesis of the capsule. Therefore, the presence of *magA* in *K. pneumoniae* strains confers resistance to phagocytosis. Serum complement factors cannot easily reach the bacterial cell membrane, making phagocytosis difficult. Consequently, the disruption of the *magA* gene results in complete loss of resistance to complement serum factors and phagocytosis, so it acts as an important virulence determinant in infections caused by *K. pneumoniae* serotype K1 (Paczosa and Mecsas 2016).

CPS biosynthesis is dependent on multiple genes in the *cps locus* (*K locus*). This region contains genes involved in the synthesis of CPS repeating units and capsular polymerase (*wzy*), important for the assembly of capsular subunits on the bacterial surface (Opoku-Temeng, Kobayashi, and DeLeo 2019). Differences in nucleotide sequence and gene number contributed to differences in *K. pneumoniae* capsule types (Shu *et al.* 2009; Paczosa and Mecsas 2016; Opoku-Temeng, Kobayashi, and DeLeo 2019).

The CPS is described as an important *K. pneumoniae* virulence factor. Several different functions of the capsule have been described in *K. pneumoniae* virulence. During infection, the capsule prevents activation of the early immune response and protects against the host immune response by inhibiting phagocytosis (Paczosa and Mecsas 2016). Unencapsulated *K. pneumoniae* strains are more susceptible to complement-mediated killing and phagocytosis, thus show decreased pathogenicity in *in vivo* models (Opoku-Temeng, Kobayashi, and DeLeo 2019).

4.2.6 Lipopolysaccharide (LPS)

Lipopolysaccharide (LPS, O antigen, endotoxin) is an amphiphilic molecule found on the cell surface of Gram-negative bacteria, consisting of a polysaccharide and a lipid part (Figure 3). It is an integral component of the outer membrane of the bacterial cell envelope, essential for the functioning and survival of bacteria. LPS protects the bacterial cell against host defence mechanisms (e.i. complement), bile acids and antibiotics. Additionally, it acts as a bacterial virulence factor in the case of inflammation, sepsis and septic shock caused by Gram-negative bacteria.

Lipopolysaccharide is composed of three regions: lipid A, core oligosaccharide (core OS) and O-specific polysaccharide (a polymer of repeating units, OPS). Due to the morphology of bacterial colonies, we distinguish smooth (S) and rough (R) bacterial strains. The differences in phenotype between these forms of bacteria are caused by the complexity of the LPS structure found on the surface of the bacterial cell. Smooth forms are characterized by the presence of a complete LPS structure with the OPS. Rough forms are characterized by the lack of the OPS region and are present in bacteria with mutations in the genes encoding proteins responsible for the biosynthesis of the chain of repeating units (RU). Rough-smooth (RS) strains are also defined having core oligosaccharide substituted by one oligosaccharide unit, for example *Bordetella pertussis* (Holst and Brade 1992).

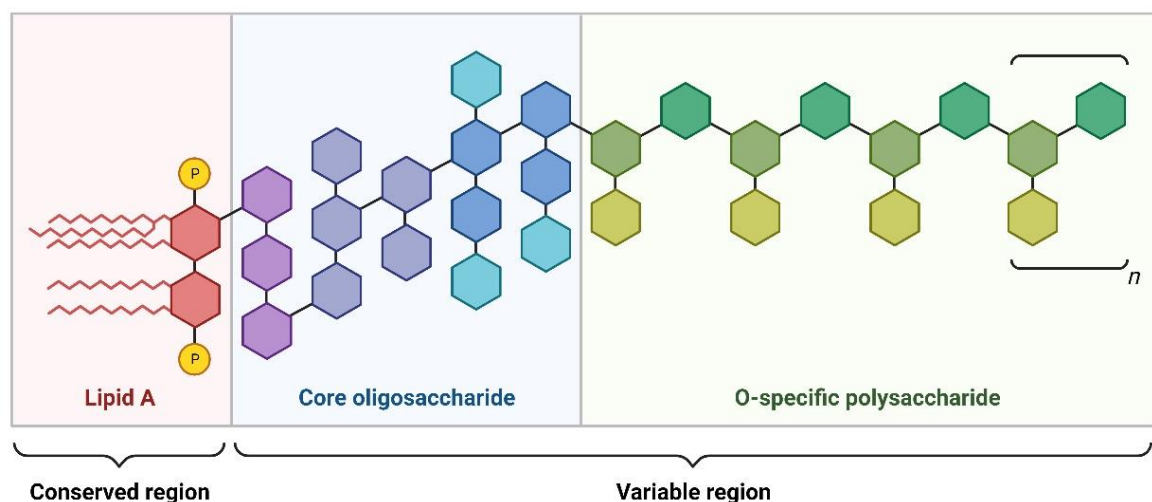


Figure 3. Scheme of the lipopolysaccharide.

4.2.6.1 Lipid A

Lipid A is a fragment that anchors the LPS molecule in the outer membrane of the bacterial cell. Lipid A contains a sugar skeleton, fatty acids (e.g. 3-hydroxy-tetradecanoic acid [14:0(3-OH)], 3-hydroxy-dodecanoic acid [12:0(3-OH)]) and phosphate groups (P) with substituents (4-amino-4-deoxy-L-arabinose (L-Arap4N), phosphoethanolamine (PEtn), phosphate, D-arabinose in the form of furanose (D-Araf), glucosamine (Glc_pN), galacturonic acid (Gal_pA)).

The Kdo residue of the core oligosaccharide connects to lipid A via an α -(2→6) ketosidic linkage. This bond is broken as a result of acid hydrolysis, which allows the separation of lipid A from the core oligosaccharide substituted by the OPS. The term lipid A was introduced in 1954 by Westphal and Luderitz, indicating the hydrophobic nature of the lipid A precipitate formed during the acid hydrolysis of the ketosidic bond within the LPS (Westphal, Lüderitz, and Bister 1952).

In terms of structural variability, lipid A is the most conserved part of LPS. Its heterogeneity most often concerns the degree of substitution of the sugar skeleton with phosphate groups, polar groups and the distribution of fatty acids. Lipid A possesses a unique structure characterized by a β -(1→6)-linked amino sugar disaccharide backbone which in most cases comprises two 2-amino-2-deoxy-D-glucopyranose (D-Glc_pN) residues (e.i. *K. pneumoniae*, *Escherichia coli*). However, lipids A of several species of bacteria contain one or two 2,3-diamino-2,3-dideoxy-D-glucopyranose (D-Glc_pN₃N) residues instead (e.i. *Oligotropha carboxidovorans*, *Phyllobacterium trifolii*, *Bradyrhizobium japonicum*) (Choma *et al.* 2014; Komaniecka *et al.* 2014; Zamlynska *et al.* 2017). In many cases carbohydrate backbones of lipid A are phosphorylated at positions 1 of the proximal α -Glc_pN and 4' of the distal β -Glc_pN. Both residues are acylated with 3-hydroxy fatty acids at positions 2, 2' and 3, 3' of both Glc_pN or Glc_pN₃N residues by amide and ester linkages. The core oligosaccharide is linked to the non-reducing distal amino sugar of lipid A.

In the case of the structure of *E. coli*, lipid A having Glc_pN-(1→6)-Glc_pN backbone carrying (*R*)-3-hydroxymyristoyl groups [14:0(3-OH)] as primary acyl chains in 2, 2', 3, 3' positions and lauroyl (12:0) and myristoyl group (14:0) as secondary fatty acids on the nonreducing sugar residue (Molinaro *et al.* 2015).

Lipid A of *K. pneumoniae* has a characteristic structure (Figure 4). The backbone consists of β -GlcN-(1 \rightarrow 6)- α -GlcN 1,4'-bisphosphate which is highly conserved and represents the structural principle of lipid A of all enterobacterial LPSs.

In *K. pneumoniae*, resistance is mediated by lipid A modifications with positively charged L-Arap4N and PEtn, which repel similarly charged polymyxin and thus attenuate the initial electrostatic interactions of the polymyxin molecules with the bacterial OM. The lipid A profiling results also showed modification by hydroxylation (-OH) (Lu *et al.* 2023).

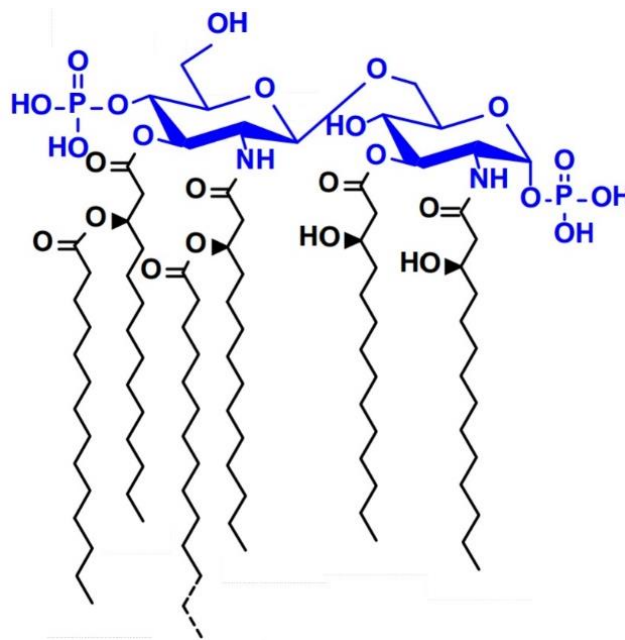


Figure 4. The lipid A structure from *K. pneumoniae*. Modifications with positively charged L-Arap4N and PEtn (Li *et al.* 2016).

Lipid A is a region of endotoxin, recognized by receptors on the surface of cells activated by LPS such as the TLR4/MD-2 (Toll-like receptor 4/Myeloid Differentiation factor 2) complex or intracellular receptors (Molinaro *et al.* 2015). This is therefore the centre of LPS biological activity. During a bacterial infection, components of the host's immune system are activated by LPS released by Gram-negative bacteria into the bloodstream. Activation of macrophages, neutrophils and endothelial cells of blood vessels often leads to effective inhibition of infection, but sometimes there is an overproduction of mediators such as Tumour Necrosis Factor α (TNF- α), interleukin 1 (IL-1), IL-6, IL-8) and septic shock, often ending with the death of the host (Rietschel *et al.* 1994; Molinaro *et al.* 2015).

4.2.6.2 Core oligosaccharide

The core oligosaccharide is present in all LPS of Gram-negative bacteria, regardless of substitution with O-specific polysaccharide. It is a hetero-oligosaccharide in which, due to its chemical structure, two parts can be distinguished: outer core – a region distal to lipid A; and the inner core - a region proximal to lipid A.

The outer part of the core determines the limited variability, while the inner part is a rather conservative fragment. In Enterobacteriaceae, the outer core, due to the presence of hexoses such as D-glucose (Glc) and D-galactose (Gal) in its structure, is also called the hexose part. Typical constituents of this region are also D-GlcN and D-galactosamine (GalN). The inner core is composed mainly of heptoses (L,D-*manno*-Hep or D,D-*manno*-Hep), which is why it is called the heptose region. Common components of the outer/hexose region are also sugar residues with a carboxyl group, the so-called uronic acids, such as D-galacturonic acid (GalA) and D-glucuronic acid (GlcA). The characteristic constituent of the inner core and Gram-negative bacteria is rarely occurred in nature, 2-keto-3-deoxyoctulosonic acid (Kdo) (Amor *et al.* 2000; Vinogradov *et al.* 2002).

The inner core oligosaccharide is a carrier of negative charges related to the presence of P, pyrophosphate (PP) and PEtn groups in its structure. These groups modulate LPS heterogeneity. It is very rare to find a core region without phosphate substituents in LPS structures. Only two such LPS have been described so far for *Plesiomonas shigelloides* and *K. pneumoniae*. The structure of the inner core of Enterobacteriaceae is conserved. Most members of this family contain a 5-sugar fragment. The reducing end of the core oligosaccharide is always occupied by a Kdo, which participates in the formation of an α -(2→6) glycosidic bond connecting the sugar part of LPS with lipid A. Modifications related to this part of the core concern the presence of substituents of the Kdo such as terminal Kdo, D-Gal, L-rhamnose (Rha), L-Ara4N and Hep (Severn *et al.* 1996)

The core oligosaccharide is a region of limited variability. In the case of *E. coli*, only 5 core types have been identified so far, and in *K. pneumoniae* - two types (Figure 5) (Knirel *et al.* 1992).

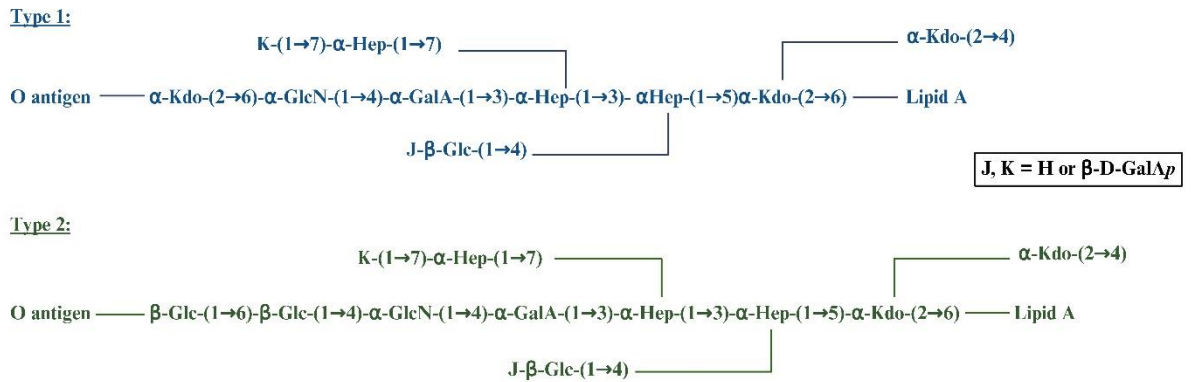


Figure 5. *K. pneumoniae* core type 1 and type 2 OS structures. Depending on the *K. pneumoniae* strain, letters J and K mark H or GalA (Fresno *et al.* 2006; Tu *et al.* 2022).

4.2.6.3 O-specific polysaccharide (OPS, O antigen)

In most cases, the OPS is a heteropolysaccharide composed of several to several dozen repeating oligosaccharide subunits containing from 2 to 8 sugar residues. The exception is LPS subunits composed of only one sugar residue – homopolymers, for example OPS of *Yersinia enterocolitica* O:3 built up of the 6-deoxy-β-L-altropyranose (Gorshkova *et al.* 1985). The OPS determines the serological specificity of LPS and this region is called O antigen. With other bacterial cell surface antigens, including capsular antigen K, mucus antigen M or ciliary antigen H, is used for bacteria serotyping (Petersson *et al.* 1997).

The structures of repeating units (RU): types of sugars, ring form (pyranose, furanose), sequence, type of bonds, substitution sites and types of substituents - show large differences between strains within the same species. Typical components of RUs include neutral sugars [Gal, mannose (Man), Glc, xylose (Xyl), ribose (Rib), Hep), amino sugars (GalN, GlcN), deoxysugars [Rha, fucose (Fuc)], uronic acids (GalA, GlcA). Additional components influencing structural variability are substituents [e.g., acetyl group (Ac), P, amino acids). Defence strategies against bacterial infections are based, among others, on the process of recognizing the components of the bacterial cell envelope. The most exposed epitope on the cell surface of Gram-negative bacteria is LPS, and above all its O-specific polysaccharide, which is the subject of the innate immune system response via a complement pathway (Morrison and Kline 1977; Petersson *et al.* 1997).

4.2.6.3.1 Structures of *K. pneumoniae* O antigens

To date only 11 *K. pneumoniae* O serotypes [O1, O2a, O2ac, O2afg, O2aeh (called also O9), O3, O4, O5, O7, O8, and O12] with two subserotypes [O3a, O3b] have been described (Figure 6) (Kelly *et al.* 1995; Hansen *et al.* 1999; Follador *et al.* 2016; Clarke *et al.* 2018; Kelly *et al.* 2019). Moreover, the O1 and O2 serotypes have at least three glycoforms (variants) each, regarding the galactan I RU structure (variants 1, v2, v3) (Szijártó *et al.* 2016; Stojkovic *et al.* 2017; Kelly *et al.* 2023). Recently, Kelly *et al.* has identified a new glycoform of the O1 antigen, defined as the O1b serotype, and composed of the previously reported O1 backbone (O1a), modified by the addition of a terminal pyruvate group by a pyruvyltransferase (WbbZ) (Kelly *et al.* 2023). A bioinformatic survey indicated that almost all *K. pneumoniae* O1 isolates, but also other bacterial species, possess genes required for the production of the O1a and O1b glycoforms (Kelly *et al.* 2023). Furthermore, O1 antigen can be modified by *O*-acetylation forming O8 serotype (Kelly *et al.* 1995). Some OPSs are capped non-stoichiometrically by a methyl group (Me) (O5), methyl phosphate (MeP) (O3, O3a, O3b), or a Kdo residues (O4, α -Kdo; O12, β -Kdo) (Vinogradov *et al.* 2002; Guachalla *et al.* 2017; Kelly *et al.* 2023).

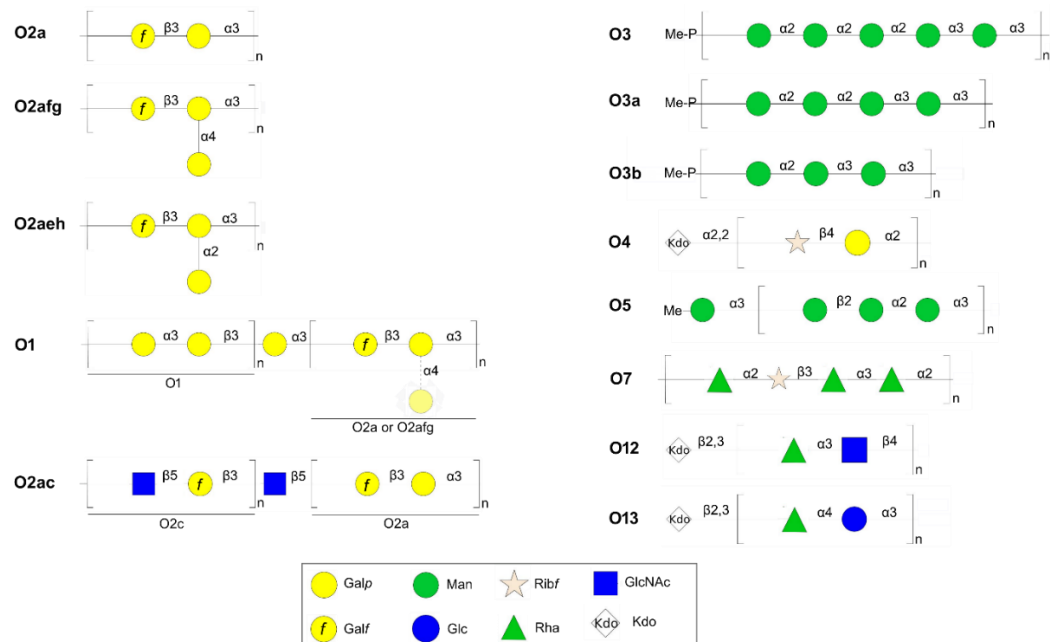


Figure 6. The known O antigen repeating unit structures of *K. pneumoniae*. The sugars contained within the square brackets are from the repeating units of the OPS. Not shown here are non-stoichiometric *O*-acetyl groups that distinguish serotypes O1 and O8 and a pyruvate group terminus for some of O1 (O1b) and O2ac antigens. Non-stoichiometric terminal Gal substitution within O1 antigen was also reported (dashed line and lighter colour of Gal symbol). The MeP and Me state for methylphosphate and methyl groups, respectively. The figure completed previously published schemes of *K. pneumoniae* O antigens (Clarke *et al.* 2018; Kelly *et al.* 2023).

4.2.6.3.2 Gene clusters required for O antigens biosynthesis

The *K. pneumoniae* OPS biosynthesis depends on the 9 *rfb* cluster (*O locus*) located between the *cps* and *hisI* loci (Clarke *et al.* 2018; Lam *et al.* 2022). Until now, twelve O antigen gene clusters have been defined (O1/O2v1, O1/O2v2, O1/O2v3, O3/O5, O4, O8, O12, OL101, OL102, OL103, OL104, and O105) (Figure 7); however, antigens specified by the OL101, OL102, OL103, OL104 and OL105 loci have not had chemical structures examined so far (Follador *et al.* 2016; Lam *et al.* 2022). Irrespective of the O serotype, the *O locus* genes encode enzymes responsible for the OPS assembly and transport (e.g., ABC transporters), monosaccharide synthesis (glycosylhydrolases) and OPS RU synthesis (glycosyltransferases) (Follador *et al.* 2016).

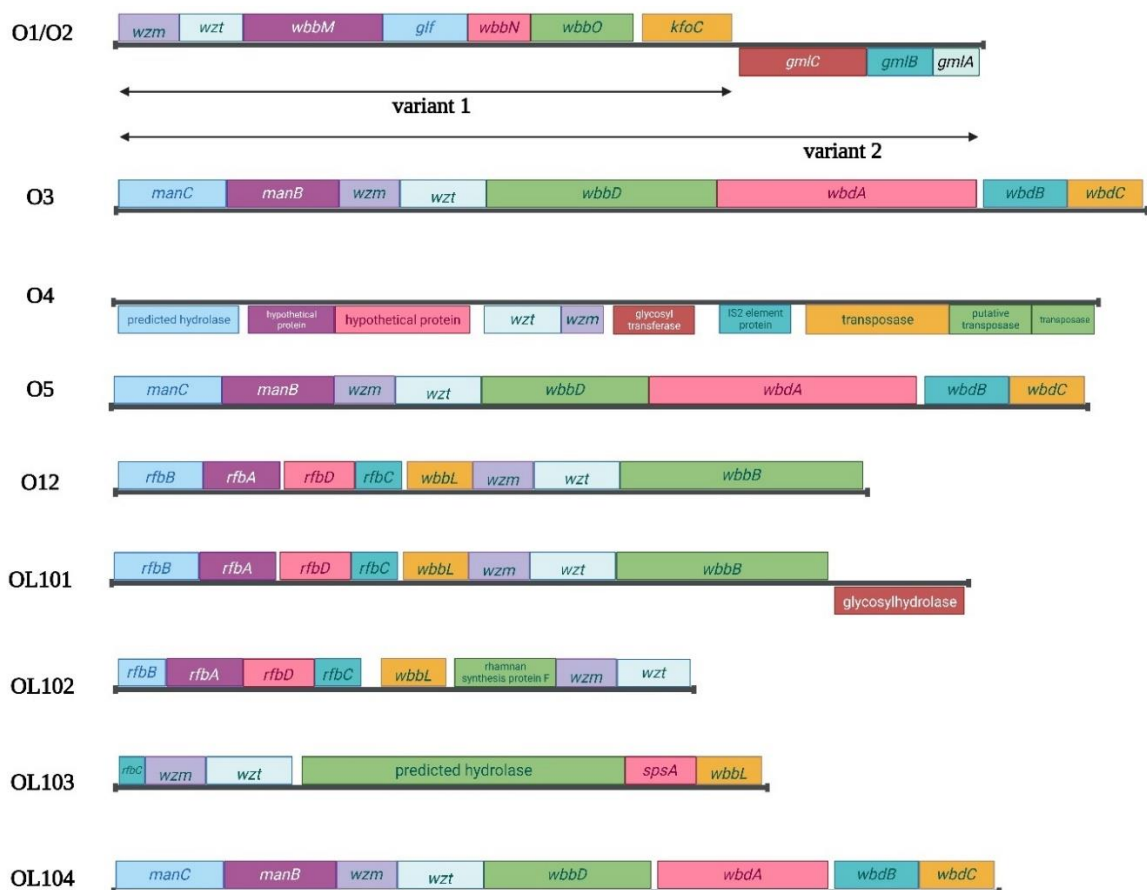


Figure 7. *rfb* gene clusters. Genes above or below the line according to the coding strand and the direction of transcription (Follador *et al.* 2016).

The O1/O2 *rfb locus* includes the following essential genes: (a) *wzm* and *wzt*, encoding transmembrane and nucleotide-binding domains of an ABC transporter; (b) *wbbMNO*, encoding glycosyltransferases; (c) *glf*, encoding the UDP-galactopyranose

mutase which converts UDP-Galp to UDP-Galf; and (d) *kfoC*, which is of unknown function (Fang *et al.* 2016; Follador *et al.* 2016; Szijártó *et al.* 2016; Clarke *et al.* 2018). Moreover, the O1/O2 antigen gene cluster occurs in two variants: the O1/O2 variant 1 (v1; encoding O2a, O2ac serotypes), characterized by the presence of D-galactan I encoded by the mandatory *wzm-wbbO* genes (Clarke and Whitfield 1992; Follador *et al.* 2016; Szijártó *et al.* 2016). The second and third variants (v2 encoding O2afg; v3, O2aeh serotypes) carries an additional three genes, *gmlABC* (*gmlABD*, in the case of the O2aeh serotype) (Figure 7) (Clarke *et al.* 2018; Stojkovic *et al.* 2017; Szijártó *et al.* 2016). These genes are in opposite orientation and located downstream of *rfb* cluster (Figure 7). The *gmlABC* products are three putative glycosyltransferases, which modify D-galactan I $\{\rightarrow 3\}$ - β -D-Galf-(1 \rightarrow 3)- α -D-Galp-(1 \rightarrow) (O2a; O2v1) to branched polymer of $\{\rightarrow 3\}$ - β -D-Galf-(1 \rightarrow 3)-[α -D-Galp-(1 \rightarrow 4)]- α -D-Galp-(1 \rightarrow) (O2afg; O2v2) or $\{\rightarrow 3\}$ - β -D-Galf-(1 \rightarrow 3)-[α -D-Galp-(1 \rightarrow 2)]- α -D-Galp-(1 \rightarrow) (O2aeh, O2v3) disaccharides, respectively (Follador *et al.* 2016; Szijártó *et al.* 2016). OPS of O1 contains D-galactan I and D-galactan II, built of $\{\rightarrow 3\}$ - β -D-Galf-(1 \rightarrow 3)- α -D-Galp-(1 \rightarrow) and $\{\rightarrow 3\}$ - α -D-Galp-(1 \rightarrow 3)- β -D-Galp-(1 \rightarrow) disaccharide RUs, respectively. D-galactan I in O1 antigen also occurs as two variants, v1 and v2. D-galactan II is encoded by a separate operon containing two genes located outside the *rfb locus*, *wbbYZ* (Fang *et al.* 2016; Follador *et al.* 2016; Szijártó *et al.* 2016). The O2 serotype is devoid of D-galactan II.

Similar to serotype O2, O3 is also a group of related structures. A recent study of O3 antigens identified three subtypes: O3, O3a and O3b (Vinogradov *et al.* 2002; Guachalla *et al.* 2017). The basis of variability of these serotypes are mutations in multidomain glycosyltransferase. The OPS termination is coupled to chain polymerization in the WbdA-WbdD complex. WbdA is a multidomain mannosyltransferase; various domains of this enzyme catalyse the addition of mannose molecules at different bonds.

The O3 (penta-mannose form) and O3b (tri-mannose form) *loci* differ in the sequences of the *wbdA* and *wbdD* genes. The O3 (penta-mannose form) and O3a (tetra-mannose form) *loci* differ in a single point mutation in *wbdA* (C80R) (Figure 7) (Guachalla *et al.* 2017; Wick *et al.* 2018; Kelly *et al.* 2023).

The composition of the O3 and O5 *rfb* clusters are identical (Figure 7). The O3 and O5 biosynthesis enzymes differ in the sequence of mannosyltransferase (WbdA) and methyltransferase (WbdD). Complex WbdA-WbdD regulates the length of the mannose chain by capping the growing chain with a methyl group in O5 (Follador *et al.* 2016).

The O4 *rfb locus* is the only *rfb* cluster in *K. pneumoniae* containing a transposase (Follador *et al.* 2016). It is also composed of the following essential genes: i) *wzm* and *wzt* encoding the transmembrane and nucleotide-binding domains of an ABC transporter, ii) glycosyl transferase, and iii) IS2 element protein (Figure 7).

The OL101 *rfb locus* is composed of the following essential genes: i) *wzm* and *wzt* encoding the transmembrane and nucleotide-binding domains of an ABC transporter, ii) *rfbBADC* specifying the enzymatic pathway converting glucopyranose-1-phosphate to dTDP-Rhap, iii) *wbbL* providing the rhamnosyltransferase for the OPS polymerization, iv) *wbbB* coding a trifunctional protein responsible for OPS polymerization, termination (i.e. adding a nonreducing terminal residue) and chain-length regulation, and v) glycosylhydrolase gene of unknown function, located downstream of the *rfb* cluster in the opposite orientation (Figure 7) (Follador *et al.* 2016; Williams *et al.* 2017).

Based on their gene content and order, OL101, OL102, and O12 have been previously described as having closely related *rfb* clusters, differing by the presence and absence of single genes. Compared with the O12 cluster, OL101 includes the additional glycosylhydrolase gene, while OL102 lacks the terminal glycosyltransferase gene but possesses an additional rhamnan synthesis gene (Figure 7) (Follador *et al.* 2016; Williams *et al.* 2017).

The subject of research in this work is more detailed genomic analyses of three of the above-mentioned *rfb* regions: O1/O2, O4 and OL101.

4.3 Therapeutic strategies against *K. pneumoniae* infections

Some of the new therapeutic approaches under development are based on concepts of active and passive immunization against major surface antigens of *K. pneumoniae* – the K antigen and lipopolysaccharide (LPS, endotoxin, O antigen) (Paczosa and Meccas 2016; Adamo and Margarit 2018). Synthetic vaccines targeting the *K. pneumoniae* sugar arsenal are currently in the preclinical phase. Several vaccine strategies targeting the most prevalent K antigens of *Klebsiella* have been developed, particularly against the most virulent serotypes - K1 and K2 (Edelman *et al.* 1994; Hegerle *et al.* 2018; Feldman *et al.* 2019).

Contrary to highly variable K-antigens (more than 80 types), LPS represents the less variable antigen and have been used for *K. pneumoniae* typing. It is an important

virulence factor, triggering the TLR4-dependent immune response, and four of O antigens (O1, O2, O3, O5) predominate among clinically relevant strains. Therefore, they were proposed as potential targets for active or passive immunization as an alternative to antibiotic treatment (Follador *et al.* 2016; Szijártó *et al.* 2016). Recently, promising bactericidal and neutralising monoclonal antibodies targeting the most common *K. pneumoniae* O serotypes (O1, O2, O3, O5) have been developed (Szijártó *et al.* 2016; Guachalla *et al.* 2017; Stojkovic *et al.* 2017; Rollenske *et al.* 2018). The most clinically relevant O1, O2, O3, and O5 antigens are considered as a part of prototype glycoconjugate vaccine that would ensure an 80% coverage against *Klebsiella pneumoniae* infections (Follador *et al.* 2016). The idea has forced synthetic chemists to develop technology of *K. pneumoniae* O antigen synthesis and proof of concept research on novel glycoconjugate vaccines against *Klebsiella* (Wang, Zhang, and Ning 2003; Chhibber, Rani, and Yadav 2005; Clements *et al.* 2008; Zhu and Yang 2012; Verkhnyatskaya, Krylov, and Nifantiev 2017; Hegerle *et al.* 2018; Argunov *et al.* 2019; Geissner *et al.* 2019; Del Bino *et al.* 2022). Therefore, the preclinical development of vaccines is mostly focused on these four serotypes (Del Bino *et al.* 2022).

Majority of commercially available vaccines that target surface carbohydrates are based on the chemical conjugation of capsular polysaccharides with carriers and due to their ability to control various bacterial infections, more vaccines targeting unmet medical needs are either in development at the preclinical level or in clinical trials. Thus, in the case of *K. pneumoniae*, two synthetic vaccines aimed at the sugar arsenal are currently in the preclinical phase – the LMTB-GSK and Alopexx vaccines (Del Bino *et al.* 2022). The LMTB-GSK contains nontoxic recombinant *Pseudomonas aeruginosa* exotoxin A as carrier protein and O1, O2 oligosaccharides (Chhibber, Rani, and Yadav 2005; Del Bino *et al.* 2022) The Alopexx contains tetanus toxoid as carrier protein and O3, O3b, O5 synthetic oligosaccharides (Del Bino *et al.* 2022).

The extracted O antigens were tested in animal models as a stand-alone polysaccharide antigen or conjugated with tetanus toxoid or *Pseudomonas aeruginosa* flagellin FlaA and FlaB. The synthesis of O antigens belonging to the O1 and O2 serotypes has been also described. Synthesized di-, tetra-, and hexasaccharides of both serotypes were biotinylated to immobilize a streptavidin-coated glycan matrix plate for screening sera for O1 of serotypes K1, K2, and K16. All sera tested showed significant concentration-dependent reactivity with a hexasaccharide of Gal, while interaction with

tetrasaccharide of Gal was observed and the disaccharide was poorly recognized (Clements *et al.* 2008; Del Bino *et al.* 2022)

O1 and O2 oligosaccharides (and related subtypes O2afg and O2ac) ranging from one to eight RUs in length have recently been synthesized using a highly convergent strategy based on the key building blocks of the galactose disaccharide. Oligosaccharides from these serotypes were conjugated to a carrier protein CRM₁₉₇ (Cross-Reactive Material 197) and tested *in vivo* in rabbits and mice. After immunization, functional antibodies against homologous oligosaccharides used for immunization and against natural O polysaccharides of *K. pneumoniae* serotypes O1 and O2 carbapenem-resistant ST258 were detected, indicating the potential of these vaccines to provide protection against *K. pneumoniae* infections (Geissner *et al.* 2019; Del Bino *et al.* 2022).

Klebsiella pneumoniae O3, O3b and O5 polysaccharides are other clinically important antigens that should be included in the vaccine to ensure adequate effectiveness against circulating strains. The syntheses of oligosaccharides O3 to O3b to O5 have recently been described. Different length poly- and oligosaccharides of OPS were conjugated to CRM₁₉₇ and bovine serum albumin (BSA). Sera from immunized rabbits cross-reacted in ELISA with the corresponding O antigens, whereas sera from mice were able to selectively recognize homologous LPS (Geissner *et al.* 2019; Del Bino *et al.* 2022).

Even though vaccines targeting O antigens are promising idea, there is limited evidence of restricted access of Ab due to the layer of CPS (Wantuch *et al.* 2023). Wantuch *et al.*, reported the production of two bioconjugate vaccines. One targeting the K2 capsular serotype and the other targeting the O1 O antigen. These studies determined the induction of specific antibody responses that recognize K2:O1 *K. pneumoniae* strains. Both *cKp* and *hvKp* strains exhibited decreased O-antibody binding in the presence of capsule. Further, O1 antibodies demonstrated decreased killing in serum bactericidal assays with encapsulated strains, suggesting that the presence of *K. pneumoniae* capsule blocks O1-antibody binding and function. Moreover, the K2 vaccine outperformed the O1 vaccine against both *cKp* and *hvKp*. These data suggest that capsule-based vaccines may be superior to O antigen vaccines for targeting *hvKp* and some *cKp* strains, due to capsule blocking the O antigen (Wantuch *et al.* 2023).

5. Aims of the study

Klebsiella pneumoniae is a Gram-negative bacterium, being a part of the human microbiota but also a frequent cause of nosocomial and community-acquired infections in new-borns, the elderly and immunocompromised patients. The global emergence of multidrug-resistant (MDR) strains, especially extended-spectrum β -lactamases (ESBL) and carbapenemase-producing *K. pneumoniae* (CPKP), has become an ultimate challenge for public health. Treatment options against CPKP are sparse and usually limited to only last-line antibiotics if at all.

The discovery of new O serotypes or their modifications are essential for development of therapies against *K. pneumoniae* infections, alternative to antibiotics, such as vaccines and passive immunization. Nowadays, only four serotypes (O1, O2, O3, and O5) out of known O antigens of *K. pneumoniae* have been selected as clinically relevant for development of glycoconjugates of carrier proteins with O-specific polysaccharides as vaccines or antigens for production of protective antibodies. However, the screening of available genomes and recent O serotype prediction data indicated higher diversity of O antigen structures among clinical isolates.

The overall goal of the PhD project was examination of the eleven clinical isolates of *K. pneumoniae* (BIDMC 7B, ABC152, ABC122, BC738, BC13-986, 3936/19, Kp164, Kp165, Kp166, Kp174, Kp177) for the presence of novel O serotypes (O antigen structures) or novel structural modifications of O antigens. Strains were selected from a previous large-scale serotyping study (unpublished results), due to the inconsistency between the lack of LPS reactivity with available specific monoclonal antibodies against known O antigen structures (data not shown) or observed differences in O *locus*. Preliminary data suggested the presence of (a) modification of known O serotype, or (b) presence of new O serotype, or (c) the loss of O antigen biosynthesis ability.

Main goals to achieve by the studies were covered by the following tasks:

1. **Determination of complete structures of O antigens isolated from selected eleven strains of *K. pneumoniae*.** Clinical isolates were cultured, and bacterial mass used to extract LPS, and O antigen structures further analysed by chemical methods and NMR spectroscopy, including HR-MAS NMR analysis of native LPSs.
2. **Whole-genome sequencing (WGS) and analysis of O loci encoding O antigens by bioinformatics tools.** WGS sequences were provided with *rfb* gene clusters and all unlinked genes responsible for O antigen biosynthesis to complete the knowledge

concerning O loci determining structures of studied O serotypes. O loci of strains were predicted by Kaptive tool and compared with phenotypes of analysed O antigens.

3. Determination of the frequency of new O serotypes or their modifications based on available databases of *K. pneumoniae* genomes.

The general outline of the study is presented in Figure 8.

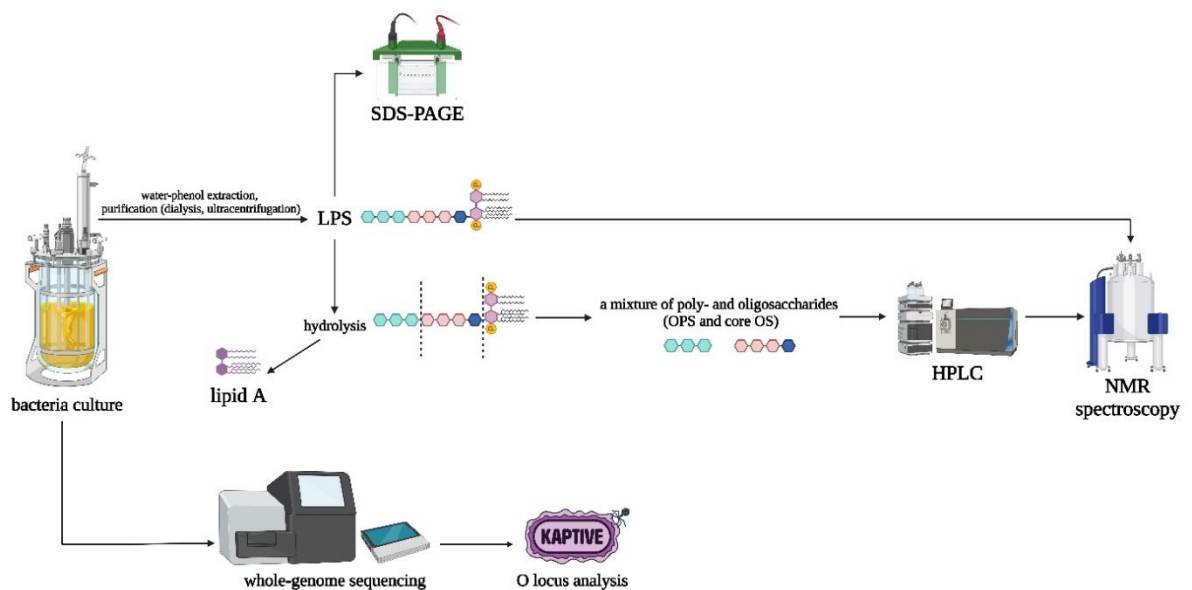


Figure 8. General outline of the experiments performed in this study.

Structural analysis was carried out using preparative (size exclusion chromatography, high performance liquid chromatography) and analytical techniques dedicated to carbohydrate chemistry (NMR spectroscopy, mass spectrometry, chemical analysis). Molecular analysis was performed using WGS and bioinformatics tools such as Kaptive and MLST analysis.

6. Materials and Methods

6.1 Media and buffers

Table 1. Compositions of buffers and medium used in this work.

Denaturing buffer for SDS-PAGE	Separation buffer, pH 8.3 for SDS-PAGE
65 mM Tris	25 mM Tris
30% glycine	192 mM glycine
2% SDS	0,1% SDS
0,6 M DTT	
0,001% bromophenol blue	
Phosphate buffered saline (PBS) (pH 7,4)	Brain Heart Infusion (BHI) Broth
NaCl 8 g/l	animal tissue hydrolysate 27,5 g/l
KCl 0,2 g/l	D-glucose 2 g/l
Na ₂ HPO ₄ 1,44 g/l	NaCl 5 g/l
KH ₂ PO ₄ 0,24 g/l	Na ₂ HPO ₄ 2,5 g/l

6.2 Bacteria and growth conditions

Twelve *Klebsiella pneumoniae* isolates used in this study are showed and characterised in Table 2. Agnes Sonnevend-Pal and Tibor Pal from the United Arab Emirates University are kindly acknowledged for ABC152, ABC122, BC738, and BC13-986 clinical isolates. Jesús Mingorance from the Hospital Universitario La Paz in Madrid is kindly acknowledged for Kp164, Kp165, and Kp166 strains. Marek Gniadkowski from the NMI in Warsaw is acknowledged for 3936/19 strain.

Table 2. *Klebsiella pneumoniae* isolates used in this study.

Isolate no./ original no.	Original source	Epidemiological data
Kp164/KpQ3	blood culture, the outbreak strain from the burn ICU at the Hospital Universitario La Paz, Madrid, Spain (2008)	ESBL; MDR (Tobes <i>et al.</i> 2013; López-Camacho <i>et al.</i> 2014)
Kp165/KpQ15	urinary tract infection isolate, the outbreak strain from the burn ICU at the Hospital Universitario La Paz, Madrid, Spain (October 2008)	ESBL; CRE; MDR (López-Camacho <i>et al.</i> 2014)
Kp166/KpQ24	urine culture, the outbreak strain from the burn ICU at the Hospital Universitario La Paz, Madrid, Spain (2009)	ESBL; CRE; MDR and colistin-resistant (López-Camacho <i>et al.</i> 2014)
BIDMC 7B	urine isolate, BEI Resources, NIAID, NIH: “ <i>Klebsiella pneumoniae</i> , strain BIDMC 7B, NR-41923”, as a reagent bought as a part of the Klebsicure-Eurostars project (no. E!7563)	CRE
Kp174	clinical isolate, hospital in the UAE	CRE
ABC122	bloodstream isolate, hospital in the UAE	MDR, NDM-1
Kp177	clinical isolate, hospital in the UAE	CRE
ABC152	urine isolate, Abu Dhabi Hospital, UAE	CRE
BC738	urinary tract isolate, hospital in the UAE	MDR
BC13-986	skin isolate, hospital in the UAE	MDR
3936/19	skin isolate, cultured in 2019 in a hospital in northern Poland; collected by the National Medicines Institute (Warsaw Poland)	MDR, VIM-1
PCM15	O4:K15, capsule prototype strain identified as the result of this study as <i>Enterobacter cloacae</i>	(Hansen <i>et al.</i> 1999)

CRE – carbapenem-resistant; ESBL – extended spectrum β -lactamase; ICU – Intensive Patient Care Unit; MDR – multidrug-resistant; NDM-1 – New Delhi metallo- β -lactamase 1; NIAID – National Institute of Allergy and Infectious Diseases, NIH – National Institutes of Health; UAE – United Arab Emirates; VIM-1 – Verona integron-encoded metallo- β -lactamase 1

Strains were verified to the species level as *K. pneumoniae* (score values of ≥ 2.000) by matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) with the MALDI Biotyper software (Bruker Daltonics, Bremen, Germany) (section 6.8).

Strains were preserved by two different methods, freeze-drying in water mQ (resistance at 25°C 18,2 MΩ × cm, total organic carbon TOC less than 5 ppb) and freezing as a suspension of 20% glycerol in -70°C.

Bacteria were grown on Trypcase Soy Agar plates in 37°C for 24 h (UV Multi-Gas Incubator, Panasonic Corporation, Japan). For the semi-preparative scale LPS isolation, the strains were cultured for 24-48 h in 4 l of brain heart infusion (BHI) broth (Merck, Darmstadt, Germany) (Table 1) in Biostat A Plus (5 l) bioreactor (Sartorius Stedim, Göttingen, Germany), with agitation (100 rpm) at 37°C with 5 l/min air flow and 1% of inoculum. Then, bacteria were inactivated with 3% phenol at 60°C for 2 h and harvested by centrifugation three times (2268 × g, 20 min, 4°C), washed each time with phosphate-buffered saline (PBS) and freeze-dried.

6.3 Lipopolysaccharide preparation

Lipopolysaccharides of *K. pneumoniae* strains were isolated from bacterial mass by hot phenol/water extraction (Westphal, Lüderitz, and Bister 1952) and purified by dialysis and ultracentrifugation with modification by Lukasiewicz *et al.* (Lukasiewicz *et al.* 2010).

Briefly, freeze-dried bacteria were suspended in 45% phenol solution (2 g/50 ml) and then, incubated at 65°C with intermittent stirring for 15 min. The suspension was cooled down to the temperature below 10°C and centrifuged (5000 × g, 25 min, 4°C, Sorvall Lynx 6000, F9-6x1000 LEX rotor). Water phase and interphase were collected, and water was added to compensate for the collected volume of water and interphase. The cycle of the heating and cooling was repeated. The pooled water phases were dialyzed against deionized water (4°C) to remove residual phenol, centrifuged (5000 × g, 25 min, 4°C, Sorvall Lynx 6000, F9-6x1000 LEX rotor), and then filtered through Whatmann 3 paper filters, and freeze-dried. The crude LPS was dissolved in mQ water and purified by ultracentrifugation (105000 × g, 6 h, 4°C, Sorvall WX+, T-865 rotor). The pellet was collected and water was added to LPS, and centrifugation was repeated twice.

6.4 O-specific polysaccharide preparation

Bacterial OPSs were isolated as previously reported (Szijártó *et al.* 2016) with slight modification. Briefly, approximately 7,5-35 mg of LPS were degraded by

treatment with 1,5% acetic acid at 100°C for 45 min, and then centrifuged (3 min, 13000 × g, 4°C) to remove lipid A precipitate. The supernatants obtained were fractionated by size-exclusion chromatography on the TSKgel®G2500PW column (7,5 × 60 cm) (Tosoh Bioscience, Tessenderlo, Belgium) equilibrated with 0,05 M acetic acid as an eluent (1 ml/min). Fractions were monitored with a refractive index detector (RI 102; Shodex) (Showa Denko Europe GmbH, Munich, Germany) and UV detector (190 nm), collected, pooled, and freeze-dried. OPS fractions were analysed by ¹H,¹³C nuclear magnetic resonance (NMR) spectroscopy (section 6.6) and constituents were identified by sugar (section 6.5.1) and methylation (section 6.5.2) analyses.

6.5 Compositional analysis

6.5.1 Sugar analysis

Monosaccharides were examined as their alditol acetates by gas chromatography coupled to mass spectrometry (GC-MS) (Pettersson *et al.* 1997). Sugar analysis was performed by hydrolysing the sample (~0,5 mg) in a 2 M CF₃COOH (1 ml) in a closed test tube for 2 h at 120°C. The trifluoroacetic acid (TFA) solution was evaporated with a stream of air, and the released sugars were converted into alditol acetates by reducing them with an aqueous NaBH₄ solution overnight (~8 mg/0,5 ml, to maintain pH 8-9). The reaction mixture was acidified with concentrated CH₃COOH to pH 5-6, and the resulting boric acid was removed by dissolving the material in 0,5 ml a methanol/acetic acid solution (9:1) and evaporating the volatile trimethyl borate (CH₃O)₃B with a stream of air. The procedure was repeated five times. The dried material was acetylated with acetic acid anhydride in pyridine (1:1, 1 ml, 100°C, 1 h), cooled, evaporated to dryness, washed with ethanol. The obtained alditol acetates dissolved in ethyl acetate (100 µl) were analysed in 1 µl volume by GC-MS using Thermo Scientific ITQ system with a Zebron™ ZB-5HT GC Capillary Column (30 m × 0,25 mm × 0,25 µm) (Thermo Fisher Scientific, Waltham, MA, USA) over the temperature program gradient from 150 to 270°C at 8°C/min gas flow. The identification of sugar derivatives was performed by comparing the retention times (t_R) of sample peaks with alditol acetates of carbohydrate standards.

6.5.2 Methylation analysis

To identify monosaccharide substitution position (glycosidic bond) partially methylated alditol acetates were analysed by GC-MS, according to the method described previously (Ciucanu and Kerek 1984). O-specific polysaccharide (~0,4 mg) in penicillin bottle was dissolved in anhydrous DMSO (0,5 ml), closed tightly and placed in an ultrasonic bath for 15 min. NaOH powder (3-4 mg) was added to the sample suspension and the bottle was placed in an ultrasonic bath for 5 min. Then the mixture was cooled to -20°C and 1 ml of methyl iodide was added. The methylation reaction was carried out for 1 h in an ultrasonic bath. The permethylated sample was extracted five times in the water/chloroform/1 M acetic acid (7,5:10:2,5) solution. The collected chloroform fraction was evaporated. Further stages of obtaining volatile sugar derivatives together with GC-MS analysis were identical to those in the sugar analysis procedure (section 6.5.1), except for the reduction step, where sodium borodeuteride (NaBD_4) was used instead of sodium borohydride. GC-MS analysis was described in section 6.5.3.

6.5.3 Gas chromatography coupled to mass spectrometry (GC-MS)

Sugar derivatives (alditol acetates and partially methylated alditol acetates) were analysed using a Thermo ITQ™ 1100 GC-Ion Trap mass spectrometer coupled with a Trace™ 1310 gas chromatograph (Thermo Scientific™). For analysis, the samples were dissolved in 30 μl of ethyl acetate, from which 1 μl of the solution was placed on a Zebtron™ ZB-5HT GC Capillary Column (30 m \times 0,25 mm \times 0,25 μm) (Thermo Fisher Scientific, Waltham, MA, USA). Helium was used as the carrier gas at a constant flow of 1,8 ml/min (temperature program 150°C – 270°C , $8^{\circ}\text{C}/\text{min}$). Chromatograms and mass spectra were analysed using Xcalibur software (Agilent Technologies, Santa Clara, USA).

6.6 NMR spectroscopy of LPS and OPS

All NMR spectra were obtained at 298 K using an Avance III 600 MHz spectrometer (Bruker, Billerica, MA, USA), equipped with a PH HR-MAS probe (LPS samples) or 5 mm QCI cryoprobe with z-gradient (OPS and oligosaccharides samples). NMR spectra of isolated OPSs were obtained in $^2\text{H}_2\text{O}$, processed and analysed as described previously (Szijártó *et al.* 2014). The samples were first exchanged three times

with $^2\text{H}_2\text{O}$ (99,95%, Sigma). For high-resolution magic angle spinning (HR-MAS) NMR spectroscopy, LPS (~3 mg) was suspended in $^2\text{H}_2\text{O}$ (100 μl) and freeze-dried, and the cycle of freeze-drying in $^2\text{H}_2\text{O}$ was repeated twice. For NMR analysis the sample was suspended in 30 μl of $^2\text{H}_2\text{O}$ (99,95%, Sigma) and placed into the ZrO_2 rotor. HR-MAS NMR experiments were carried out at a spin rate of 4-5 kHz. Acetone was used as an internal reference ($\delta_{\text{H}}/\delta_{\text{C}}$ 2,225/31,05 ppm) for OPS spectra (Stojkovic *et al.* 2017). Spectra were recorded with water suppression using excitation sculpting or wet techniques within the pulse programs. The data were acquired and processed using TopSpin software (Bruker, Billerica, MA, USA). The processed spectra (^1H , ^{13}C HSQC-DEPT, HMBC, HSQC-TOCSY, ^1H , ^1H COSY, NOESY and TOCSY) were assigned with the use of NMRFAM-SPARKY program (v1.2, NMRFAM, Madison, WI, USA) (Lee, Tonelli, and Markley 2015). The $J_{\text{C-1,H-1}}$ constants were determined based on the ^1H , ^{13}C HMBC spectra.

6.7 Absolute configuration of sugar components analysis by NMR spectroscopy

The absolute configurations of sugars were determined using ^1H nuclear magnetic resonance (NMR) spectroscopy by converting the polysaccharide hydrolysate components and relevant monosaccharide standards into *O*-(*S*)-2-methyl butyrate derivatives, as described by York *et al.* (York *et al.* 1997).

Briefly, sample of the OPS (~0,2 mg) was hydrolysed in 2 M trifluoroacetic acid (121°C, 90 min), and the solvent was evaporated. (*S*)-(+)-2-Methylbutyric anhydride (SMB) (100 μl) and anhydrous pyridine (100 μl) were added to the LPS sample and standards of monosaccharides (D- and L-sugars, 2 mg) and samples were incubated at 121°C for 4 h in a tube with a screw cap. The reagents were evaporated by compressed air for approximately 8 h. After that, toluene (300 μl) was added and evaporated. The residues were dissolved in 1 ml CH_2Cl_2 and extracted with 2 M Na_2CO_3 (1:1, 1 ml) and then with mQ water (1:2, 2 ml). The extraction was repeated three times. During the final water extraction, the lower phase (with CH_2Cl_2), containing the SMB derivatives was transferred to a clean glass tube. The sample was concentrated to syrup by evaporation and dried after adding of 2-propanol (300 μl) by evaporation. For the NMR analyses, the samples were dissolved in acetone-*d*6 (~160 μl) and transferred to 3 mm diameter NMR tubes for ^1H NMR spectra collection. The spectra were calibrated to the internal reference of acetone-*d*6 (δ 2,05 ppm of internal acetone-*d*5). NMR spectra of samples were

compared in the range of 5,5–6,5 ppm for *O*-(*S*)-2-methyl butyrate derivatives of monosaccharide standards and *K. pneumoniae* OPS hydrolysate to discriminate between D- and L- configuration of sugar constituents.

6.8 Matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS)

Identification of bacteria at the species level was performed according to the MALDI Biotyper Tube Extraction (formic acid extraction) procedure. Briefly, the single colony of bacteria was suspended in mQ water (300 μ l). The sample was mixed using vortex shaker thoroughly. Then, ethanol (96%) was added (900 μ l), and the sample was mixed and centrifuged at $13000 \times g$ for 3 min at 10°C. After removing the excess of ethanol with pipette, 70% formic acid was added and the sample mixed using vortex shaker thoroughly, followed by addition of equal volume of acetonitrile (50 μ l) and mixing. Then, the sample was centrifuged at $13000 \times g$ for 2 min and 1 μ l of supernatant was placed onto steel target and dried on air. The sample was overlaid with 1 μ l of HCCA (α -cyano-4-hydroxycinnamic acid) matrix and dried at room temperature. The Bacterial Test Standard (BTS) solution was placed on steel plate and dried at room temperature. The BTS spot was overlaid with 1 μ l of HCCA matrix solution and drying at room temperature. MALDI-TOF MS analyses were performed in linear mode, in positive polarization, in the mass range from 2000 to 20000 m/z . MALDI-TOF MS measurements were performed using the UltrafleXtreme device (Bruker Daltonics, Bremen, Germany) with a TOF detector.

6.9 Electrophoresis under denaturing conditions (SDS-PAGE)

Electrophoresis was performed in a 12% polyacrylamide gel with dimensions of $8,6 \times 6,7 \times 0,1$ cm (Bio Rad Ltd, Hercules, USA) with 4% stacking gel according to Laemmli, with modification (Laemmli 1970) Briefly, LPS solutions (1 mg/ml in mQ water) were incubated in denaturing buffer (0,05 M Tris-HCl, pH 6,8, containing 4% SDS, 4% glycerol, and 0,001% bromophenol blue) in the ratio 1:1 for 20 min in 100°C, then 10 μ g of LPS was applied per well in a volume of ≤ 20 -40 μ l. Electrophoresis was performed for approximately 30-40 min at a current of 30 mA and 40 mA per plate for stacking and resolving gel, respectively. The separated LPSs were silver stained according to the procedure described by Tsai and Frasch (section 6.10) (Tsai and Frasch 1982).

6.10 Silver staining

The LPS samples were separated as described above. Silver staining was performed as published previously (Tsai and Frasch 1982). Briefly, following overnight fixation of the gel in 25% isopropanol with 7% acetic acid solution, the gel was oxidized in 0,7% periodic acid in 40% ethanol and 5% acetic acid with shaking. Washing of the gel in mQ water was repeated three times. Then, the gel was stained with 0,8% silver nitrate in 1,4% ammonium hydroxide and 200 mM sodium hydroxide solution (10 ml) for 10 min and developed with 0,019% formaldehyde in 0,005% citric acid (10 ml) with shaking. MQ water (10 ml) was used to stop the reaction and the gel was washed three times with mQ water before digital archiving.

6.11 Genomics analyses

6.11.1 DNA isolation

Genomic DNA of the 10 *K. pneumoniae* strains was extracted from overnight cultures with the Genomic Mini kit (A&A Biotechnology, Gdynia, Poland), according to the manufacturer's protocol in the Laboratory of Microbial Immunochemistry and Vaccines, Institute of Immunology and Experimental Therapy (Wrocław, Poland). DNA from 3936/19 strain was isolated in the Department of Molecular Microbiology, National Medicines Institute (Warsaw, Poland).

6.11.2 Whole-genome sequencing

Libraries preparation and sequencing of DNA fragments was ordered at Genomed sequencing laboratory (Warsaw, Poland).

6.11.3 Genomic sequence analysis

Obtained reads from ABC122 and 3936/19 isolates were trimmed by Cutadapt 1.16 (Martin 2011) in the Department of Molecular Microbiology, National Medicines Institute (Warsaw, Poland). Reads from all other 9 isolates were trimmed by Trimmomatic v0.39 (Bolger, Lohse, and Usadel 2014) in the Laboratory of Genomics and Bioinformatics, Institute of Immunology and Experimental Therapy (Wrocław, Poland). Genomes of all strains were assembled using Spades 3.11.1. (Bankevich *et al.* 2012), and analysed by the Kaptive Web algorithm (<https://github.com/katholt/Kaptive>) for the *in silico* O serotyping (Lam *et al.* 2021, 2022).

All the strains were typed by 7-loci multi-locus sequence typing (MLST) with Multi-Locus Sequence Typing 2.0 (<http://cge.cbs.dtu.dk/services/MLST>) (Larsen *et al.* 2012).

Sequences of the *rfb* gene clusters of the ABC122 (GenBank accession number: GCA_031734875.1) and 3936/19 (GenBank accession number: GCA_030912885.1) isolates were compared with the *K. pneumoniae* OL101 reference strain (GenBank accession number: LT174596.1). Sequences of the *rfb* gene clusters of the Kp164, Kp165, Kp166 sequences were compared with the *K. pneumoniae* O4 reference strain (GenBank accession number: LT174605). Sequences of the *rfb* gene clusters of the Kp174, Kp177 isolates were compared with the *K. pneumoniae* O1v1 reference strain (GenBank accession number: LT174601). Sequences of the *rfb* gene clusters of the BIDMC 7B (GenBank accession number: JCNG000000000.1) and ABC152 (GenBank accession number: JACENF000000000) were compared with those of the reference strains *K. pneumoniae* NTUH-K2044 (serotype O1v2) and *K. pneumoniae* 441 (O1/O2v2) (GenBank accession numbers, AB117611 and LT174602, respectively) using the CLC Main Workbench version 20 software (<https://digitalinsights.qiagen.com>; Qiagen, Hilden, Germany).

6.11.4 *In silico* studies for O loci prevalence among publicly available *K. pneumoniae* genomes

All *K. pneumoniae* whole-genome sequences available from GenBank (n = 8130, on December 3, 2019) were downloaded and screened using Kaptive for the O1v2 and O2v2 serotype predicted isolates and for the *rfb* region (O locus) length discrepancies (excess of ≥ 700 bp and ≥ 400 bp, respectively). Identification of ISs was performed by ISfinder (<http://www-is.biotoul.fr>) (Siguier *et al.* 2006). Subsequently, all putative O1v2 and O2v2 genomes with *rfb* length discrepancies were aligned against the genome of ABC152 strain, inferring SNP-based phylogeny with Parsnp v.1.2. (Treangen *et al.* 2014). The phylogenetic tree was visualized with iTOL (<https://itol.embl.de>) (Letunic and Bork 2019) in Department of Molecular Microbiology, National Medicines Institute (Warsaw, Poland).

Moreover, all *K. pneumoniae* genomes available at GenBank (n = 71377, on July 27, 2023) were downloaded and screened for the OL101 predicted serotype using Kaptive.

6.11.5 Data availability

The sequence of the BIDMC 7B strain is available in GenBank as the whole genome shotgun sequencing project under accession number JCNG000000000.1. The sequence of the ABC152 strain was submitted to the GenBank database as the whole genome shotgun sequencing project under accession number JACENF000000000. The sequences of the strains is available in GenBank under accession numbers GCA_031734875.1 (ABC122 strain) and GCA_030912885.1 (3936/19 strain). The data have been deposited with links to BioProject accession number PRNJA1004904 (ABC122 strain) and PRJNA1005055 (3936/19 strain) in the NCBI BioProject database (<https://www.ncbi.nlm.nih.gov/bioproject/>).

6.12 Software used in this study

- Biorender – <https://app.biorender.com/> (illustration software)
- CLC Main Workbench version 20 – Qiagen, Hilden, Germany (sequences alignment)
- Inkscape 1.2 – Inkscape Team (graphics software)
- Kaptive – <https://kaptive-web.erc.monash.edu/> (O serotype prediction)
- MALDI Biotyper software – Bruker, Billerica, MA, USA (species prediction)
- NMRFAM-SPARKY v1.2 – NMRFAM, Madison, WI, USA (NMR spectra interpretation)
- TopSpin – Bruker, Billerica, MA, USA (NMR spectra acquisition and processing)
- Xcalibur – Agilent Technologies, Santa Clara, USA (GC-MS data)

7. Results

7.1 Isolation and preliminary analysis of lipopolysaccharides

The project has started with 10 nontypeable *K. pneumoniae* strains and has been supplemented with additional strain 3936/19 (See section 7.3). Lipopolysaccharides were isolated by phenol-water extraction described by Westphal and Jann (Westphal and Jann 1965). All LPS were derived from the aqueous phase. The phenol phase did not contain LPS (trace amounts detected only for a few strains). The list of LPSs and extraction yields are shown in Table 3.

Table 3. Efficiency of LPS isolation from dry bacterial mass^{a)}

Strain symbol	LPS yield
BIDMC 7B	0,32%
ABC152	0,70%
ABC122	0,06%
BC738	0,23%
BC13-986	3,38%
Kp164	0,19%
Kp165	0,35%
Kp166	0,28%
Kp174	0,64%
Kp177	0,56%

^{a)} The LPS yields were calculated relative to the dry bacterial mass (0,5 – 7,5 g) and represented average values for all preparations (2-4 extractions). LPSs having identical OPS structure based on ¹H HR-MAS NMR spectra (Figure 9) are marked by different colour and were assigned to 4 groups.

The ¹H HR-MAS NMR spectroscopy was used for the preliminary structural analysis of LPSs. The spectra of selected LPSs were compared with the spectra of available *K. pneumoniae* LPSs of known O serotypes and their subtypes (data of the Laboratory of Microbial Immunochemistry and Vaccines), which allowed to assign analysed LPSs to 4 groups (Figure 9). The O antigen library contained spectra of 11 O antigens and was devoid of O2aeh spectrum only. For smooth LPS, the HR-MAS NMR spectroscopy enables to observe OPS signals within the native LPS molecule and signals from core OS and lipid A are not observed or have very low intensities. Comparison of ¹H NMR spectra of LPSs showed that no selected LPS represented known *K. pneumoniae* O antigen used for comparison except BIDMC 7B and ABC152 (O2v1), and some had identical OPS structures. As the result of this preliminary analysis selected

LPSs formed four groups considering the OPS structure: i) group 1 (strains BIDMC 7B, ABC152), ii) group 2 (strains ABC122, BC738, BC13-986), iii) group 3 (strains Kp164, Kp165, Kp166); and iv) group 4 (strains Kp174, Kp177) (Table 3).

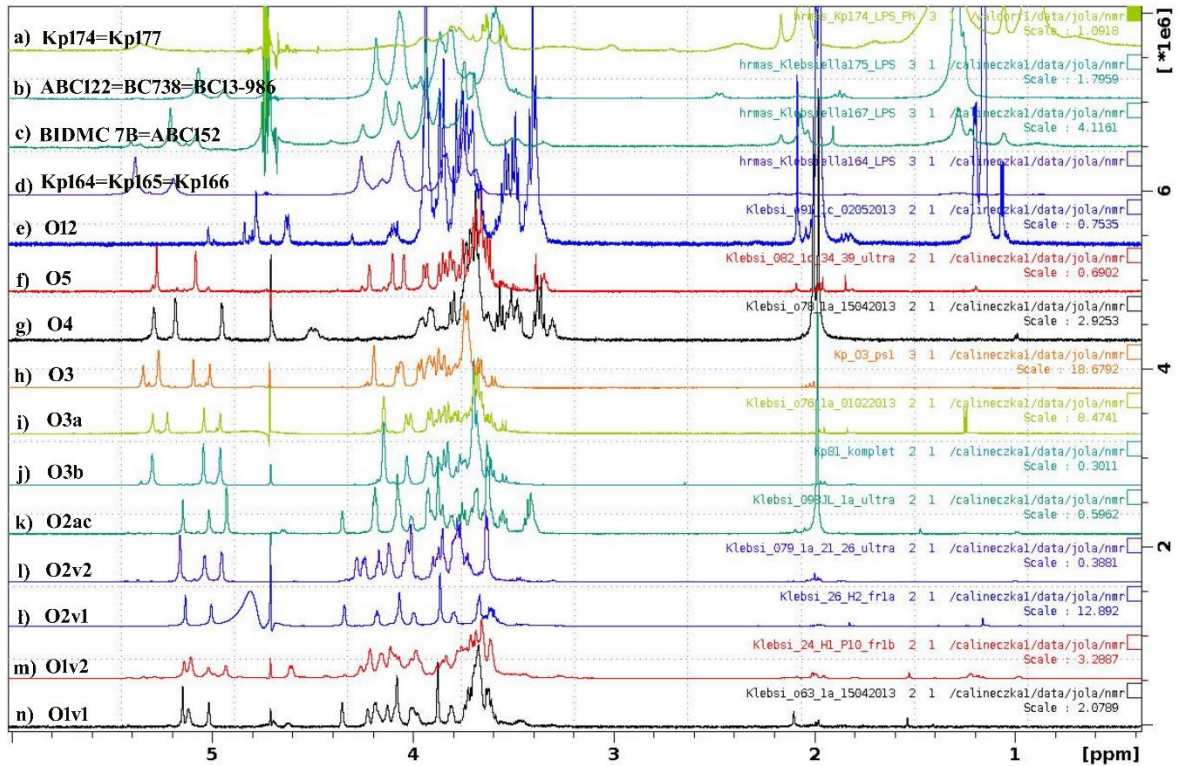


Figure 9. ^1H NMR spectra of *K. pneumoniae* O antigens. Spectra marked as O1-O12 were collected for OPS and available in the Laboratory of Microbial immunochemistry and Vaccines. Spectra collected for native LPS of 10 nontypeable strains by HR-MAS NMR (a-d).

Most of studied LPS samples (at least one representative per group) were analysed by SDS-PAGE and silver stained (Figure 10). Most profiles of LPSs showed the presence of a low molecular weight fraction of lipid A substituted with a core OS and higher molecular weight LPS fractions with different length O-specific chains composed of RUs. Results indicated the presence of smooth LPS containing OPS RUs (Figure 10).

The structural analysis of *K. pneumoniae* O antigens is effectively supported by molecular biology and bioinformatics tools regarding O locus prediction. One such tool is Kaptive (<https://kaptive-web.erc.monash.edu/>), analysing clusters of genes (O loci) specifying individual O serotypes determined by the corresponding OPS structures. DNA was isolated from all 10 nontypeable *K. pneumoniae* isolates and sequenced. The contigs obtained after assembly were analysed by Kaptive. Based on analyses of the presence of genes characteristic for different O loci, a predicted O serotype was assigned to each

isolate. Results of Kaptive-based analyses including interpretation of SDS-PAGE profiles of LPSs, are presented in Table 4.

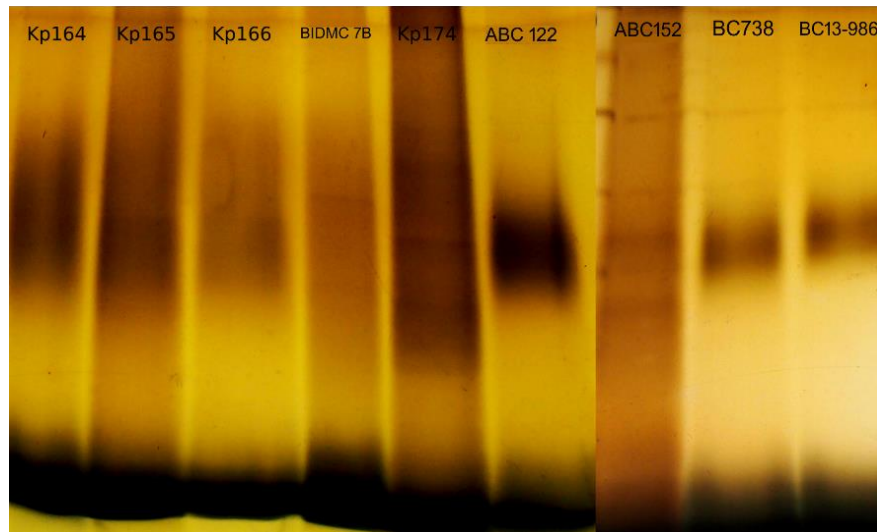


Figure 10. LPS profiles obtained by SDS-PAGE analysis of 9 nontypeable *K. pneumoniae* isolates.

Table 4. Summary of the O serotype analyses for ten nontypeable *K. pneumoniae* isolates^{a)}.

Strain	Sequence Type (ST)	LPS profile (SDS-PAGE)	O locus according to Kaptive
BIDMC 7B	ST258	S-LPS	O2v2?*
ABC152	ST147	S-LPS	O2v2?*
ABC122	ST11	S-LPS	OL101*
BC738	ST485		OL101?*
BC13-986	ST3658		OL101?*
Kp164	ST37	S-LPS	O4
Kp165			
Kp166			
Kp174	ST14	R-LPS	O1v1?-
Kp177	ST14	R-LPS	O1v1?-

^{a)} S-LPS – smooth LPS containing O-specific polysaccharide (OPS); R-LPS – rough LPS devoid of OPS. Smooth or rough character of LPS was analysed by silver stained SDS-PAGE. Colour marks the group of isolates characterised by identical O serotype according to NMR data (Figure 9). The “*” character indicates that one or more of the expected genes within the O locus falls below the identity threshold (default 95%). The “?” character indicates that the O locus sequence was not found in a single piece in the assembly. The “-” character indicates that one of the expected genes was not found in the gene BLAST search.

O loci of strains BIDMC 7B and ABC152 were predicted as O2v2?* and results suggested some differences in O loci in comparison with reference O2 locus for Kaptive. Strains ABC122, BC738 and BC13-986 were predicted as OL101*, OL101?*, and OL101?*, respectively. Strains Kp164, Kp165, and Kp166 were predicted as known O4

serotype. Strains Kp174 and Kp177 were predicted as O1v1?- (Table 4). According to the Kaptive manual the “*” symbol indicates that one or more of the expected O *locus* genes falls below the identity threshold (default 95%). The “?” character indicates that the O *locus* sequence was not found in a single piece in the assembly. The “-” character indicates that one of the expected genes was not found in the gene BLAST search.

Additionally, Multi-Locus Sequence Typing (MLST) analyses were performed for each isolate sequence (Table 4). The process involves sequencing highly conserved seven housekeeping genes (*rpoB*, *gapA*, *mdh*, *pgi*, *phoE*, *infB*, *tonB*) (Diancourt *et al.* 2005). The obtained sequences were compared with previously described sequences (alleles) of a given *locus* on the Centre for Genomic Epidemiology (CGE) server and the allele number was assigned, and then the appropriate sequence type (ST) was assigned (Table 4).

7.2 O antigen and O loci analysis of *K. pneumoniae* BIDMC 7B and ABC152

7.2.1 Structural analysis of BIDMC 7B and ABC152 O antigens

Lipopolysaccharides were extracted from bacteria and O antigens chemical structures of *K. pneumoniae* BIDMC 7B and ABC152 clinical isolates were characterized by ¹H, ¹³C HR-MAS NMR spectroscopy of native LPSs (Figure 11a, b, e).

Proton and carbon signals of the OPS region of studied LPS predominated in the NMR spectra. The ¹H NMR spectra were recorded for BIDMC 7B (Figure 11a) and ABC152 (Figure 11b) indicated identity of O antigens structures. The complete assignments of ¹H and ¹³C resonances for both OPS (Table 5) were performed by interpretation of the ¹H, ¹³C HSQC-DEPT NMR spectra (Figure 11e, the spectrum of BIDMC 7B LPS), including comparison of the spectrum with spectra and data for the reference OPS structures of *K. pneumoniae* Kp26 (O2v1) (Figure 11c) and PCM27 (O2v2) isolates (Figure 11d) (Szijártó *et al.* 2016; Artyszuk *et al.* 2020).

The phenotype analysis supported by NMR spectroscopy demonstrated that the OPS of the BIDMC 7B LPS had an O2v1 serotype structure identical with Kp26 OPS (Szijártó *et al.* 2016) characterised by the [→3)-β-D-Galp-(1→3)-α-D-Galp-(1→] disaccharide as the non-modified D-galactan I RU (Figure 11a, e, inset structures; Table 5). The same D-galactan I structure was determined for the ABC152 LPS (Figure 11b; Table 5). The lack of a terminal α-D-Galp residue, characteristic for the O2v2 serotype

represented by PCM27 LPS (Figure 11d, inset structure, C1 signal), resulted in O2v1 phenotype was confirmed for both isolates in contradiction with Kaptive predictions for O locus (O2v2). However, Kaptive indicated some undefined differences between analysed and reference O loci by assigning predicted O loci as O2v2?* (Artyszuk *et al.* 2020).

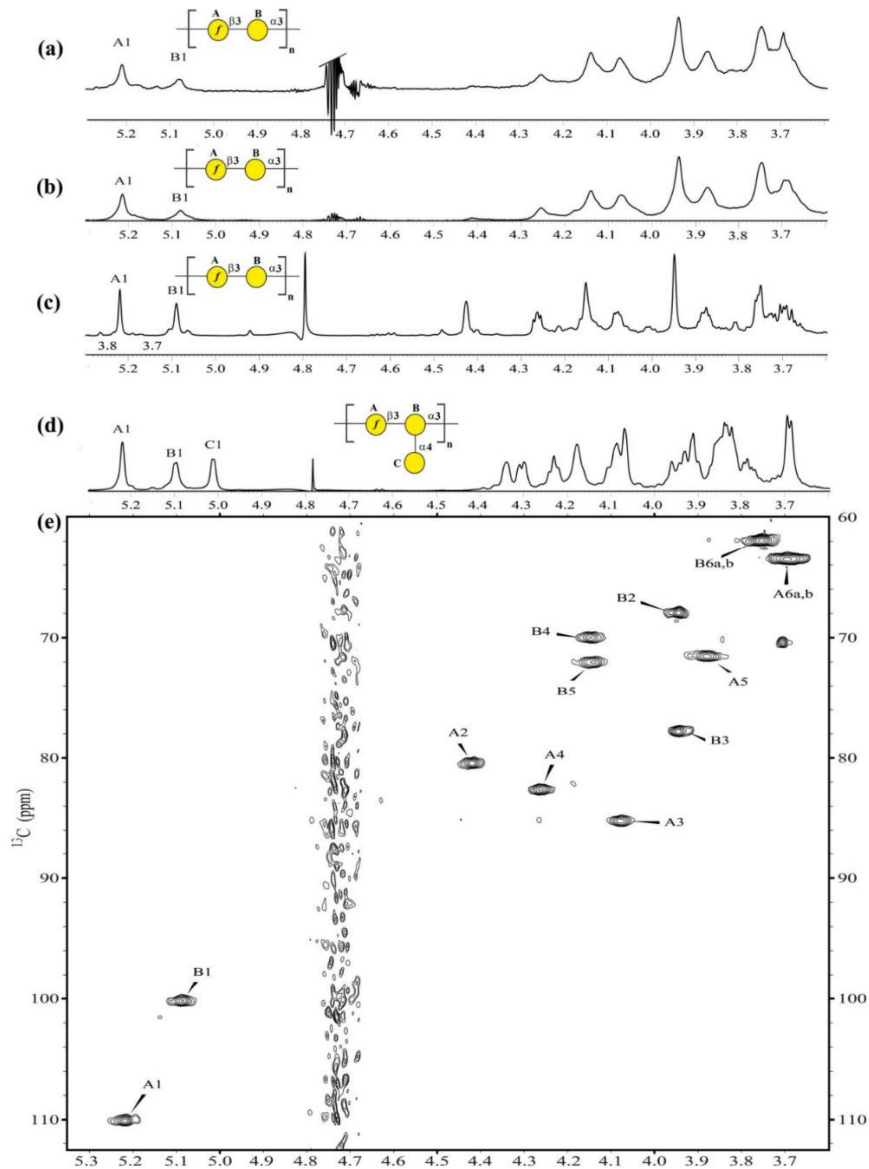




Figure 11. Comparative NMR analysis of *K. pneumoniae* O2v2 and O2v1 O antigens. ¹H HR-MAS NMR spectra of the lipopolysaccharides from (a) BIDMC 7B and (b) ABC152; and (c) the O-specific polysaccharides of *K. pneumoniae* Kp26 strain (O2v1) and (d) the PCM27 strain (O2v2); (e) ¹H,¹³C HSQC-DEPT NMR spectrum of the BIDMC 7B LPS (O2v1). The capital letters refer to protons and/or carbons of OPS carbohydrate residues, as shown in Table 6 (Artyszuk *et al.* 2020). The Symbol Nomenclature for Graphical Representation of Glycans is used for OPS visual representation:  Galactose;  Galactofuranose to show O-specific polysaccharide repeating units: $\rightarrow 3$)- β -D-Galp-(1 \rightarrow 3)- α -D-Galp-(1 \rightarrow (D-galactan I, O2 variant 1) and $\rightarrow 3$)- β -D-Galp-(1 \rightarrow 3)-[α -D-Galp-(1 \rightarrow 4)]- α -D-Galp-(1 \rightarrow (O2v2) (Varki *et al.* 2015).

Table 5. ¹H and ¹³C NMR chemical shifts of O antigens isolated from *K. pneumoniae* BIDMC 7B and ABC152 (Artyszuk *et al.* 2020)^{a)}.

Chemical shift (ppm)								
Strain	Residue	Description	H1	H2	H3	H4	H5	H6a,b
			CI	C2	C3	C4	C5	C6
BIDMC 7B	A	→3)-β-D-Galf-(1→	5.22 110.1	4.42 80.5	4.07 85.2	4.26 82.6	3.88 71.5	3.69 ^{nr} 63.5
	B	→3)-α-D-Galp-(1→	5.09 100.2	3.94 67.8	3.94 77.8	4.14 70.1	4.14 72.0	3.75 ^{nr} 61.9
ABC152	A	→3)-β-D-Galf-(1→	5.22 110.0	4.42 80.4	4.07 85.2	4.26 82.6	3.88 71.5	3.69 ^{nr} 63.5
	B	→3)-α-D-Galp-(1→	5.09 100.1	3.94 67.9	3.94 77.8	4.14 70.0	4.14 72.0	3.75 ^{nr} 61.9

^{a)} nr – not resolved; H1,C1=175 Hz and =170 Hz for residue A and B, respectively

7.2.2 Bioinformatics analysis of BIDMC 7B and ABC152 O loci

Kaptive-based O serotyping was performed with whole-genome sequences of both strains (Table 4). Contrary to the structural analysis, the O serotypes were predicted to be O2v2? with high match confidence, according to the Kaptive measures of match quality (Figure 12). However, the BIDMC 7B and ABC152 *rfb* clusters (O loci) demonstrated an increased size (by 777 bp) when compared to the reference sequences of O2 locus in Kaptive database. The *gmlABC* genes showed 100% (*gmlA*), 90,79% (*gmlB*), and 97,33% (*gmlC*) identity, respectively, to those in the Kaptive reference database for O1/O2v2 locus (Figure 12) (Artyszuk *et al.* 2020).

The high identity was detected for other genes in O locus for BIDM 7B strain: *wzm* – 100%, *wzt* – 99,19%, *wbbM* – 98,58%, *glf* – 98,70%, *wbbN* – 98,99%, whereas *wbbO* and *kfoC* genes showed 99,47% and 99,01% identity, respectively, to the reference sequence.

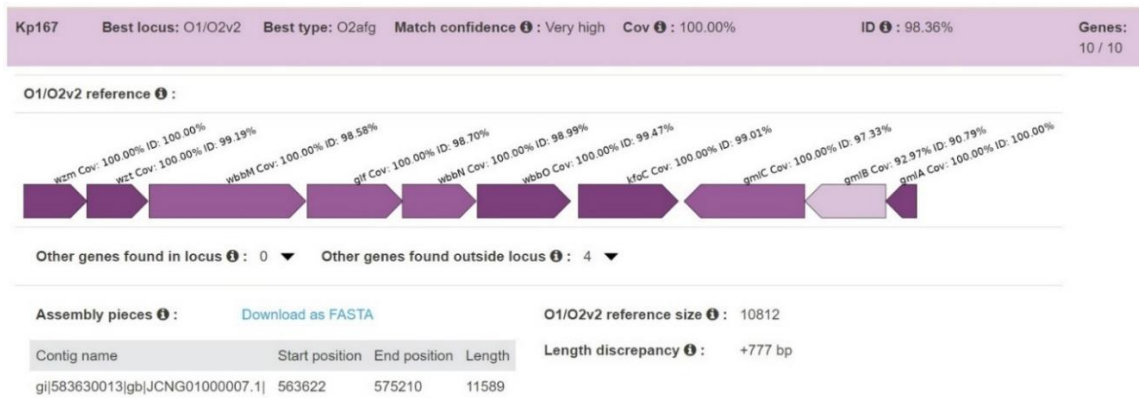


Figure 12. Screenshot with Kaptive Web results for *K. pneumoniae* BIDMC 7B O locus prediction.

Molecular analysis of BIDMC 7B and ABC152 was performed to explain the discrepancies observed between the O antigen phenotype (O2v1) and the Kaptive-predicted O serotype (O2v2). The alignment of the *gmlB* genes of BIDMC 7B, ABC152, and two O1/O2v2 reference *K. pneumoniae* NTUH-K2044 and 441 strains are shown in Figure 13. This comparison shows the disruption of *gmlB* genes in both strains by an identical insertion sequence (IS) element (777 bp), whereas other genes in the *rfb* locus were intact, in comparison to the reference O loci. The IS element was identified as the ISR1 by ISfinder (section 6.11.4). These results indicated that the ISR1 disruption completely inactivated the gene encoding GmlB glycosyltransferase, resulting in biosynthesis of the O2v1 (devoid of terminal-Galp residue) instead of O2v2 structure, and thus, being the likely reason for the discrepancy between the actual O antigen phenotype and Kaptive-based predictions (Artyszuk *et al.* 2020). The *gmlB* gene is a part of a three-component cassette *gmlABC*, resembling other systems that direct the periplasmic modification of glycan backbones in a variety of contexts, including the addition of glucosyl side groups to OPSs such terminal α -Galp in case of the O2v2 antigen (Szijártó *et al.* 2014; Clarke *et al.* 2018).

gmlB BIDMC 7B	ATGACAACCT	CAACTGATAT	AAAAAGCACT	CCITCTTTAG	CTATTGTGGT	ACCTTGCTAT	AATGAACAAG	AGGCTTTTCC	TTTCTGTCTC	GAAAAGCTTT	100
gmlB ABC152	100
gmlB NTUH-K2044	100
gmlB 441	100
Conservation	100
gmlB BIDMC 7B	CGAATGTACT	AAATTCATTG	ATAGCAAGAA	ATAAAATTA	CAACAATAGT	TATCTTTTGT	TTGTCGATGA	TGGTAGTCGT	GACAATACTT	GGGCACAAAT	200
gmlB ABC152	200
gmlB NTUH-K2044	200
gmlB 441	200
Conservation	200
gmlB BIDMC 7B	IAAAGAIGCC	TCGACCGCII	ATCACIAIGT	GCGAGGAAIA	AAATTAICAA	GAAAIAAAGG	ACAIFCAAAT	GCCITAAIGG	CAGGGTIACG	CICGGICGAI	300
gmlB ABC152	300
gmlB NTUH-K2044	300
gmlB 441	300
Conservation	300
gmlB BIDMC 7B	ACAGACGTAA	CCATTAGCAT	CGATGCGGAT	CTACAAGATG	ATGTAATTTG	CATCGAAAAA	ATGATTGACG	CTTACAGCCA	AGGATATGAC	ATAGTATACG	400
gmlB ABC152	400
gmlB NTUH-K2044	400
gmlB 441	400
Conservation	400
gmlB BIDMC 7B	GCGTAAGATG	TAATCGAGAG	AGTGACACTT	TTTTAAACG	TACAACAGCT	AATGCATTTT	ACGCAATAAT	GTCCACATG	GGTGTAAATC	AAACTCCAAA	500
gmlB ABC152	500
gmlB NTUH-K2044	500
gmlB 441	500
Conservation	500
gmlB BIDMC 7B	TCATGCAGAT	TATCGATTAT	TAAGTAATCG	TGCGTTGGAG	GCTCTFAAAC	AATATAAAGA	GCAAATATA	TATTTACGTG	GATTAGTGCC	TCTTGTGGGA	600
gmlB ABC152	600
gmlB NTUH-K2044	600
gmlB 441	600
Conservation	600
gmlB BIDMC 7B	TATCCCTCGA	TCGAGGTGCA	ATATAGCCGT	GAAGAAAGAA	TTGCAGGTGA	ATCAAAATAT	CCAATTAATA	AAATGCTTGC	GCTGGCTCTC	GAGGGAAATA	700
gmlB ABC152	700
gmlB NTUH-K2044	700
gmlB 441	700
Conservation	700
gmlB BIDMC 7B	CCTCATTATC	AGTTACACCG	TTACGAATTA	TAGCTATGAC	AGGTTTTATA	ACTTGCATCA	TATCTACCAT	CGCTGCGATT	TATGCTTTAA	TTCAAAAAAC	800
gmlB ABC152	800
gmlB NTUH-K2044	800
gmlB 441	800
Conservation	800
gmlB BIDMC 7B	AACAGGTAAC	ACAGTGGGTG	GATGCTGCCA	ACTTACTGAT	TAGTGTATG	ATGGTGTIT	TGAGGTGCTC	CAGTGGCTTC	TGTTTCTATC	AGCTGTCCCT	900
gmlB ABC152	900
gmlB NTUH-K2044	900
gmlB 441	900
Conservation	900
gmlB BIDMC 7B	CCTGTTTACG	TACTGACGGG	GTGGTGCCTA	ACGGCAAAAG	CACCGCCGGA	CATCAGCGCT	ATCTCTGCTC	TCACTGCCGT	AAAACATGGC	AACTGACGTT	1000
gmlB ABC152	1000
gmlB NTUH-K2044	1000
gmlB 441	1000
Conservation	1000
gmlB BIDMC 7B	CACTTACACC	GCTTCTCAAC	CCGGTACGCA	CCAGAAATC	ATTGATATGG	CCATGAATGG	CGTTGGATGC	CGGGCAACCG	CCCACATTAT	GGGCGTTGGC	1100
gmlB ABC152	1100
gmlB NTUH-K2044	1100
gmlB 441	1100
Conservation	1100
gmlB BIDMC 7B	CTCAACACGA	TTTTCCGCCA	TTAAAAAAC	TCAGGCGCCA	GTGGTAAACC	TCGGCATAAC	AGCCGGGCAG	TGACGTCATC	GTCTGCGCGG	AAATGGACGA	1200
gmlB ABC152	1200
gmlB NTUH-K2044	1200
gmlB 441	1200
Conservation	1200
gmlB BIDMC 7B	ACAGTGGGGA	TACGTGCGGG	CTAAATCGCG	CCAGCGCTGG	CTGTTTACG	CGTATGACAG	GCTCCGGGAG	ACGGTTGTTG	CGCACGTATT	CGGTGAACGC	1300
gmlB ABC152	1300
gmlB NTUH-K2044	1300
gmlB 441	1300
Conservation	1300
gmlB BIDMC 7B	ACTATGGCGA	CGCTGGGGCG	TCTTATGAGC	CTGCTGTAC	CCTTTGCAGT	GGTGATATGG	ATGACGGATG	GCTGGCCGCT	GTATGAATCC	CGCTGAAGG	1400
gmlB ABC152	1400
gmlB NTUH-K2044	1400
gmlB 441	1400
Conservation	1400
gmlB BIDMC 7B	GAAAGCTGCA	CGTAATCAGC	AAGCGATATA	CGACGCAAT	TGAGCGGCAT	AACCTGAATC	TGAGGCAGCA	CCTGGCACGG	CTGGGACGGA	AGTCGCTGTC	1500
gmlB ABC152	1500
gmlB NTUH-K2044	1500
gmlB 441	1500
Conservation	1500
gmlB BIDMC 7B	GTTCTCAAAA	TCGGTGGAGC	TGCATGACAA	AGTCATCGGG	CATTATCTGA	ACATAAAACA	CTATCAATAA	GTTGGAGTCA	TTACCCTACA	GTTGAGGGAT	1600
gmlB ABC152	1600
gmlB NTUH-K2044	1600
gmlB 441	1600
Conservation	1600
gmlB BIDMC 7B	GGACATCAGT	CATGATCGGT	ATATTCTTTC	TTGGCGGTGT	GCAAAATGCTT	TCTTTAGGTA	TTATAGGAGA	ATATGTCGGG	AAAATTTATA	TAGAGACGAA	1700
gmlB ABC152	1700
gmlB NTUH-K2044	1700
gmlB 441	1700
Conservation	1700
gmlB BIDMC 7B	AAATAGACCT	AAATAITTTCA	TTGATGAAAG	CGTAGGTAAT	GATAGCAATG	GAAAAATA	1758
gmlB ABC152	1758
gmlB NTUH-K2044	981
gmlB 441	981
Conservation	981

Figure 13. The alignment of the *gmlB* genes of *K. pneumoniae* BIDMC 7B, ABC152 with *gmlB* genes of two reference strains (NTUH-K2044 and 441). Grey areas mark regions of differences in nucleotide sequence (777 bp). The alignment was performed using the CLC Main Workbench. Symbols Kp181 and Kp167 stand for ABC152 and BIDMC 7B sequences, respectively (Artyszuk *et al.* 2020).

7.2.3 Insertion sequences occurrence in *K. pneumoniae* O2v2 and O1v2 loci — *in silico* study of publicly available genomes

In order to assess the occurrence of IS elements in *rfb* loci of *K. pneumoniae* O1 and O2, 8130 genome sequences available in the public domain (December 3, 2019) were analysed (Appendix, Table S1, available on-line at <https://www.mdpi.com/1422-0067/21/18/6572/s1>).

Based on the Kaptive results of the O serotyping (Appendix, Table S1, column B, see link to the supplementary data file <https://www.mdpi.com/1422-0067/21/18/6572/s1>), 2281 isolates (~28%) were predicted to be O2v2, and 839 isolates (~10%) to be O1v2. For O2v2, 55 genomes (~2,40%) revealed a significant difference in length of the *rfb* region (≥ 400 bp), of which 49 genomes were of sufficient quality for further analysis (Table 6). The presence of different ISs (e.g., ISR1, IS903B, ISKpn14, or ISKpn26) were identified in several genes of these loci; namely, *gmlBC*, *kfoC*, *wbbMNO*, *glf*, *wzm*, and *wzt*. In several isolates, the same or two different ISs interrupted two genes; namely, *gmlB* or *gmlC* and *wbbO*, or *wbbM* and *wbbO* (Table 6) (Artyszuk *et al.* 2020). The functions of genes were outlined in accordance with the current understanding in Section 4.2.6.3.2.

As with O2 serotype, O1 antigen is built of galactan I available in two versions, v2 and v1 encoded by reference O1/O2 loci (reviewed in Section 4.2.6.3.2), the screening for IS sequences was also performed for O1v2-predicted genomes. Among the 839 Kaptive-identified O1v2 isolates out of 8130 genomes, significant length discrepancies (≥ 700 bp) occurred in the *rfb* region of six isolates (~0,7%), one of which was excluded due to the low quality of reads. Selected *rfb* genes of these isolates, namely *gmlABC*, *wbbM*, and *wzm*, were interrupted by IS5, IS102, IS903B, or ISKpn14 (Table 7). For ASM275277 and ASM275281 isolates, and for ASM296687 isolate, two and four genes, respectively, were disrupted simultaneously (Artyszuk *et al.* 2020).

Table 6. Kaptive analysis results extracted for 49 *K. pneumoniae* O2 variant 2-predicted genomes characterised by IS occurrence (Artyszuk *et al.* 2020)^a.

Isolate	Assembly accession number	Sequence type	Insertion sequence	Gene	Position of the IS element (from the first nucleotide of CDS)
BIDMC 7A	GCF_000567825.1_Kleb_pneu_BIDMC_7A_V2_genomic	ST258	IS1R	<i>gmlB</i>	818 bp
BIDMC 7B	GCF_000567425.1_Kleb_pneu_BIDMC_7B_V2_genomic	ST258	IS1R	<i>gmlB</i>	818 bp
BIDMC 42a	GCF_000567205.1_Kleb_pneu_BIDMC_42a_V1_genomic	ST258	IS1R	<i>gmlB</i>	818 bp
BIDMC 42b	GCF_000567185.1_Kleb_pneu_BIDMC_42b_V1_genomic	ST258	IS1R	<i>gmlB</i>	818 bp
BWH 36	GCF_000567565.1_Kleb_pneu_BWH_36_V1_genomic	ST258	IS1R	<i>gmlB</i>	818 bp
UC175	GCF_001034165.1_Kleb_pneu_UC175_V1_genomic	ST258	IS1R	<i>kfoC</i>	656 bp
UC181	GCF_001034215.1_Kleb_pneu_UC181_V1_genomic	ST258	IS1R	<i>kfoC</i>	656 bp
UC191	GCF_001034245.1_Kleb_pneu_UC191_V1_genomic	ST258	IS1R	<i>kfoC</i>	656 bp
UC193	GCF_001034295.1_Kleb_pneu_UC193_V1_genomic	ST258	IS1R	<i>kfoC</i>	656 bp
UC133	GCF_000566865.1_Kleb_pneu_UC1_33_V1_genomic	ST258	IS1R	<i>kfoC</i>	656 bp
CHS163	GCF_001032195.1_Kleb_pneu_CHS163_V1_genomic	ST258	IS1R	<i>wbbM</i>	1881 bp
CHS139	GCF_001031785.1_Kleb_pneu_CHS139_V1_genomic	ST258	IS1R	<i>wbbM</i>	1881 bp
CHS91	GCF_001030945.1_Kleb_pneu_CHS91_V1_genomic	ST258	IS1R	<i>wbbO</i>	165 bp
CHS134	GCF_001031705.1_Kleb_pneu_CHS134_V1_genomic	ST258	ISKpn26	<i>gmlB</i>	4 bp
CHS57	GCF_000694075.1_Kleb_pneu_CHS_57_V1_genomic	ST258	ISKpn26	<i>gmlB</i>	4 bp
UC138	GCF_000566805.1_Kleb_pneu_UC1_38_V1_genomic	ST258	ISKpn26	<i>gmlB</i>	453 bp
			IS1294	<i>wbbO</i>	900 bp
BIDMC 16	GCF_000474905.1_Kleb_pneu_BIDMC_16_V1_PacBio_genomic	ST258	ISKpn26	<i>gmlB</i>	453 bp
			ISKpn26	<i>wbbO</i>	490 bp
BIDMC 13	GCF_000567345.1_Kleb_pneu_BIDMC_13_V1_genomic	ST258	ISKpn26	<i>gmlB</i>	453 bp
			ISKpn26	<i>wbbO</i>	490 bp
BIDMC 32	GCF_000567265.1_Kleb_pneu_BIDMC_32_V1_genomic	ST258	ISKpn26	<i>gmlB</i>	453 bp
			ISKpn26	<i>wbbO</i>	490 bp
CHS 71	GCF_000694295.1_Kleb_pneu_CHS_71_V1_genomic	ST258	ISKpn26	<i>wbbO</i>	1080 bp
CHS235	GCF_001033335.1_Kleb_pneu_CHS235_V1_genomic	ST258	ISKpn26	<i>wbbO</i>	641 bp
ASM147162	GCF_001471625.1_ASM147162v2_genomic	ST258	ISKpn26	<i>wbbO</i>	1014bp
			ISKpn26	<i>wbbM</i>	96 bp
CHS105	GCF_001031225.1_Kleb_pneu_CHS105_V1_genomic	ST258	ISKpn26	<i>wbbM</i>	1548 bp
ASM386117	GCF_003861175.1_ASM386117v1_genomic	ST258	ISKpn26	<i>wbbM</i>	45 bp
			ISKpn26	<i>wbbM</i>	2156 bp
CHS165	GCF_001032265.1_Kleb_pneu_CHS165_V1_genomic	ST258	ISKpn26	<i>kfoC</i>	158 bp
MGH51	GCF_000694435.1_Kleb_pneu_MGH_51_V1_genomic	ST258	ISKpn26	<i>wzm</i>	315 bp
CHS99	GCF_001031105.1_Kleb_pneu_CHS99_V1_genomic	ST258	ISKpn26	<i>wzt</i>	470 bp
ASM205647	GCF_002056475.1_ASM205647v1_genomic	ST258	ISKpn14	<i>kfoC</i>	782 bp
ASMS0749	GCF_000807495.1_ASM80749v2_genomic	ST258	ISKpn14	<i>kfoC</i>	440 bp
18174_7_5	GCF_900515885.1_18174_7_5_genomic	ST258	ISKpn14	<i>wbbO</i>	377 bp
BIDMC 54	GCF_000692935.1_Kleb_pneu_BIDMC_54_V1_genomic	ST258	IS1294	<i>kfoC</i>	1186 bp
CHS46	GCF_000693875.1_Kleb_pneu_CHS_46_V1_genomic	ST258	IS1F	<i>wbbM</i>	1138 bp
ASM195291	GCF_001952915.1_ASM195291v1_genomic	ST258	ISVsa5	<i>wbbN</i>	133 bp
27097_7_178-2	GCF_900776535.1_27097_7_178-2_genomic	ST11	IS903	<i>wbbO</i>	228 bp
kpneu028	GCF_900607955.1_kpneu028_genomic	ST11	IS1R	<i>wbbM</i>	47 bp
ASM893134	GCF_008931345.1_ASM893134v1_genomic	ST12	IS903B	<i>kfoC</i>	267 bp
UC175	GCF_000492535.1_Kleb_pneu_UCI_7_V1_genomic	ST17	ISKpn26	<i>wbbO</i>	1014 bp
ASM303004	GCF_003030045.1_ASM303004v1_genomic	ST512	ISKpn26	<i>wbbO</i>	1014 bp
ASM966157	GCF_009661575.1_ASM966157v1_genomic	ST23	IS1R	<i>wbbM</i>	1420 bp
ASM170423	GCF_001704235.1_ASM170423v1_genomic	ST34	IS903B	<i>gmlC</i>	7956 bp in <i>rfb</i> (CDS <i>gmlC</i> from 7996 bp)
ASM187487	GCF_001874875.1_ASM187487v1_genomic	ST34	IS903B	<i>gmlC</i>	7956 bp in <i>rfb</i> (CDS <i>gmlC</i> from 7996 bp)
ASM320446	GCF_003204465.1_ASM320446v1_genomic	ST34	IS903B	<i>gmlC</i>	7956 bp in <i>rfb</i> (CDS <i>gmlC</i> from 7996 bp)
ASM307130	GCF_003071305.1_ASM307130v1_genomic	ST34	IS903B	<i>wbbO</i>	130 bp
			IS903B	<i>gmlC</i>	7956 bp in <i>rfb</i> (CDS <i>gmlC</i> from 7996 bp)
ASM307331	GCF_003073315.1_ASM307331v1_genomic	ST34	IS903B	<i>wbbO</i>	130 bp
			IS903B	<i>gmlC</i>	7956 bp in <i>rfb</i> (CDS <i>gmlC</i> from 7996 bp)
ASM366018	GCF_003660185.1_ASM366018v1_genomic	ST105	IS903B	<i>wbbO</i>	182 bp
BIDMC 55	GCF_000692955.1_Kleb_pneu_BIDMC_55_V1_genomic	ST105	IS903B	<i>glf</i>	105 bp
ABC152	GCA_014433645.1	ST147	IS1R	<i>gmlB</i>	818 bp
UCI 8	GCF_000492515.1_Kleb_pneu_UCI_8_V1_genomic	ST1198	IS9033	<i>wbbM</i>	1667 bp
IS39	GCF_000529425.1_IS39v1_genomic	unknown	IS102	<i>gmlC</i>	7933 bp in <i>rfb</i> (CDS <i>gmlC</i> from 7996 bp)

^a) Data was acquired on December 3, 2019. Each colour indicates closely related strains characterised by genetic similarity according to method SNP-based phylogeny (section 6.11.4) and as presented in Figure 14. Not related isolates were not coloured. CDS – Coding Sequence.

Table 7. Sequence types and location of IS elements in the *rfb* clusters of *K. pneumoniae* isolates selected on the basis of Kaptive-based O1v2 predictions (Artyszuk *et al.* 2020)^a.

Isolate	Assembly accession number	Sequence type	Insertion sequence	Gene	Position of the IS element (from the first nucleotide of CDS)
ASM492431	GCF_004924315.1_ASM492431v1_genomic	ST23	IS102	<i>gmlC</i>	1217 bp
ASM275277	GCF_002752775.1_ASM275277v1_genomic	ST29	IS102	<i>gmlB</i>	472 bp
			IS903B	<i>gmlC</i>	526 bp
ASM275281	GCF_002752815.1_ASM275281v1_genomic	ST29	IS102	<i>gmlB</i>	472 bp
			IS903B	<i>gmlC</i>	526 bp
ASM296687	GCF_002966875.1_ASM296687v1_genomic	ST34	IS5	<i>wzm</i>	315 bp
			IS903B	<i>wbbM</i>	465 bp
			IS903B	<i>gmlB</i>	926 bp
			IS903B	<i>gmlC</i>	143 bp
ASM290977	GCF_002909775.1_ASM290977v2_genomic	ST231	ISKpn14	<i>gmlA</i>	95 bp

^a) Data was acquired on December 3, 2019. Blue colour indicates the same strain characterised by genetic similarity according to method SNP-based phylogeny (section 6.11.4). CDS – Coding Sequence.

In both the O2v2 and O1v2-predicted isolates, the same ISs were observed at the same positions of the same genes in several isolates, suggesting their close genetic relatedness. The *gmlB*:ISR1 (nt 818) disruption of the studied isolates BIDMC 7B and ABC152 (O2v1 phenotype) was found in four other genomes (Table 6).

In order to sort out the approximate number of independent IS insertions into the *rfb* loci of the available *K. pneumoniae* O2v2 and O2v1 genomes, clonality (MLST) and phylogenetic analyses for O2v2 were performed on the isolates using the ABC152 strain as a reference (Tables 6; Figure 14). These confirmed that some individual disruptions within the *rfb* locus have spread in *K. pneumoniae* populations clonally with specific lineages, indicating single IS insertion events at their origins. This was demonstrated by clusters of O2v2 ST258 isolates with *kfoC*:ISR1 (nt 656) or *wbbM*:ISR1 (nt 1881) disruptions, ST258 with double *gmlB*:ISKpn26 (nt 453) plus *wbbO*:ISKpn26 (nt 490) disruptions, or ST34 with *gmlC*:IS903B (nt 7956). In some cases, an additional IS insertion likely marked on-going diversification within a lineage, such as *wbbO*:IS903B (nt 130) in ST34 with *gmlC*:IS903B (nt 7956). An interesting case was the *gmlB*:ISR1 (nt 818) disruption in the study isolates BIDMC 7B and ABC152, which was observed also in four others. BIDMC 7B together with the four others formed a closely related cluster of ST258 isolates. ABC152 was of a non-related ST147, suggesting a horizontal transfer and recombination event. A similar case was represented by the disruption *wbbO*:ISKpn26 (nt 1014), present in two ST258 and ST512 close relatives, as well as a non-related ST17 isolate (Table 6).

Based on these results, it may be assumed that IS disruptions within the *rfb* loci in *K. pneumoniae* O2v2 and O2v1 genomes might have occurred at least ~35 and ~10 times, respectively (Table 6) (Artyszuk *et al.* 2020). To reach definitive conclusions, the O antigen phenotype must be determined by structural analyses to support IS-driven structural modifications.

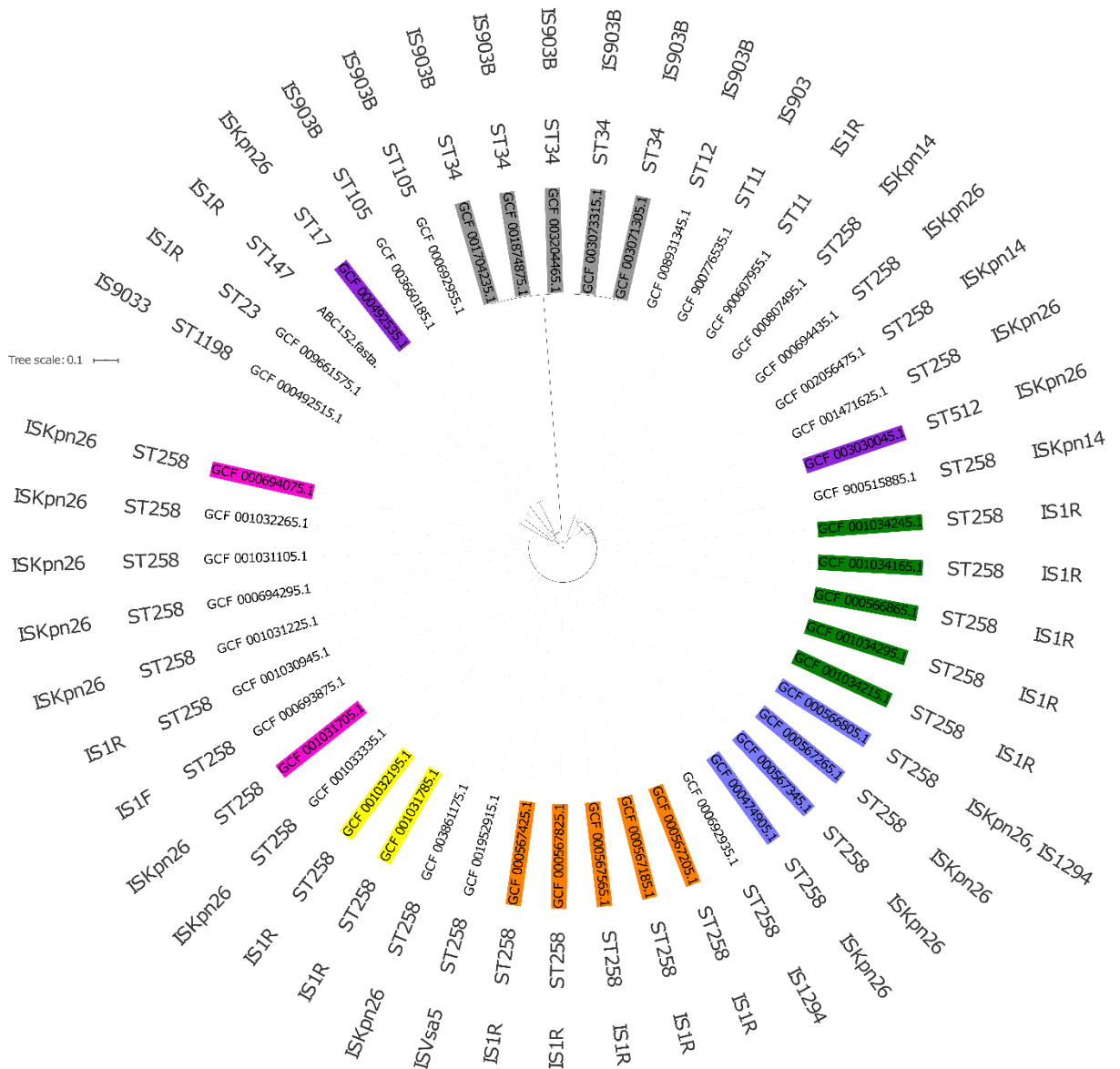


Figure 14. Results of clonality (MLST) and phylogenetic analyses of O2v2 *K. pneumoniae* genomes characterized by ST and insertion sequences distribution. Each colour indicates closely related strains characterised by genetic similarity, shown in detail in Table 6. Not related isolates are not coloured (Artyszuk *et al.* 2020).

7.3 O antigen and O loci analysis of *K. pneumoniae* ABC122, BC738, BC13-986 and 3936/19

7.3.1 Structural analysis of *K. pneumoniae* ABC122, BC738, BC13-986 and 3936/19 O antigens

For the preliminary characterization of the O antigen structure, the OPS regions were analysed directly in isolated LPSs by the HR-MAS NMR spectroscopy. In order to confirm the relationship between OL101 *locus* and chemical structure of selected O antigens, the group of ABC122, BC738 and BC13-986 strains has been supplemented by one more clinical isolate predicted by Kaptive with high match confidence as OL101 (Figure 27c, Table 11). To identify available strain the local database of genomes of *K. pneumoniae* isolates from polish hospitals (the collection of the National Medicines Institute, Warsaw, Poland) were searched for OL101 *loci* by Kaptive and the clinical isolate 3936/19 was found according to the Kaptive measures of match quality as OL101 with high match confidence with 8 genes characteristic for OL101 *locus* (Table 11).

The comparison of the ¹H HR-MAS NMR spectra indicated identity of the O antigen structures for all of the isolates (Figure 15a-d) (Vinogradov *et al.* 2002). Besides anomeric (a1, b1, b'1) and ring protons and carbons, characteristic signals of deoxy protons H3 of the β-3-deoxy-D-manno-oct-2-ulosonic acid (Kdo) residue (k) were identified in each LPS (Figure 15a-d). Considering the published data, the characteristic Kdo signals described above, and the lack of N-acetylated sugars, a novel *K. pneumoniae* O antigen was identified and further analysed to elucidate its chemical structure (Artyszuk *et al.* 2024).

To elucidate the O antigen structure, the LPS of the ABC122 strain was chosen as a representative and hydrolysed with 1,5% acetic acid to obtain poly- and oligosaccharides, followed by fractionation of the resulting hydrolysate on TSKgel®G2500PW column. Gel permeation chromatography yielded four fractions (Figure 16). The ¹H NMR spectroscopy identified fraction 1 as the OPS (Fig. 15e), fraction 2 and 3 as the core oligosaccharide (data not shown), and fraction 4 as degradation products of the hydrolysis of labile regions of the inner core oligosaccharide (Kdo residues; data not shown).

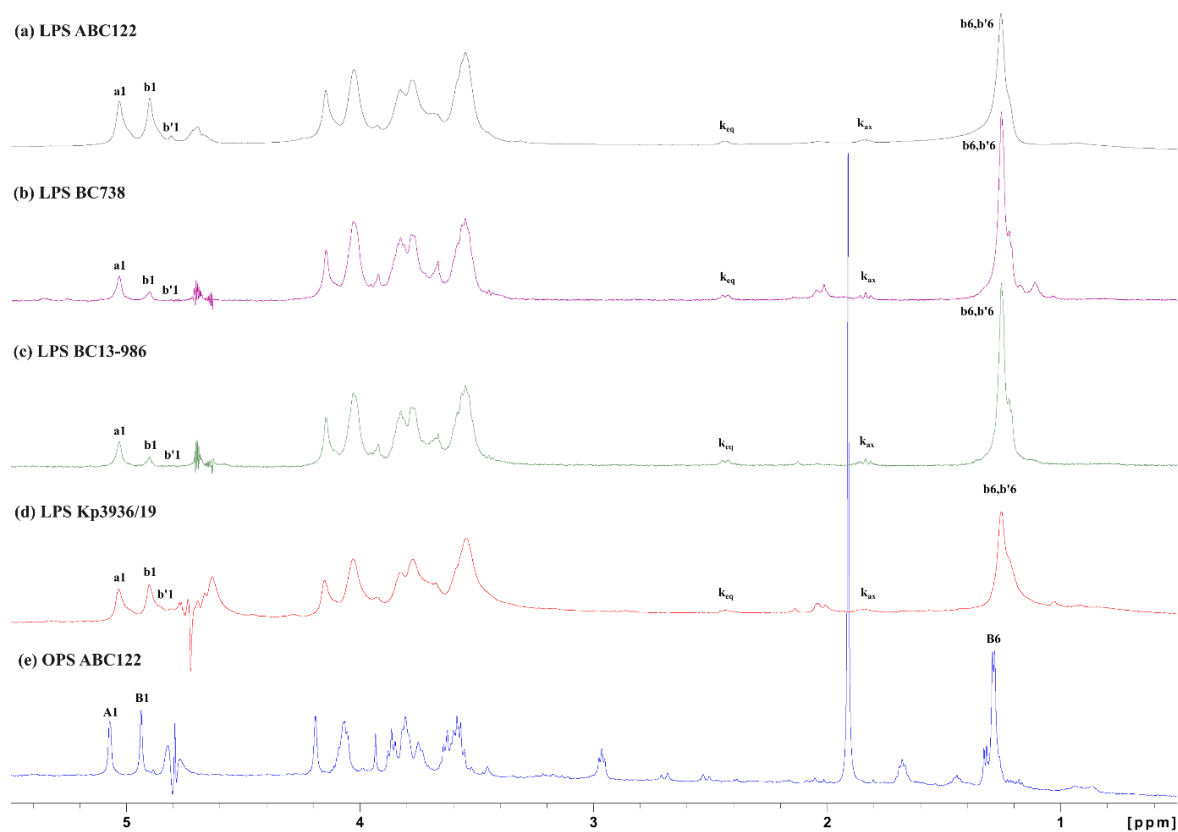


Figure 15. Comparative NMR analysis of the *K. pneumoniae* ABC122, BC738, BC13-986 and 3936/19 O antigens. ^1H HR-MAS NMR spectra of the LPSs isolated from (a) ABC122, (b) BC738, (c) BC13-986, and (d) 3936/19 strains. (e) ^1H NMR spectrum of the O-specific polysaccharide isolated from ABC122 LPS. Capital letters refer to protons of O antigen carbohydrate residues, as shown in Table 8. Letter b' represents a variant of the residue b ($\rightarrow 3$)-L- α -Rhap-(1 \rightarrow) substituted by terminal β -Kdo (Artyszuk *et al.* 2024).

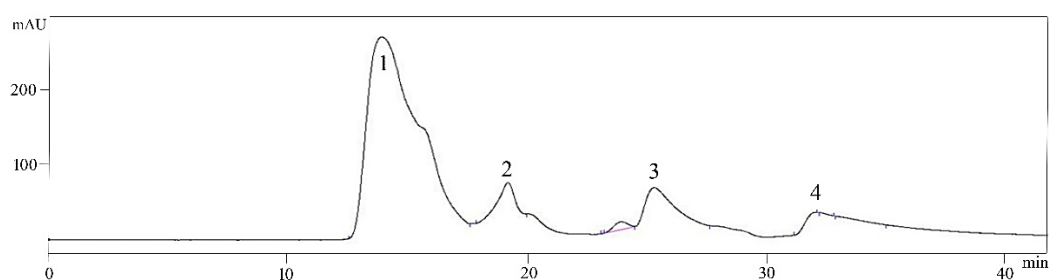


Figure 16. Chromatogram of poly- and oligosaccharides obtained by hydrolysis of ABC122 LPS (60 mg) and fractionated on TSKgel[®]G2500PW column (flow: 1 ml/min). The OPS was identified in fraction 1.

Further, the OPS structure was analysed by sugar and methylation analyses, and by NMR spectroscopy.

The sugar composition of the OPS was determined based on sugar analysis performed using GC-MS. The retention times of the obtained volatile derivatives of sugar

components, alditol acetates, were compared with standards for Glc and Rha (data not shown). Based on the analyses, the presence of Rha ($t_R=9,67$ min) and Glc ($t_R=11,82$ min) was identified in the OPS (Figure 17).

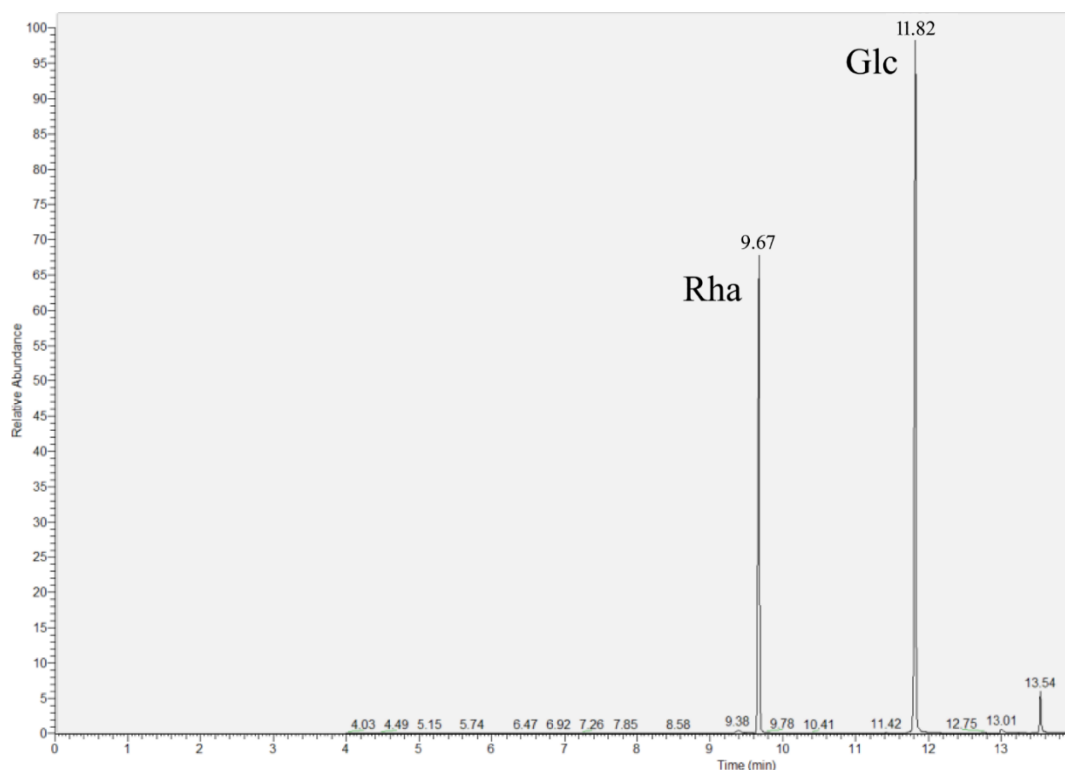


Figure 17. Chromatogram of alditol acetates of sugar components obtained as a result of sugar analysis of *K. pneumoniae* ABC122 OPS. Derivatives of Rha and Glc were identified based of retention time of standards.

The substitution positions (hydroxyl groups forming glycosidic bonds) in Rha and Glc residues were determined by methylation analysis. The chromatogram obtained for partially methylated alditol acetates revealed the presence of two derivatives (Figure 18a). Derivative 1,3,5-tri-*O*-Ac-2,4-di-*O*-Me-rhamnitol ($t_R=8,52$ min) was identified as 3-substituted Rha_p based on the detected mass spectrum (Figure 18b). Derivative 1,4,5-tri-*O*-Ac-2,3,6-tri-*O*-Me-glucitol ($t_R=9,81$ min) was identified as 4-substituted Glc_p based on the detected mass spectrum (Figure 18c). No terminal sugar residue (OPS nonreducing end) was observed in the methylation analysis.

The absolute configuration of the monosaccharide components of the OPS was analysed by the method described by York *et al.* (York *et al.* 1997) (Figure 19). The *O*-(*S*)-2-methyl butyrate (SMB) derivatives of D- and L-monosaccharides are diastereomeric and can be differentiated by comparing their chemical shift and coupling patterns observed in the NMR spectra. The OPS was hydrolysed with 2 M TFA and the

resulting monosaccharides were converted into SMB derivatives, and then their absolute configurations were analysed by the ^1H NMR spectroscopy (Artyszuk *et al.* 2024).

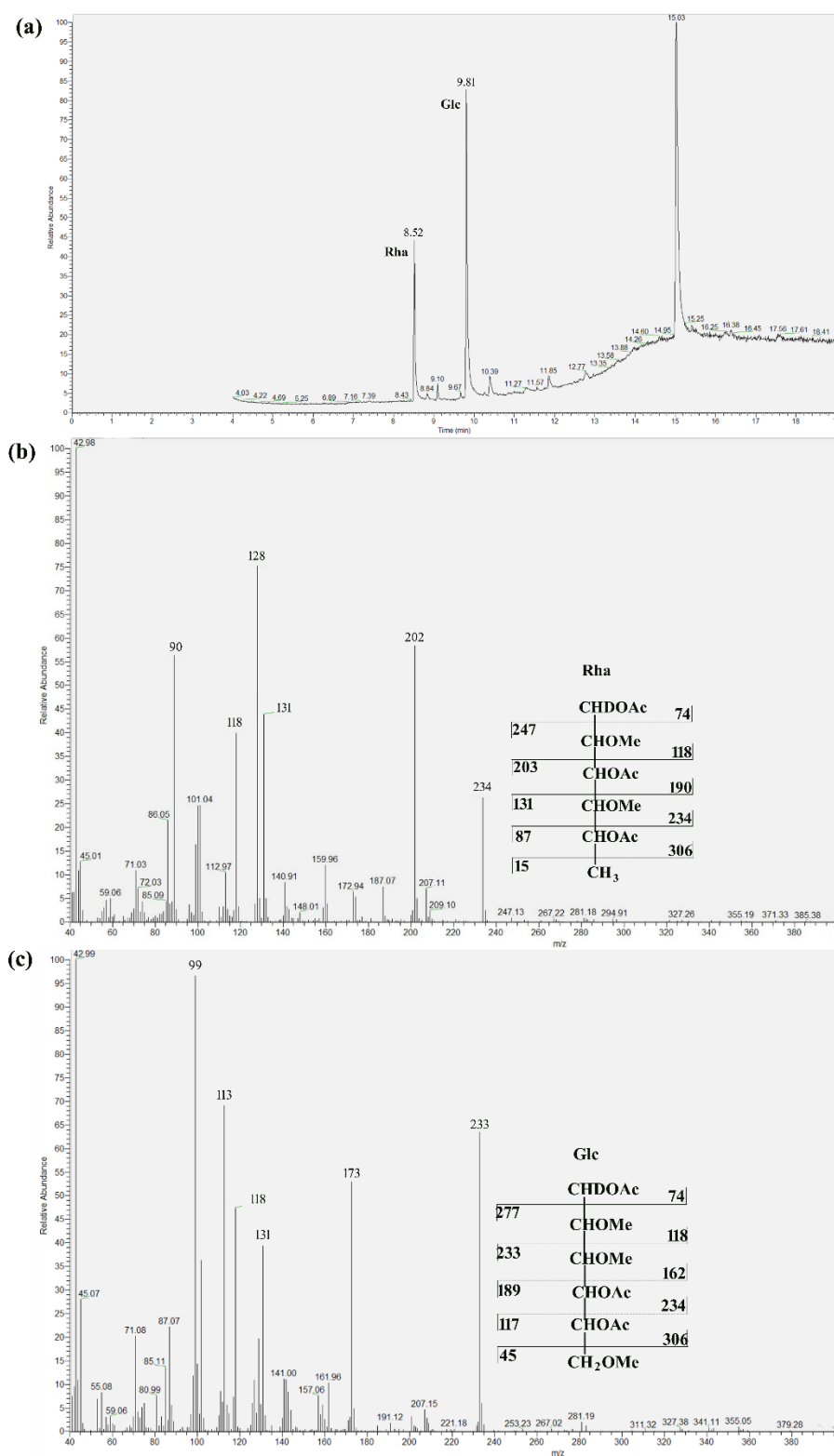


Figure 18. Results of methylation analysis of the *K. pneumoniae* ABC122 OPS. (a) chromatogram of sugar derivatives; (b) mass spectrum of the 1,3,5-tri-*O*-Ac-2,4-di-*O*-Me-rhamnitol, the derivative of the \rightarrow 3)-Rhap; (c) mass spectrum of the 1,4,5-tri-*O*-Ac-2,3,6-tri-*O*-Me-glucitol, the derivative of the \rightarrow 4)-Glc.

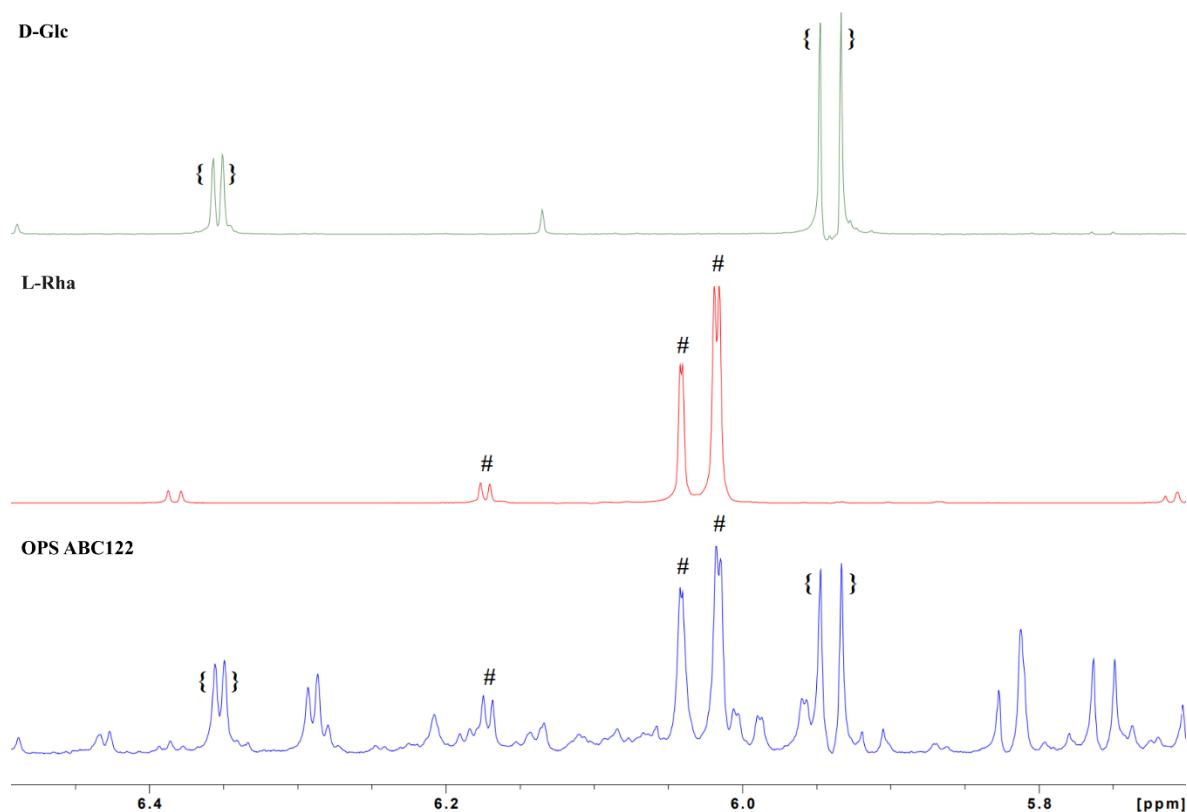


Figure 19. Comparison of ^1H NMR spectra in the range of 5.5–6.5 ppm for *O-(S)*-2-methyl butyrate derivatives of monosaccharide standards and the *K. pneumoniae* ABC122 OPS hydrolysate. The signs “{”, “#” indicate proton signals of standards (D-Glc, L-Rha) that were also identified in the OPS spectrum. The spectra were obtained for acetone- d_6 solutions at 25°C. (Artyszuk *et al.* 2024).

The standards (D-Glc and L-Rha derivatives) and the sample were dissolved in deuterated acetone (acetone- d_6) and all spectra were calibrated to the internal reference of acetone- d_6 containing internal acetone- d_5 (δ_{H} 2,05 ppm). The comparison of the ^1H NMR profile of the SMB-modified OPS hydrolysate and the SMB-derivatives of the selected monosaccharides indicated the presence of L-Rha and D-Glc in the RU (Figure 19).

The complete assignments of ^1H and ^{13}C resonances for the ABC122 OPS were performed by interpretation of one- and two-dimensional NMR spectra, including ^1H , ^1H ^1H , ^{13}C HSQC-DEPT (Figure 20), COSY (data not shown), TOCSY (Figure 21), NOESY (Figure 21), HSQC-TOCSY (data not shown) and HMBC (Figure 20, Table 8) (Artyszuk *et al.* 2024).

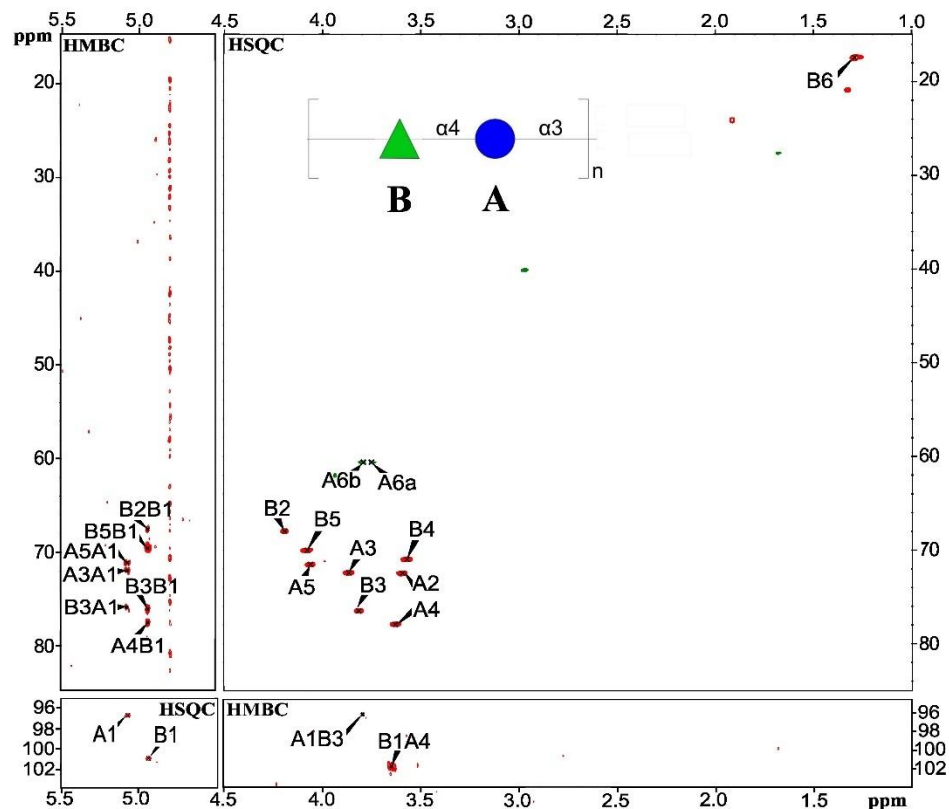


Figure 20. Selected regions of ^1H , ^{13}C HSQC-DEPT and HMBC spectra of the *K. pneumoniae* ABC122 OPS. Capital letters refer to carbohydrate residues, as shown in the inset structures and Table 9. Arabic numerals refer to protons and carbons in respective residues (Artyszuk *et al.* 2024). The Symbol Nomenclature for Graphical Representation of Glycans is used for OPS visual representation:
 ● Glucose; ▲ Rhamnose to show O-specific polysaccharide repeating units:
 $[-\rightarrow 3)-\alpha\text{-L-Rhap-(1}\rightarrow 4)-\alpha\text{-D-Glcp-(1}\rightarrow)_n$ (Varki *et al.* 2015).

Chemical shift values for the ^1H and ^{13}C resonances (Table 8, Figure 20) were compared with the data reported previously (Bock and Pedersen 1974; R. F. Kelly *et al.* 1995; Vinogradov *et al.* 2002).

Table 8. The ^1H and ^{13}C NMR chemical shifts and selected inter-residue correlations from ^1H , ^1H NOESY and ^1H , ^{13}C HMBC spectra of the OPS of *K. pneumoniae* ABC122 (Artyszuk *et al.* 2024)^{a)}.

Residue ^(a)	Chemical Shifts (ppm)						Selected Inter-Residue NOE and $^3J_{\text{H,C}}$ Connectivities		
	H1 C1	H2, C2	H3 (H3ax,H3eq) C3	H4 C4	H5 C5	H6a, H6b (H6) C6	H1/C1 (C2) Connectivities to	Inter-Residue Atom/ Residue	
ABC122 OPS									
A	$\rightarrow 4)-\alpha\text{-D-Glcp-(1}\rightarrow$	5.07 96.3	3.59 72.5	3.87 72.4	3.63 77.9	4.06 71.6	3.75, 3.79 60.6	3.82 76.5	H3, C3 of B
B	$\rightarrow 3)-\alpha\text{-L-Rhap-(1}\rightarrow$	4.94 101.2	4.19 68.0	3.82 76.5	3.57 71.0	4.08 70.0	1.29 17.4	3.63 77.9	H4, C4 of A

^{a)} The $J_{\text{C1,H1}}$ constants = 172 and 173 Hz for A and B residues, respectively.

Residue **A** with the H1/C1 signals at 5.07/96.3 ppm, $J_{C1,H1} \sim 172$ Hz was recognised as the 4-substituted α -D-Glcp residue based on the downfield shift of the C4 signal (77,9 ppm) and the strong vicinal couplings between H2, H3, H4, and H5. Residue **B** with the H1/C1 signals at δ 4,94/101,2 ($J_{C1,H1} \sim 173$ Hz) was recognized as the 3-substituted α -L-Rhap based on the high chemical shifts of the C3 (76,5 ppm) and the characteristic shift for an exocyclic CH₃ group (1,29 ppm, 17,4 ppm) (Table 8).

The sequence of sugar residues in the ABC122 OPS was identified by ¹H,¹³C HMBC (Figure 20, Table 8) and ¹H,¹H NOESY (Figure 21, Table 8) experiments, revealing the OPS RU structure to be $\rightarrow 3$)- α -L-Rhap-(1 \rightarrow 4)- α -D-Glcp-(1 \rightarrow disaccharide (Figure 20). Figure 21 represents the overlay of the NOESY and TOCSY spectra of the ABC122 OPS (Artyszuk *et al.* 2024).

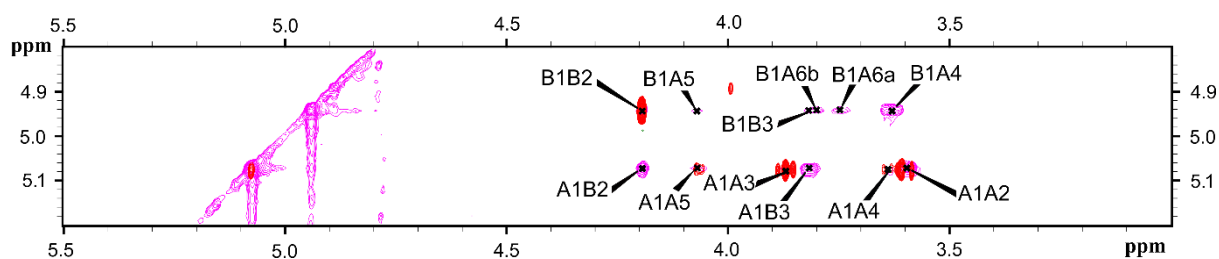


Figure 21. The TOCSY (red) and NOESY (magenta) NMR spectra of *K. pneumoniae* ABC122 OPS. The key inter-residual correlations were indicated; letters are as in Table 8 (Artyszuk *et al.* 2024).

In order to elucidate the Kdo position within the O antigen structure, the complete assignments of ¹H and ¹³C resonances in HR-MAS NMR spectra of the intact ABC122 LPS has been performed indicating the presence of residue **a** recognised as the $\rightarrow 4$)- α -D-Glcp, residue **b** as $\rightarrow 3$)- α -L-Rhap, and a residue **k** as terminal Kdo residue (Figure 22, Table 9). Additionally **b'** residue was identified as $\rightarrow 3$)- α -L-Rhap (Artyszuk *et al.* 2024).

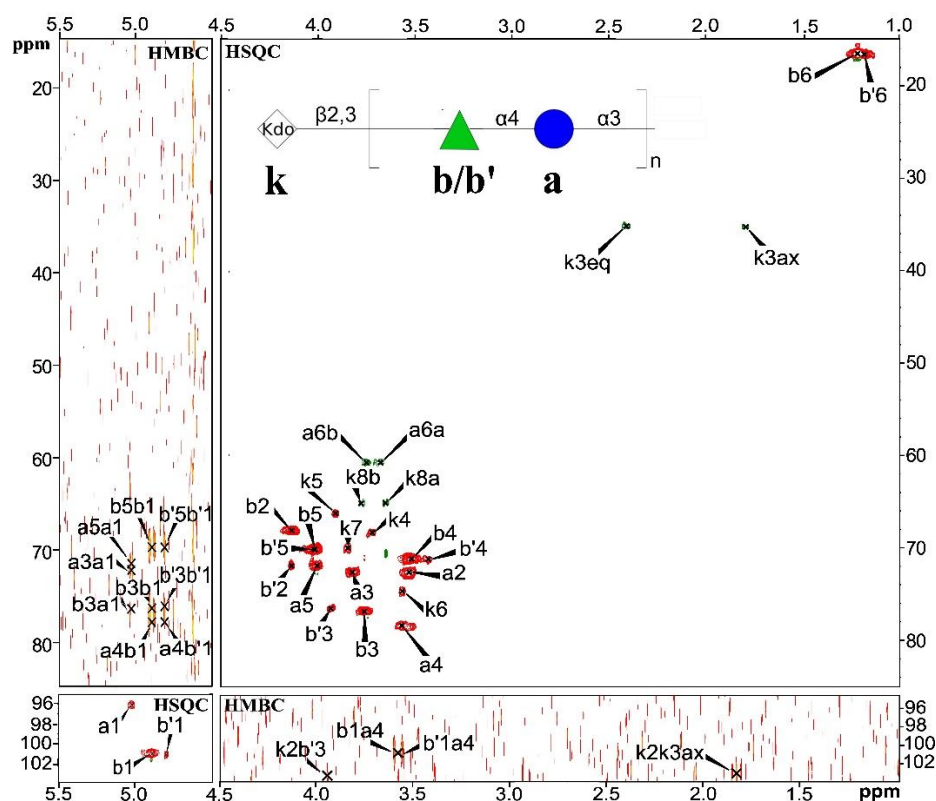


Figure 22. Selected regions of ^1H , ^{13}C HSQC-DEPT and HMBC spectra of the *K. pneumoniae* ABC122 LPS. LPS were analysed by the ^1H , ^{13}C HR-MAS NMR spectroscopy. The β -Kdop (HMBC) substitution crosspeak (k2b'3 signal) distinct from noise on the basis of signal shape/pattern. Capital letters refer to carbohydrate residues, as shown in the inset structures and Table 9. Arabic numerals refer to protons and carbons in respective residues. Residue b' observed in the ABC122 LPS is a variant of residue $\rightarrow 3$ -L- α -Rhap-(1 \rightarrow substituted by terminal β -Kdop (k) capping the O antigen (Artyszuk *et al.* 2024). The Symbol Nomenclature for Graphical Representation of Glycans is used for OPS visual representation: \bullet Glucose; \blacktriangle Rhamnose; \diamond Kdo to present O antigen structure: β -Kdop-[$\rightarrow 3$]- α -L-Rhap-(1 $\rightarrow 4$)- α -D-Glcp-(1 \rightarrow)_n (Varki *et al.* 2015).

Table 9. The ^1H and ^{13}C NMR chemical shifts and selected inter-residue correlations from ^1H , ^1H NOESY and ^1H , ^{13}C HMBC spectra of the LPS of *K. pneumoniae* ABC122 (Artyszuk *et al.* 2024)^a.

Residue ^(a)	Chemical Shifts (ppm)								Selected Inter-Residue NOE and $^3J_{\text{H-C}}$ Connectivities		
	H1 C1	H2, C2	H3 (H3ax, H3eq) C3	H4 C4	H5 C5	H6a, H6b (H6) C6	H7 C7	H8a, H8b C8	H1/C1 (C2) Connectivities to	Inter-Residue Atom/Residue	
ABC122 LPS											
a	$\rightarrow 4$ - α -D-Glcp-(1 \rightarrow)	5.03 96.3	3.54 72.6	3.83 72.5	3.58 78.1	4.02 71.7	3.69, 3.76 60.8		3.58 76.7	H3, C3 of b	
b	$\rightarrow 3$ - α -L-Rhap-(1 \rightarrow)	4.90 101.4	4.14 68.1	3.77 76.7	3.53 71.2	4.03 70.1	1.25 17.7		3.58 78.1	H4, C4 of a	
b'	$\rightarrow 3$ - α -L-Rhap-(1 \rightarrow)	4.81 101.4	4.14 71.7	3.94 76.3	3.45 71.2	4.03 70.1	1.22 17.7		3.58 78.1	H4, C4 of a	
k	β -Kdop-(2 \rightarrow)	174.6	103.5	(1.83, 2.44) 35.9	3.74 68.3	3.92 66.3	(3.58) 74.5	3.86 70.0	3.66, 3.79 65.2	(3.94) -	H3 of b'

^a) The $J_{\text{C1,H1}}$ constants =172 and 173 Hz for a and b, b' residues, respectively.

The sequence of the a and b/b' residues within the RU was in agreement with OPS results and was confirmed by $^1\text{H},^{13}\text{C}$ HMBC (Figure 22, Table 9) and $^1\text{H}, ^1\text{H}$ NOESY (Figure 23) analyses. Figure 24 represents the overlay of the NOESY and TOCSY spectra of the ABC122 LPS with crosspeaks supporting the sequence. No inter-residue NOEs were observed for β -Kdo (k).

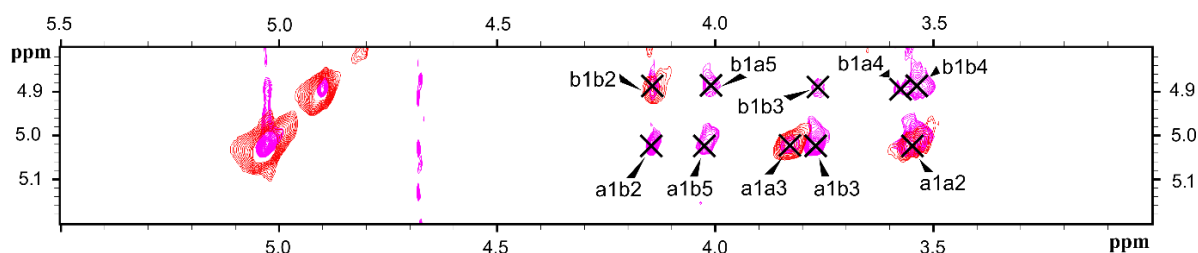


Figure 23. The TOCSY (red) and NOESY (magenta) NMR spectra of *Klebsiella pneumoniae* ABC122 LPS. The key inter-residual correlations were indicated; letters are as in Table 9. LPS spectrum was recorded by HR-MAS NMR spectroscopy (Artyszuk *et al.* 2024).

The Kdo residue was identified as terminal β -3-deoxy-D-manno-oct-2-ulosonic acid and showed a β anomeric configuration based on the chemical shifts of axial and equatorial H3 protons (δ_{H} 1,83 and 2,44 ppm), and C1 (δ_{C} 174,6 ppm) and C2 (δ_{C} 103,5 ppm) observed at the HMBC spectrum (Figure 22) (Birnbaum *et al.* 1987; Vinogradov *et al.* 2002). The identified β -Kdop-(1 \rightarrow) (k) residue is not a part of the O antigen RU and represents terminal modification of the O antigen (Figure 22, the inset structure). The attachment position of the β -Kdop-(1 \rightarrow) was found on the basis of the observed HMBC correlation between C2 of k (δ_{C} 103,5 ppm) and H3 of b' residue \rightarrow 3)- α -L-Rhap (δ_{H} 3,94 ppm) (Figure 22). The terminal residue of β -Kdo is thus linked to O-3 of the Rha residue, the same position as subsequent RU is linked. This β -Kdop-(1 \rightarrow 3)- α -L-Rhap substitution is further supported by the downfield shift of H3 of Kdo substituted L-Rhap of δ_{H} 3,94 ppm from 3,77 ppm observed for the α -D-Glcp substituted α -L-Rhap moiety similar as data published by Vinogradov *et al.* (H3 shift from 3,75 ppm to 4,01 ppm) (Vinogradov *et al.* 2002) as well as no other possible substitution signals observed (Figure 22). Finally, the following structure of the novel *K. pneumoniae* O antigen was elucidated: β -Kdop-[- \rightarrow 3)- α -L-Rhap-(1 \rightarrow 4)- α -D-Glcp-(1 \rightarrow)]_n (Figure 22, inset structure). The novel O serotype was assigned as O13 according to the current numerical order of the *K. pneumoniae* O serotypes characterised to date (Clarke and Whitfield 1992; Kelly *et al.* 2023; Artyszuk *et al.* 2024).

To confirm identity of the ABC122 O antigen structure with the BC738, BC13-986, and 3936/19 O antigens, $^1\text{H},^{13}\text{C}$ HSQC-DEPT spectra were compared (Figure 24-26).

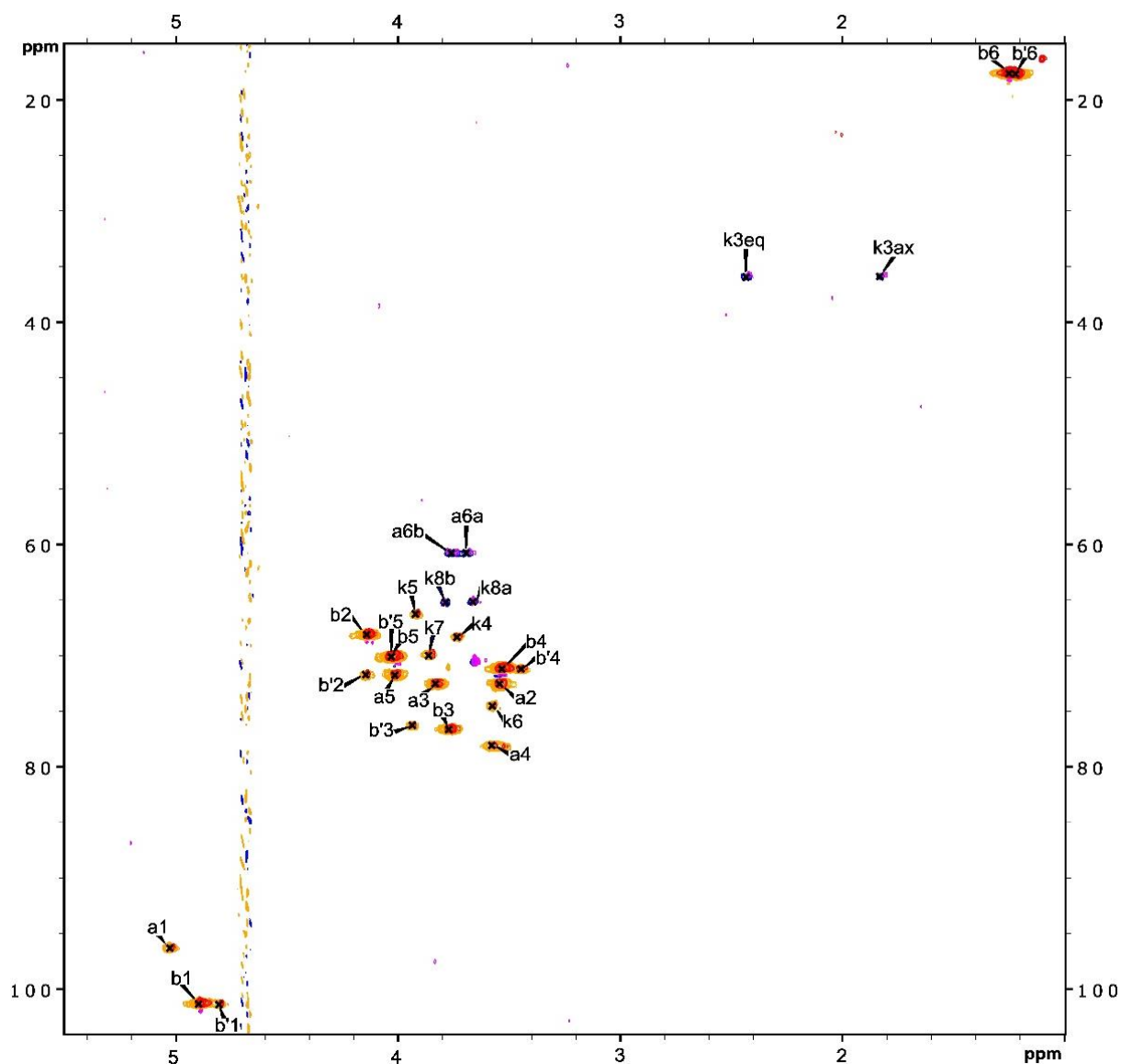


Figure 24. The HR-MAS $^1\text{H},^{13}\text{C}$ HSQC-DEPT NMR spectra overlay for ABC122 LPS (orange and blue) and BC738 LPS (red and magenta). Capital letters and Arabic numerals refer to carbohydrate residues and protons and carbons, as in the Table 9 for ABC122 LPS. Chemical shifts for BC738 LPS are presented in Table 10 (Artyszuk *et al.* 2024).

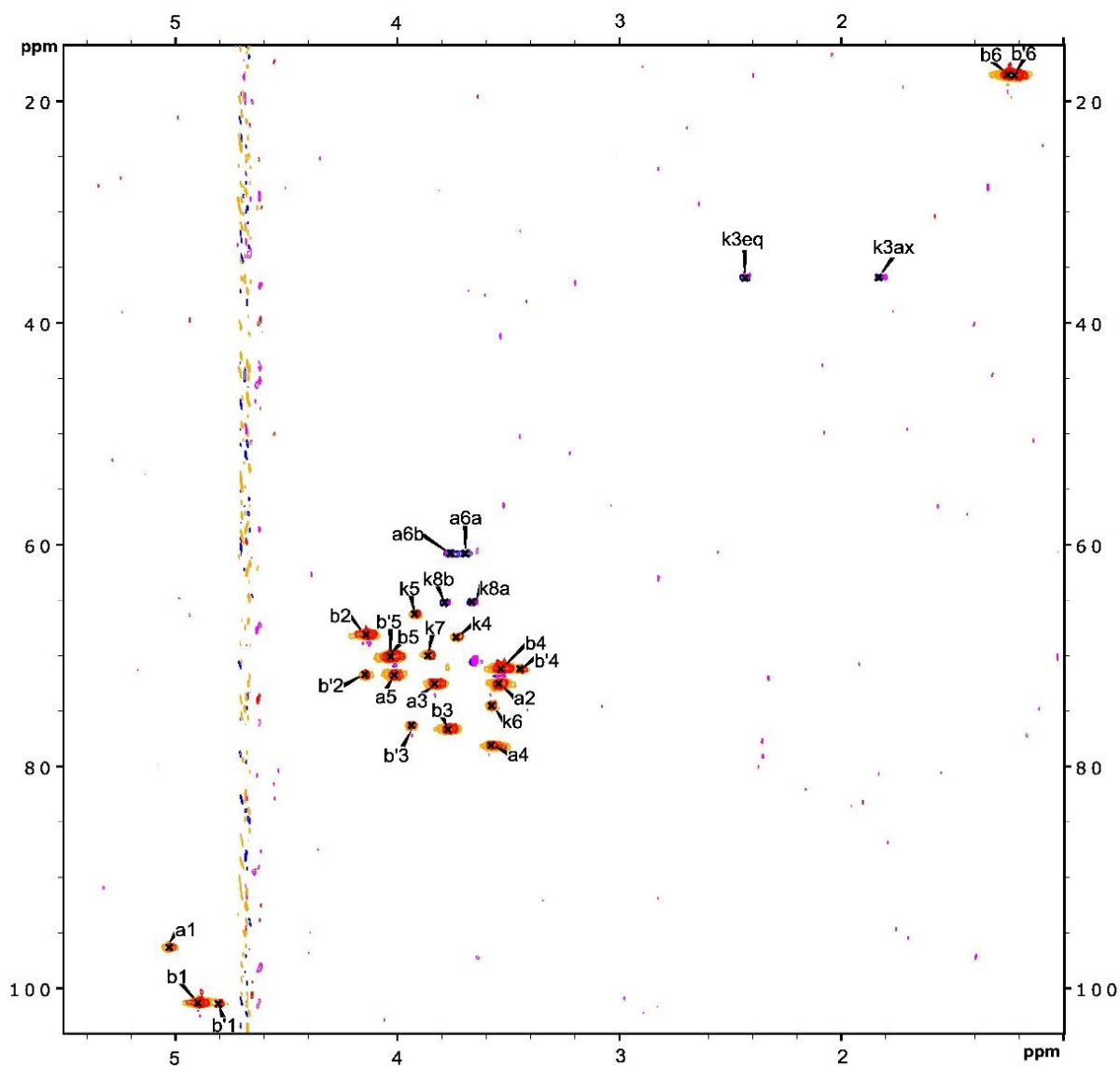


Figure 25. The HR-MAS ^1H , ^{13}C HSQC-DEPT NMR spectra overlay for ABC122 LPS (orange and blue) and BC13-986 LPS (red and magenta). Capital letters and Arabic numerals refer to carbohydrate residues and protons and carbons, as in the Table 9 for ABC122 LPS. Chemical shifts for BC13-986 LPS are presented in Table 10 (Artyszuk *et al.* 2024).

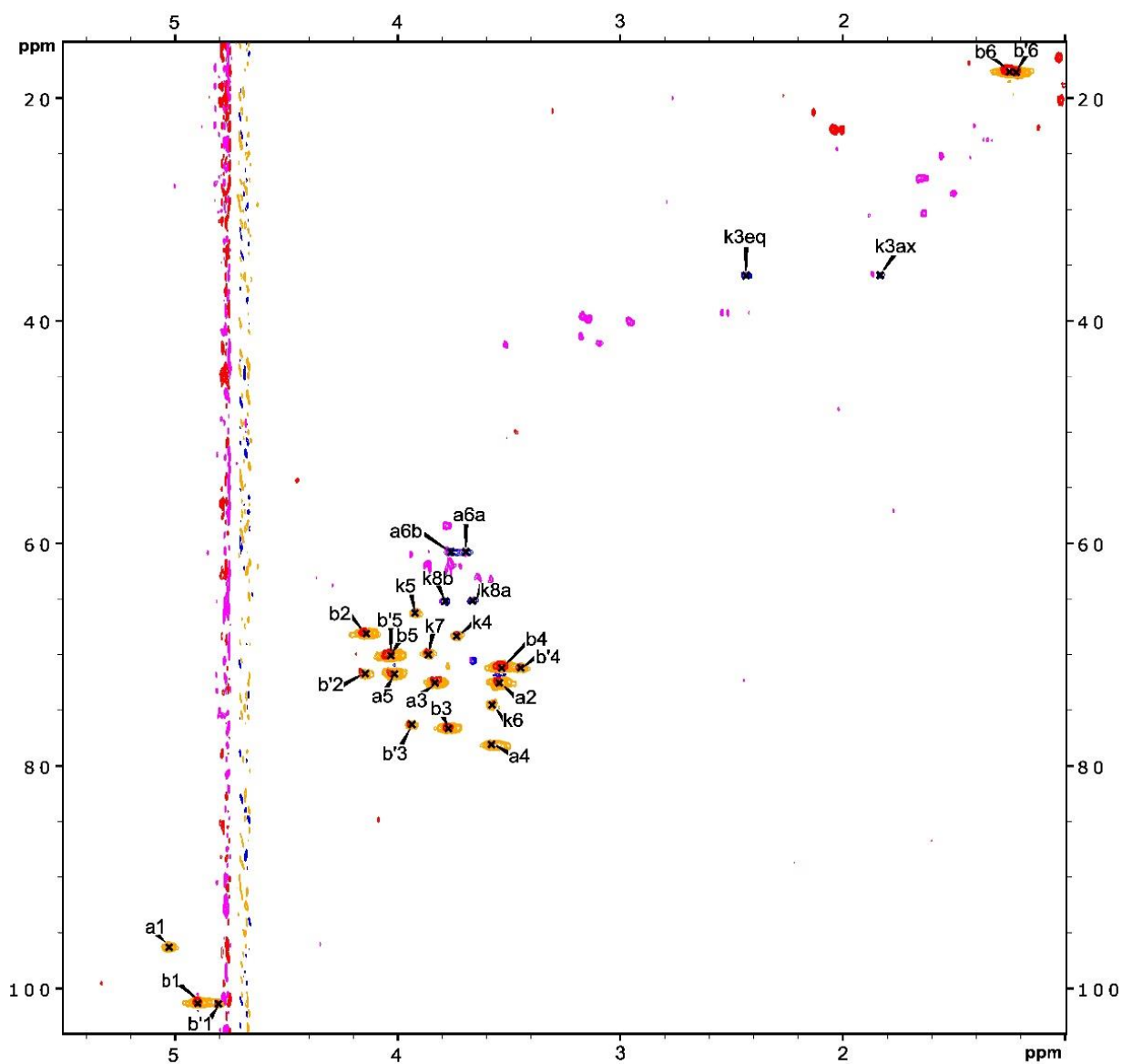


Figure 26. The HR-MAS ^1H , ^{13}C HSQC-DEPT NMR spectra overlay for ABC122 LPS (orange and blue) and 3936/19 LPS (red and magenta). Capital letters and Arabic numerals refer to carbohydrate residues and protons and carbons, as in the Table 9 for ABC122 LPS. Chemical shifts for 3936/19 LPS are presented in Table 10 (Artyszuk *et al.* 2024).

Identified proton and carbon chemical shifts of O antigen constituents of BC738, BC13-986, and 3936/19 were identical to ABC122 LPS within the measurement errors (Table 10).

Table 10. The ^1H and ^{13}C HR-MAS NMR chemical shifts of *K. pneumoniae* BC738, BC13-986, and 3936/19 O antigens (Artyszuk *et al.* 2024^a).

Residue ^{a)}		Chemical Shifts (ppm)							
		H1 C1	H2, C2	H3 (H3ax,H3eq) C3	H4 C4	H5 C5	H6a, H6b (H6) C6	H7 C7	H8a, H8b C8
BC738 LPS									
a	$\rightarrow 4)\text{-}\alpha\text{-D-Glcp-(1}\rightarrow$	5.02 96.3	3.54 72.4	3.83 72.5	3.57 78.0	4.01 71.7	3.68, 3.76 60.7		
b	$\rightarrow 3)\text{-}\alpha\text{-L-Rhap-(1}\rightarrow$	4.89 101.3	4.13 68.1	3.76 76.6	3.53 71.1	4.02 70.0	1.24 17.6		
b'	$\rightarrow 3)\text{-}\alpha\text{-L-Rhap-(1}\rightarrow$	4.79 101.3	4.14 71.6	3.93 76.2	3.44 71.2	4.02 70.0	1.21 17.6		
k	$\beta\text{-Kdop-(2}\rightarrow$	nd	nd	1.82, 2.42 35.8	3.73 68.2	3.91 66.2	3.57 74.5	3.85 69.9	3.67, 3.79 65.2
BC13-986 LPS									
a	$\rightarrow 4)\text{-}\alpha\text{-D-Glcp-(1}\rightarrow$	5.02 96.3	3.54 72.5	3.82 72.5	3.57 78.1	4.01 71.7	3.70, 3.77 60.8		
b	$\rightarrow 3)\text{-}\alpha\text{-L-Rhap-(1}\rightarrow$	4.89 101.3	4.13 68.1	3.76 76.6	3.52 71.2	4.02 70.1	1.24 17.6		
b'	$\rightarrow 3)\text{-}\alpha\text{-L-Rhap-(1}\rightarrow$	4.80 101.4	4.14 71.7	3.93 76.3	3.44 71.2	4.02 70.1	1.21 17.7		
k	$\beta\text{-Kdop-(2}\rightarrow$	nd	nd	1.82, 2.42 35.9	3.72 68.3	3.91 66.2	3.57 74.5	3.85 70.0	3.65, 3.77 65.2
3936/19 LPS									
a	$\rightarrow 4)\text{-}\alpha\text{-D-Glcp-(1}\rightarrow$	5.03 96.3	3.55 72.4	3.83 72.4	3.57 77.9	4.04 71.7	3.71, 3.78 60.6		
b	$\rightarrow 3)\text{-}\alpha\text{-L-Rhap-(1}\rightarrow$	4.90 101.1	4.15 68.0	3.77 76.5	3.53 71.0	4.04 69.9	1.25 17.5		
b'	$\rightarrow 3)\text{-}\alpha\text{-L-Rhap-(1}\rightarrow$	4.81 101.3	4.17 71.6	3.95 76.2	3.45 71.1	4.04 69.9	1.22 17.6		
k	$\beta\text{-Kdop-(2}\rightarrow$	nd	nd	1.86, 2.44 35.8	3.73 68.2	3.93 66.2	3.57 74.4	3.87 69.7	3.66, 3.81 65.1

^{a)} Spectra were recorded for LPS. Chemical shifts for protons and carbons were determined based on HR-MAS NMR ^1H , ^{13}C HSQC-DEPT spectra; nd – not determined.

7.3.2 Clonality of the isolates and genetic analysis of the OL101 *loci*

The MLST analysis revealed that all of the isolates were genetically diverse, each representing another sequence type (ST), namely ST11 (ABC122), ST485 (BC738), ST1427 (3936/19), ST3658 (BC13-986). ABC122 revealed the same structure of the *rfb* cluster and high match confidence between its all nine *rfb* genes and those of the OL101 reference strain (GenBank accession number: LT174596.1) (Figure 27a, b). BC738, BC13-986 and 3936/19 strains were predicted as OL101 with identity of 95,05%; 92,04% and 91,79%, respectively (Table 11). For BC13-986 and 3936/19 isolates, 8 out of 9 genes of OL101 reference *locus* were found (Artyszuk *et al.* 2024).

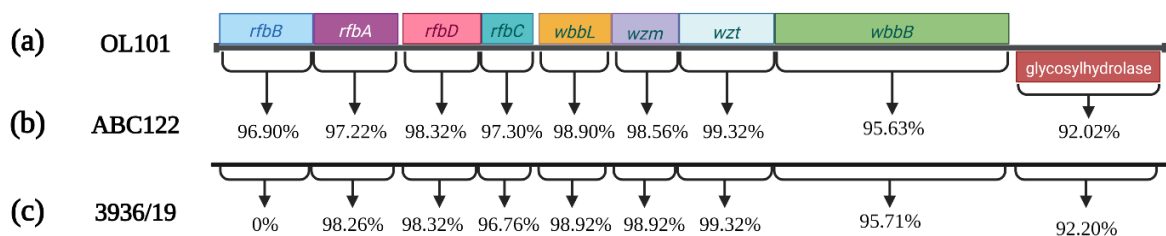


Figure 27. Organization of the O antigen biosynthesis gene cluster (*rfb*, O locus) from the *K. pneumoniae* OL101 reference strain (a); and comparison to ABC122 O locus (b); and comparison to 3936/19 O locus based on Kaptive results (c) (Artyszuk *et al.* 2024).

Table 11. Kaptive-based *O locus* prediction results for BC738, BC13-986 and 3936/19 *K. pneumoniae*^{a)}.

Assembly	Best match locus	Match confidence	Problems	Coverage	Identity	Length discrepancy	Expected genes in locus	Expected genes in locus, details	Missing genes
BC738	OL101	Good	?*	100.00%	95.05%	n/a	9 / 9	OL101_01_ <i>rfbB</i> ,97.465%;OL101_02_ <i>rfbA</i> ,97.917%;OL101_03_ <i>rfbD</i> ,98.571%;OL101_04_ <i>rfbC</i> ,96.757%;OL101_05_ <i>wbbL</i> ,98.917%;OL101_06_ <i>wzm</i> ,98.921%;OL101_07_ <i>wzt</i> ,99.324%;OL101_08_ <i>wbbB</i> ,95.542%;OL101_09_ <i>glycosylhydrolase</i> ,92.199%	-
BC13-986	OL101	Good	?-*	100.00%	92.04%	n/a	8 / 9	OL101_01_ <i>rfbB</i> ,97.183%;OL101_03_ <i>rfbD</i> ,98.316%;OL101_04_ <i>rfbC</i> ,97.297%;OL101_05_ <i>wbbL</i> ,98.917%;OL101_06_ <i>wzm</i> ,98.921%;OL101_07_ <i>wzt</i> ,99.324%;OL101_08_ <i>wbbB</i> ,95.711%;OL101_09_ <i>glycosylhydrolase</i> ,92.157%	OL101_02_ <i>rfbA</i>
3936/19	OL101	Good	?-*	100.00%	91.79%	n/a	8 / 9	OL101_02_ <i>rfbA</i> ,98.264%;OL101_03_ <i>rfbD</i> ,98.316%;OL101_04_ <i>rfbC</i> ,96.757%;OL101_05_ <i>wbbL</i> ,98.917%;OL101_06_ <i>wzm</i> ,98.921%;OL101_07_ <i>wzt</i> ,99.324%;OL101_08_ <i>wbbB</i> ,95.711%;OL101_09_ <i>glycosylhydrolase</i> ,92.199%	OL101_01_ <i>rfbB</i>

^{a)} The “*” character indicates that one or more of the expected genes falls below the identity threshold (default 95%). The “?” character indicates that the *O locus* sequence was not found in a single piece in the assembly. The “-” character indicates that one of the expected genes was not found in the gene BLAST search. Missing genes and larger discrepancies between the *rfb* region of the analysed strains and the *rfb* region of the reference isolate are marked in red. n/a – not available

According to the Kaptive prediction, for ABC122, genes *rfbBADC* showed 96,90-98,32% identity scores, *wbbL* – 98,90%, *wzm* – 98,56%, *wzt* – 99,32%, whereas *wbbB* and glycosylhydrolase genes showed 95,63% and 92,02% identity, respectively, to the reference strain (Figure 27b). The 3936/19 strain revealed almost the same homology levels in eight genes, except for *rfbB* which has not been detected by Kaptive at all (Figure 27c) (Artyszuk *et al.* 2024). The functions of OL101 *locus* genes were outlined in accordance with the current understanding in Section 4.2.6.3.2.

However, more detailed sequence alignment of 3936/19 sequence to the reference OL101 showed the presence of a part of the *rfb* gene (517 bp out of 1065 bp) in the study isolate (nucleotide positions from 1 to 517) (Figure 28) (Artyszuk *et al.* 2024)

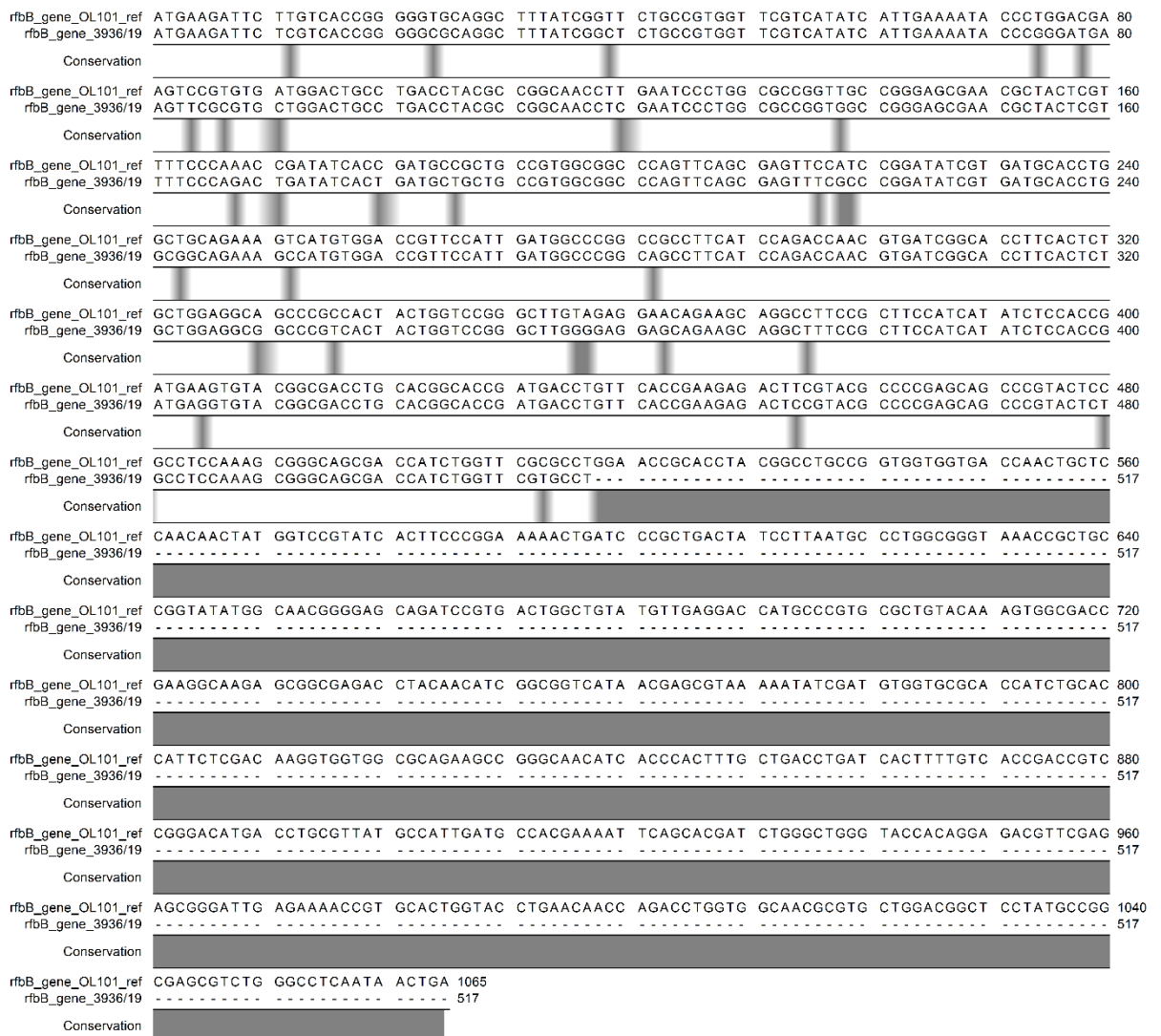


Figure 28. The alignment of the *rfbB* gene of *K. pneumoniae* 3936/19 and OL101 *locus*. Grey areas mark regions of differences in nucleotide sequence (Artyszuk *et al.* 2024).

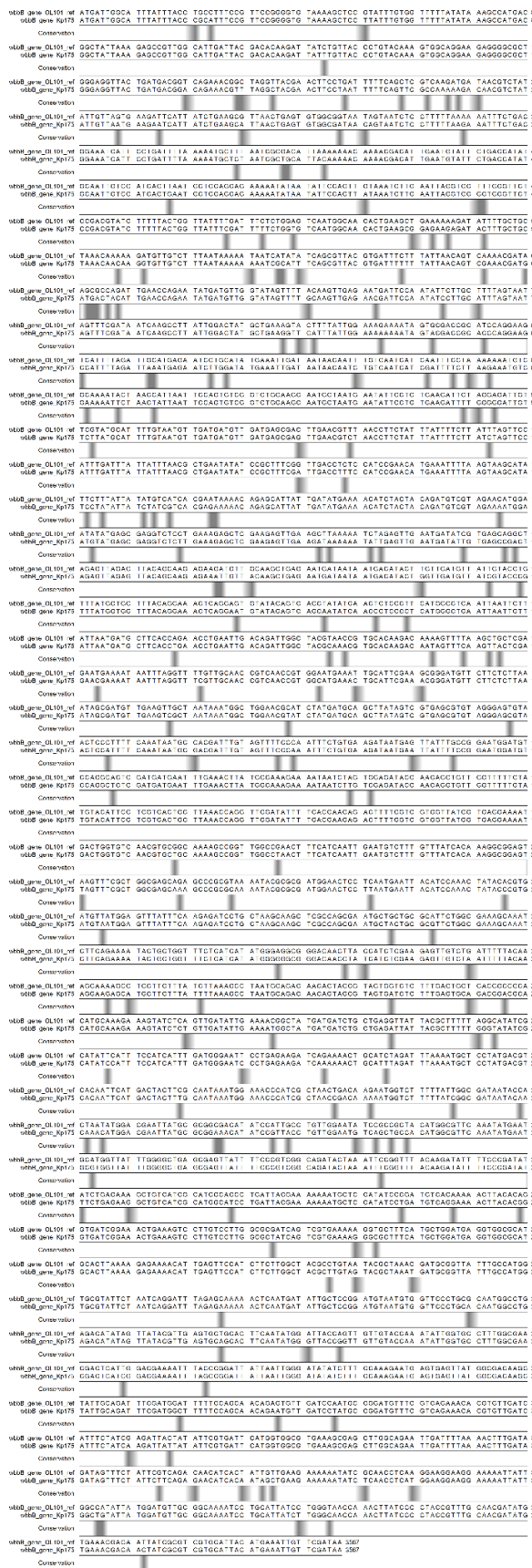


Figure 29. The alignment of the *wbbB* gene of *K. pneumoniae* ABC122 and OL101 locus. Grey areas mark regions of differences in nucleotide sequence. Symbol Kp175 stand for ABC122 sequence.

The more detailed comparison of *rfb* clusters showed the numerous nucleotide differences of *wbbB* gene in ABC122 isolate in comparison to the reference OL101 strain (Figure 29). The results indicated that the changes in *wbbB* gene may lead to altered activity of the WbbB protein, resulting in different polymerization of the OPS in OL101 (See section 4.2.6.3.2).

However, the alignment of the ABC122 *rfb locus* also showed the glycosylhydrolase gene occurrence. This gene was identical (100% identity) in comparison to the reference strain (Figure 30), although function of this glycosylhydrolase is still unknown and to date no OPS structure of *K. pneumoniae* OL101 reference strain has been published.

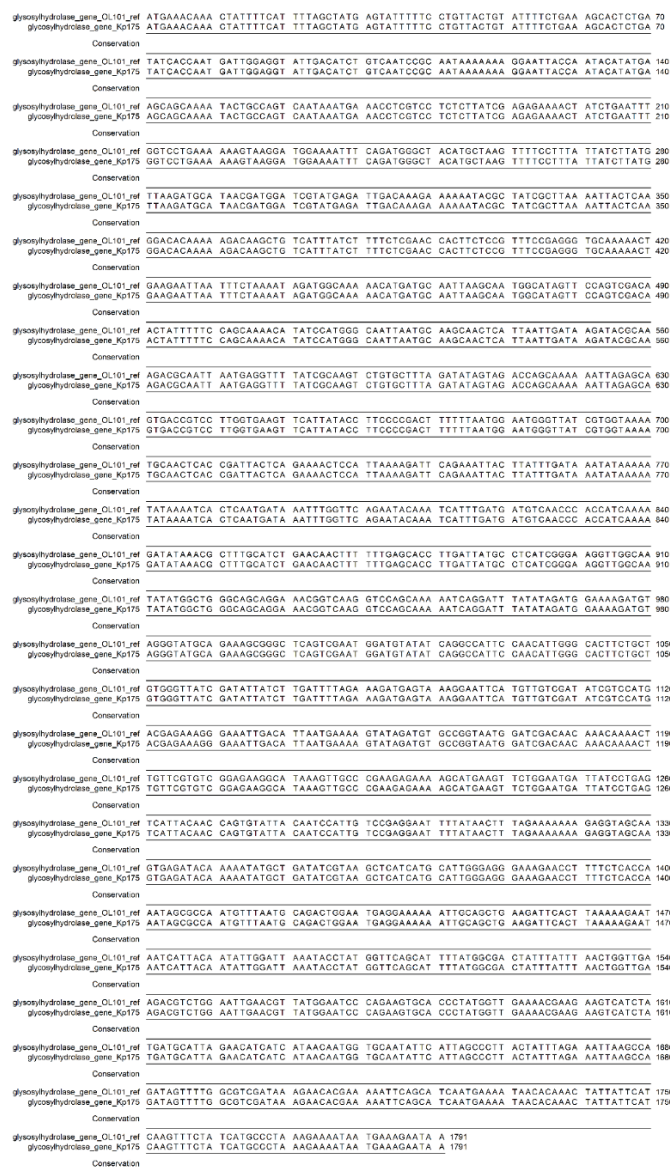


Figure 30. The alignment of the glycosylhydrolase gene of *K. pneumoniae* ABC122 and OL101 *locus*. Symbol Kp175 stand for ABC122 sequence.

7.3.3 OL101 prevalence among the *K. pneumoniae* isolates with publicly available genomes

The large-scale analysis of all 71377 *K. pneumoniae* genomes available in public databases (July 27, 2023) was performed in order to assess the prevalence of OL101 *loci* among the *K. pneumoniae* isolates sequenced. Based on the Kaptive O-serotyping results, 4674 isolates (6,55%) were predicted to contain the OL101 *loci* (Figure 31; Appendix, Table S2, available on-line at <https://www.sciencedirect.com/science/article/pii/S0144861723010469?via%3Dihub#ec000>) (Artyszuk *et al.* 2024).

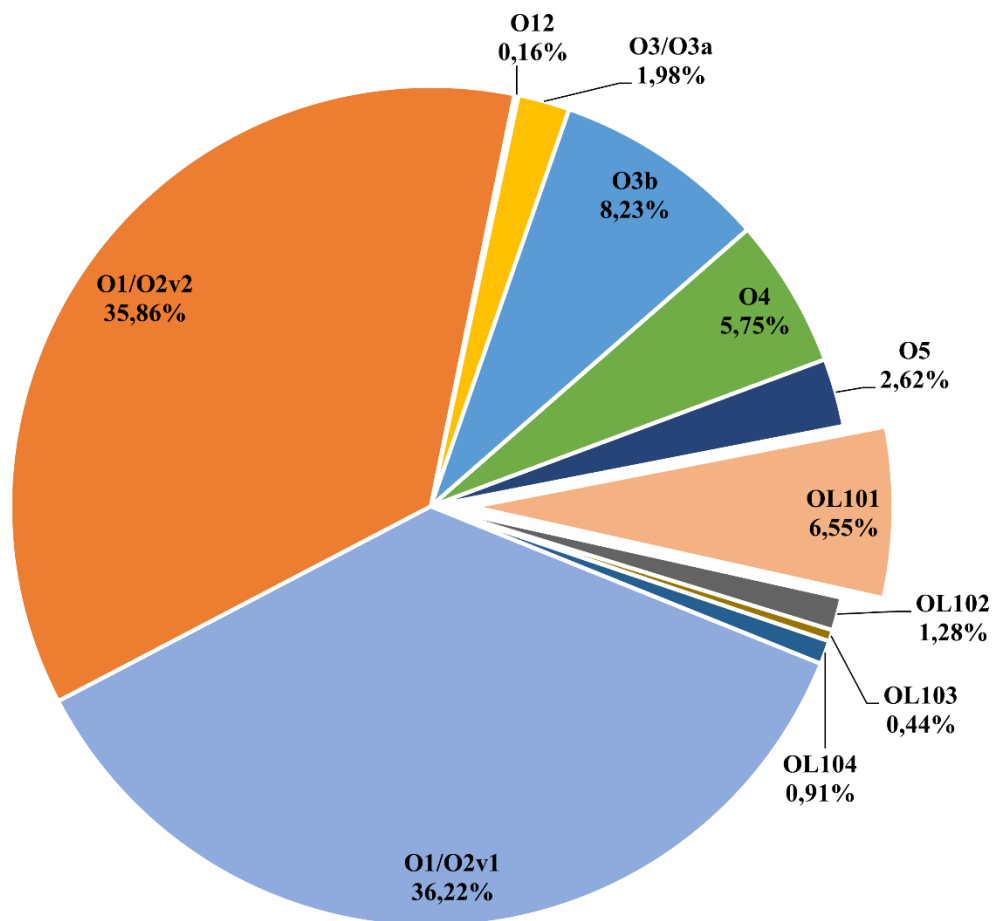


Figure 31. Distribution of *K. pneumoniae* O serotypes, based on genomes available in NCBI database for July 27, 2023 (Artyszuk *et al.* 2024).

7.4 O antigen and O loci analysis of *K. pneumoniae* Kp164, Kp165 and Kp166

7.4.1 Structural analysis

Strains Kp164, Kp165 and Kp166 have been selected as nontypeable by comparison of their O loci with O locus of reference strain PCM15 O4:K15 (data not available; personal communication; analyses were performed by Arsanis GmbH upon implementation of the EUROSTARS Klebsicure project). ^1H , ^{13}C HR-MAS NMR spectra of Kp164 LPS (Figure 9d) and reference PCM15 LPS (Figure 9g) indicated structural differences. Moreover the ^1H HR-MAS NMR spectrum of PCM15 LPS (Figure 9g) showed four anomeric signals, more than expected for the RU of O4 serotype, $\alpha\text{-Kdop-(2}\rightarrow[\rightarrow 2)\text{-}\beta\text{-D-Ribf-(1}\rightarrow 4)\text{-}\alpha\text{-D-Galp-(1}\rightarrow]$ (Vinogradov *et al.* 2002). Results raised doubts about correct classification of the PCM15 reference strain.

To elucidate the O antigen structure of the group 3 of LPSs, the LPS of the Kp164 strain was chosen as a representative, and hydrolysed to obtain poly- and oligosaccharides, followed by the fractionation on TSKgel[®]G2500PW column. This yielded polysaccharide fraction 1 (Figure 32).

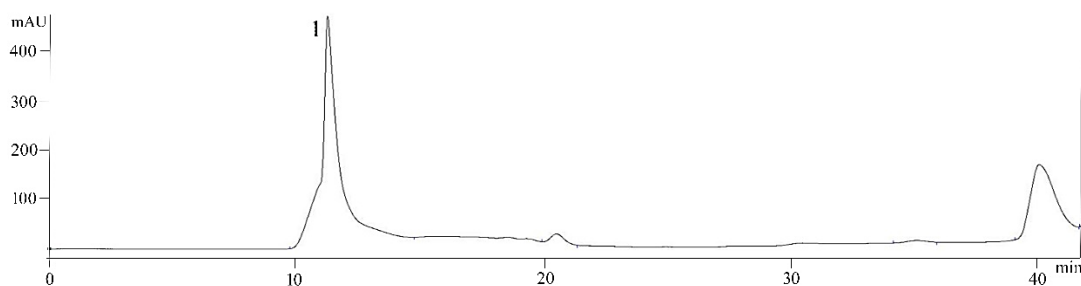


Figure 32. Chromatogram of poly- and oligosaccharides obtained by hydrolysis of Kp164 LPS (20 mg) and fractionated on TSKgel[®]G2500PW column (flow: 1 ml/min). The OPS was identified in fraction 1.

Fraction 1 and native Kp164 LPS were analysed by NMR spectroscopy. Interpretation of the NMR spectra of both samples showed the known O4 serotype when compared to spectra and data published for O4 antigen (Björndal *et al.* 1973; Vinogradov *et al.* 2002) (Figure 33). Chemical shifts observed in the ^1H , ^{13}C HSQC-DEPT spectrum of the OPS were similar to published data with characteristic spin system of $\rightarrow 2)\text{-}\beta\text{-D-Ribf-(1}\rightarrow$ (residue A) containing H1/C1 5,39/107,7 ppm) and the downfield shift of the C2 signal (80,1 ppm) indicating substitution at position 2 (Figure 33, Table 12). Besides $\rightarrow 4)\text{-}\alpha\text{-D-Galp-(1}\rightarrow$ (residue B), ^1H HR-MAS NMR spectrum of the Kp164 LPS showed

the presence of α -Kdop terminus (residue K), a feature of O4 serotype (Fig. 33, Table 12). Characteristic H3 protons of Kdo were not detected by 2D HR-MAS ^1H , ^{13}C HSQC-DEPT NMR, however H3ax proto (K3_{ax}) was detected by ^1H NMR (Figure 33). Figure 33 shows the HSQC-DEPT NMR spectra overlay for Kp164 LPS and Kp164 OPS supplemented by their 1D ^1H NMR spectra. As chemical shifts of identified compounds were in agreement with published data and further Kaptive O serotype prediction indicated O4 interpretation of NMR spectra has been limited to comparative analysis with published data (Björndal *et al.* 1973; Vinogradov *et al.* 2002).

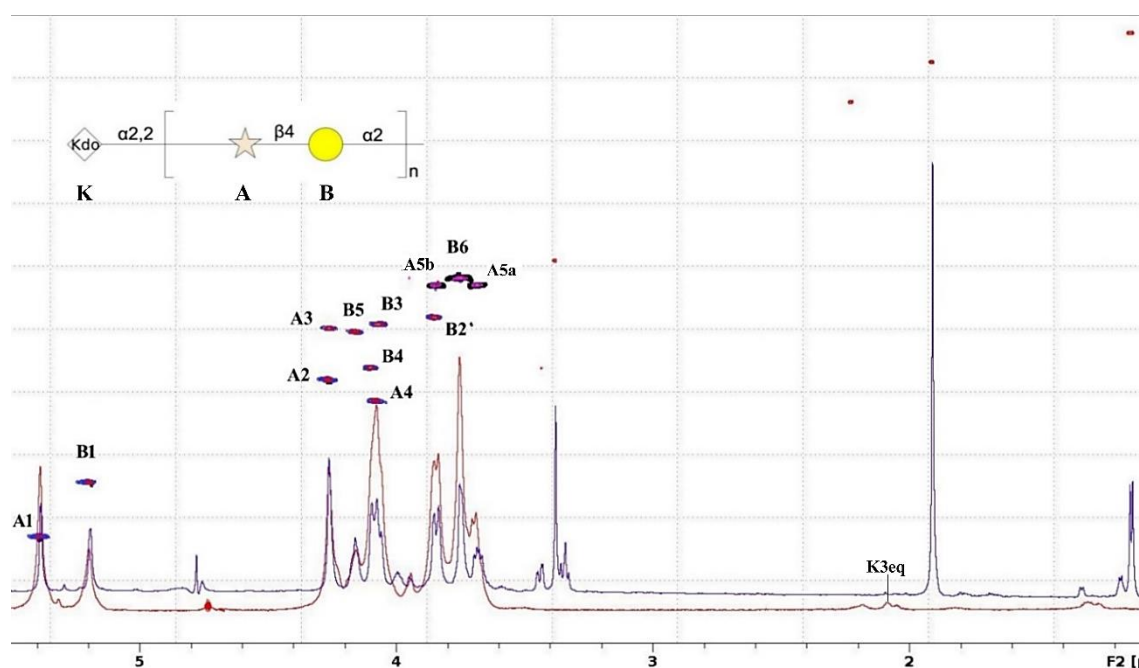


Figure 33. The NMR spectra overlay for Kp164 LPS (blue – positive signals and black – negative signals) and Kp164 OPS (red – positive signals and magenta – negative signals). LPS was analysed by ^1H , ^{13}C HSQC-DEPT HR-MAS NMR. ^1D ^1H HR-MAS NMR spectra of the Kp164 OPS (blue) and LPS (red). Capital letters and Arabic numerals refer to carbohydrate residues and protons and carbons, as in the Table 13. The inset structure of Kp164 O antigen. The Symbol Nomenclature

for Graphical Representation of Glycans is used for OPS visual representation: ● Galactose; ★ Ribose; ◊ Kdo to show O-antigen structure: α -Kdop-(2→[→2)- β -D-Ribf-(1→4)- α -D-Galp-(1→) $_n$ (Varki *et al.* 2015)

Table 12. The ^1H and ^{13}C NMR chemical shifts of *K. pneumoniae* Kp164 LPS and O-specific polysaccharide^{a)}.

Residue ^{a)}		Chemical Shifts (ppm)							
		H1 C1	H2, C2	H3 (H3ax,H3eq) C3	H4 C4	H5 C5	H6a, H6b (H6) C6	H7 C7	H8a, H8b C8
Kp164 LPS									
A	→2)-β-D-Ribf-(1→	5.39 107.7	4.26 80.1	4.26 70.9	4.08 83.8	3.84,3.68 63.4			
B	→4)-α-D-Galp-(1→	5.20 98.1	3.85 69.1	4.06 70.2	4.10 77.9	4.16 71.7	3.75 ^{nr} 62.1		
K	β-Kdop-(2→	nd	nd	nd, 2.18 nd	nd nd	nd nd	nd nd	nd nd	nd nd
Kp164 OPS									
A	→2)-β-D-Ribf-(1→	5.39 107.7	4.26 80.1	4.26 70.9	4.08 83.8	3.84,3.68 63.3	-		
B	→4)-α-D-Galp-(1→	5.19 98.2	3.85 69.1	4.06 70.3	4.10 77.9	4.16 71.6	3.75 ^{nr} 62.2		

^{a)} Spectra were recorded for LPS. Chemical shifts for protons and carbons were determined based on HR-MAS NMR ^1H , ^{13}C HSQC-DEPT spectra; nd – not determined; nr – not resolved.

7.4.2 Genetic analysis of the *rfb* regions

DNA isolated from 3 isolates (Kp164, Kp165 and Kp166) was sequenced. The obtained genomes in the form of assembled contigs were subjected to further analyses. Kaptive-based O serotyping was performed with whole-genome sequences of three strains (Wick *et al.* 2018). The O serotypes were predicted to be O4 for all isolates. These strains revealed the same structure of the *O loci* and high match confidence between their all six *rfb* genes and those of the O4 reference strain of *K. pneumoniae* (GenBank accession number: LT174605), according to the Kaptive measures of match quality (Table 13). The alignments of the *rfb* clusters from Kp164, Kp165, Kp166 strains are shown in Figure 34. All 3 isolates have complete *rfb* region. Moreover, the *rfb* sequences of these isolates are identical. Bioinformatics analyses confirm the results of structural analyses and the representation of the known O4 serotype in this group.

Table 13. Kaptive-based O serotype prediction for *K. pneumoniae* Kp164, Kp165 and Kp166 isolates.

Assembly	Best match <i>locus</i>	Match confidence	Coverage	Identity	Length discrepancy	Expected genes in <i>locus</i>	Expected genes in <i>locus</i> , details
Kp164	O4	Very high	100.00%	99.71%	+1 bp	6 / 6	O4_01_hydrolase,99.499%;O4_02_hypothetical,99.748%;O4_03_hypothetical,99.858%;O4_04_wzt,100.0%;O4_05_wzm,98.77%;O4_06_glycosyltransferase,99.738%
Kp165	O4	Very high	100.00%	99.71%	+1 bp	6 / 6	O4_01_hydrolase,99.499%;O4_02_hypothetical,99.748%;O4_03_hypothetical,99.858%;O4_04_wzt,100.0%;O4_05_wzm,98.77%;O4_06_glycosyltransferase,99.738%
Kp166	O4	Very high	100.00%	99.71%	+1 bp	6 / 6	O4_01_hydrolase,99.499%;O4_02_hypothetical,99.748%;O4_03_hypothetical,99.858%;O4_04_wzt,100.0%;O4_05_wzm,98.77%;O4_06_glycosyltransferase,99.738%

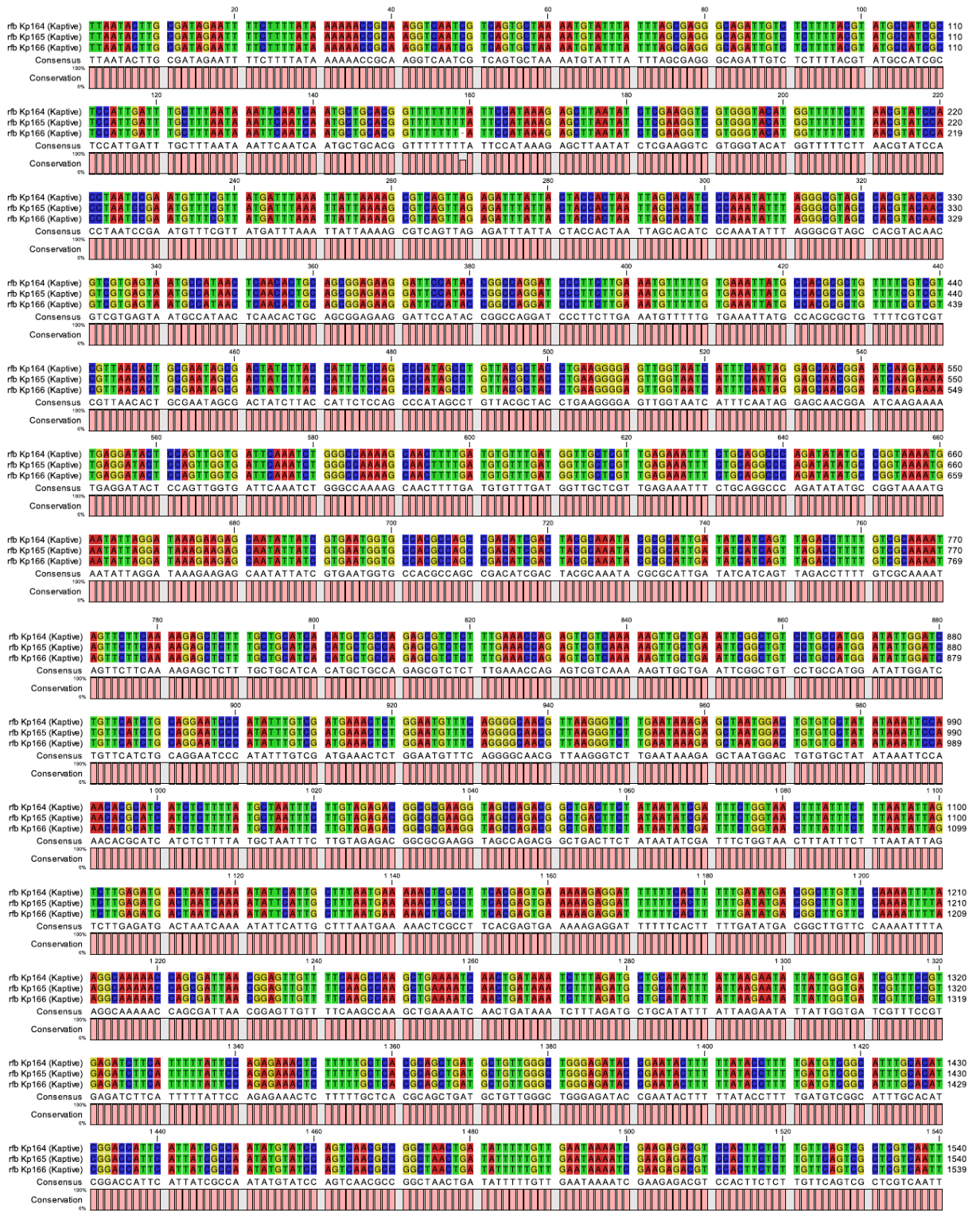


Figure 34. Part of the Kp164, Kp165 and Kp166 *rfb* regions alignment (from 1 to 1540 bp) to demonstrate *rfb* regions identity. The rest part of *rfb* regions also revealed identity (data not shown).

7.4.3 MALDI-TOF Biotyper analysis of *K. pneumoniae* PCM15 reference strain

To explain identification of the O4 serotype for Kp164, Kp165 and Kp166 assigned previously as nontypeable, the species of the PCM15 strain was verified by MALDI Biotyper. Contrary to the PCM description as *K. pneumoniae*, the PCM15 strain was identified as *Enterobacter cloacae* (Figure 35).

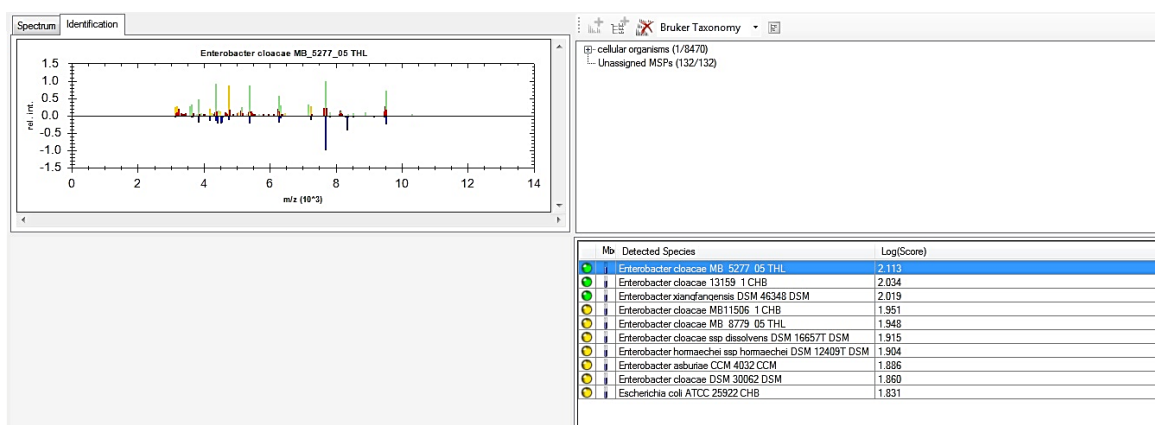


Figure 35. Screenshot with MALDI Biotyper result for PCM15 strain.

7.5. O antigen and O loci analysis of Kp174 and Kp177

7.5.1 Structural analysis

For the preliminary characterization of the O antigen structure, the OPS regions of all of the study strains were analysed directly in isolated LPSs by the HR-MAS NMR spectroscopy. The comparison of the ¹H HR-MAS NMR spectra of LPS (Figure 9a) indicated identity of the structures in all of the isolates.

To elucidate the complete O antigen structure, the LPS of the Kp174 strain was chosen as a representative, and hydrolysed to obtain poly- and oligosaccharides, followed by the fractionation by TSKgel®G2500PW. This yielded two high molecular weight fractions (Figure 36), collected as fraction 1 and 2.

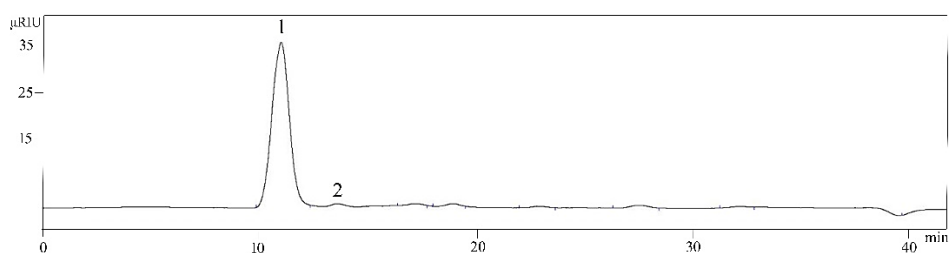


Figure 36. Chromatogram of poly- and oligosaccharides obtained by hydrolysis of Kp174 LPS (35 mg) and fractionated on TSKgel®G2500PW column (flow: 1 ml/min).

The ^1H NMR spectroscopy did not show any sugar components in the fraction 1 and 2a (Figure 36 and 37).

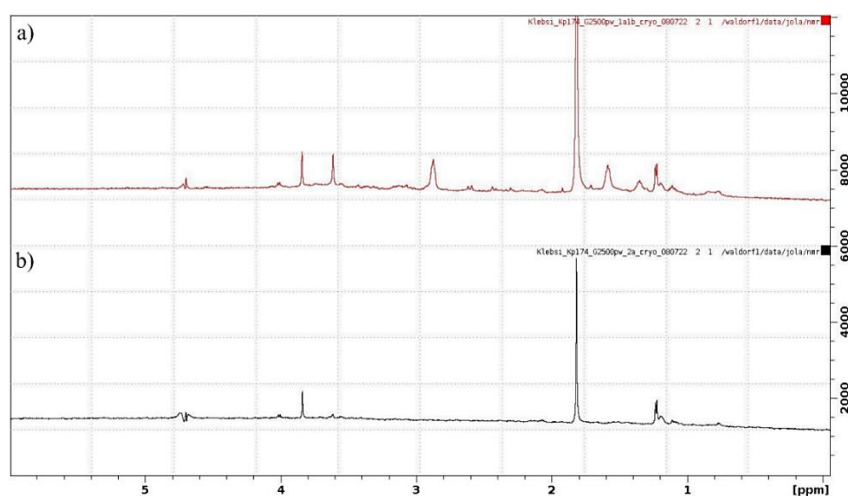


Figure 37. Comparative NMR analysis of the *K. pneumoniae* Kp174 fractions. ^1H NMR spectra of the (a) fraction 1 and (b) fraction 2.

We noted the absence of the OPS fraction, even though high molecular weight fractions were visible in SDS-PAGE analysis (Figure 10). Observed discrepancy between SDS-PAGE results and structural analyses may be due to the presence of capsular antigen contaminating LPS sample. To support rough character of group 4 LPSs, Kp177 LPS was also isolated and hydrolysed, giving identical results suggesting the rough character of its LPS (data not shown).

7.5.2 Genetic analysis of the *rfb* regions

Kaptive-based O serotyping was performed with whole-genome sequences of both strains (Wick *et al.* 2018). Contrary to the structural analysis, the O serotypes were predicted to be O1v1? for Kp174 and Kp177 isolate (Table 15). But the sequence of the *rfb* region is located on a few contigs – Kaptive was unable to serotype these isolates correctly.

The more detailed comparison of *rfb* clusters with the reference O locus showed the lack of *wbbO* gene in both the Kp174 and the Kp177 isolate in comparison to the reference O1v1 strain. *wbbMNO* genes in O1/O2 serotypes encodes glycosyltransferases responsible for the OPS repeating unit synthesis. Therefore, the results may suggest that lack of WbbO protein may prevent the production of OPS. However, it is worth noting that the results obtained with Kaptive may indicate the reason for the truncated *rfb* cluster located on several contigs. This might be the effect of technical obstacles in short-read sequencing.

Table 14. Kaptive results for Kp174 and Kp177 isolates^{a)}.

Assembly	Best match locus	Match confidence	Problems	Coverage	Identity	Length discrepancy	Expected genes in locus	Expected genes in <i>locus</i> , details	Missing expected genes
Kp174	O1v1	Good	?-	100.00%	99.98%	n/a	6 / 7	O1/O2v1_01_wzm,100.0%; O1/O2v1_02_wzt,100.0%; O1/O2v1_03_wbbM,100.0%; O1/O2v1_04_glf,100.0%; O1/O2v1_05_wbbN,100.0%; O1/O2v1_07_kfoC,100.0%	O1/O2v1_06_wbbO
Kp177	O1v1	Good	?-	100.00%	99.98%	n/a	6 / 7	O1/O2v1_01_wzm,100.0%; O1/O2v1_02_wzt,100.0%; O1/O2v1_03_wbbM,100.0%; O1/O2v1_04_glf,100.0%; O1/O2v1_05_wbbN,100.0%; O1/O2v1_07_kfoC,100.0%	O1/O2v1_06_wbbO

^{a)} The “?” character indicates that the *O locus* sequence was not found in a single piece in the assembly. The “-” character indicates that one of the expected genes was not found in the gene BLAST search. Missing genes and larger discrepancies between the *rfb* region of the analysed strains and the *rfb* region of the reference isolate are marked in red. n/a – not available

8. Discussion

Klebsiella pneumoniae is a Gram-negative bacterium belonging to the family Enterobacteriaceae. It is known for its clinical significance as an opportunistic pathogen causing a range of infections in humans. It is the pathogen pointed out by the WHO as top priority pathogen regarding multi-drug resistance problem and limited options of infection treatment (Holt *et al.* 2015; Tacconelli *et al.* 2018).

Strains of *K. pneumoniae* are known for their resistance to multiple antibiotics, including ESBLs and carbapenemases, posing challenges in clinical management. Due to the increasing prevalence of antibiotic-resistant strains, treatment options for *K. pneumoniae* infections can be limited. Surveillance is critical for monitoring the prevalence of drug-resistant strains and implementing strategies to weaken their spread. Therefore novel therapeutic strategies against infections and development of tools facilitating monitoring of clinical isolates are necessary (Grundmann *et al.* 2017; Wyres, Lam, and Holt 2020).

Some of the new therapeutic approaches under development are based on concepts of active and passive immunization against major surface antigens of *K. pneumoniae* – the K antigen and O antigen. Several vaccine strategies targeting the most prevalent K antigens of *Klebsiella* have been developed, particularly against the most virulent serotypes – K1 and K2 (Edelman *et al.* 1994; Hegerle *et al.* 2018; Feldman *et al.* 2019).

K. pneumoniae LPS (O antigen) was also proposed as potential target for active or passive immunization as an alternative to antibiotic treatment (Follador *et al.* 2016; Szijártó *et al.* 2016). Recently, promising bactericidal and neutralising monoclonal antibodies targeting the most common *K. pneumoniae* O serotypes (O1, O2, O3, O5) have been developed (Szijártó *et al.* 2016; Guachalla *et al.* 2017; Stojkovic *et al.* 2017; Rollenske *et al.* 2018). Moreover, the most clinically relevant O1, O2, O3, and O5 antigens are considered as a part of prototype glycoconjugate vaccines that would ensure an 80% coverage against *Klebsiella pneumoniae* infections (Follador *et al.* 2016). Therefore, most of the preclinical development of vaccines focused on these four serotypes (Del Bino *et al.* 2022).

All commercially available vaccines are based on the chemical conjugation of capsular polysaccharides with carrier proteins and due to their ability to control various

bacterial infections, more vaccines targeting the unmet the medical needs are either in development at the preclinical level or in clinical trials. Thus, in the case of *K. pneumoniae*, synthetic vaccines aimed at the sugar arsenal are currently in the preclinical phase as was summarised in section 4.3 of the Introduction (Del Bino *et al.* 2022).

If newly developed O antigen-based strategies against *Klebsiella* infections (therapeutic monoclonal antibodies or vaccines) are going to be effective, the knowledge about novel O serotypes and their structure diversity among clinical isolates is mandatory. Nowadays, laborious structural analysis may be effectively supported by O serotype (O *locus*) prediction by bioinformatic tools such Kaptive used in these studies. Owing to their universality, reproducibility, varied resolution, and standardized high-throughput protocols, molecular biology methods have become an excellent tool for pathogen characterization, finding wide application in microbiology diagnostics and surveillance. In recent years, these have been revolutionized by WGS, an increasingly common approach used in public health laboratories for the control of antimicrobial resistance or bacterial genotyping. At present, WGS is also successfully used to complement laborious structural chemical analyses, such as those of microbial surface antigens, being the key pathogenicity factors as well as critical targets for vaccines and therapeutic strategies (Quainoo *et al.* 2017; Artyszuk and Wołkowicz 2018; Wołkowicz 2018)].

As the molecular genetics of the *K. pneumoniae* O and K antigen biosynthesis has been well-elucidated, new O-genotyping techniques have been demonstrated to be useful for O serotype prediction. Thus, the major message from WGS genotyping studies is an indication of new *rfb loci* encoding LPS O antigen that differ in their composition from O *loci* encoding known 11 O serotypes of *K. pneumoniae*. Nowadays, the *K. pneumoniae* O antigen diversity may be traced easily using online bioinformatic tools such as Kaptive (Wick *et al.* 2018) and Kleborate (Lam *et al.* 2022), using rapidly growing global deposits of sequenced genomes. The increase of *K. pneumoniae* genomes number was observed even in the studies reported herein. In December 2019, 8130 genomes were available in public databases for IS survey (Artyszuk *et al.* 2020), whereas for OL101 *locus* identification, a total of 71377 genomes of *K. pneumoniae* were available in July 2023 (Artyszuk *et al.* 2024).

Klebsiella pneumoniae O serotype prediction has been based on variability of *rfb* gene clusters within the genomic O *loci*, seven of which have been correlated with known

antigen chemical structures so far (O1/O2v1, O1/O2v2, O1/O2v3, O3/O5, O4, O8, O12) (Follador *et al.* 2016; Wick *et al.* 2018; Lam *et al.* 2022). There are several useful examples of tracking O or K antigen diversity among *K. pneumoniae* isolates (Fang *et al.* 2016; Follador *et al.* 2016; Wick *et al.* 2018). For example, Fang *et al.* used a PCR-based O genotyping approach to explore the distribution of the O antigen genetic determinants in 87 clinical *K. pneumoniae* strains, showing a high prevalence of O1 (~57%), followed by the O2a, O3, and O5 O genotypes (Fang *et al.* 2016). Moreover, researchers identified 5% of nontypeable strains within 87 genomes of clinical isolates (Fang *et al.* 2016). Follador *et al.* analysed over 500 whole-genome sequences and reached a similar conclusion: that O1, O2, and O3 serotypes were the most common, with approximately 80% of all isolates. Moreover, Follador *et al.* analysed WGS data for global collection of 216 strains and identified 2,4-17% nontypeable strains (Follador *et al.* 2016). Finally, Wick *et al.* presented the user-friendly Kaptive Web, an online tool for the rapid typing of *K. pneumoniae* surface polysaccharide *loci* and demonstrated its utility using more than 500 *K. pneumoniae* genomes (Wick *et al.* 2018).

Recently, several O1 and O2 subtypes have been associated with specific combinations of O1/O2 *loci* with or without additional genes characteristic for O2a, O2afg, O2ac, and O2aeh antigens (galactan I) (Clarke *et al.* 2018; Kelly *et al.* 2019) and as a consequence Kaptive's algorithm for O serotype prediction has been updated recently (Lam *et al.* 2022). Even though, to date no information has been available about O antigen structures associated with five novel O *loci* identified recently by the bioinformatic approach (OL101, OL102, OL103, OL104, and OL105).

Nontypeable *Klebsiella pneumoniae* represent a distinct and challenging subset within the *Klebsiella* genus. The clinical significance of nontypeable *K. pneumoniae* is an important, global problem, with an increasing number of studies associating nontypeable *K. pneumoniae* with a spectrum of infections. Nontypeable strains have been implicated in a range of infections, including respiratory tract infections, urinary tract infections, bloodstream infections, and wound infections. The elusive nature of nontypeable *K. pneumoniae* for traditional serological typing methods poses diagnostic challenges, necessitating the development of alternative diagnostic strategies for accurate identification and surveillance (Fang *et al.* 2016; Follador *et al.* 2016).

The overall goal of this PhD research was examination of the eleven selected clinical isolates of *K. pneumoniae* for the presence of novel O serotypes or novel structural modifications of O antigens. In course of the EUROSTARS Klebsicure project

(2012-2015) 10 nontypeable clinical isolates of *K. pneumoniae* have been identified that could not be typed by available methods (serological or genetic). Therefore, the analyses undertaken in this doctoral thesis were:

1. Determination of complete structures of O antigens isolated from selected strains of *K. pneumoniae*.
2. Whole-genome sequencing and genomic analysis of O loci encoding O antigens by bioinformatic tools.
3. Determination of the frequency of occurrence of new O13 serotype or IS-dependent variations among publicly available genomes of *K. pneumoniae*.

As results of the analyses summarised in Table 16, it was shown that:

1. **For *K. pneumoniae* BIDMC 7B and ABC152 isolates, structural analyses showed O2v1 serotype, while Kaptive-based genotyping results showed O2v2 serotype.** The ISR1 disruption completely inactivated the GmlB glycosyltransferase gene, resulting in biosynthesis of the O2v1 instead of O2v2 structure, and being the reason for the discrepancy between the O antigen phenotype and O locus prediction (Table 15). A variety of ISs have been identified in *K. pneumoniae*, including the common elements ISR1, ISKpn14, ISKpn26, and IS903B. Moreover, the large-scale *in silico* analysis showed ~2,40% O2v2 isolates revealed a significant difference in length of the *rfb* region. Among the Kaptive-identified O1v2 isolates, significant length discrepancies occurred in the *rfb* region of ~0,7% isolates (Artyszuk *et al.* 2020).
2. **ABC122, BC738, BC13-986, and 3936/19 isolates are characterised by new O13 serotype.** The OL101 locus has been deciphered by elucidation of O antigen chemical structure as β -Kdop-(2→[→3)- α -L-Rhap-(1→4)- α -D-Glcp]_n numbered as O13 according to valid numbering system (Table 15) (Artyszuk *et al.* 2024). The study has also indicated O13 to be notably represented in the species (6,55% of 71377 genomes), including clinically relevant isolates.
3. **Both structural and genetic analyses of isolates Kp164, Kp165 and Kp166 revealed previously known serotype O4** (Hansen *et al.* 1999) (Table 15). These isolates, although initially classified as nontypeable, ultimately turned out to belong to the O4 serotype. It was the effect of incorrect assignment of the reference strain obtained from the Polish Collection of Microorganisms, PCM15 as O4:K15. In fact, this strain was identified as *Enterobacter cloacae* by MALDI Biotyper analysis.

4. **Kp174 and Kp177 isolates have rough LPS (devoid of OPS)** (Table 15). Loss of OPS is a natural process occurring in bacteria for example as a result of multiple passages of cultures in a laboratory. Although the original cultures obtained from the primary one sample were used in this studies, it is not known how and how many times the bacteria were previously passaged.

Table 15. Summary of the O serotype analyses for eleven nontypeable *K. pneumoniae* isolates^{a)}.

Strain	Sequence Type (ST)	LPS profile (SDS-PAGE)	O serotype (NMR)	O locus (Kaptive)	O serotype
BIDMC 7B	ST258	S-LPS	O2v1	O2v2?*	O2v1
ABC152	ST147				
ABC122	ST11	S-LPS	O13	OL101*	O13
3936/19	ST1427				
BC738	ST485				
BC13-986	ST3658			OL101?*	
Kp164	ST37	S-LPS	O4	O4	O4
Kp165					
Kp166					
Kp174	ST14	SR-LPS	R-LPS	O1v1?-	lack of OPS
Kp177					

^{a)} S-LPS, smooth LPS containing O-specific polysaccharide (OPS); SR-LPS, smooth-rough LPS containing short OPS, usually one RU; R-LPS, rough LPS (lack of OPS). Smooth or rough character of LPS was analysed by silver stained SDS-PAGE. Colour marks the group of isolates characterised by identical O serotype according to NMR data. The “*” character indicates that one or more of the expected genes falls below the identity threshold (default 95%). The “?” character indicates that the O locus sequence was not found in a single piece in the assembly. The “-” character indicates that one of the expected genes was not found in the gene BLAST search.

As a result of the research, three interesting phenomena were observed:

1. The impact of ISs on phenotypes of O2 and O1 antigens.
2. Novel O13 serotype identified.
3. Rough character of clinically relevant strains of *K. pneumoniae* (CRE isolates).

All of findings, if implemented into bioinformatic algorithms may improve predictions made by high-throughput methods of pathogen identification (for example O serotype prediction).

Roughness of clinically relevant strains was previously reported (Moran, Lindner, and Walsh 1997; Janusch *et al.* 2002). Moran *et al.* described rough character of clinical isolate of *Helicobacter pylori* LPS (Moran, Lindner, and Walsh 1997). Janusch *et al.* characterized the rough isolate of *Salmonella enterica* sv. Minnesota (Janusch *et al.*

2002). The rough phenotype may increase the immunogenicity of minor surface antigens, such as porins and lipoprotein antigens, which may improve cross-protection against heterologous bacteria (Nagy *et al.* 2008). Moreover, K antigen was reported as possibly inhibitor of O antigen interaction with antibodies directed against O-antigen. Furthermore, it has been reported that the K antigen is largely inhibitor of antibodies raised against the O antigen in vaccines (Wantuch *et al.* 2023). Hence, the conclusion that perhaps long LPS is not always crucial for bacterial virulence. Expression of rough LPS is caused by crucial gene disruptions. For example in rough Kp174 and Kp177 the *wbbO* genes were missing and strains were predicted as O1v1 serotype. Broadening the knowledge concerning relationships between O *locus* elements and O antigen phenotype will improve future bioinformatic tools used for predictions. However, as a prerequisite functions of all genes have to be deciphered first, since some of them were not characterized to date (See section 4.2.6.3.2 in the Introduction). Moreover, novel O *loci* has to be defined by structural analyses of encoded O antigens.

Exact fast prediction will enable monitoring of *K. pneumoniae* antigen drift that is vulnerable to the selective pressure of therapies and vaccines, such as these targeted O antigens. The emergence of pandrug-resistant strains of *Klebsiella* highlights the critical need for implementation of effective strategies for prevention and transmission of *Klebsiella* infections.

Modifications of described *K. pneumoniae* serotypes may be one of the reasons for the inability to obtain a clear serotyping result. These modifications may be the result of mutations in the genes encoding proteins responsible for the biosynthesis of the specific O chain.

In this study, two cases of genotype–phenotype discrepancies for O antigens in the *K. pneumoniae* clinical isolates BIDMC 7B and ABC152 have been presented (Artyszuk *et al.* 2020). The phenotype was determined as O2v1, whereas Kaptive predicted O2v2 serotype. However, the tool provided an alert concerning length disruption within the *rfb* region and recommended further analyses. As for BIDMC 7B and ABC152 isolates, the structural analysis by the HR-MAS NMR spectroscopy demonstrated the O antigen structure to be O2v1 (Figure 9).

The ISR1 element was identified in the *gmlB* gene, one of the three responsible for the D-galactan I conversion from v1 [\rightarrow 3)- β -D-Galf-(1 \rightarrow 3)- α -D-Galp-(1 \rightarrow)] to v2 [\rightarrow 3)- β -D-Galf-(1 \rightarrow 3)-[α -D-Galp-(1 \rightarrow 4)]- α -D-Galp-(1 \rightarrow)]. The presence of IS, identified O antigen structures, and the lack of other obvious differences between the

analysed genomes and genome of O2v1 reference strain indicated that the IS disruption was the reason for the difference between the O antigen phenotype and the Kaptive-based prediction. The large-scale *in silico* analysis of publicly available genomes of *K. pneumoniae* O2v2 and O1v2 clearly showed that various insertions have occurred in several *rfb* fragments, possibly causing similar divergences between the O serotype prediction and phenotype. A variety of ISs have been identified, including the common elements ISR1, ISKpn14, ISKpn26, and IS903B. As only structural verification in each strain could provide definite proof of the O antigen phenotype, the *in silico* survey only suggested the influence of IS on the O antigen chemical structure. By analogy with the BIDMC 7B and ABC152 strains, fourteen O2 strains (e.g., ASM170423, CHS57, and IS39) revealed ISs in the *gmlABC* region with higher prevalence of *gmlB* and *gmlC* disruptions, likely representing similar genotype–phenotype discrepancies that may result in O2v1 phenotype (the lack of terminal Gal residue in the RU). Three O2 strains with an IS in the *gmlABC* region had additional disruption within the *wzm–wbbO* region, suggesting failure of the O antigen biosynthesis and the rough form of LPS (i.e., ASM307130, UCI 38, and BIDMC 13). Other identified cases also suggested O antigen biosynthesis failure, including O1v2 isolates.

Regarding the genetic background of O1 and O2 antigen biosynthesis, the presence of an IS in the O *locus* may influence the O antigen phenotype by: (i) O2v2 to O2v1 or O1v2 to O1v1 conversion; or (ii) conversion from smooth to rough LPS. It is noteworthy that the results obtained from Kaptive depend on the IS location. In the case of gene disruption or frameshift mutation, the results indicate the lack of an enzyme specific for the analysed O serotype, that may contribute to the false serotype prediction by the algorithm. For instance, the Kaptive results for the IS39 isolate show the absence of the *wbbM* and *gmlABC* genes responsible for encoding glycosyltransferase as well as encoding three putative glycosyltransferases, that modify D-galactan I (O2a; O2v1) to branched polymer of O2afg (O2v2) or O2aeh (O2v2), respectively (Table 7).

Kaptive-supported differentiation between *K. pneumoniae* O1/O2v1 and O1/O2v2 serotypes based on two steps: First, Kaptive recognizes the serotype by searching for the D-galactan-II-encoding genes (*wbbY*, *wbbZ*) characteristics of the O1 serotype. Next, the O1 and O2 serotypes are distinguished by the analysis of genes found in the *rfb* cluster. Finally, these are reported as variant v1 or v2 (Wick *et al.* 2018). As the final result, the tool prediction is accompanied by length discrepancy information, which may indicate the possibility of some rearrangements in the *rfb* region.

For instance, the Kaptive results for the isolate IS39 indicated the absence of the *wbbM* and *gmlABC* genes. Detailed analysis of the *rfb* region has shown the presence of these genes with a *gmlC* IS disruption and point mutations in the other three genes (Table 7). Although Kaptive suggests the possibility of the presence of IS by reporting differences in length discrepancy, it is worth analysing the nucleotide sequence of the *rfb* region more precisely, in order to exclude falsely predicted serotypes based on errors occurring during O antigen genotyping (O locus prediction). Sequence analysis of isolates from the database confirmed the occurrence of many IS insertions in the *rfb* region; however, the frequency of such events is hard to evaluate. Although there were no previous data on IS disruptions within the O locus, in general, these elements are common in *K. pneumoniae* genomes (Adams, Bishop, and Wright 2016).

The hyperepidemic clone ST258, characterized by the notorious production of KPC-type carbapenemases and extensive drug resistance, has more ISs than an *K. pneumoniae* isolates of other ST (Adams, Bishop, and Wright 2016; Baraniak *et al.* 2017). Several IS types are notably frequent in ST258, such as ISKpn26 (Adams, Bishop, and Wright 2016; Lev *et al.* 2017). In our study, the majority of O2 isolates identified belonged to ST258 (~70%), and ISKpn26 was commonly found in these analysed group of genomes (~50%). These data further emphasized the impact of ISs on the evolution of *K. pneumoniae* ST258; however, one must also consider the over-representation of ST258 genomes in public databases, resulting from the high clinical and epidemiological relevance of these pathogens.

According to Adams *et al.*, 94% of *K. pneumoniae* strains have at least one IS in their genome, where transposition of these elements within the genome causes rearrangements and may create new genotypes (Adams, Bishop, and Wright 2016). One consequence of an IS disruption of the *rfb* cluster genes may be the protection of bacteria against the host immune system. Although the presence of ISs in *gmlABC* genes may cause the phenotype–genotype discrepancy discussed above, the disruption of the genes determining D-galactan I elements may abolish the O antigen synthesis, as in the case of the *wzm* or *wzt* genes, coding for ABC transporters (Follador *et al.* 2016; Clarke *et al.* 2018). Insertion sequence elements in the *rfb* and/or *wbbYZ* operon can either inhibit the expression of the O antigen on the surface of the bacterial cell (Fang *et al.* 2016) (resulting in the rough phenotype) or cause the switch from one serotype to another (Artyszuk *et al.* 2020). In both cases, the change in phenotype can alter or impair the virulence of the bacteria (Lukacova, Barák, and Kazár 2008).

Structural large-scale analysis of *K. pneumoniae* isolates could determine the consequences of ISs in the *rfb* region and their effects on bacterial antigenicity and host interactions. Such changes can significantly affect the ability of bacteria to survive during antibacterial therapies; for example, by changing the surface antigens and virulence in regard to reactivity with the complement, antibodies, or phage resistance (Siguier *et al.* 2006). The antigenic drift of LPS could present possible pathway for bacteria to avoid the immune system or bactericidal antibodies. This is the case, for example, in the *Salmonella* species, the O antigen composition of which affects the host–pathogen interactions during infection. Strains belonging to one serovar can have a different repertoire of O antigen-modifying genes. Moreover, their expression is different depending on the phase variations. The *gtrABC* operon acquired by horizontal gene transfer is a set of genes responsible for modification of the *Salmonella* O antigen. These genes encode proteins showing functional homology to the glycosyltransferases encoding by the *gmlABC* genes cluster in *K. pneumoniae* (Mann and Whitfield 2016; Kintz *et al.* 2017). This study showed that some *K. pneumoniae* isolates, flagged by Kaptive as having length discrepancy within the *rfb* locus and advised for further analysis, may be basically mis-O-serotyped by the tool.

Results of our studies may improve algorithms used for O serotype prediction by providing a knowledge about new O serotype of *K. pneumoniae*. To date no information has been available about O antigen structures associated with five novel O loci identified recently by the bioinformatic analyses (OL101-OL105). Of these, OL101 has been of particular interest, having occurred in ~6,55% of isolates in the original dataset screened, which suggested the likely new O serotype to be of notable clinical and/or epidemiological importance (Follador *et al.* 2016). A significant correlation was found between the OL101 and the asymptomatic *K. pneumoniae* carriage (Follador *et al.* 2016).

In this work four *K. pneumoniae* strains were predicted by Kaptive to carry the OL101 *rfb* locus. Classified into different STs, the isolates were clonally non-related, and all they were MDR, including two MBL-type carbapenemase producers. In order to resolve their O antigen chemical structure, two types of NMR datasets had to be combined, namely for the isolated OPS (to identify RU) and for the complete LPS molecules (to identify Kdo terminus). For all the strains these analyses revealed the novel disaccharide RU structure capped by the terminal β -Kdop residue as the O chain nonreducing terminus: β -Kdop-[\rightarrow 3)- α -L-Rhap-(1 \rightarrow 4)- α -D-Glcp-(1 \rightarrow)]_n. Since the

linkage between β -Kdo and $\rightarrow 3$)- α -L-Rhap is labile in the acidic condition, the use of the ^1H , ^{13}C HR-MAS NMR spectroscopy allowed for its observation in the intact LPS molecule.

Regarding the OPS terminus (β -Kdop), the novel antigen resembles the O12 structure: β -Kdop-(2 \rightarrow 3)[\rightarrow 4)- α -L-Rhap-(1 \rightarrow 4)- α -D-GlcNAcp]_n (Vinogradov *et al.* 2002), with a few structural differences within the RU. These included 4-substituted Glcp being a part of the O13 RU instead of GlcpNAc, and the Rhap residue being substituted at position 3 within the O13 RU. In the new O antigen the terminal residue of β -Kdo is linked to O-3 of the Rha residue, the same position as in the RU. In contrast to O12 (Vinogradov *et al.* 2002), we have not identified even minor amounts of O antigens devoid of the Kdo terminus. However, such nonstoichiometric substitution cannot be excluded since, like in the O12 LPS, it may only apply to a minority of OPS molecules, and thus might be nondetectable during the HR-MAS NMR analysis and mass spectra of methylation analysis. Considering all of the above, the structural analyses performed for the four *K. pneumoniae* strains have deciphered the structure specified by the OL101 locus as the novel O13 serotype of this species.

The Kaptive-based comparative analysis revealed high and very similar sequence identity scores between the genes of the ABC122 and 3936/19 isolates and the reference OL101 locus from the Kaptive database, with relatively lower homology levels in the glycosylhydrolase (~92%) and *wbbB* (~95%) genes. The difference between ABC122 and 3936/19 was that more than 50% of the *rfbB* gene was missing in sequence assemblies of the 3936/19 isolate. However, this might be a result of technical obstacles in short-read sequencing. As it was mentioned above, the chemical analysis demonstrated the presence of the O13 antigen structure in this isolate as well, indicating functionality of its O locus.

Based on their gene content and order, OL101, OL102, and O12 have been previously described as having closely related *rfb* clusters, differing by the presence and absence of single genes (Follador *et al.* 2016; Williams *et al.* 2017). Compared with the O12 cluster, OL101 includes the additional glycosylhydrolase gene, while OL102 lacks the terminal glycosyltransferase gene but possesses an additional rhamnan synthesis gene (Figure 7). Despite these similarities, antigen structures encoded by OL101 and O12 loci differed a lot indicating that structural analysis combined with functional analysis of individual genes is necessary in order to explain the genotype-phenotype relationship for the O12 and O13 antigens.

In this study we have comprehensively analysed four *K. pneumoniae* clinical isolates, identified by the Kaptive as having the new OL101-type O antigen-encoding *loci*. The O antigen structure turned out to be unique, and was numbered O13, according to the current classification of the *K. pneumoniae* O serotypes (Kelly *et al.* 1995; Hansen *et al.* 1999; Szijártó *et al.* 2016; Clarke *et al.* 2018; Follador *et al.* 2016; Kelly *et al.* 2019, 2023). The Kaptive-based O *locus* screening of all 71377 *K. pneumoniae* genomes available in July 2023, revealed that ~6,55% of these were predicted to be OL101. This value was consistent with the results of Follador *et al.* (Follador *et al.* 2016), and confirmed that the new O serotype should be considered in workflows relevant to new therapeutic strategies based on protective antibodies. It should also be underlined that the study O13 strains were of high clinical importance. The presented data provide further insight into structural modifications of *K. pneumoniae* LPS which may influence binding of therapeutic or diagnostic antibodies and enhance O antigen typing and development of novel *Klebsiella* control strategies. Despite supportive role for O and K antigen-based therapies, bioinformatic tools such as Kaptive or Kleborate are widely used for studies of relationships between O/K serotypes and other virulence factors of *K. pneumoniae* such as siderophores (Wyres, Lam, and Holt 2020; Biedrzycka *et al.* 2021; Lam *et al.* 2021).

As a methodology for the identification of an actual O antigen phenotype is not broadly available, we assume that the development of the Kaptive algorithm in described direction would increase quality of its predictions and usefulness, particularly for inexperienced users. Further development might be based on broader studies of isolates with non-clear O serotyping results or identified O genotype–phenotype discrepancies, with the use of structural analysis to precisely elucidate the O genotype–phenotype relationships. As O and K antigens represent target molecules for therapeutic strategies against *Klebsiella* infections, it is important to broaden our knowledge about genotype–phenotype relationships. Filling all detected gaps will improve serotype predictions based on bioinformatic tools.

9. Conclusions

1. The results indicated that the ISR1 disruption completely inactivated the GmlB glycosyltransferase gene, resulting in biosynthesis of the O2v1 instead of O2v2 structure, and being the reason for the discrepancy between the O antigen phenotype and genotype (Artyszuk *et al.* 2020).
2. The large-scale *in silico* analysis of publicly available genomes of *K. pneumoniae* O2v2 and O1v2 showed that various insertions have occurred in different *rfb* genes, possibly causing similar divergences between the Kaptive-predicted O serotype and O antigen phenotype (Artyszuk *et al.* 2020).
3. A variety of ISs have been identified in *K. pneumoniae* O1 and O2, including the common elements ISR1, ISKpn14, ISKpn26, and IS903B and suggested as possible modulators of O antigen phenotypes (Artyszuk *et al.* 2020).
4. In the O2 isolates identified belonged to the hyperepidemic clone ST258, ISKpn26 was commonly found IS (~50%) (Artyszuk *et al.* 2020). These data further emphasized the impact of ISs on the evolution of *K. pneumoniae* ST258.
5. The OL101 *locus* has been assigned to O antigen chemical structure: β -Kdop-[\rightarrow 3)- α -L-Rhap-(1 \rightarrow 4)- α -D-Glcp]_n and described as a new O13 serotype (Artyszuk *et al.* 2024).
6. The study has also indicated O13 to be notably represented in the species (~6,55%), including clinically relevant isolates (Artyszuk *et al.* 2024). Even though nowadays O serotypes characterised by such low prevalence are not qualified as relevant antigens for protective antibodies, trends of personalized medicine may change such strategy in future.

10. References

1. Adamo R. and Margarit I. 2018. “Fighting antibiotic-resistant *Klebsiella pneumoniae* with “sweet” immune targets.” *mBio*. 2018;9.
2. Adams, Mark D, Brian Bishop, and Meredith S Wright. 2016. “Quantitative Assessment of Insertion Sequence Impact on Bacterial Genome Architecture.” *Microbial Genomics* 2(7).
3. Alcántar-Curiel, María D. *et al.* 2013. “Multi-Functional Analysis of *Klebsiella pneumoniae* Fimbrial Types in Adherence and Biofilm Formation.” *Virulence* 4(2): 129–38.
4. Amor, Karen *et al.* 2000. “Distribution of Core Oligosaccharide Types in Lipopolysaccharides from *Escherichia coli*.” *Infection and Immunity* 68(3): 1116–24.
5. Argunov, Dmitry A. *et al.* 2019. “Convergent Synthesis of Oligosaccharides Structurally Related to Galactan I and Galactan II of *Klebsiella pneumoniae* and Their Use in Screening of Antibody Specificity.” *European Journal of Organic Chemistry* 2019(26): 4226–32.
6. Artyszuk, Daria *et al.* 2020. “The Impact of Insertion Sequences on O-Serotype Phenotype and Its O-Locus-Based Prediction in *Klebsiella pneumoniae* O2 and O1.” *International Journal of Molecular Sciences* 21(18): 1–15.
7. Artyszuk, Daria *et al.* 2024. “The OL101 O Antigen Locus Specifies a Novel *Klebsiella pneumoniae* Serotype O13 Structure.” *Carbohydrate Polymers* 326: 121581.
8. Artyszuk, Daria, and Tomasz Wołkowicz. 2018. “Zastosowanie sekwencjonowania pełnogenomowego do genotypowania bakterii.” *Postępy Mikrobiologii*: 15.
9. Bach, S., de Almeida, A. and Carniel, E. 2000. The *Yersinia* high-pathogenicity island is present in different members of the family Enterobacteriaceae. *FEMS Microbiology Letters*, 183(2), 289–294.
10. Bankevich, Anton *et al.* 2012. “SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing.” *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* 19(5): 455–77.
11. Baraniak, Anna *et al.* 2017. “Multiregional Dissemination of KPC-Producing *Klebsiella pneumoniae* ST258/ST512 Genotypes in Poland, 2010–14.” *Journal of Antimicrobial Chemotherapy* 72(6): 1610–16.

12. Biedrzycka, M. *et al.* 2021. "Dissemination of *Klebsiella pneumoniae* ST147 NDM-1 in Poland, 2015-19." *The Journal of Antimicrobial Chemotherapy* 76(10): 2538–45.
13. Del Bino, Linda *et al.* 2022. "Synthetic Glycans to Improve Current Glycoconjugate Vaccines and Fight Antimicrobial Resistance." *Chemical Reviews* 122(20): 15672–716.
14. Birnbaum, George I., Roy Rene, Jean Robert Brisson, and Harold J. Jennings. 1987. "Conformations of Ammonium 3-Deoxy-D-Manno-2-Octulosonate (KDO) and Methyl α - and β -Ketopyranosides of KDO: X-Ray Structure and ¹H NMR Analyses." *Journal of Carbohydrate Chemistry* 6(1): 17–39.
15. Björndal, Håkan *et al.* 1973. "Structural Studies of the *Klebsiella* Type 47 Capsular Polysaccharide." *Carbohydrate Research* 27(2): 373–78.
16. Bock, Klaus, and Christian Pedersen. 1974. "A Study of ¹³CH Coupling Constants in Hexopyranoses." *Journal of the Chemical Society, Perkin Transactions 2* (3): 293–97.
17. Bolger, Anthony M, Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics (Oxford, England)* 30(15): 2114–20.
18. Chang, De, Lokesh Sharma, Charles S. Dela Cruz, and Dong Zhang. 2021. "Clinical Epidemiology, Risk Factors, and Control Strategies of *Klebsiella pneumoniae* Infection." *Frontiers in Microbiology* 12: 750662.
19. Chhibber, S., Mamta Rani, and Vanashree Yadav. 2005. "Immunoprotective Potential of Polysaccharide-Tetanus Toxoid Conjugate in *Klebsiella pneumoniae* Induced Lobar Pneumonia in Rats." *Indian Journal of Experimental Biology* 43(1): 40–45.
20. Choma, Adam *et al.* 2014. "Occurrence of an Unusual Hopanoid-Containing Lipid A Among Lipopolysaccharides from Bradyrhizobium Species." *Journal of Biological Chemistry* 289: 35644–55.
21. Ciucanu, Ionel, and Francisc Kerek. 1984. "A Simple and Rapid Method for the Permethylolation of Carbohydrates." *Carbohydrate Research* 131(2): 209–17.
22. Clarke, Bradley R, and C Whitfield. 1992. "Molecular Cloning of the *rfb* Region of *Klebsiella pneumoniae* Serotype O1:K20: The Rfb Gene Cluster Is Responsible for Synthesis of the D-Galactan I O Polysaccharide." *Journal of Bacteriology* 174(14): 4614–21.

23. Clarke, Bradley R *et al.* 2018. “Molecular Basis for the Structural Diversity in Serogroup O2-Antigen Polysaccharides in *Klebsiella pneumoniae*.” *The Journal of Biological Chemistry* 293(13): 4666–79.
24. Clements, Abigail *et al.* 2008. “Targeting Subcapsular Antigens for Prevention of *Klebsiella pneumoniae* Infections.” *Vaccine* 26(44): 5649–53.
25. Corsaro, Maria Michela *et al.* 2005. “¹H and ¹³C NMR Characterization and Secondary Structure of the K2 Polysaccharide of *Klebsiella pneumoniae* Strain 52145.” *Carbohydrate Research* 340(13): 2212–17.
26. Diancourt, Laure *et al.* 2005. “Multilocus Sequence Typing of *Klebsiella pneumoniae* Nosocomial Isolates.” *Journal of Clinical Microbiology* 43(8): 4178.
27. Edelman, Robert *et al.* 1994. “Phase 1 Trial of a 24-Valent *Klebsiella* Capsular Polysaccharide Vaccine and an Eight-Valent *Pseudomonas* O-Polysaccharide Conjugate Vaccine Administered Simultaneously.” *Vaccine* 12(14): 1288–94.
28. European Centre for Disease Prevention and Control. Antimicrobial resistance surveillance in Europe 2011. Annual Report of the European Antimicrobial Resistance Surveillance Network (EARS-Net). Stockholm: ECDC; 2012.
29. Fang, Chi-Tai, Yun-Jui Shih, Cheng-Man Cheong, and Wen-Ching Yi. 2016. “Rapid and Accurate Determination of Lipopolysaccharide O-Antigen Types in *Klebsiella pneumoniae* with a Novel PCR-Based O-Genotyping Method.” *Journal of Clinical Microbiology* 54(3): 666–75.
30. Feldman, Mario F. *et al.* 2019. “A Promising Bioconjugate Vaccine against Hypervirulent *Klebsiella pneumoniae*.” *Proceedings of the National Academy of Sciences of the United States of America* 116(37): 18655–63.
31. Follador, Rainer *et al.* 2016. “The Diversity of *Klebsiella pneumoniae* Surface Polysaccharides.” *Microbial Genomics* 2(8): e000073.
32. Fresno, Sandra *et al.* 2006. “The Ionic Interaction of *Klebsiella pneumoniae* K2 Capsule and Core Lipopolysaccharide.” *Microbiology* 152(Pt 6): 1807–18.
33. Geissner, Andreas *et al.* 2019. “Microbe-Focused Glycan Array Screening Platform.” *Proceedings of the National Academy of Sciences of the United States of America* 116(6): 1958–67.
34. Gorshkova, Raisa P., *et al.* 1985. “Structural Studies on O-Specific Polysaccharides of Lipopolysaccharides from *Yersinia enterocolitica* Serovars O:1,2a,3, O:2a,2b,3 and O:3.” *European Journal of Biochemistry* 150(3): 527–31.

35. Grundmann, Hajo *et al.* 2017. “Occurrence of Carbapenemase-Producing *Klebsiella pneumoniae* and *Escherichia coli* in the European Survey of Carbapenemase-Producing Enterobacteriaceae (EuSCAPE): A Prospective, Multinational Study.” *The Lancet. Infectious Diseases* 17(2): 153–63.
36. Guachalla, Luis M. *et al.* 2017. “Discovery of Monoclonal Antibodies Cross-Reactive to Novel Subserotypes of *K. pneumoniae* O3.” *Scientific Reports* 7(1).
37. Hafiz, Taghreed A. *et al.* 2023. “*Klebsiella pneumoniae* Bacteraemia Epidemiology: Resistance Profiles and Clinical Outcome of King Fahad Medical City Isolates, Riyadh, Saudi Arabia.” *BMC Infectious Diseases* 23(1): 1–9.
38. Hansen, Dennis S. *et al.* 1999. “*Klebsiella pneumoniae* Lipopolysaccharide O Typing: Revision of Prototype Strains and O-Group Distribution among Clinical Isolates from Different Sources and Countries.” *Journal of Clinical Microbiology* 37(1): 56–62.
39. Hegerle, Nicolas *et al.* 2018. “Development of a Broad Spectrum Glycoconjugate Vaccine to Prevent Wound and Disseminated Infections with *Klebsiella pneumoniae* and *Pseudomonas aeruginosa*.” *PloS one* 13(9).
40. Holst, O. and Brade H. (1992) in: Bacterial endotoxic lipopolysaccharides (Morrison, D.C. and Ryan, J.L., Eds.), pp. 135–170, CRC, Boca Raton, FL.
41. Holt, Kathryn E *et al.* 2015. “Genomic Analysis of Diversity, Population Structure, Virulence, and Antimicrobial Resistance in *Klebsiella pneumoniae*, an Urgent Threat to Public Health.” *Proceedings of the National Academy of Sciences of the United States of America* 112(27): E3574-3581.
42. Hsieh, Pei-Fang *et al.* 2008. “Serum-Induced Iron-Acquisition Systems and TonB Contribute to Virulence in *Klebsiella pneumoniae* Causing Primary Pyogenic Liver Abscess.” *The Journal of Infectious Diseases* 197(12):1717–1727.
43. Janusch, Holger *et al.* 2002. “Structural and Biological Characterization of Highly Purified Hepta-Acyl Lipid A Present in the Lipopolysaccharide of the *Salmonella enterica* Sv. Minnesota Re Deep Rough Mutant Strain R595.” *Journal of Endotoxin Research* 8(5): 343–56.
44. Kelly, R F, M B Perry, L L MacLean, and C Whitfield. 1995. “Structures of the O-Antigens of *Klebsiella* Serotypes O2 (2a,2e), O2 (2a,2e,2h), and O2 (2a,2f,2g), Members of a Family of Related D-Galactan O-Antigens in *Klebsiella* Spp.” *Journal of Endotoxin Research* 2(2): 131–40.

45. Kelly, Steven D. *et al.* 2019. “*Klebsiella pneumoniae* O1 and O2ac Antigens Provide Prototypes for an Unusual Strategy for Polysaccharide Antigen Diversification.” *The Journal of Biological Chemistry* 294(28): 10863–76.
46. Kelly, Steven D. *et al.* 2023. “Identification of a Second Glycoform of the Clinically Prevalent O1 Antigen from *Klebsiella pneumoniae*.” *Proceedings of the National Academy of Sciences of the United States of America* 120(29): e2301302120.
47. Kintz, Erica *et al.* 2017. “*Salmonella enterica* Serovar Typhi Lipopolysaccharide O-Antigen Modification Impact on Serum Resistance and Antibody Recognition.” *Infection and Immunity* 85(4).
48. Knirel, Yuriy A. *et al.* 1992. “Structure of the Polysaccharide Chains of *Pseudomonas pseudomallei* Lipopolysaccharides.” *Carbohydrate Research* 233(C): 185–93.
49. Köhler, W. and Mochmann, H. 1987. “Carl Friedländer (1847–1887) and the discovery of the *Pneumococcus*—in memory of the centenary of his death.” *Zeitschrift für arztliche Fortbildung und Qualitätssicherung* 81, 615–618.
50. Komaniecka, Iwona *et al.* 2014. “Occurrence of an Unusual Hopanoid-Containing Lipid A Among Lipopolysaccharides from *Bradyrhizobium* Species.” *Journal of Biological Chemistry* 289(51): 35644–55.
51. Laemmli, U. K. 1970. “Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4.” *Nature* 227(5259): 680–85.
52. Lam, Margaret M.C. *et al.* 2021. “A Genomic Surveillance Framework and Genotyping Tool for *Klebsiella pneumoniae* and Its Related Species Complex.” *Nature Communications* 2021 12:1 12(1): 1–16.
53. Lam, Margaret M.C. *et al.* 2022. “Kaptive 2.0: Updated Capsule and Lipopolysaccharide Locus Typing for the *Klebsiella pneumoniae* Species Complex.” *Microbial Genomics* 8(3).
54. Larsen, Mette V *et al.* 2012. “Multilocus Sequence Typing of Total-Genome-Sequenced Bacteria.” *Journal of Clinical Microbiology* 50(4): 1355–61.
55. Lee, Woonghee, Marco Tonelli, and John L. Markley. 2015. “NMRFAM-SPARKY: Enhanced Software for Biomolecular NMR Spectroscopy.” *Bioinformatics (Oxford, England)* 31(8): 1325–27.
56. Letunic, Ivica, and Peer Bork. 2019. “Interactive Tree Of Life (ITOL) v4: Recent Updates and New Developments.” *Nucleic Acids Research* 47(W1).

57. Lev, Anastasia I *et al.* 2017. "Identification of IS1R and IS10R Elements Inserted into Ompk36 Porin Gene of Two Multidrug-Resistant *Klebsiella pneumoniae* Hospital Strains." *FEMS Microbiology Letters* 364(10).
58. Li, Yanyan *et al.* 2016. "Identification of Two Genes Encoding for the Late Acyltransferases of Lipid A in *Klebsiella pneumoniae*." *Current Microbiology* 73(5): 732–38.
59. Lin, Yi Tsung *et al.* 2012. "Seroepidemiology of *Klebsiella pneumoniae* Colonizing the Intestinal Tract of Healthy Chinese and Overseas Chinese Adults in Asian Countries." *BMC Microbiology* 12(1): 1–7.
60. Liu, Chao, and Jun Guo. 2019. "Hypervirulent *Klebsiella pneumoniae* (Hypermucoviscous and Aerobactin Positive) Infection over 6 Years in the Elderly in China: Antimicrobial Resistance Patterns, Molecular Epidemiology and Risk Factor." *Annals of Clinical Microbiology and Antimicrobials* 18(1): 1–11.
61. López-Camacho, Elena *et al.* 2014. "Genomic Analysis of the Emergence and Evolution of Multidrug Resistance during a *Klebsiella pneumoniae* Outbreak Including Carbapenem and Colistin Resistance." *Journal of Antimicrobial Chemotherapy* 69(3): 632–36.
62. Lu, Jing *et al.* 2023. "Lipid A Modification and Metabolic Adaptation in Polymyxin-Resistant, New Delhi Metallo- β -Lactamase-Producing *Klebsiella pneumoniae*." *Microbiology Spectrum* 11(4).
63. Lukacova, Magda, I. Barák, and J. Kazár. 2008. "Role of Structural Variations of Polysaccharide Antigens in the Pathogenicity of Gram-Negative Bacteria." *Clinical Microbiology and Infection* 14(3): 200–206.
64. Lukasiewicz, Jolanta *et al.* 2010. "Structural Analysis of the Lipid A Isolated from *Hafnia alvei* 32 and PCM 1192 Lipopolysaccharides." *Journal of Lipid Research* 51(3): 564–74.
65. Lundberg, Urban *et al.* 2013. "Identification and Characterization of Antigens as Vaccine Candidates against *Klebsiella pneumoniae*." *Human vaccines and Immunotherapeutics* 9(3): 497–505.
66. Mann, Evan, and Chris Whitfield. 2016. "A Widespread Three-Component Mechanism for the Periplasmic Modification of Bacterial Glycoconjugates." *Canadian Journal of Chemistry* 94(11): 883–93.
67. Martin, Marcel. 2011. "Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads." *EMBnet.Journal* 17(1): 10–12.

68. Molinaro, Antonio *et al.* 2015. “Chemistry of Lipid A: At the Heart of Innate Immunity.” *Chemistry – A European Journal* 21(2): 500–519.
69. Moran, Anthony P., Buko Lindner, and Evelyn J. Walsh. 1997. “Structural Characterization of the Lipid A Component of *Helicobacter pylori* Rough- and Smooth-Form Lipopolysaccharides.” *Journal of Bacteriology* 179(20): 6453–63.
70. Morrison, David C., and Lois F. Kline. 1977. “Activation of the Classical and Properdin Pathways of Complement by Bacterial Lipopolysaccharides (LPS).” *The Journal of Immunology* 118(1): 362–68.
71. Nagy, Gábor *et al.* 2008. “‘Gently Rough’: The Vaccine Potential of a Salmonella Enterica Regulatory Lipopolysaccharide Mutant.” *The Journal of Infectious Diseases* 198(11): 1699–1706.
72. Nassif, N. J. 1989. “A Brief Self-Administered Questionnaire for Craniomandibular Disorders (CMD): Rationale, Patient Complaints, and Craniomandibular Symptoms.” *Cranio: The Journal of Craniomandibular Practice* 7(1): 63–70.
73. Ni, Rui Ting *et al.* 2020. “The Role of RND-Type Efflux Pumps in Multidrug-Resistant Mutants of *Klebsiella pneumoniae*.” *Scientific Reports* 10(1).
74. Ogawa, Wakano *et al.* 2012. “Functional Study of the Novel Multidrug Efflux Pump KexD from *Klebsiella pneumoniae*.” *Gene* 498(2): 177–82.
75. Opoku-Temeng, Clement, Scott D. Kobayashi, and Frank R. DeLeo. 2019. “*Klebsiella pneumoniae* Capsule Polysaccharide as a Target for Therapeutics and Vaccines.” *Computational and Structural Biotechnology Journal* 17: 1360–66.
76. Paczosa, Michelle K, and Joan Mecsas. 2016. “*Klebsiella pneumoniae*: Going on the Offense with a Strong Defense.” *Microbiology and Molecular Biology Reviews: MMBR* 80(3): 629–61.
77. Padilla, Emma *et al.* 2010. “*Klebsiella pneumoniae* AcrAB Efflux Pump Contributes to Antimicrobial Resistance and Virulence.” *Antimicrobial Agents and Chemotherapy* 54(1): 177.
78. Petersson, Carl *et al.* 1997. “Structural Studies of the O-Specific Polysaccharide of *Hafnia Alvei* Strain PCM 1206 Lipopolysaccharide Containing D-Allothreonine.” *European Journal of Biochemistry* 244(2): 580–86.
79. Quainoo, Scott *et al.* 2017. “Whole-Genome Sequencing of Bacterial Pathogens: The Future of Nosocomial Outbreak Analysis.” *Clinical Microbiology Reviews* 30(4):1015–63.

80. Rietschel, Ernst T. *et al.* 1994. "Bacterial Endotoxin: Molecular Relationships of Structure to Activity and Function." *The FASEB Journal* 8(2): 217–25.
81. Rollenske, Tim *et al.* 2018. "Cross-Specificity of Protective Human Antibodies against *Klebsiella pneumoniae* LPS O-Antigen." *Nature Immunology* 19(6): 617–24.
82. Russo, Thomas A., and Candace M. Marr. 2019. "Hypervirulent *Klebsiella pneumoniae*." *Clinical Microbiology Reviews* 32(3).
83. Schroll, C. *et al.* 2010. "Role of type 1 and type 3 fimbriae in *Klebsiella pneumoniae* biofilm formation." *BMC Microbiology* 10, 179.
84. Schurtz Sebghati, Tricia A., Timo K. Korhonen, Douglas B. Hornick, and Steven Clegg. 1998. "Characterization of the Type 3 Fimbrial Adhesins of *Klebsiella* Strains." *Infection and Immunity* 66(6): 2887.
85. Severn, Wayne B., Robert F. Kelly, James C. Richards, and Chris Whitfield. 1996. "Structure of the Core Oligosaccharide in the Serotype O8 Lipopolysaccharide from *Klebsiella pneumoniae*." *Journal of Bacteriology* 178(6): 1731–41.
86. Shu, Hung Yu *et al.* 2009. "Genetic Diversity of Capsular Polysaccharide Biosynthesis in *Klebsiella pneumoniae* Clinical Isolates." *Microbiology* 155(12): 4170–83.
87. Siguier, P *et al.* 2006. "ISfinder: The Reference Centre for Bacterial Insertion Sequences." *Nucleic Acids Research* 34(suppl_1): D32–36.
88. Silhavy, Thomas J., Daniel Kahne, and Suzanne Walker. 2010. "The Bacterial Cell Envelope." *Cold Spring Harbor Perspectives in Biology* 2(5).
89. Srinivasan, Vijaya Bharathi, and Govindan Rajamohan. 2013. "KpnEF, a New Member of the *Klebsiella pneumoniae* Cell Envelope Stress Response Regulon, Is an SMR-Type Efflux Pump Involved in Broad-Spectrum Antimicrobial Resistance." *Antimicrobial Agents and Chemotherapy* 57(9): 4449–62.
90. Stojkovic, Katarina *et al.* 2017. "Identification of D-Galactan-III As Part of the Lipopolysaccharide of *Klebsiella pneumoniae* Serotype O1." *Frontiers in Microbiology* 8: 684.
91. Struve, Carsten, Martin Bojer, and Karen Angeliki Krogfelt. 2008. "Characterization of *Klebsiella pneumoniae* Type 1 Fimbriae by Detection of Phase Variation during Colonization and Infection and Impact on Virulence." *Infection and Immunity* 76(9): 4055.

92. Sugawara, Etsuko, Seiji Kojima, and Hiroshi Nikaido. 2016. “*Klebsiella pneumoniae* Major Porins OmpK35 and OmpK36 Allow More Efficient Diffusion of β -Lactams than Their *Escherichia coli* Homologs OmpF and OmpC.” *Journal of Bacteriology* 198(23): 3200.
93. Sun, Jiawei, Steven T. Rutherford, Thomas J. Silhavy, and Kerwyn Casey Huang. 2021. “Physical Properties of the Bacterial Outer Membrane.” *Nature Reviews Microbiology* 2021 20:4 20(4): 236–48.
94. Szijártó, Valéria *et al.* 2014. “Diagnostic Potential of Monoclonal Antibodies Specific to the Unique O-Antigen of Multidrug-Resistant Epidemic *Escherichia coli* Clone ST131-O25b:H4” ed. C J Papasian. *Clinical and Vaccine Immunology* 21(7): 930–39.
95. Szijártó, Valéria *et al.* 2016. “Both Clades of the Epidemic KPC-Producing *Klebsiella pneumoniae* Clone ST258 Share a Modified Galactan O-Antigen Type.” *International Journal of Medical Microbiology : IJMM* 306(2): 89–98.
96. Tacconelli, Evelina *et al.* 2018. “Discovery, Research, and Development of New Antibiotics: The WHO Priority List of Antibiotic-Resistant Bacteria and Tuberculosis.” *The Lancet. Infectious Diseases* 18(3): 318–27.
97. Tobes, Raquel *et al.* 2013. “Genome Sequence of *Klebsiella pneumoniae* KpQ3, a DHA-1 β -Lactamase-Producing Nosocomial Isolate.” *Genome Announcements* 1(1).
98. Treangen, Todd J., Brian D. Ondov, Sergey Koren, and Adam M. Phillippy. 2014. “The Harvest Suite for Rapid Core-Genome Alignment and Visualization of Thousands of Intraspecific Microbial Genomes.” *Genome Biology* 15(11): 1–15.
99. Tsai, Chao Ming, and Carl E. Frasch. 1982. “A Sensitive Silver Stain for Detecting Lipopolysaccharides in Polyacrylamide Gels.” *Analytical Biochemistry* 119(1): 115–19.
100. Tsai, Yu Kuo *et al.* 2011. “*Klebsiella pneumoniae* Outer Membrane Porins OmpK35 and OmpK36 Play Roles in Both Antimicrobial Resistance and Virulence.” *Antimicrobial Agents and Chemotherapy* 55(4): 1485.
101. Tu, I. Fan *et al.* 2022. “Structural and Biological Insights into *Klebsiella pneumoniae* Surface Polysaccharide Degradation by a Bacteriophage K1 Lyase: Implications for Clinical Use.” *Journal of Biomedical Science* 29(1): 1–17.

102. Vandhana, V., K. Vishwas Saralaya, *et al.* 2022. “Characterization of Hypervirulent *Klebsiella pneumoniae* (Hv-Kp): Correlation of Virulence with Antimicrobial Susceptibility.” *International Journal of Microbiology* 2022.
103. Varki, Ajit *et al.* 2015. “Symbol Nomenclature for Graphical Representations of Glycans.” *Glycobiology* 25(12): 1323–24.
104. Verkhnyatskaya, Stella A., Vadim B. Krylov, and Nikolay E. Nifantiev. 2017. “Pyranoside-into-Furanoside Rearrangement of 4-Pentenyl Glycosides in the Synthesis of a Tetrasaccharide-Related to Galactan I of *Klebsiella pneumoniae*.” *European Journal of Organic Chemistry* 2017(3): 710–18.
105. Vinogradov, Evgeny *et al.* 2002. “Structures of Lipopolysaccharides from *Klebsiella pneumoniae*.” *Journal of Biological Chemistry* 277(28): 25070–81.
106. Wang, Hairong, Guohua Zhang, and Jun Ning. 2003. “First Synthesis of β -D-Galp-(1 \rightarrow 3)-D-Galp - The Repeating Unit of the Backbone Structure of the O-Antigenic Polysaccharide Present in the Lipopolysaccharide (LPS) of the Genus *Klebsiella*.” *Carbohydrate Research* 338(10): 1033–37.
107. Wantuch, Paeton L. *et al.* 2023. “Capsular Polysaccharide Inhibits Vaccine-Induced O-Antigen Antibody Binding and Function across Both Classical and Hypervirulent K2:O1 Strains of *Klebsiella pneumoniae*.” *PLoS pathogens* 19(5).
108. Westphal, O. and Jann, K. (1965) Bacterial Lipopolysaccharides Extraction with Phenol-Water and Further Applications of the Procedure. *Methods in Carbohydrate Chemistry*, 5, 83-91.
109. Westphal, Otto, Otto Lüderitz, and Fritz Bister. 1952. “Über Die Extraktion von Bakterien Mit Phenol/Wasser.” *Zeitschrift für Naturforschung B* 7(3): 148–55.
110. Wick, Ryan R., Eva Heinz, Kathryn E. Holt, and Kelly L. Wyres. 2018. “Kaptive Web: User-Friendly Capsule and Lipopolysaccharide Serotype Prediction for *Klebsiella* Genomes.” *Journal of Clinical Microbiology* 56(6). /
111. Williams, Danielle M. *et al.* 2017. “Single Polysaccharide Assembly Protein That Integrates Polymerization, Termination, and Chain-Length Quality Control.” *Proceedings of the National Academy of Sciences of the United States of America* 114(7): E1215–23.
112. Wolkowicz, Tomasz. 2018. “The Utility and Perspectives of NGS-Based Methods in BSL-3 and BSL-4 Laboratory – Sequencing and Analysis Strategies.” *Briefings in Functional Genomics* 17(6): 471–76.

113. Wyres, Kelly L., Margaret M.C. Lam, and Kathryn E. Holt. 2020. "Population Genomics of *Klebsiella pneumoniae*." *Nature Reviews Microbiology* 2020 18:6 18(6): 344–59.
114. Yang, Feng-Ling *et al.* 2011. "Structure and immunological characterization of the capsular polysaccharide of a pyrogenic liver abscess caused by *Klebsiella pneumoniae*: activation of macrophages through Toll-like receptor 4." *The Journal of biological chemistry* 86(24):21041-51.
115. Yong, Dongeun *et al.* 2009. "Characterization of a New Metallo- β -Lactamase Gene, BlaNDM-1, and a Novel Erythromycin Esterase Gene Carried on a Unique Genetic Structure in *Klebsiella pneumoniae* Sequence Type 14 from India." *Antimicrobial Agents and Chemotherapy* 53(12): 5046.
116. York, William S., Stephen Hantus, Peter Albersheim, and Alan G. Darvill. 1997. "Determination of the Absolute Configuration of Monosaccharides by ^1H NMR Spectroscopy of Their Per-O-(S)-2-Methylbutyrate Derivatives." *Carbohydrate Research* 300(3): 199–206.
117. Zamlynska, Katarzyna *et al.* 2017. "Studies on Lipid A Isolated from *Phyllobacterium trifolii* PETP02T Lipopolysaccharide." *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology* 110(11): 1413–33.
118. Zamze S. *et al.* 2002. "Recognition of bacterial capsular polysaccharides and lipopolysaccharides by the macrophage mannose receptor." *Journal of Biological Chemistry*. 277:41613–41623
119. Zhu, San Yong, and Jin Song Yang. 2012. "Synthesis of Tetra- and Hexasaccharide Fragments Corresponding to the O-Antigenic Polysaccharide of *Klebsiella pneumoniae*." *Tetrahedron* 68(20): 3795–3802.
120. Zwama, Martijn, and Akihito Yamaguchi. 2018. "Molecular Mechanisms of AcrB-Mediated Multidrug Export." *Research in Microbiology* 169(7–8): 372–83.

11. Appendix

Table S1. Kaptive Web analysis results for the 8130 assemblies of *K. pneumoniae* isolates. Isolates belonging to the O2v2 serotype are marked in green. Isolates belonging to the O1v2 serotype are marked in blue.

Due to the very large amounts of data contained in the table, the analysed sequences from the public database are available in the paper (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7556023/>) in Supplementary Materials at <https://www.mdpi.com/1422-0067/21/18/6572/s1> under the name Table S1).

Table S2. Kaptive Web analysis results for the 71377 assemblies of *K. pneumoniae* isolates.

Due to the very large amounts of data contained in the table, the analysed sequences from the public database are available in the paper (<https://www.sciencedirect.com/science/article/pii/S0144861723010469?via%3Dihub#ec0005>) in Supplementary Materials at <https://www.sciencedirect.com/science/article/pii/S0144861723010469?via%3Dihub#ec0005> under the name Table S1).

12. List of figures

Figure 1. Major virulence factors of <i>K. pneumoniae</i> ; the scheme of the cell envelope of Gram-negative bacteria.....	24
Figure 2. Schematic representation of the repeating unit of the K1 and K2 capsular polysaccharide.....	29
Figure 3. Scheme of the lipopolysaccharide.....	31
Figure 4. The lipid A structure from <i>K. pneumoniae</i>	33
Figure 5. <i>K. pneumoniae</i> core type 1 and type 2 OS structures.....	35
Figure 6. The known O antigen repeating unit structures of <i>K. pneumoniae</i>	36
Figure 7. <i>rfb</i> gene clusters.....	37
Figure 8. General outline of the experiments performed in this study.....	43
Figure 9. ¹ H NMR spectra of <i>K. pneumoniae</i> O antigens.....	55
Figure 10. LPS profiles obtained by SDS-PAGE analysis of 9 nontypeable <i>K. pneumoniae</i> isolates.....	56
Figure 11. Comparative NMR analysis of <i>K. pneumoniae</i> O2v2 and O2v1 O antigens..	58
Figure 12. Screenshot with Kaptive Web results for <i>K. pneumoniae</i> BIDMC 7B O locus prediction.....	60
Figure 13. The alignment of the <i>gmlB</i> genes of <i>K. pneumoniae</i> BIDMC 7B, ABC152 with <i>gmlB</i> genes of two reference strains (NTUH-K2044 and 441).....	61
Figure 14. Results of clonality (MLST) and phylogenetic analyses of O2v2 <i>K. pneumoniae</i> genomes characterized by ST and insertion sequences distribution.....	65
Figure 15. Comparative NMR analysis of the <i>K. pneumoniae</i> ABC122, BC738, BC13-986 and 3936/19 O antigens.....	67
Figure 16. Chromatogram of poly- and oligosaccharides obtained by hydrolysis of ABC122 LPS (60 mg) and fractionated on TSKgel®G2500PW column	67
Figure 17. Chromatogram of alditol acetates of sugar components obtained as a result of sugar analysis of <i>K. pneumoniae</i> ABC122 OPS.....	68
Figure 18. Results of methylation analysis of the <i>K. pneumoniae</i> ABC122 OPS.....	69
Figure 19. Comparison of ¹ H NMR spectra in the range of 5.5–6.5 ppm for <i>O</i> -(<i>S</i>)-2-methyl butyrate derivatives of monosaccharide standards and the <i>K. pneumoniae</i> ABC122 OPS hydrolysate.....	70
Figure 20. Selected regions of ¹ H, ¹³ C HSQC-DEPT and HMBC spectra of the <i>K. pneumoniae</i> ABC122 OPS.....	71

Figure 21. The TOCSY and NOESY NMR spectra of <i>K. pneumoniae</i> ABC122 OPS....	72
Figure 22. Selected regions of ¹ H, ¹³ C HSQC-DEPT and HMBC spectra of the <i>K. pneumoniae</i> ABC122 LPS.....	73
Figure 23. The TOCSY and NOESY NMR spectra of <i>Klebsiella pneumoniae</i> ABC122 LPS.....	74
Figure 24. The HR-MAS ¹ H, ¹³ C HSQC-DEPT NMR spectra overlay for ABC122 LPS and BC738 LPS.....	75
Figure 25. The HR-MAS ¹ H, ¹³ C HSQC-DEPT NMR spectra overlay for ABC122 LPS and BC13-986 LPS.....	76
Figure 26. The HR-MAS ¹ H, ¹³ C HSQC-DEPT NMR spectra overlay for ABC122 LPS and 3936/19 LPS.....	77
Figure 27. Organization of the O antigen biosynthesis gene cluster (<i>rfb</i> , O locus) from the <i>K. pneumoniae</i> OL101 reference strain and comparison to ABC122 O locus; and comparison to 3936/19 O locus based on Kaptive results.....	79
Figure 28. The alignment of the <i>rfbB</i> gene of <i>K. pneumoniae</i> 3936/19 and OL101 locus.....	81
Figure 29. The alignment of the <i>wbbB</i> gene of <i>K. pneumoniae</i> ABC122 and OL101 locus.....	82
Figure 30. The alignment of the glycosylhydrolase gene of <i>K. pneumoniae</i> ABC122 and OL101 locus.....	83
Figure 31. Distribution of <i>K. pneumoniae</i> O serotypes, based on genomes available in NCBI database for July 27, 2023.....	84
Figure 32. Chromatogram of poly- and oligosaccharides obtained by hydrolysis of Kp164 LPS (20 mg) and fractionated on TSKgel®G2500PW column.....	85
Figure 33. The NMR spectra overlay for Kp164 LPS and Kp164 OPS.....	86
Figure 34. Part of the Kp164, Kp165 and Kp166 <i>rfb</i> regions alignment (from 1 to 1540 bp) to demonstrate <i>rfb</i> regions identity.....	89
Figure 35. Screenshot with MALDI Biotyper result for PCM15 strain.....	90
Figure 36. Chromatogram of poly- and oligosaccharides obtained by hydrolysis of Kp174 LPS (35 mg) and fractionated on TSKgel®G2500PW column (flow: 1 ml/min).....	90
Figure 37. Comparative NMR analysis of the <i>K. pneumoniae</i> Kp174 fractions.....	91

13. List of tables

Table 1. Compositions of buffers and medium used in this work.....	44
Table 2. <i>Klebsiella pneumoniae</i> isolates used in this study.....	45
Table 3. Efficiency of LPS isolation from dry bacterial mass.....	54
Table 4. Summary of the O serotype analyses for ten nontypeable <i>K. pneumoniae</i> isolates.....	56
Table 5. ¹ H and ¹³ C NMR chemical shifts of O antigens isolated from <i>K. pneumoniae</i> BIDMC 7B and ABC152.....	59
Table 6. Kaptive analysis results extracted for 49 <i>K. pneumoniae</i> O2 variant 2-predicted genomes characterised by IS occurrence.....	62
Table 7. Sequence types and location of IS elements in the <i>rfb</i> clusters of <i>K. pneumoniae</i> isolates selected on the basis of Kaptive-based O1v2 predictions.....	64
Table 8. The ¹ H and ¹³ C NMR chemical shifts and selected inter-residue correlations from ¹ H, ¹ H NOESY and ¹ H, ¹³ C HMBC spectra of the OPS of <i>K. pneumoniae</i> ABC122.....	71
Table 9. The ¹ H and ¹³ C NMR chemical shifts and selected inter-residue correlations from ¹ H, ¹ H NOESY and ¹ H, ¹³ C HMBC spectra of the LPS of <i>K. pneumoniae</i> ABC122.....	73
Table 10. The ¹ H and ¹³ C HR-MAS NMR chemical shifts of <i>K. pneumoniae</i> BC738, BC13-986, and 3936/19 O antigens.....	78
Table 11. Kaptive-based O <i>locus</i> prediction results for BC738, BC13-986 and 3936/19 <i>K. pneumoniae</i>	80
Table 12. The ¹ H and ¹³ C NMR chemical shifts of <i>K. pneumoniae</i> Kp164 LPS and O-specific polysaccharide.....	87
Table 13. Kaptive-based O serotype prediction for <i>K. pneumoniae</i> Kp164, Kp165 and Kp166 isolates.....	88
Table 14. Kaptive results for Kp174 and Kp177 isolates.....	92
Table 15. Summary of the O serotype analyses for eleven nontypeable <i>K. pneumoniae</i> isolates.....	97

14. Scientific achievements

Publications:

1. **Artyszuk D**, Wołkowicz T. Implementation of whole genome sequencing for bacteria genotyping. *Post. Mikrobiol.* 2018, 57, 2, 179–193. IF: 0.354
2. **Artyszuk D**, Izdebski R, Maciejewska A, Kaszowska M, Herud A, Szijártó V, Gniadkowski M, Łukasiewicz J. The Impact of Insertion Sequences on O-Serotype Phenotype and Its O-Locus-Based Prediction in *Klebsiella pneumoniae* O2 and O1. *Int J Mol Sci.* 2020, Sep 8;21(18):6572. IF: 5.923
3. **Artyszuk D**, Jachymek W, Izdebski R, Gniadkowski M, Łukasiewicz J. The OL101 O antigen locus specifies a novel *Klebsiella pneumoniae* serotype O13 structure. *Carbohydr. Polym.* 2024, 326 121581. doi: 10.1016/j.carbpol.2023.121581. IF: 11.2

Scientific conferences:

Oral presentations:

1. 9th Biennial Meeting on Microbial Carbohydrates (BMMC), 27-29 July 2022, Naples, Italy – oral presentation „Morphology and carbohydrate surface antigens of *Flavonifractor plautii* PCM 3108”; Artyszuk D., Jachymek W., Migdał P., Łukasiewicz J.
2. 9th Biennial Meeting on Microbial Carbohydrates (BMMC), 27-29 July 2022, Naples, Italy – oral presentation „O-antigen variability among clinical isolates of *Klebsiella pneumoniae* – structural and genetic analyses”; Artyszuk D., Jachymek W., Maciejewska A., Kaszowska M., Łukasiewicz J.
3. 21st European Carbohydrate Symposium in Paris (Eurocarb 21), 9-13 July 2023, Paris, France – oral presentation „O-antigen variability among clinical isolates of *Klebsiella pneumoniae*”; Artyszuk D., Jachymek W., Maciejewska A., Kaszowska M., Łukasiewicz J.
4. Paths of Glycobiology. Jerzy Kościelak Memorial Conference, 28-29 September 2023, Wrocław, Poland – oral presentation „Morphology and carbohydrate surface

antigens of *Flavonifractor plautii* PCM 3108”; Artyszuk D., Jachymek W., Migdał P., Pyclik M., Łukasiewicz J.

Flash communications and posters:

5. 21st European Carbohydrate Symposium in Paris (Eurocarb 21), 9-13 July 2023, Paris, France – poster + flash communication „Morphology and carbohydrate surface antigens of *Flavonifractor plautii* PCM 3108”; Artyszuk D., Jachymek W., Migdał P., Pyclik M., Łukasiewicz J.

Participation in research projects:

1. PhD student-scholarship at the Laboratory of Immunochemistry of Microorganisms and Vaccines of the Institute of Immunology and Experimental Therapy of the Polish Academy of Sciences in Wrocław (project entitled: "Nontypeable antigens of *Klebsiella pneumoniae* - structure and seroepidemiology", financed by the National Science Center OPUS 16 UMO-2018/31/B/NZ7/ 04002)