Manuscript Details

Manuscript number	IJBIOMAC_2018_7171_R2
Title	Structure of the capsular polysaccharide of the KPC-2-producing Klebsiella pneumoniae strain KK207-2 and assignment of the glycosyltransferases functions
Article type	Research Paper

Abstract

Klebsiella pneumoniae strain KK207-2 was isolated in 2010 from a bloodstream infection of an inpatient at an Italian hospital. It was previously found to produce the KPC-2 carbapenemase and to belong to clade 1 of sequence type 258. Genotyping of the conserved wzi and wzc genes from strain KK207-2 yielded contrasting results: the wzc-based method assigned the cps207-2 to a new K-type, while the wzi-based method assigned it to the known K41 K-type. In order to resolve this contradiction, the capsular polysaccharide of K. pneumoniae KK207-2 was purified and its structure determined by using GLC-MS of appropriate carbohydrate derivatives, ESI-MS of both partial hydrolysis and Smith degradation derived oligosaccharides, and NMR spectroscopy of oligosaccharides, and the lithium degraded, native and de-O-acetylated polysaccharide. All the collected data demonstrated the following repeating unit for the K. pneumoniae KK207-2 capsular polysaccharide: OAc I 6 [3)-β-D-Gal-(1-4)-β-D-Glc-(1-]n 4 I 1 β-D-Glcp-(1-6)-α-D-Glcp-(1-4)-β-D-GlcpA-(1-6)-α-D-Glcp The polysaccharide contains about 0.60 acetyl groups per repeating unit on C6 of the Gal residue. The reactions catalysed by each glycosyltransferase in the cpsKK207-2 gene cluster were assigned on the basis of structural homology with other Klebsiella K antigens.

Keywords	Klebsiella pneumoniae Sequence Type 258 strain KK207-2; capsular polysaccharide structure; glycosyltransferases
Manuscript category	Carbohydrates, Natural Polyacids and Lignins

Corresponding Author	Paola Cescutti
Order of Authors	Barbara Bellich, Neil Ravenscroft, Roberto Rizzo, Cristina Lagatolla, Marco D'Andrea, Gian Maria Rossolini, Paola Cescutti

Suggested reviewers	Joanna Kubler-Kielb, Kathryn E. Holt, Andrzej Gamian, Yuriy Knirel, Domenico Garozzo

Submission Files Included in this PDF

File Name [File Type]

Cover letter2.docx [Cover Letter]

Responses to Reviewers.docx [Response to Reviewers]

Abstract.docx [Abstract]

manuscript revised Cescutti_2.docx [Manuscript File]

Fig 1.pptx [Figure]

Fig 2.tif [Figure]

Fig 3.pptx [Figure]

Fig 4.pptx [Figure]

Fig 5.pptx [Figure]

Fig 6.pptx [Figure]

Fig 7.pptx [Figure]

Supplementary material 190218.pdf [e-Component]

Dear Editor,

I carefully revised the manuscript according to the reviewers’ comments. I accepted all their suggestions and changed the manuscript accordingly.

I also moved the “Materials and Methods” section before the “Results and discussion” one, as you suggested. I changed the numbering of the references accordingly, as well as those of Figs and Tables. Changes regarding the order of Figs and Tables were also made in the Supplementary file.

Thank you for your work

Regards

Paola

Responses to Reviewers

Reply to Reviewer 1
Thank you for all your comments and suggestions. I moved Table 1 to the supplementary material and numbered the remaining Tables accordingly, changed “Superdex G30” with “Superdex 30 prep grade” and checked all the DOI links.

Reply to Reviewer 2
Thank you for all your comments and suggestions. I inserted all suggested corrections in the text.

Regarding Fig 1, I forgot to “group” all the components of the figure, and for this reason there were no labels in it. Now I have corrected it. I also corrected the Figure legend by deleting “native **CPS KK207-2** (d)” whose 1H NMR spectrum was not present in Fig. 1.

Regarding your comment: line 165, …per-ethylated?
We used ethyl iodide to derivatise the sample in order not to have ambiguous results. In fact, the difference in mass between an Hex and an HexA is 14. The same mass value corresponds to substitution of an hydrogen atom with a methyl group. Using ethyl iodide this ambiguity is avoided.

Abstract

Klebsiella pneumoniae strain KK207-2 was isolated in 2010 from a bloodstream infection of an inpatient at an Italian hospital. It was previously found to produce the KPC-2 carbapenemase and to belong to clade 1 of sequence type 258. Genotyping of the conserved *wzi* and *wzc* genes from strain KK207-2 yielded contrasting results: the *wzc*-based method assigned the $cps_{207\text{-}2}$ to a new K-type, while the *wzi*-based method assigned it to the known K41 K-type. In order to resolve this contradiction, the capsular polysaccharide of *K. pneumoniae* KK207-2 was purified and its structure determined by using GLC-MS of appropriate carbohydrate derivatives, ESI-MS of both partial hydrolysis and Smith degradation derived oligosaccharides, and NMR spectroscopy of oligosaccharides, and the lithium degraded, native and de-O-acetylated polysaccharide. All the collected data demonstrated the following repeating unit for the *K. pneumoniae* KK207-2 capsular polysaccharide:

OAc
|
6
[3)-β-D-Gal-(1→4)-β-D-Glc-(1→]$_n$
4
↑
1
β-D-Glc*p*-(1→6)-α-D-Glc*p*-(1→4)-β-D-Glc*p*A-(1→6)-α-D-Glc*p*

The polysaccharide contains about 0.60 acetyl groups per repeating unit on C6 of the Gal residue. The reactions catalysed by each glycosyltransferase in the $cps_{KK207\text{-}2}$ gene cluster were assigned on the basis of structural homology with other *Klebsiella* K antigens.

Structure of the capsular polysaccharide of the KPC-2-producing *Klebsiella pneumoniae* strain KK207-2 and assignment of the glycosyltransferases functions

Barbara Bellich[a], Neil Ravenscroft[b], Roberto Rizzo[a], Cristina Lagatolla[a], Marco Maria D'Andrea[c,d] Gian Maria Rossolini[e,f] and Paola Cescutti[a*]

[a] Department of Life Sciences, University of Trieste, via L. Giorgieri 1, Bdg C11, 34127 Trieste, Italy

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

[c] Department of Medical Biotechnologies, University of Siena, Siena, Italy

[d] Department of Biology, University of Rome "Tor Vergata", Rome, Italy

[e] Department of Experimental and Clinical Medicine, University of Florence, Florence, Italy

[f] Microbiology and Virology Unit, Florence Careggi University Hospital, Florence, Italy

Corresponding author at: Department of Life Sciences, University of Trieste, via L. Giorgieri 1, Bdg C11, 34127 Trieste, Italy.

E-mail address: pcescutti@units.it (Paola Cescutti)

Abstract

Klebsiella pneumoniae strain KK207-2 was isolated in 2010 from a bloodstream infection of an inpatient at an Italian hospital. It was previously found to produce the KPC-2 carbapenemase and to belong to clade 1 of sequence type 258. Genotyping of the conserved *wzi* and *wzc* genes from strain KK207-2 yielded contrasting results: the *wzc*-based method assigned the $cps_{207\text{-}2}$ to a new K-type, while the *wzi*-based method assigned it to the known K41 K-type. In order to resolve this contradiction, the capsular polysaccharide of *K. pneumoniae* KK207-2 was purified and its structure determined by using GLC-MS of appropriate carbohydrate derivatives, ESI-MS of both partial hydrolysis and Smith degradation derived oligosaccharides, and NMR spectroscopy of oligosaccharides, and the lithium degraded, native and de-O-acetylated polysaccharide. All the collected data demonstrated the following repeating unit for the *K. pneumoniae* KK207-2 capsular polysaccharide:

```
                                                    OAc
                                                     |
                                                     6
                                          [3)-β-D-Gal-(1→4)-β-D-Glc-(1→]n
                                                     4
                                                     ↑
                                                     1
β-D-Glcp-(1→6)-α-D-Glcp-(1→4)-β-D-GlcpA-(1→6)-α-D-Glcp
```

The polysaccharide contains about 0.60 acetyl groups per repeating unit on C6 of the Gal residue. The reactions catalysed by each glycosyltransferase in the $cps_{KK207\text{-}2}$ gene cluster were assigned on the basis of structural homology with other *Klebsiella* K antigens.

Keywords: *Klebsiella pneumoniae* Sequence Type 258; strain KK207-2; capsular polysaccharide structure; NMR; ESI-MS; glycosyltransferases.

1. Introduction

Klebsiella pneumoniae is a Gram-negative encapsulated rod which is mostly responsible for nosocomial infections in immunocompromised patients, but can also cause severe community-acquired infections [1]. The clinical scenario has worsened during recent years, with the global emergence and dissemination of *K. pneumoniae* strains resistant to carbapenems (CR-Kp), which are among the last-resort antibiotics for treatment of infections caused by multiresistant Gram-negative rods. Production of carbapenemase of different types (e. g. KPC, VIM, NDM, OXA-48) is the major carbapenem resistance mechanism among CR-Kp. KPC-type carbapenemases are among the most prevalent and challenging carbapenemases, since they can degrade virtually all β-lactam antibiotics including carbapenems [2]. Dissemination of KPC-producing strains of *K. pneumoniae* (KPC-Kp) has been sustained by the expansion of various clones, with members of clonal group (CG) 258 being likely the most successful and widespread. CG258 includes strains of the sequence type (ST) 258 and some related variants [2]. Infections caused by KPC-Kp strains are challenging in healthcare settings, where they spread rapidly and are associated with significant morbidity and mortality [2].

K. pneumoniae strains can be classified based on serotyping of two different types of antigens exposed on the bacterial surface: O antigen, the outer polysaccharide chain of the lipopolysaccharide, and the K antigen, the capsule. *K. pneumoniae* produces approximately 80 structurally different capsular polysaccharides (CPS, or K antigen) which are recognized as virulence factors and confer different antigenic properties within the same species, with relevant consequences to bacterial virulence. Although serotyping of K antigen has been extensively used to classify *K. pneumoniae* strains, genotyping of the conserved *wzi* [3] and *wzc* [4] genes has recently been developed to establish the K-type. Moreover, Pan *et al*. [5] carried out an extended analysis of the sequences of the *cps* clusters from strains belonging to 79 different K types, identifying more than 1500 different genes that were grouped into 361 homology groups. Despite this high number of different CPSs expressed by *K. pneumoniae* strains, capsular variations are limited among KPC-Kp isolates, due to their clonal origin.

Genotyping of KPC-Kp isolates belonging to ST258 [6, 7] identified the presence of at least two lineages, indicated as clade I and clade II, which differ from each other primarily in the *cps* gene cluster (named *cps-1* and *cps-2* gene cluster, respectively) and are novel with respect to those described by Pan *et al* [5]. Regarding *cps-1*, the results obtained with the two genotyping methods were not in agreement with each other: the *wzc*-based method [4] assigned *cps-1* to a new K-type, while the *wzi*-based method [3] associated it to the already known K41 K-type. However, the *cps-1* cluster does not contain the genes for the biosynthesis of rhamnose, thus suggesting a chemical structure at least partially different from K41.

The present work reports the primary structure of the CPS produced by *K. pneumoniae* KK207-2 [6], a clade I representative, determined by using GLC-MS of appropriate carbohydrate derivatives, ESI-MS of both partial hydrolysis and Smith degradation derived oligosaccharides, and NMR spectroscopy of oligosaccharides, lithium degraded CPS, and the native and de-O-acetylated CPS. Moreover, taking advantage of the sequenced $cps_{207\text{-}2}$ gene cluster [6], and of the structural knowledge gained in the present investigation, each glycosyltransferase (GT) in the CPS gene cluster was assigned to the corresponding catalyzed reaction.

2. Materials and methods

2.1. Bacterial strain, capsular polysaccharide production and purification

The strain of *Klebsiella pneumoniae* KK207-2 was isolated in 2010 from a bloodstream infection of an inpatient at an Italian hospital, and was previously described to produce the KPC-2 carbapenemase and to belong to ST258 [6]. Bacterial cells were grown on 25 Worfel-Ferguson agar plates for 4 days at 30°C. The bacterial lawn was collected with 0.9 % NaCl (about 3 mL per dish), gently stirred at 10 °C for 2 h, centrifuged at 22400 x g at 4 °C for 30 min to separate the cells from the supernatant which was precipitated with 4 vol of cold ethanol. The precipitated material was recovered by centrifugation at 1900 x g at 4 °C for 30 min, dissolved in water, dialyzed first against 0.1 M NaCl and then water, taken to pH = 6.9, filtered (Millipore membranes 0.45 μm) and lyophilized. The bacterial cells recovered in the pellet of the first centrifugation were suspended in a 2% phenol solution at 10 °C for 2 h, centrifuged at 22400 x g at 4 °C for 30 min and the supernatant was treated as the one reported above. The purified capsular polysaccharide was named **CPS KK207-2**.

2.2. General procedures

Native and permethylated oligo- and polysaccharides were hydrolysed with 2 M trifluoroacetic acid (TFA) at 125 °C for 1 h. Alditol acetates were prepared as previously described [8]. TMS methyl glycosides were obtained by derivatization with the reagent Sylon™ HTP (Sigma) after methanolysis of the polysaccharide [9]. In order to detect the possible presence of uronic acids, different methanolysis procedures were applied using 2 M HCl in methanol: i) at 85 °C for 16 h; ii) hydrolysis with 2 M TFA at 100 °C for 1 h, followed by methanolysis at 85 °C for 16 h; iii) at 121 °C for 5 min and iv) at 121 °C for 15 min in a microwave oven (CEM- Discover® SP Microwave). Permethylation of the CPS and oligosaccharides was achieved following the protocol by Harris [10]. Integration values of the areas of the partially methylated alditol acetates (PMAA) were corrected by the effective carbon response factors [11]. De-O-acetylation was obtained by stirring the CPS (100 mg) in 10 mM NaOH (1 mg/mL) for 5 h under a stream of nitrogen. Carboxyl reduction was performed on 20 mg of CPS using carbodiimide [12] and following the modification described by Osman *et al.*[13]. Size exclusion chromatography of **deOAc CPS KK207-2** was performed on a Sephacryl S-400 column (1.6 cm i.d. × 90 cm) using 0.05 M $NaNO_3$ as eluent, and a flow rate of 7.2 mL/h. Fractions were collected at 15 min intervals. Elution was monitored using a refractive index detector (Knauer, RI detector K-2301, Lab-Service Analitica) which was interfaced with a computer via PicoLog software. The head, core and tail fractions of the eluted peak were pooled together, dialyzed and used for ^{1}H NMR spectroscopy.

Oligosaccharides were separated on a Bio Gel P2 column (1.6 cm i.d. x 90 cm) using the same conditions and set up reported above, except the collection of fractions performed at 12 min interval.

Oligosaccharides were desalted on a Superdex 30 prep grade column (1.0 cm i.d. x 82 cm) using water as eluent at a flow rate of 1.5 mL/min. Elution was monitored using a refractive index detector (Knauer, RI detector Smartline 2300, Lab-Service Analitica) which was interfaced with a computer via Clarity software. Fractions were collected at 30 sec interval.

Analytical GLC was performed on a Perkin-Elmer Autosystem XL gas chromatograph equipped with a flame ionization detector and using He as carrier gas. An HP-1 capillary column (Agilent Technologies, 30 m) was used to separate alditol acetates (temperature program: 3 min at 150 °C, 150–270 °C at 3 °C/min, 2 min at 270 °C), PMAA (temperature program: 1 min at 125 °C, 125–240 °C at 4 °C/min, 2 min at 240 °C), TMS methyl glycosides (temperature program: 1 min at 150 °C, 150–280 °C at 3 °C/min, 2 min at 280 °C), and TMS (+)-2-butyl glycosides, for the determination of the absolute configuration of the sugar residues [14], (temperature program: 1 min at 50 °C, 50-130 °C at 45 °C/min, 1 min at 130 °C, 130–200 °C at 1 °C/min, 10 min at 200 °C). GLC-MS analyses were carried out on an Agilent Technologies 7890A gas chromatograph coupled to an Agilent Technologies 5975C VL MSD, using the same temperature programs reported above.

*2.3. Partial acid hydrolysis of the **deOAc CPS KK207-2**, purification and characterization of the oligosaccharides obtained*

deOAc CPS KK207-2 (33 mg) was hydrolyzed with 32 mL of 0.5 M TFA at 100 °C for 2 h. The sample was roto-evaporated to dryness under reduced pressure, followed by three washes with water. After dissolution in 10 mL of water, the sample was taken to pH = 6.8, and recovered by lyophilization. Separation of the products was achieved by size-exclusion chromatography on a Bio Gel P2 column with a flow rate of 6.8 mL/h. Fractions belonging to the same peak were pooled together and desalted on a Superdex 30 prep grade column. De-salted homogenous fractions were combined after determining their composition by ESI-MS, and structurally characterized by ESI-MS, NMR spectroscopy and GLC-MS of the PMAA derivatives.

*2.4. Smith degradation of the **deOAc CPS KK207-2**, purification and characterization of the oligosaccharides obtained*

deOAc CPS KK207-2 was subjected to Smith degradation [15]: 33 mg of the polysaccharide were dissolved in 6 mL of water, 0.7 mL of a 0.29 M $NaIO_4$ solution were added (0.8 moles of $NaIO_4$/1 mole of oxidizable diol) and the oxidation was let to proceed for 1 h in the dark at 10 °C. After addition of glycerol to react with excess of periodate, reduction of the aldehyde groups was achieved by incubation with $NaBH_4$ at room temperature for 16 h. Excess of borohydride was destroyed with 50% aqueous acetic acid and the solution was dialyzed. The recovered material was hydrolysed with 0.5 M TFA for 6 days at room temperature, rotoevaporated to dryness, washed three times with water to remove TFA, taken to pH = 7.0 and recovered by lyophilization. The products were separated on a Bio Gel P2 column with a flow rate of 7.2 mL/h. Fractions belonging to the same peak were pooled together and desalted on a Superdex 30 prep grade column. De-salted homogenous fractions were combined after determining their composition by ESI-MS, and structurally characterized by ESI-MS, NMR spectroscopy and GLC-MS of the PMAA derivatives.

*2.5. Lithium degradation of the **CPS KK207-2**, purification and characterization of the products obtained*

CPS KK207-2 was subjected to degradation with lithium metal following the protocol by Lau et al. [16]. The CPS (20 mg) was dissolved in 4 mL of ethylenediamine and treated with lithium wire; it was then purified according to literature procedures [16], except that the resin used was Amberlite IR H^+120. After the ion exchange procedure, the sample was lyophilized and separated by size exclusion chromatography on a Bio Gel P2 column (Fig. S1). The peak eluting at the exclusion volume of the column was dialyzed (Spectra Por 6 membrane, CO 1000 Da), and recovered by freeze-drying.

2.6. ESI mass spectrometry

ESI mass spectra were recorded on a Bruker Esquire 4000 ion trap mass spectrometer connected to a syringe pump for the injection of the samples. The instrument was calibrated using a tune mixture provided by Bruker. Oligosaccharides were dissolved in 50% aqueous methanol - 11 mM NH_4OAc; permethylated oligosaccharides were dissolved in a 1:1 chloroform : methanol mixture, 11mM NH_4OAc. Samples were injected at 180 µL/h. Detection was performed in the positive ion mode.

2.7. NMR experiments

The **CPS KK207-2** and **deOAc CPS KK207-2** samples were dissolved in water (about 3 mg/mL) and sonicated using a Branson sonifier equipped with a microtip at 2.8 Å, in order to decrease their molecular masses. The samples were cooled in an ice bath and sonicated using 10 bursts of 1 min each, separated by 1 min intervals. The polysaccharides (~5 mg) were exchanged twice with 99.9% D_2O by lyophilization and then dissolved in 0.6 mL of 99.96% D_2O and introduced into a 5 mm NMR tube for data acquisition. 1D ^{1}H and ^{13}C and 2D, COSY, TOCSY, NOESY, HSQC, HMBC and hybrid HSQC-TOCSY and HSQC-NOESY spectra were obtained using a Bruker Advance III 600 MHz NMR spectrometer equipped with a BBO Prodigy cryoprobe and processed using standard Bruker software (Topspin 3.2). The probe temperature was set at 343 K. 2D TOCSY experiments were performed using mixing times of 180 ms and the 1D variants using mixing times up to 200 ms. The 2D NOESY experiment and 1D variants were performed using a mixing time of 300 ms. The HSQC (with multiplicity editing) experiment was optimized for J = 145 Hz (for directly attached ^{1}H-^{13}C correlations), and the HMBC experiment optimized for a coupling constant of 8 Hz (for long-range ^{1}H-^{13}C correlations). HSQC-TOCSY and HSQC-NOESY NMR spectra were recorded using mixing times of 120 and 300 ms, respectively. 2D experiments were recorded using non-uniform sampling: 50% for homonuclear and 25% for heteronuclear experiments. Spectra were referenced relative to H-2/C-2 of β-Glc: ^{1}H at 3.34 ppm and ^{13}C at 74.0 ppm, established using acetone as an internal standard.

Oligosaccharides were exchanged twice with 99.9% D_2O by lyophilization and then dissolved in 0.6 mL of 99.96% D_2O. Spectra were recorded on a 500 MHz Varian spectrometer operating at 333

K. 2D experiments were performed using standard pulse sequences and pulsed field gradients for coherence selection when appropriate. HSQC spectra were recorded using 145 Hz (for directly attached ^{1}H–^{13}C correlations). TOCSY spectra were acquired using 150 ms spin-lock time and a 1.2 s relaxation time. NOESY experiments were recorded with 200 ms mixing time and a 1.2 s relaxation time. NMR spectra were processed using MestreNova software. Chemical shifts are expressed in ppm using acetone as internal reference (2.225 ppm for ^{1}H and 31.07 ppm for ^{13}C).

3. Results and Discussion

3.1. Purification and composition analysis of ***CPS KK207-2***

The strain *Klebsiella pneumoniae* KK207-2 was grown on carbohydrate rich Worfel-Ferguson agar medium for 4 days at 30°C. The bacterial lawn was first collected with 0.9 % NaCl and subsequent purification gave 224 mg of polysaccharide, while treatment of the remaining bacterial cells with a 2% phenol solution followed by purification resulted in 152 mg of polysaccharide. ^{1}H NMR spectroscopy showed that the two samples were identical and the polysaccharide was named **CPS KK207-2**. ^{1}H NMR spectroscopy determined about 0.60 O-acetyl groups per repeating unit. An aliquot of de-acetylated sample (**deOAc CPS KK207-2**) eluted as a single broad peak upon size exclusion chromatography on a Sephacryl S-400 column (data not shown); the head, core and tail fractions of the peak were pooled and analyzed by ^{1}H NMR spectroscopy which showed that the three fractions were structurally homogenous, indicating the production of a single polysaccharide.

Composition analysis as alditol acetate derivatives revealed Gal and Glc in the molar ratios 1.0 : 2.5, while analysis of methyl-glycosides trimethylsilyl (TMS) derivatives confirmed Gal and Glc in variable molar ratios depending on the experimental conditions used (from 1.0 : 3.1 to 1.0 : 5.3), but did not reveal the presence of any uronic acids. The position of the glycosidic linkages was determined by GLC and GLC-MS analysis after derivatization of the sugar components to partially methylated alditol acetates. GLC analysis on a HP-1 column showed four peaks attributed to t-Glc, 4-linked Glc, 6-linked Glc and 3,4-linked Gal, all in the pyranose configuration. Integration of the respective peak areas, after correction for the effective carbon response factors [11], gave the relative molar ratios reported in Table S1. In order the verify the presence of uronic acids, **CPS KK207-2** was reduced with carbodiimide [12,13], and linkage analysis revealed double the amount of 4-linked Glc (Table S1), thus indicating a 4-linked Glc*p*A residue in the native CPS. These data also showed that **CPS KK207-2** has a branched repeating unit made of six monosaccharides. The absolute configuration was established to be D for all residues.

*3.2. Production and characterization of the oligosaccharides obtained by partial acid hydrolysis of **deOAc CPS KK207-2***

deOAc CPS KK207-2 was treated with 0.5 M TFA and the products were separated by size exclusion chromatography on a Bio Gel P2 column. The elution profile is reported in Fig. S2a. Fractions belonging to the same peak were pooled together and, after desalting, subjected to ESI-MS and ^{1}H NMR spectroscopy to identify peaks suitable for further structural investigation. The partial hydrolysis treatment gave only one reasonably pure peak (**PH-2**) which was identified by ESI-MS as the disaccharide HexA-Hex; only 6-linked Glc was detected by GLC-MS analysis of the PMAA derivatives. The ^{1}H NMR spectrum (Fig. 1a) gave the three signals in the anomeric region at 5.23, 4.65 and 4.50 ppm (with integration values of 0.3, 0.5, and 1.0, respectively), assigned to H-1 of the reducing end 6-Glc-α and 6-Glc-β, and of t-Glc*p*A-β, respectively, thus identifying **PH-2** as the aldobiuronic acid β-D-Glc*p*A-(1→6)-D-Glc-OH.

*3.3. Production and characterization of the oligosaccharides obtained by Smith degradation of **deOAc CPS KK207-2***

deOAc CPS KK207-2 was subjected to partial periodate oxidation, because, due to the type of glycosidic linkages, a complete oxidation would have left only the branched Gal residue intact, resulting in little structural information. After oxidation, the sample was subjected to Smith degradation [15] and the oligosaccharides produced separated by size exclusion chromatography on a Bio Gel P2 column (Fig. S2b). Fractions belonging to the same peak were pooled together, desalted and subjected to ESI-MS and ^{1}H NMR spectroscopy to identify peaks containing mainly only one oligosaccharide and thus suitable for further structural investigation. In this way two peaks were selected, named **SD-1** and **SD-2**.

The most intense ion in the ESI mass spectrum of the sample **SD-2** (Fig. 2a) at 483.2 *m/z* corresponded to the sodium adduct of a compound constituted of two hexoses and one 2,3,4-trihydroxybutanoic acid (THBA) moiety at the reducing end, the latter deriving from oxidation of the 4-linked GlcA residue. MS^{2} of the ion at 483.2 *m/z* (Fig. 2b) gave the sequence Hex-Hex-THBA. A less intense ion at 637.2 *m/z* was present in the ESI mass spectrum (data not shown) and it was attributed to the oligosaccharide found as main component in **SD-1** (see next paragraph). **SD-2** was per-methylated and subjected to ESI-MS and MS^{2} which confirmed both its composition and sequence. After hydrolysis and derivatization to alditol acetates, GLC-MS analysis showed t-Glc and 6-Glc in the relative molar ratio 1.0 : 0.8. The ^{1}H NMR spectrum (Fig. 1b) gave three signals in the anomeric region: two resonances at 5.15 and 4.48 ppm (with integration values of 1.0, and 1.6, respectively), were attributed to H-1 of 6-linked α-D-Glc*p* and t-β-D-Glc*p* respectively, while the less intense signal at 5.47 ppm, with integration value of 0.3, was assigned to a residue in the oligosaccharide **SD-1** (Fig. 1c), in agreement with ESI-MS data. Homo- and hetero-nuclear 2D NMR experiments (data not shown) gave all the chemical shifts for each spin system (Table S2) and established the following structure:

β-D-Glc*p*-(1→6)-α-D-Glc*p*-(1→O-CH-(CH_2OH)-CHOH-COOH.

The residue α-D-Glc*p* is linked to C3 of the THBA, as expected for a 4-linked Glc*p*A.

ESI-MS of the oligosaccharides **SD-1** (Fig. 2c) indicated two hexoses and a hexuronic acid with glycerol at the reducing end; MS^2 of the ion at 615.1 *m/z* (Fig. 2d) gave the sequence Hex-Hex-HexA-Gro. The sample was per-ethylated and subjected to ESI-MS and MS^2 which confirmed both the composition and the sequence. After hydrolysis and derivatization to alditol acetates, GLC-MS identified two main peaks with t-Glc and 6-Glc in the relative molar ratio 1.0 : 1.7. The ^{1}H NMR spectrum (Fig. 1c) contained two anomeric signals at 5.47 and 4.49 ppm, the latter with integration values corresponding to two residues. Based on the chemical shifts assignments for **PH-2** and **SD-2**, and on the 2D NMR data, the signal at 5.47 ppm was attributed to α-D-Glc*p*, and those at 4.49 ppm to β-D-Glc*p*A and β-D-Glc*p*. Homo- and hetero-nuclear 2D NMR experiments (data not shown) determined the chemical shifts for each spin system (Table S3). NOESY plot showed inter-residue connectivities between H-1 (4.49 ppm) of t-β-D-Glc*p* to H-6 (3.89 ppm) of 6-linked α-D-Glc*p*, and H-1 (5.47 ppm) of 6-linked α-D-Glc*p* to H-4 (3.81 ppm) of 4-linked β-D-Glc*p*A, thus establishing the following structure β-D-Glc*p*-(1→6)-α-D-Glc*p*-(1→4)-β-D-Glc*p*A-(1→1)-Gro, in agreement with ESI-MS data. The deshielded chemical shift of Gro C-1 at 71.93 ppm is due to its linkage to C-1 of β-D-Glc*p*A, thus indicating that the uronic acid is linked to C-6 of another 6-linked α-D-Glc*p* in the CPS.

The fractions eluted in the void volume of the Bio Gel P2 column (**SD-Vo**) were subjected to linkage analysis and the results (Table S1) showed a significant change in the relative molar ratios with the additional presence of 3-linked Gal, deriving from the original 3,4-linked Gal. By calculating the molar ratios relative to the sum of the values for 3-linked Gal and 3,4-linked Gal, since only this residue is not susceptible to oxidation (column IV, Table S1), it was clear that 4-Glc had not been oxidised, whereas about half of the branched galactose had become 3-linked Gal, suggesting that the side chain is linked to C-4 of the Gal residue.

The experimental findings obtained from analyzing the Smith degradation products suggested that **CPS KK207-2** has a disaccharide backbone and the following side chain:

β-D-Glc*p*-(1→6)-α-D-Glc*p*-(1→4)-β-D-Glc*p*A-(1→6)-α-D-Glc*p*-(1→

3.4. NMR assignments for ***CPS KK207-2*** *repeating unit*

The repeating unit (RU) structure of **deOAc CPS KK207-2** was investigated at 600 MHz. The ^{1}H NMR spectrum contains six main anomeric signals designated **A** to **F** (Fig. 1d). Residues **A** and **E** gave sharp peaks at 5.48 and 4.49 ppm and resulted in strong crosspeaks in the 2D experiments; these were attributed to the flexible terminal disaccharide β-D-Glc*p*-(1→6)-α-D-Glc*p*-(1→ identified in **SD-1**. The remaining residues gave broad anomeric signals and 2D crosspeaks of lower intensity characteristic of units with restricted motion and branch points. The hexasaccharide RU spin systems were elucidated using a combination of 1D and 2D ^{1}H-^{1}H correlation experiments with correlations established from the six anomeric protons. The anomeric region of COSY, shown as an overlay with TOCSY (Fig. 3), gave H-1 to H-2 for each residue. TOCSY (180 ms) gave additional correlations for each of the spin systems depending on the coupling constants: H-1 to H-6 for α-Glc (residue **A** at 5.48 ppm) and β-Glc (residues **C** and **E** at 4.69 and 4.49 ppm), H-1 to H-4 for α-Glc (residue **B** at 4.92 ppm) and β-GlcA (residue **F** at 4.48 ppm), and H-1 to H-3 for β-Gal (residue **D** at 4.59 ppm). These assignments were confirmed and additional correlations established using 1D TOCSY experiments recorded using a mixing time of 200 ms (Fig. S3): H-1 to H-4 for residue **D** (β-Gal), H-1 to H-5 for residue **F** (β-GlcA) and H-1 to H-6 for residue **B** (α-Glc). Through-space correlations, indicated by negative peaks in the 1D TOCSY spectra, were revealed in the anomeric region of the NOESY experiment (Fig. 4). NOESY (300 ms) gave intra-residue correlations, mainly H-1 to H-2 for the α-sugars and H-1 to H-3 and H-5 for the β-sugars including the 3,4-linked β-Gal (**D**). NOESY also gave inter-residue correlations to the neighbouring residue and the linkage site proton for most units: H-1 α-Glc (**A**) to H-4 of GlcA (**F**), H-1 α-Glc (**B**) to H-5 of β-Gal (**D**), H-1 β-Glc (**C**) to H-3 (and H-5) of β-Gal (**D**), H-1 β-Gal (**D**) to H-4 of β-Glc (**C**), and H-1 β-Glc (**E**) to H-6 and H-6' of α-Glc (**A**). Small crosspeaks were detected at a lower level between H-1 α-Glc (**B**) to H-4 of β-Gal (**D**) at 4.27 ppm and H-1 β-GlcA (**F**) to H-6' of α-Glc (**B**) at 3.95 ppm, thus elucidating all the linkage positions in the **CPS KK207-2** RU. The assignments and sequence of residues in the RU were corroborated by ^{1}H-^{13}C correlation experiments: edHSQC, HSQC-TOCSY, HSQC-NOESY and HMBC. This permitted full assignment of the ^{1}H and ^{13}C chemical shifts for each spin system which are collected in Table 1 and the ^{13}C assignments are shown in Fig. 5. Downfield displacements of the signals for C-6 of **A** and **B**, C-4 of **C** and C-3 and C-4 of **D** compared to their shifts in the spectra of the corresponding non-substituted monosaccharides [17], demonstrated the glycosylation pattern of the RU. This was in agreement with the NOESY data and confirmed by inter-residue ^{1}H and ^{13}C correlations established using HMBC and HSQC-NOESY. Some key HMBC correlations are shown in Fig. 6, displayed as an overlay with HSQC: H-1 of α-Glc (**A**) to C-4 of GlcA (**F**) at 77.4 ppm, H-1 of β-Glc (**C**) to C-3 of β-Gal (**D**) at 82.0 ppm, H-1 β-Glc (**E**) to C-6 of α-Glc (**A**) at 68.7 ppm and H-4 of β-Gal (**D**) to C-1 of α-Glc (**B**) at 99.5 ppm. Thus a combination of ^{1}H-^{1}H and ^{1}H-^{13}C experiments confirmed the linkage positions, sequence and structure of the hexasaccharide RU with the main chain disaccharide →4)-β-D-Glc-(1→3)-β-D-Gal-(1→ and the tetrasaccharide side chain β-D-Glc-(1→6)-α-D-Glc-(1→4)-β-D-GlcA-(1→6)-α-D-Glc→ linked to C-4 of β-D-Gal.

The ^{1}H NMR spectrum of native **CPS KK207-2** contains a singlet at 2.16 ppm due to ~60% O-acetylation. 2D NMR experiments identified that the O-acetylation was on a single site: C-6 of the 3,4-linked β-D-Gal (unit **D**). Comparison of the HSQC spectra for **deOAc CPS KK207-2** and **CPS KK207-2** showed a decrease in the intensity of the C-6 Gal peak at 61.0 ppm and the presence of a new peak at 63.6 ppm with attached protons at 4.35 and 4.36 ppm attributed to the CH_2OAc group of β-D-Gal6Ac.

3.5. Characterization of the products obtained by lithium degradation of the ***CPS KK207-2***

CPS KK207-2 was treated with lithium metal [16], in order to destroy the glucuronic acid residue, and subsequently purified on a Bio Gel P2 column; the elution profile is shown in Fig. S1. The fractions eluting in the void volume of the column, were grouped and named **Li CPS KK207-2** were subjected to 1D and 2D NMR spectroscopy in order to confirm the sequence in the polysaccharide backbone. The ^{1}H NMR spectrum (Fig. 1e) showed three main anomeric signals at 4.97, 4.71, and 4.56 ppm, which were assigned to t-α-Glc, 4-linked β-Glc and 3,4-linked β-Gal, respectively, based on previous findings. A less intense signal at 4.49 ppm was attributed to residual GlcA which had not been destroyed completely by the lithium treatment. Homo-and heteronuclear 2D NMR experiments (data not shown) determined the chemical shifts for each spin system (Table S4). Some key HMBC correlations are shown in Fig. S4, displayed as an overlay with HSQC: H-1 of α-Glc to C-4 of β-Gal at 76.63 ppm, H-1 of β-Glc to C-3 of β-Gal at 81.62 ppm, and H-1 β-Gal to C-4 of β-Glc at 80.16 ppm. These inter-residue correlations confirmed the linkages and sequence resulting in the following repeating unit structure:

[4)-β-D-Glc*p*-(1→3)-β-D-Gal*p*-(1→]$_n$
4
↑
1
α-D-Glc*p*

In conclusion, all the experimental data collected demonstrated that the capsular polysaccharide produced by *Klebsiella pneumoniae* clinical strain KK207-2, belonging to the sequence type ST258, has a repeating unit comprised of a disaccharide backbone and a tetrasaccharide side chain, with the following structure:

```
                                                OAc
                                                 |
                                                 6
                                         [3)-β-D-Gal-(1→4)-β-D-Glc-(1→]n
                                                 4
                                                 ↑
                                                 1
β-D-Glcp-(1→6)-α-D-Glcp-(1→4)-β-D-GlcpA-(1→6)-α-D-Glcp
```

This structure is a novel one among the *K. pneumoniae* K antigens, in agreement with the original structure of the *cps*$_{207\text{-}2}$ gene cluster based on the *wzc*-based genotyping method [4,6]. **CPS KK207-2** shares structural similarities with other K antigens. In fact, it has the same backbone of K22 [18], including the O-acetyl position on C-6 of the Gal residue, K25 [19], and K37 [20]. Furthermore, part of its side chain is also present in K18 [21], K22, K23 [22], K26 [23], K37 and K41 [24] polysaccharides (Fig. S5).

3.6. Prediction of the reactions catalyzed by the cps$_{207\text{-}2}$ gene cluster glycosyltransferases

Taking advantage of the sequenced *cps*$_{207\text{-}2}$ gene cluster [6], and of the structural knowledge gained in the present investigation, the assignment of each glycosyltransferase (GT) to the corresponding catalyzed reaction was undertaken. The *cps* gene cluster in *Klebsiella* consists of some highly conserved genes together with a less conserved region [5]. Among the highly conserved are six genes located at the 5’ end and mainly involved in the export and polymerization of the repeating unit (*galF, cpsACP, wzi, wza, wzb, wzc)*, and two genes located at the 3’ end and encoding for glucose-6-phosphate dehydrogenase and UDP-glucose dehydrogenase (*gnd* and *ugd*). The central region is the less conserved one and is responsible for K-type variation. In the *cps*$_{207\text{-}2}$ cluster such region contains mainly genes coding for GTs. The structural similarity of **CPS KK207-2** with K22 [18] and K37 [20] polysaccharides agrees with the high homology detected within their corresponding *cps* gene clusters [6]. In fact, the gene content of *cps*$_{207\text{-}2}$ is the same of *cps*$_{K22}$ except for a transposon insertion that caused the substitution of three genes in the central region of the cluster. *cps*$_{207\text{-}2}$ and *cps*$_{K22}$ share three glycosyltransferases that can be related to the formation of the following glycosidic bonds: Gal(β1-4)Glcβ in the backbone, and Glc(α1-4)Galβ and GlcA(β1-

6)Glcα in the side chain (Fig. S5). The precise reaction catalyzed by each of these three GTs can be inferred by comparison with other strains with which **CPS KK207-2** and K22 share only one glycosidic bond and therefore, one GT in their gene clusters [5]. More precisely, they share the Gal(β1-4)Glcβ structure and the WcuW GT with K25 [5,19], and the GlcA(β1-6)Glcα structure and the WckA GT with K23 [5,22]. By exclusion, WcmA GT, the third common GT between $cps_{207\text{-}2}$ and cps_{K22}, can be predicted to catalyze the addition of the first sugar residue of the side chain to the backbone, Glc(α1-4)Galβ (Fig. 7).

Two genes carried by the transposon insertion in $cps_{207\text{-}2}$ (*orf13* and *orf15*) code for GTs not present in the cps_{K22} cluster, they can be predicted to catalyze the remaining two glycosidic bonds in the side chain. In an attempt to identify the precise reaction catalyzed by these enzymes, it was found that **CPS KK207-2** shares only one glycosidic bond with the K18 polysaccharide [5,21], namely Glc(α1-4)GlcAβ (Fig. S5). Comparison of the protein sequences of the GTs coded in the *cps* clusters of these two strains showed an acceptable homology level only between the protein coded by *orf13* of $cps_{K207\text{-}2}$ and WcuH of cps_{K18} (51% identities, 67% positives), strongly suggesting that this GT is responsible for the addition of the activated α-D-Glc*p* residue to C4 of the uronic acid. By exclusion, the GT coded by *orf15* should catalyze the addition of the terminal β-D-Glc*p* residue to C6 of the α-D-Glc*p* in the side chain. This hypothesis was confirmed by comparison with the K26 type [5,23], since the structure Glc(β1-6)Glcα is common to both CPSs (Fig. S5). Sequence analysis identified homology between the protein coded by *orf15* and the glycosyltransferase WcuU (41% identities, 57% positives).

Despite the structural similarities of **CPS KK207-2** and K41 side chains (Glc(β1-6)Glc(α1-4)GlcA(β) which might explain the cross-reactivity observed for three ST258 *K. pneumoniae* clinical strains isolated in a Greek hospital [25], comparison of the GTs coded by $cps_{KK207\text{-}2}$ and cps_{K41} did not retrieve any similarity. After comparison of common and unique genes and structures of K41 and K12 antigens, Pan et al. [5] suggested that WcpT and WcpU might be the GTs involved in the synthesis of the side chain of K41 CPS. Nevertheless, this seems unlikely for two reasons: i) the *wcpT* and *wcpU* sequences are actually part of a unique gene in a *K. oxytoca* strain genome (GenBank accession no. CP026285, locus_tag=”C2U42_08200”) and ii) in the cps_{K41} cluster the *wcpT* and *wcpU* sequences flank an IS*1* transposase insertion which might have interrupted a single gene sequence. Therefore, it is likely that some of the cps_{K41} genes have yet to be identified. The biochemical pathway for the synthesis of the **CPS KK207-2** repeating unit is reported in Fig. 7 and the data are summarized in Table 2.

Author Contributions:

P. C. and R. R. designed the study. B. B. purified the capsular polysaccharide, performed all the wet chemistry experiments and GLC-MS analyses of carbohydrate derivatives, prepared figures. N. R., P. C., B. B. and R. R. recorded and interpreted NMR spectra. C. L. grew the bacteria and assigned the glycosyltransferases functions. M. M. D. and G. M. R. contributed to genetic analysis. P. C., N. R., and C. L. wrote the main manuscript text and prepared figures. All authors critically reviewed and edited the manuscript.

Competing interests

Authors have no competing interests to declare.

Acknowledgments

Funding: This work was supported by the University of Trieste (FRA 2015).

References

[1] M.K. Paczosa, J. Mecsas, *Klebsiella pneumoniae*: Going on the Offense with a Strong Defense, Microbiol. Mol. Biol. Rev. 80(3) (2016) 629–661. https://doi.org/10.1128/MMBR.00078-15.

[2] C-R. Lee, J.H. Lee, K.S. Park, Y.B. Kim, B.C. Jeong, S.H. Lee, 2016. Global dissemination of carbapenemase-producing *Klebsiella pneumoniae*: epidemiology, genetic context, treatment options, and detection methods, Front. Microbiol. 7, 895. https://doi.org/10.3389/fmicb.2016.00895.

[3] S. Brisse, V. Passet, A.B. Haugaard, A. Babosan, N. Kassis-Chikhani, C. Struve, D. Decré, *wzi* gene sequencing, a rapid method for determination of capsular type for *Klebsiella* strains, J. Clin. Microbiol. 51 (2013) 4073–4078. https://doi.org/10.1128/JCM.01924-13.

[4] Y.J. Pan, T.L. Lin, Y.H. Chen, C.R. Hsu, P.F. Hsieh, M.C. Wu, J.T. Wang, 2013. Capsular types of *Klebsiella pneumoniae* revisited by *wzc* sequencing, PLoS ONE 8(12), e80670. https://doi.org/10.1371/journal.pone.0080670.

[5] Y.J. Pan, T.L. Lin, C.T. Chen, Y.Y. Chen, P.F. Hsieh, C.R. Hsu, M.C. Wu, J.T. Wang, 2015. Genetic analysis of capsular polysaccharide synthesis gene clusters in 79 capsular types of *Klebsiella* spp., Scientific Reports 5, 15573. https://doi.org/10.1038/srep15573.

[6] M.M. D'Andrea, F. Amisano, T. Giani, V. Conte, N. Ciacci, S. Ambretti, L. Santoriello, G.M. Rossolini, 2014. Diversity of capsular polysaccharide gene clusters in Kpc-producing *Klebsiella pneumoniae* clinical isolates of sequence type 258 involved in the italian epidemic, PLoS ONE 9(5), e96827. https://doi.org/10.1371/journal.pone.0096827.

[7] F.R. Deleo, L. Chen, S.F. Porcella, C.A. Martens, S.D. Kobayashi, A.R. Porter, K.D. Chavda, M.R. Jacobs, B. Mathema, R.J. Olsen, R.A. Bonomo, J.M. Musser, B.N. Kreiswirth, Molecular dissection of the evolution of carbapenem-resistant multilocus sequence type 258 *Klebsiella pneumoniae*, Proc. Natl. Acad. Sci. 111(13) (2014) 4988–4993. https://doi.org/10.1073/pnas.1321364111.

[8] P. Albersheim, D.J. Nevins, P.D. English, A. Karr, A method for the analysis of sugars in plant cell-wall polysaccharides by gas-liquid chromatography, Carbohydr. Res. 5 (1967) 340–345. https://doi.org/10.1016/S0008-6215(00)80510-8.

[9] K. Kakehi, S. Honda, Silyl ethers of carbohydrates, in: C.J. Biermann, G.D. McGinnis (Eds.), Analysis of carbohydrates by GLC and MS, Boca Raton (FL), CRC Press, 1989, pp. 43–85.

[10] P.J. Harris, R.J. Henry, A.B. Blakeney, B.A. Stone, An improved procedure for the methylation analysis of oligosaccharides and polysaccharides, Carbohydr. Res. 127 (1984) 59–73. https://doi.org/10.1016/0008-6215(84)85106-X.

[11] D.P. Sweet, R.H. Shapiro, P. Albersheim, Quantitative analysis by various g.l.c. response-factor theories for partially methylated and partially ethylated alditol acetates, Carbohydr. Res. 40 (1975) 217–225. https://doi.org/10.1016/S0008-6215(00)82604-X.

[12] R.L. Taylor, H.E. Conrad, Stoichiometric depolymerization of polyuronides and glycosaminoglycuronans to monosaccharides following reduction of their carbodiimide-activated carboxyl group, Biochemistry 11 (1972) 1383–1388. https://doi.org/10.1021/bi00758a009.

[13] S.F. Osman, W.F. Fett, M.L. Fishman, Exopolysaccharides of the phytopathogen *Pseudomonas syringae* pv. Glycinea, J. Bacteriol. 166(1) (1986) 66–71. https://doi.org/10.1128/jb.166.1.66-71.1986.

[14] G.J. Gerwig, J.P. Kamerling, J.F.G. Vliegenthart, Determination of the d and l configuration of neutral monosaccharides by high-resolution capillary g.l.c., Carbohydr. Res. 62 (1978) 349–357. https://doi.org/10.1016/S0008-6215(00)80881-2.

[15] I.J. Goldstein, G.W. Hay, B.A. Lewis, F. Smith, Controlled degradation of polysaccharides by periodate oxidation, reduction, and hydrolysis, Methods Carbohydr. Chem. 5, (1965) 361–370.

[16] J.M. Lau, M. McNeil, A.G. Darvill, P. Albersheim, Selective degradation of the glycosyluroic acid residues of complex carbohydrates by lithium dissolved in ethylenediamine, Carbohydr. Res. 168 (1987) 219–243. https://doi.org/10.1016/0008-6215(87)80028-9.

[17] P.E. Jansson, L. Kenne, G. Widmalm, Computer-assisted structural analysis of polysaccharides with an extended version of CASPER using ^{1}H-and ^{13}C-NMR data. Carbohydr. Res. 188 (1989) 169–191. https://doi.org/10.1016/0008-6215(89)84069-8.

[18] L.A. Parolis, H. Parolis, H. Niemann, S. Stirm, Primary structure of *Klebsiella* serotype K22 capsular polysaccharide: another glycan containing 4-O-[(S)-1-carboxyethyl]-D-glucuronic acid, Carbohydr. Res. 179 (1988) 301–314. https://doi.org/10.1016/0008-6215(88)84126-0.

[19] H. Niemann, B. Kwiatkowski, U. Westphal, S. Stirm, *Klebsiella* serotype 25 capsular polysaccharide: primary structure and depolymerization by a bacteriophage-borne glycanase, J. Bacteriol. 130 (1977) 366–374.

[20] B. Lindberg, B. Lindqvist, J. Lönngren, W. Nimmich, Structural studies of the capsular polysaccharide of *Klebsiella* type 37, Carbohydr. Res. 58 (1977) 443–451. https://doi.org/10.1016/S0008-6215(00)84371-2.

[21] G.G.S. Dutton, K.L. Mackie, M.T.Yang, Structural investigation of *Klebsiella* serotype K18 polysaccharide, Carbohydr. Res. 65 (1978) 251–263. https://doi.org/10.1016/S0008-6215(00)84317-7.

[22] G.G.S. Dutton, K.L. Mackie, A.V.Savage, M.D. Stephenson, Structural investigation of the capsular polysaccharide of *Klebsiella* serotype K23, Carbohydr. Res. 66, (1978) 125–131. https://doi.org/10.1016/S0008-6215(00)83245-0.

[23] J. Di Fabio, G.G.S. Dutton, Structural investigation of the capsular polysaccharide of *Klebsiella* serotype K26, Carbohydr. Res. 92, (1981) 287–298. https://doi.org/10.1016/S0008-6215(00)80399-7.

[24] J.P. Joseleau, M. Lapeyre, M. Vignon, G.G.S. Dutton, Chemical and n.m.r.-spectroscopic investigation of the capsular polysaccharide of *Klebsiella* serotype K41, Carbohydr. Res. 67 (1978) 197–212. https://doi.org/10.1016/S0008-6215(00)83742-8.

[25] L.S. Tzouvelekis, V. Miriagou, S.D. Kotsakis, K. Spyridopoulou, E. Athanasiou, E. Karagouni, E. Tzelepi, G.L. Daikos, KPC-producing, multidrug-resistant *Klebsiella pneumoniae* sequence type 258 as a typical opportunistic pathogen, Antimicrob. Agents Chemother. 57 (2013) 5144–5146. https://doi.org/10.1128/AAC.01052-13.

Appendix

Supplementary material

Table 1

^{1}H and ^{13}C chemical shift assignments of **CPS deOAc-KK207-2**. Spectra were recorded at 600 MHz at 343 K.

Residue	Chemical shifts (ppm)[a]					
	H-1 C-1	H-2 C-2	H-3 C-3	H-4 C-4	H-5 C-5	H-6 C-6
→6)-α-D-Glc*p*-(1→ (A)	5.48 98.9	3.52 72.5	3.72 73.7	3.52 70.0	3.87 71.8	3.90, 4.11 68.7
→6)-α-D-Glc*p*-(1→ (B)	4.92 99.5	3.52 72.6	3.78 73.4	3.69 69.7	4.32 71.2	3.95, 4.10 68.5
→4)-β-D-Glc*p*-(1→ (C)	4.69 105.1	3.25 74.2	3.66 75.5	3.50 80.9	3.58 75.7	3.77, 4.00 61.7
→3,4)-β-D-Gal*p*-(1→ (D)	4.59 103.8	3.74 71.7	3.92 82.0	4.27 76.2	3.88 76.5	~3.86 61.0
β-D-Glc*p*-(1→ (E)	4.49 103.5	3.34 74.0	3.52 76.6	3.42 70.6	3.46 76.7	3.74, 3.91 61.7
→6)-β-D-Glc*p*A-(1→ (F)	4.48 103.3	3.42 73.9	3.78 77.1	3.80 77.4	3.80 77.5	 175.4

[a] Chemical shifts are given relative to acetone (2.225 ppm for ^{1}H and 31.07 ppm for ^{13}C).

Table 2

Assignments of glycosyltransferases (GT) catalysed reactions on the basis of CPS structural homology and protein sequence homology.

GT	Glycosidic bond	Accession no.	K type	Reference	
				Structure	*cps* cluster
WcuW	β-D-Gal*p*(1→4)-β-D-Glc*p*	BAT23470.1	K25	[19]	[5]
		BA027504.1	K22	[18]	[5]
		BAT23694.1	K37	[20]	[5]
		CCI88053.1	KK207-2	This work	[6]
WcmA	α-D-Glc*p*(1→4)-β-D-Gal*p*	BA027502.1	K22	[18]	[5]
		BAT23692.1	K37	[20]	[5]
		CCI88051.1	KK207-2	This work	[6]
WckA	β-D-Glc*p*A(1→6)-α-D-Glc*p*	BAT23435.1	K23	[22]	[5]
		BA027503.1	K22	[18]	[5]
		BAT23693.1	K37	[20]	[5]
		CCI88052.1	KK207-2	This work	[6]
WcuH	α-D-Glc*p*(1→4)-β-D-Glc*p*A	BAT23379.1	K18	[21]	[5]
		CCI88056.1	KK207-2	This work	[6]
WcuU	β-D-Glc*p*(1→6)-α-D-Glc*p*	BAT23487.1	K26	[23]	[5]
		CCI88058.1	KK207-2	This work	[6]

FIGURE LEGEND

Fig. 1. ^{1}H NMR spectra of **PH-2** (a), **SD-2** (b), **SD-1** (c), **de-OAc CPS KK207-2** (d), **Li CPS KK207-2** (e).

Fig. 2. ESI mass spectrum of the oligosaccharide **SD-2** (a) and ESI MS^2 of the ion at 483.2 *m/z* (b); ESI mass spectrum of the oligosaccharide **SD-1** (c) and ESI MS^2 of the ion at 615.1 *m/z* (d). Assignments are indicated in the figures.

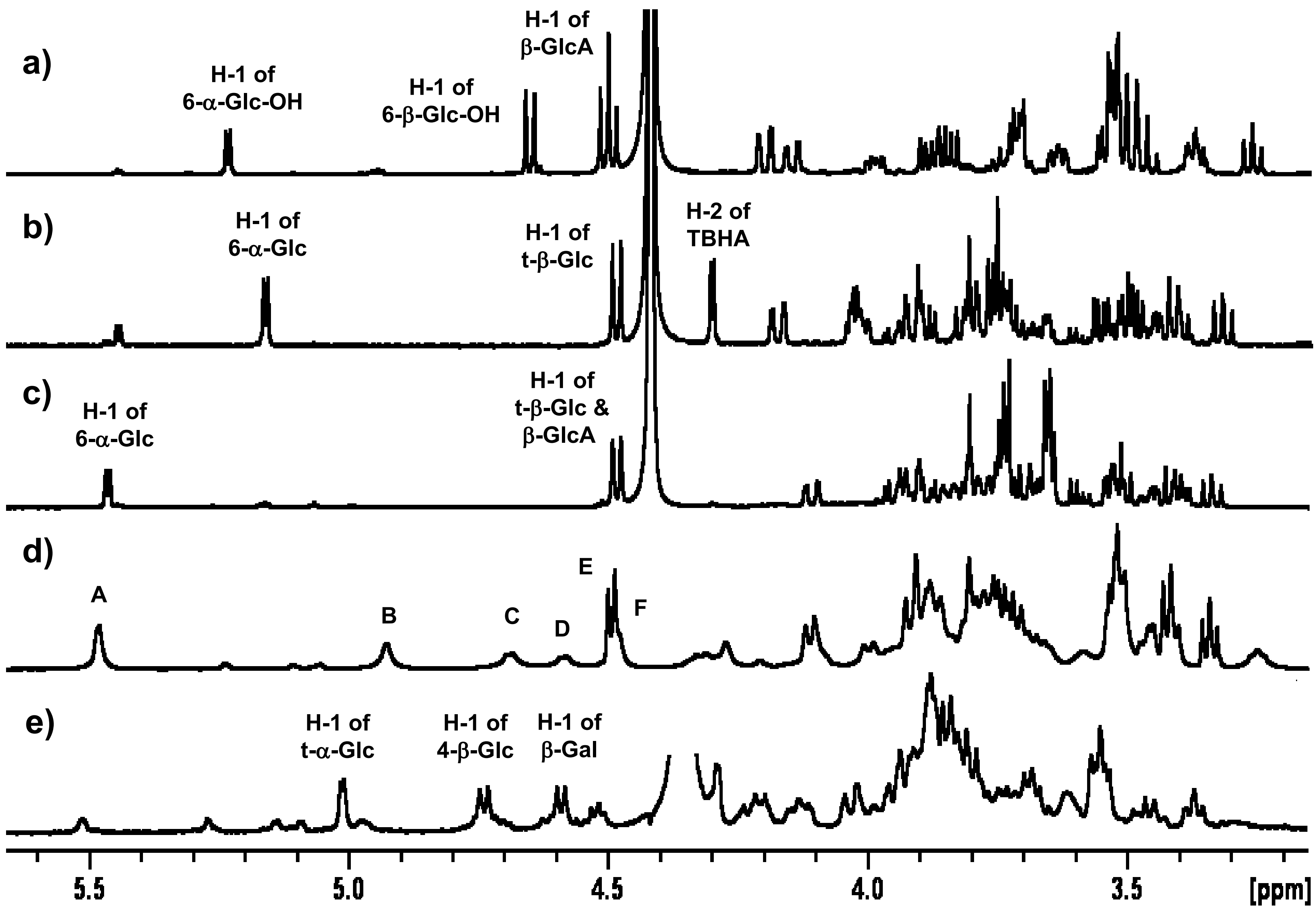
Fig. 3. Overlay of the COSY (red)/TOCSY (black) anomeric region of **deOAc CPS KK207-2** recorded at 600 MHz and 343K. The insert shows crosspeaks for residues **B**, **C** and **D** present at a lower level. Crosspeaks have been labelled according to the corresponding residue (**A** to **E**).

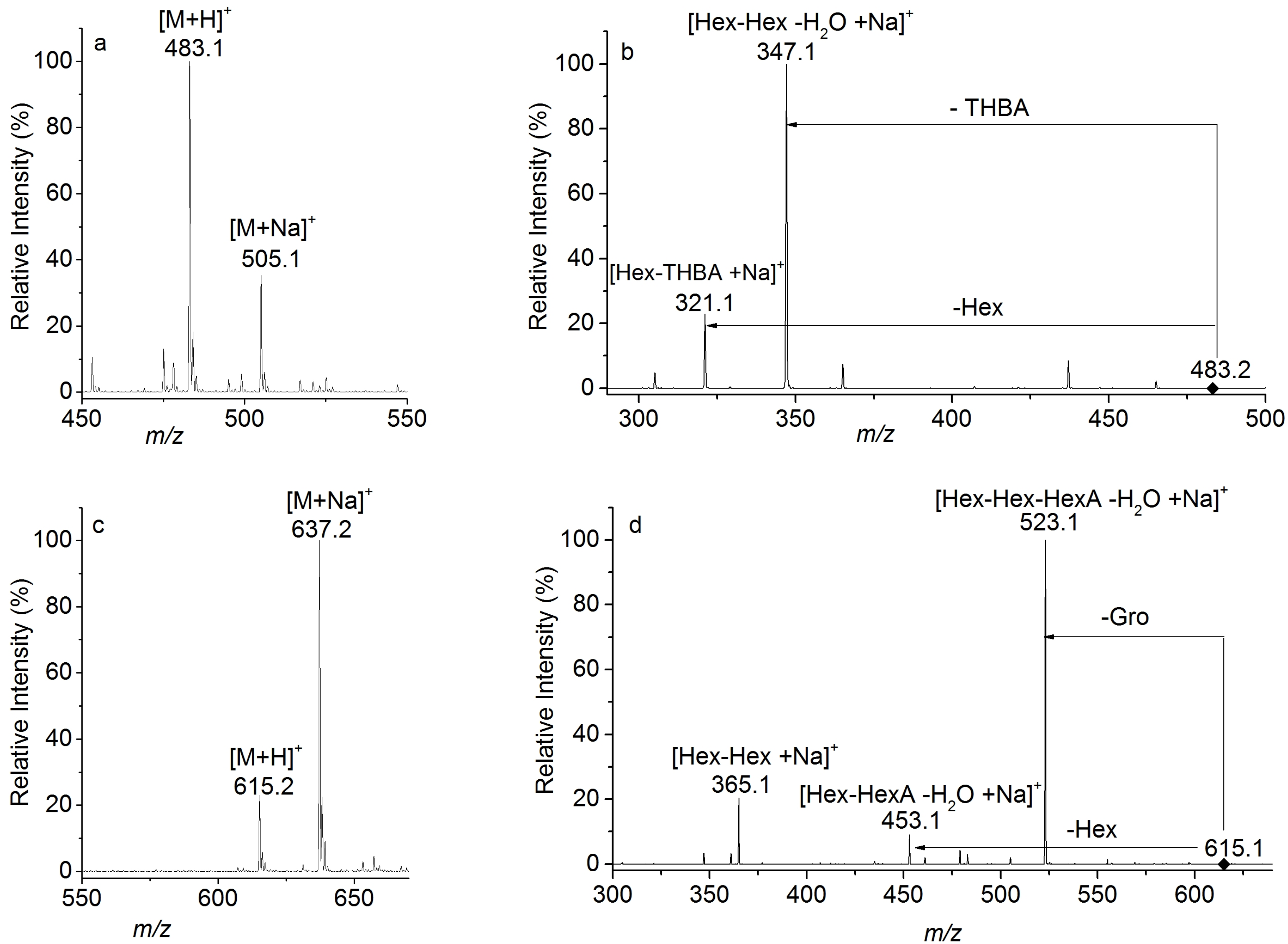
Fig. 4. Expansion of the NOESY anomeric region of **deOAc CPS KK207-2** recorded at 600 MHz and 343K. Intra- and inter-residue correlation crosspeaks have been labelled according to the corresponding residue (**A** to **E**).

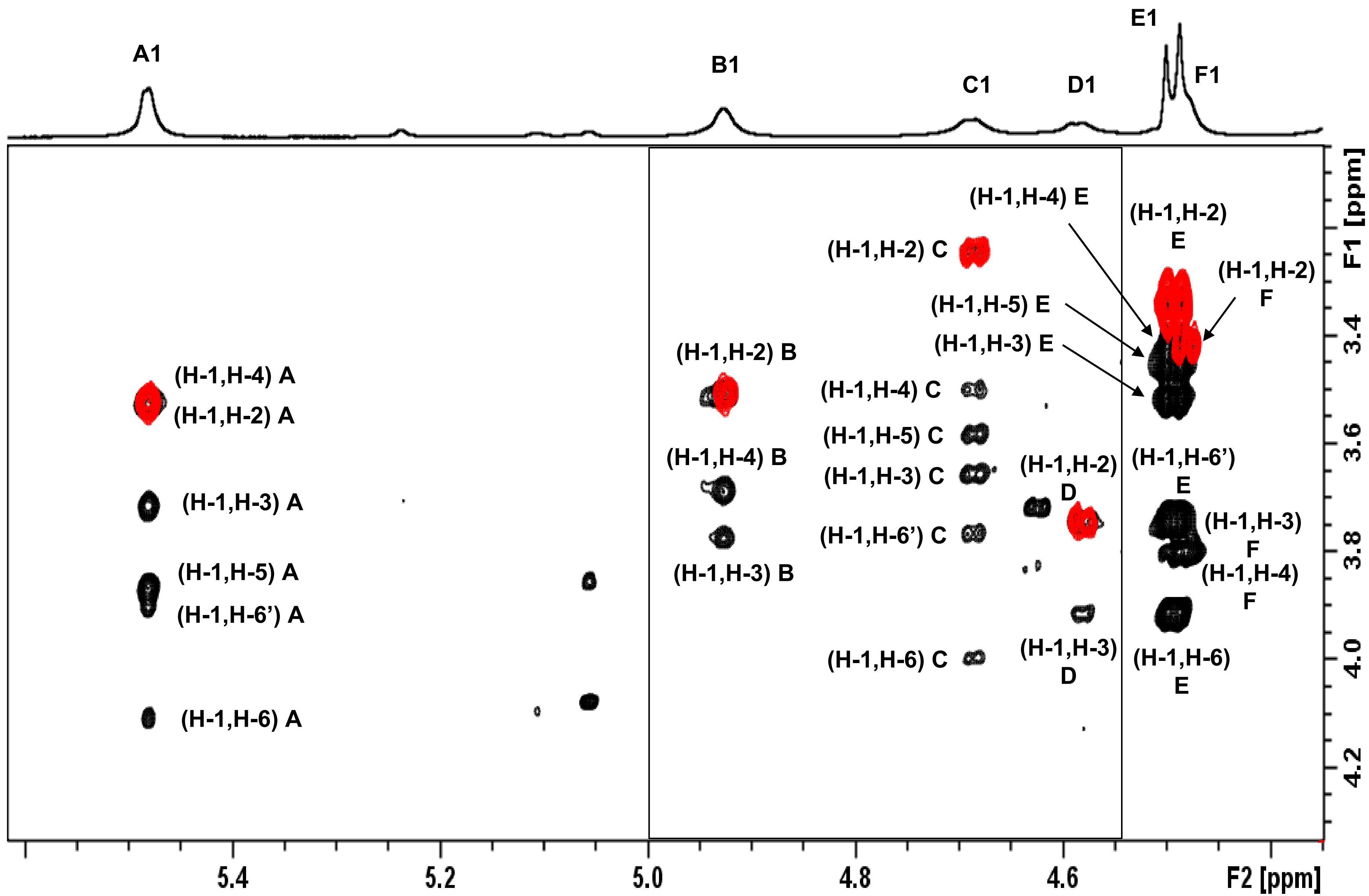
Fig. 5. Expansion of the 1D ^{13}C NMR spectrum of **deOAc CPS KK207-2** recorded at 150 MHz and 343K showing the anomeric and ring regions. Carbon peaks have been labelled according to the corresponding residue (**A** to **E**). Methylene assignments were confirmed by recording the DEPT spectrum.

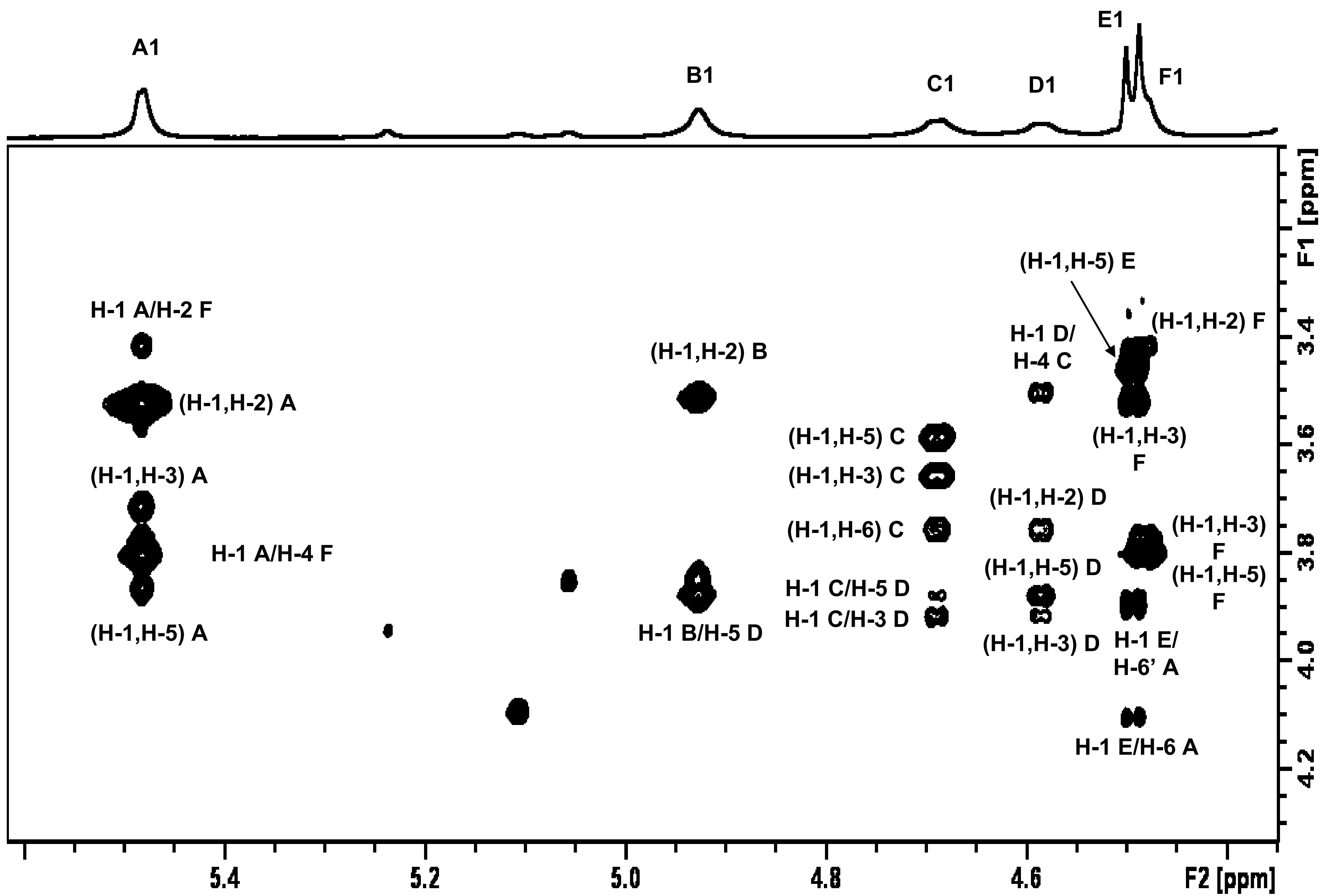
Fig. 6. Expansion overlay of the HSQC (red)/HMBC (black) spectra of **deOAc CPS KK207-2** recorded at 600 MHz and 343K. The corresponding parts of the ^{1}H (1D DOSY) and ^{13}C NMR spectra are shown along the horizontal and vertical axis, respectively. Proton/carbon crosspeaks have been labelled according to the corresponding residue (**A** to **E**).

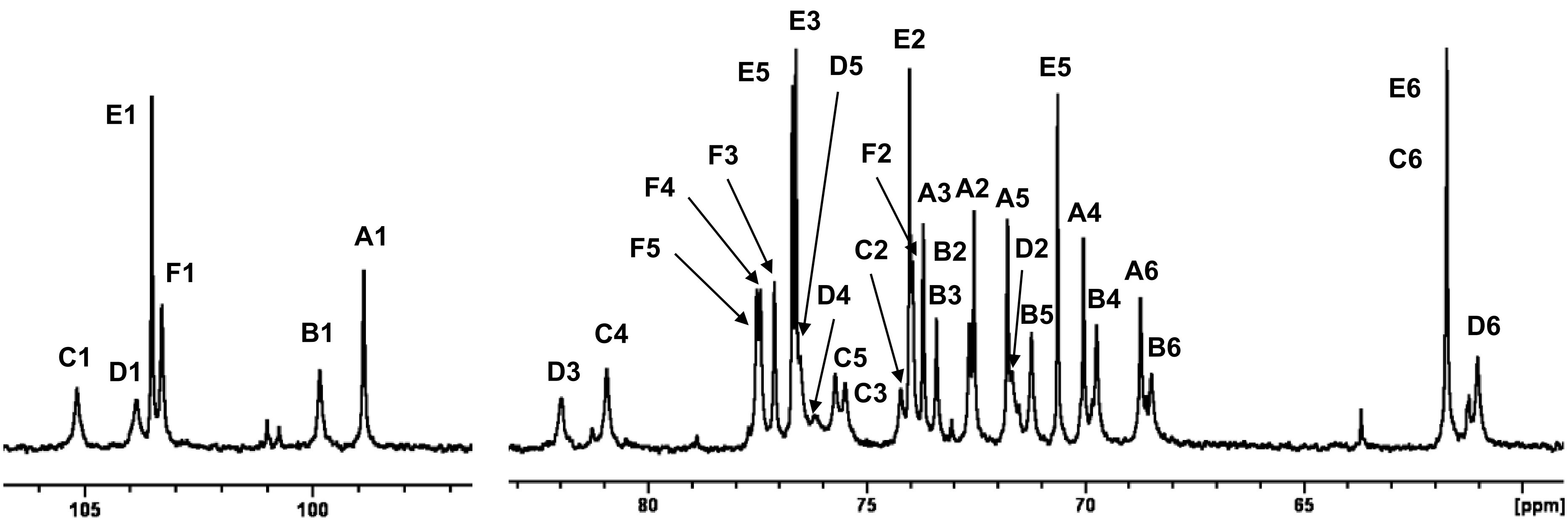
Fig. 7. Proposed glycosyltransferase and polymerase activity of the *Klebsiella pneumoniae* KK207-2 *cps* gene cluster. Glycosyltransferases responsible for each elongation step are listed above the respective glycosidic linkage. The polymerization site is marked by an arrow (a). Identification of genes coded in the central region of the $cps_{207\text{-}2}$ cluster (portion 8294-19716 of accession number HE866752). IS*Ecl1*-like and a truncated version of IS*1*-like insertion sequences are indicated by a blue and a green box, respectively (b).

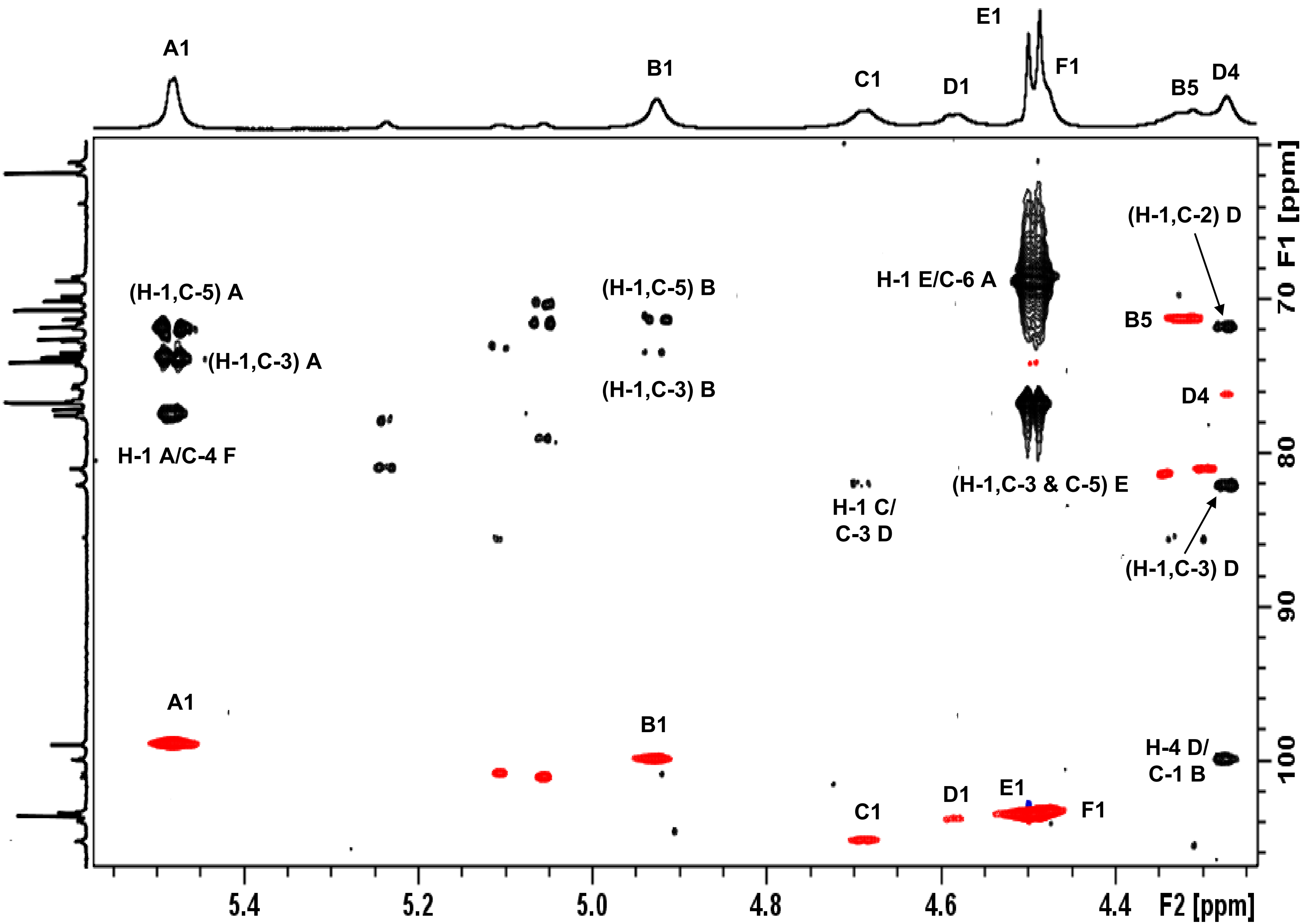

a)
H-1 of 6-α-Glc-OH
H-1 of 6-β-Glc-OH
H-1 of β-GlcA
b)
H-1 of 6-α-Glc
H-1 of t-β-Glc
H-2 of TBHA
c)
H-1 of 6-α-Glc
H-1 of t-β-Glc & β-GlcA
d)
A
B
C
D
E
F
e)
H-1 of t-α-Glc
H-1 of 4-β-Glc
H-1 of β-Gal
5.5
5.0
4.5
4.0
3.5
[ppm]

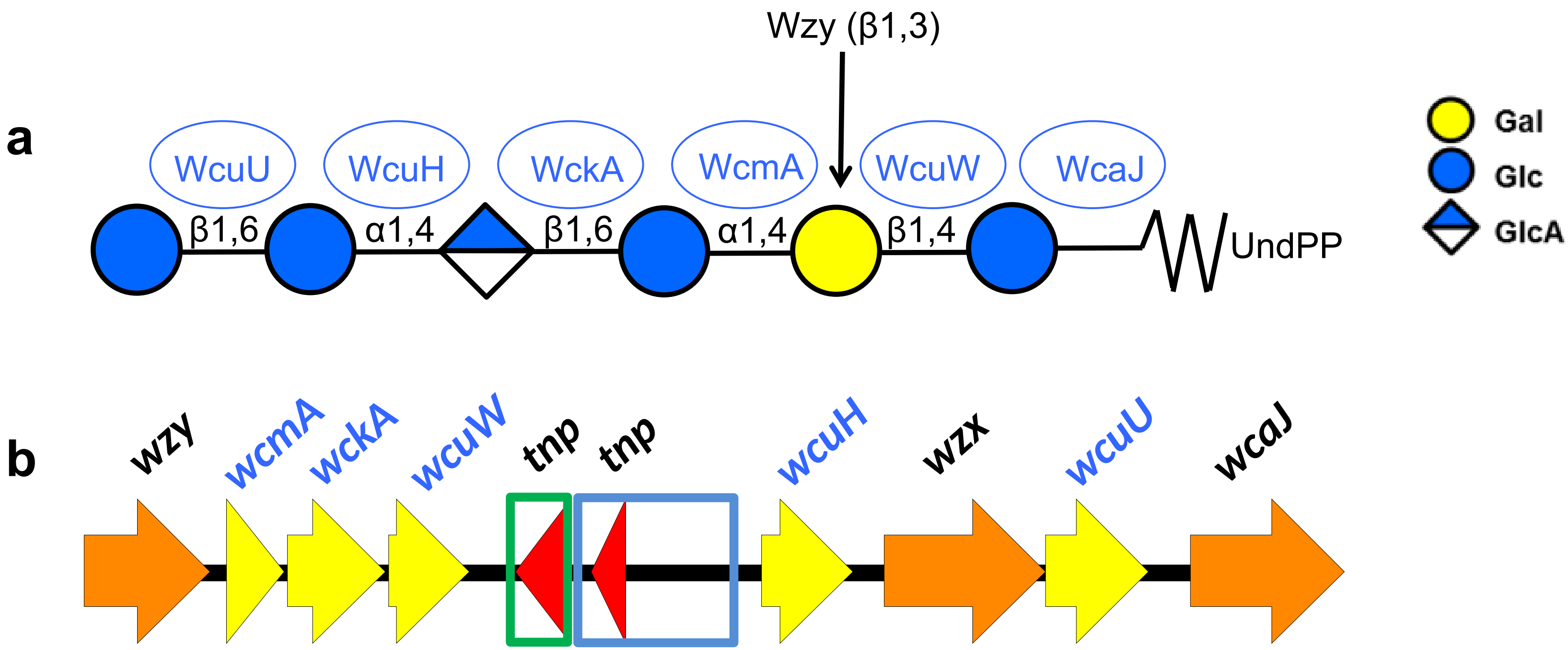

a
[M+H]+
483.1
[M+Na]+
505.1
Relative Intensity (%)
m/z
b
[Hex-Hex -H2O +Na]+
347.1
- THBA
[Hex-THBA +Na]+
321.1
-Hex
483.2
Relative Intensity (%)
m/z
c
[M+Na]+
637.2
[M+H]+
615.2
Relative Intensity (%)
m/z
d
[Hex-Hex-HexA -H2O +Na]+
523.1
-Gro
[Hex-Hex +Na]+
365.1
[Hex-HexA -H2O +Na]+
453.1
-Hex
615.1
Relative Intensity (%)
m/z


A1
B1
C1
D1
E1
F1
(H-1,H-4) A
(H-1,H-2) A
(H-1,H-3) A
(H-1,H-5) A
(H-1,H-6') A
(H-1,H-6) A
(H-1,H-2) B
(H-1,H-4) B
(H-1,H-3) B
(H-1,H-2) C
(H-1,H-4) C
(H-1,H-5) C
(H-1,H-3) C
(H-1,H-6') C
(H-1,H-6) C
(H-1,H-2) D
(H-1,H-3) D
(H-1,H-4) E
(H-1,H-2) E
(H-1,H-5) E
(H-1,H-3) E
(H-1,H-6') E
(H-1,H-6) E
(H-1,H-2) F
(H-1,H-3) F
(H-1,H-4) F
F1 [ppm]
3.4
3.6
3.8
4.0
4.2
5.4
5.2
5.0
4.8
4.6
F2 [ppm]


A1
B1
C1
D1
E1
F1
F1 [ppm]
(H-1,H-5) E
H-1 A/H-2 F
(H-1,H-2) F
H-1 D/
H-4 C
(H-1,H-2) B
(H-1,H-2) A
(H-1,H-5) C
(H-1,H-3)
F
(H-1,H-3) A
(H-1,H-3) C
(H-1,H-2) D
(H-1,H-6) C
(H-1,H-3)
F
H-1 A/H-4 F
(H-1,H-5) D
(H-1,H-5)
F
H-1 C/H-5 D
H-1 C/H-3 D
(H-1,H-5) A
H-1 B/H-5 D
(H-1,H-3) D
H-1 E/
H-6' A
H-1 E/H-6 A
3.4
3.6
3.8
4.0
4.2
5.4
5.2
5.0
4.8
4.6
F2 [ppm]


C1
D1
E1
F1
B1
A1
D3
C4
F5
F4
F3
E5
E3
D5
D4
C5
C3
C2
F2
E2
A3
B2
B3
A2
A5
D2
B5
E5
A4
B4
A6
B6
E6
C6
D6
105
100
80
75
70
65
[ppm]


A1
B1
C1
D1
E1
F1
B5
D4
(H-1,C-2) D
(H-1,C-5) A
(H-1,C-3) A
H-1 A/C-4 F
(H-1,C-5) B
(H-1,C-3) B
H-1 E/C-6 A
B5
D4
(H-1,C-3 & C-5) E
H-1 C/
C-3 D
(H-1,C-3) D
A1
B1
C1
D1
E1
F1
H-4 D/
C-1 B
F1 [ppm]
70
80
90
100
5.4
5.2
5.0
4.8
4.6
4.4
F2 [ppm]


a
Wzy (β1,3)
WcuU
WcuH
WckA
WcmA
WcuW
WcaJ
β1,6
α1,4
β1,6
α1,4
β1,4
UndPP
Gal
Glc
GlcA
b
wzy
wcmA
wckA
wcuW
tnp
tnp
wcuH
wzx
wcuU
wcaJ

SUPPLEMENTARY MATERIAL

Structure of the capsular polysaccharide of the KPC-2-producing *Klebsiella pneumoniae* strain KK207-2 and assignment of the glycosyltransferases functions

Barbara Bellich[a], Neil Ravenscroft[b], Roberto Rizzo[a], Cristina Lagatolla[a], Marco Maria D'Andrea[c,d] Gian Maria Rossolini[e,f] and Paola Cescutti[a*]

[a] Department of Life Sciences, University of Trieste, via L. Giorgieri 1, Bdg C11, 34127 Trieste, Italy

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

[c] Department of Medical Biotechnologies, University of Siena, Siena, Italy

[d] Department of Biology, University of Rome "Tor Vergata", Rome, Italy

[e] Department of Experimental and Clinical Medicine, University of Florence, Florence, Italy

[f] Microbiology and Virology Unit, Florence Careggi University Hospital, Florence, Italy

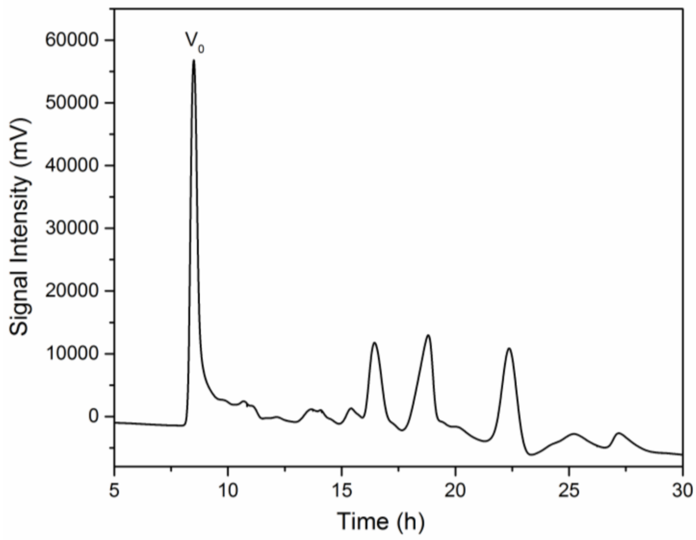


Fig. S1: Bio Gel P2 elution profile of the products obtained after lithium degradation of the **CPS-KK207-2**. Fraction eluting in the void volume of the column (Vo peak) were pooled, named **Li CPS KK207-2**, and used for recording NMR spectra.

Table S1

Determination of the glycosidic linkages in **CPS KK207-2** native and after carboxyl reduction, and **SD-Vo** by GLC-MS of PMAA derivatives.

Linkage[a]		Relative molar ratio[c]			
	RRT[b]	I	II	III	IV
t-Glc	1.00	1.00	1.00	1.00	0.41
4-Glc	1.22	1.00	1.98	2.85	1.18
3-Gal	1.23	-	-	1.07	0.44
6-Glc	1.26	1.64	1.92	0.67	0.28
3,4-Gal	1.34	1.12	1.44	1.35	0.56

[a] t-Glc= terminal non-reducing glucose; the numbers indicate the position of the glycosidic linkages, e.g. 4-Glc = 4-linked glucose; [b] Relative retention time; [c] Peak areas were corrected by the effective carbon response factor [11] and the molar ratio are expressed relative to t-Glc in columns I, II, III and to the sum of 3-Gal + 3,4-Gal in column IV; I = native **CPS KK207-2**; II = **CPS KK207-2** after –COOH reduction [12,13]; III = **SD-Vo**; IV = **SD-Vo** with molar ratios relative to the sum of 3-Gal + 3,4-Gal.

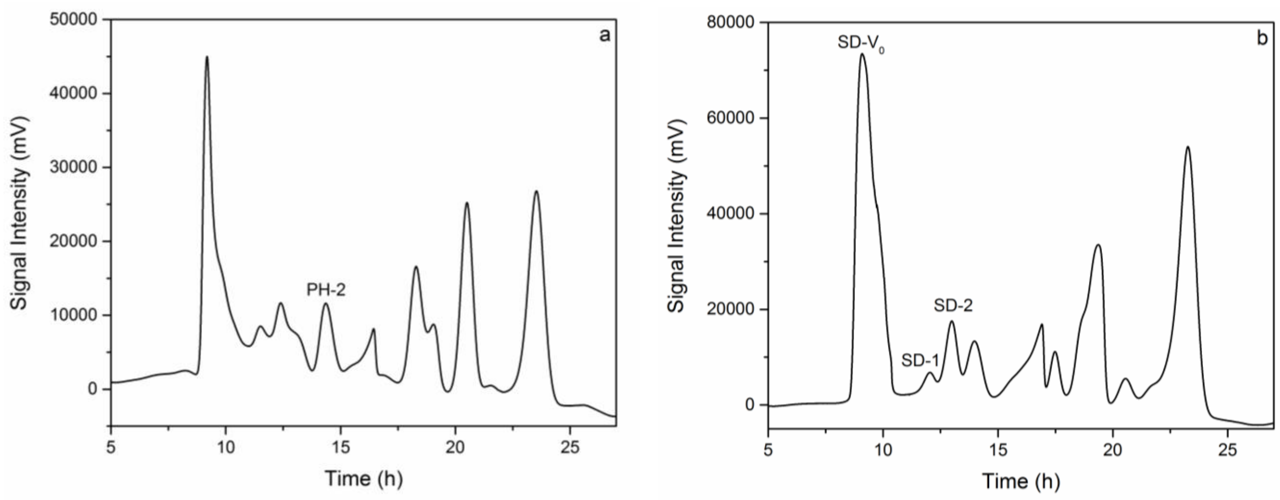


Fig. S2: Bio Gel P2 elution profiles of the products obtained after partial hydrolysis (a) and Smith degradation (b) of **deOAc CPS KK207-2**. Labelled peaks contained mainly only one oligosaccharide and were subjected to ESI-MS and ^{1}H NMR spectroscopy analyses.

Table S2

^{1}H and ^{13}C chemical shift assignments of the sample **SD-2** obtained from Smith degradation of the **CPS KK207-2**. Spectra were recorded at 500 MHz and 333 K.

Residue	Chemical shifts (ppm)[a]					
	H-1 C-1	H-2 C-2	H-3 C-3	H-4 C-4	H-5 C-5	H-6 C-6
→6)-α-D-Glc*p*-(1→	5.15	3.54	3.76	3.49	4.00	3.87, 4.16
	99.83	72.66	73.89	70.51	72.06	69.59
β-D-Glc*p*-(1→	4.48	3.31	3.49	3.40	3.44	3.73, 3.91
	103.65	74.02	76.67	70.59	76.76	61.73
→3)-2,3,4 trihydroxy-butanoic acid	-	4.29	4.02	3.79, 3.76		
	-	74.21	82.24	61.51		

[a]Chemical shifts are given relative to internal acetone (2.225 ppm for ^{1}H and 31.07 ppm for ^{13}C).

Table S3

^{1}H and ^{13}C chemical shift assignments of the sample **SD-1** obtained from Smith degradation of the **CPS KK207-2**. Spectra were recorded at 500 MHz and 333 K.

Residue	Chemical shifts (ppm)[a]					
	H-1 C-1	H-2 C-2	H-3 C-3	H-4 C-4	H-5 C-5	H-6 C-6
→6)-α-D-Glc*p*-(1→	5.47	3.53	3.71	3.53	3.85	3.89 4.11
	99.11	72.63	73.68	70.09	71.83	68.80
β-D-Glc*p*-(1→	4.49	3.34	3.51	3.41	3.45	3.74 3.92
	103.57	74.02	76.70	70.66	76.74	61.78
→4)-β-D-Glc*p*A-(1→	4.49	3.40	3.78	3.81	3.81	-
	103.57	74.02	77.16	77.65	77.65	
→1)-Gro	3.95 3.67	3.85	3.66 3.61			
	71.93	71.83	63.44			

[a]Chemical shifts are given relative to internal acetone (2.225 ppm for ^{1}H and 31.07 ppm for ^{13}C).

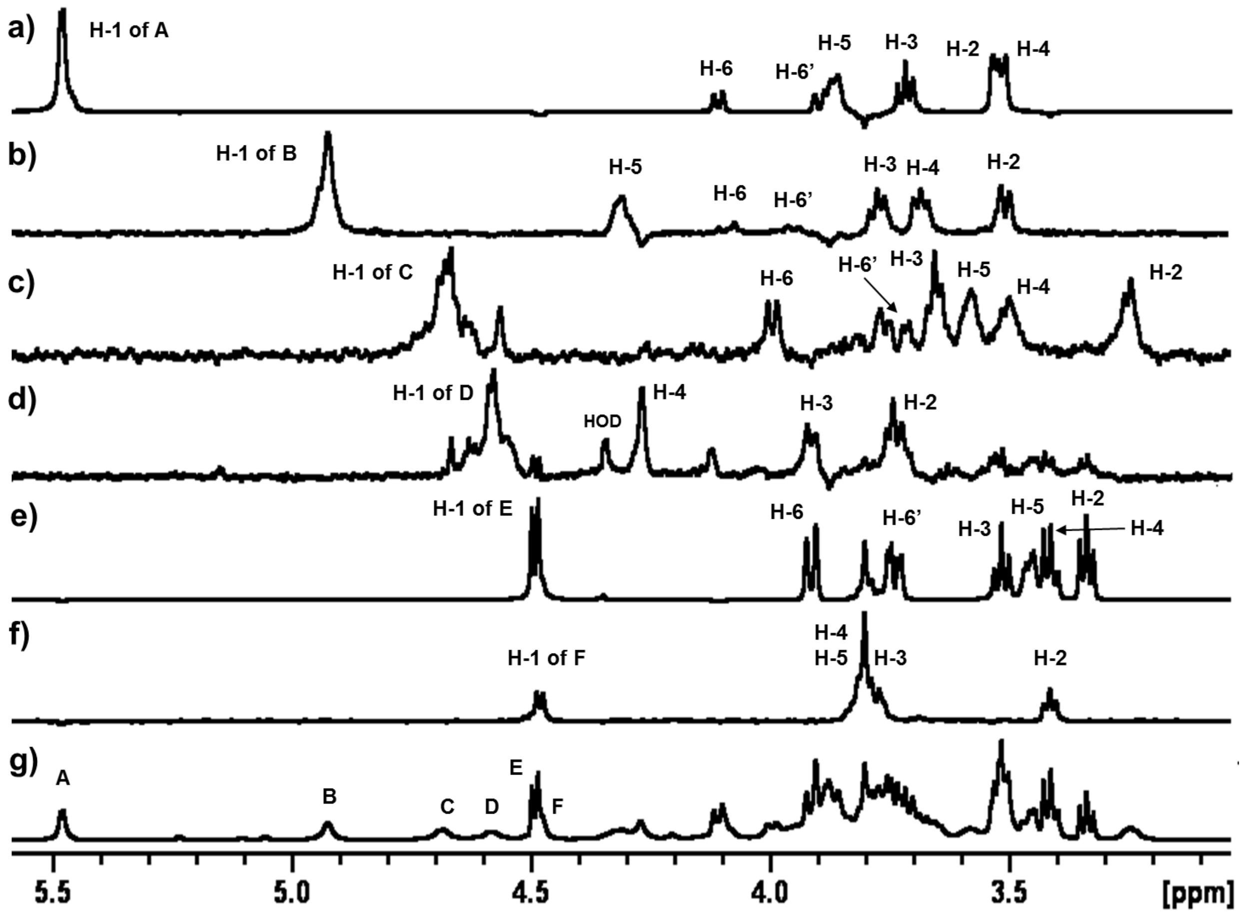


Fig. S3: ^{1}H NMR spectra of **deOAc CPS KK207-2**. Selective 1D TOCSY experiments performed by irradiation of the anomeric protons for residues **A** to **F** (a to f) overlaid with the ^{1}D DOSY spectrum of the CPS (g). Spectra were recorded at 600 MHz and 343 K using a mixing time of 200 ms.

Table S4

[1]H and [13]C chemical shift assignments of the **Li CPS KK207-2** obtained from Lithium degradation [16] of the **CPS KK207-2**. Spectra were recorded at 600 MHz and 343 K.

Residue	Chemical shifts (ppm)[a]					
	H-1 C-1	H-2 C-2	H-3 C-3	H-4 C-4	H-5 C-5	H-6 C-6
α-D-Glc*p*-(1→	4.98	3.52	3.78	3.52	4.18	~3.84
	100.20	72.74	73.53	70.26	72.45	61.21
→4)-β-D-Glc*p*-(1→	4.71	3.34	3.66	3.64	3.56	3.83, 4.00
	105.13	74.17	75.42	80.16	75.66	61.33
→3,4)-β-D-Gal*p*-(1→	4.56	3.79	3.87	4.26	3.81	-
	103.96	71.63	81.62	76.63	76.56	

[a]Chemical shifts are given relative to internal acetone (2.225 ppm for [1]H and 31.07 ppm for [13]C).

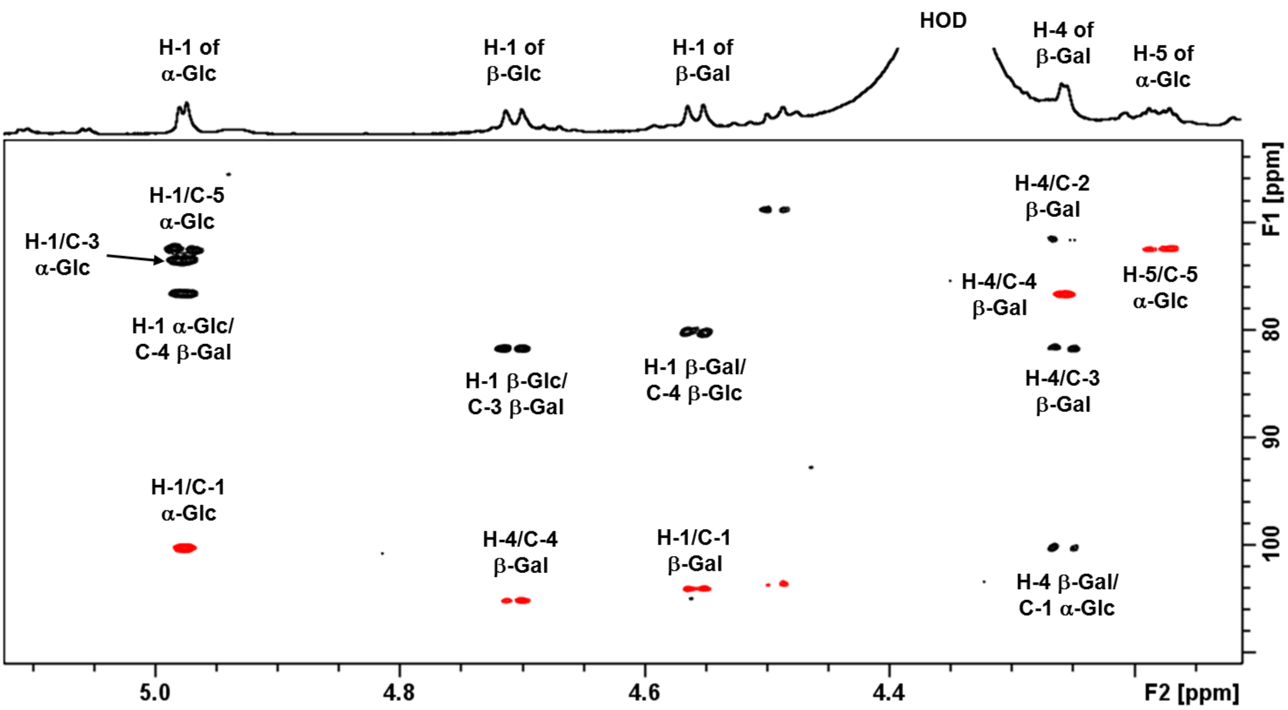


Fig. S4: Expansion overlay of the HSQC (red)/HMBC (black) spectra of **Li CPS KK207-2** recorded at 600 MHz and 343K. The proton/carbon crosspeaks have been labelled according to the corresponding residue (α-Glc, β-Glc and β-Gal); the small peaks are due to residual β-GlcA. The inter-residue correlations confirm the linkages and sequence: α-D-Glc-(1→4)-β-D-Gal-(1→ and the disaccharide backbone, →4)-β-D-Glc-(1→3)-β-D-Gal-(1→.

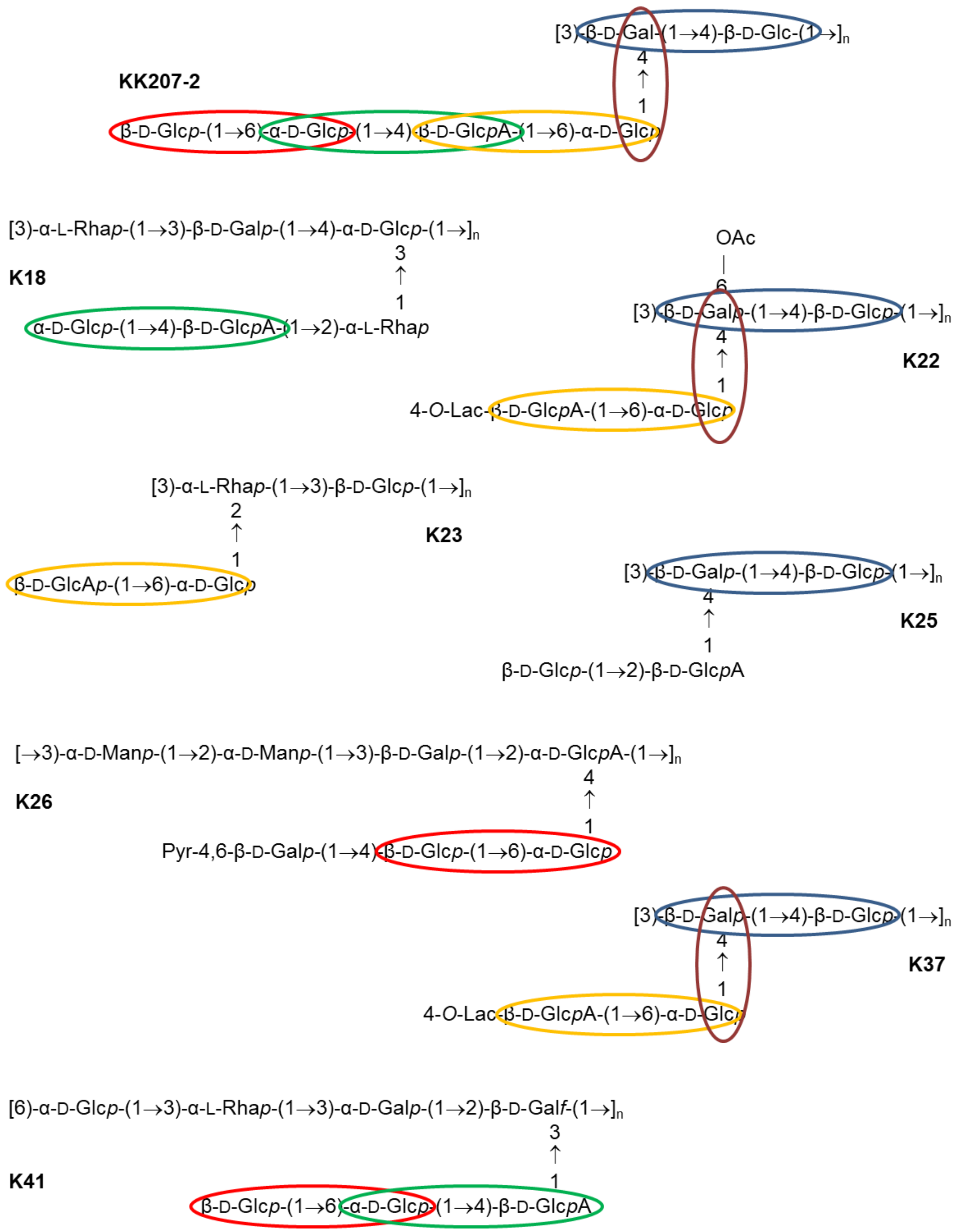


Fig. S5: Structure of the repeating units of different CPS: **KK207-2** (this article), K18 [21], K22 [18], K23 [22], K25 [19], K26 [23], K37 [20] and K41 [24] *Klebsiella pneumoniae* capsular polysaccharides. Circles of different colours highlight common glycosidic linkages. These structural similarities were used to assign the corresponding GTs genes.