

<https://doi.org/10.1038/s41522-024-00555-z>

# *Bacillus subtilis* EpsA-O: A novel exopolysaccharide structure acting as an efficient adhesive in biofilms



Iztok Dogsa<sup>1</sup>, Barbara Bellich<sup>2,3</sup>, Mojca Blaznik<sup>1</sup>, Cristina Lagatolla<sup>3</sup>, Neil Ravenscroft<sup>4</sup>, Roberto Rizzo<sup>3</sup>, David Stopar<sup>1</sup> & Paola Cescutti<sup>3</sup>✉

Extracellular polysaccharides are crucial components for biofilm development. Although *Bacillus subtilis* is one of the most characterized Gram-positive biofilm model system, the structure-function of its exopolysaccharide, EpsA-O, remains to be elucidated. By combining chemical analysis, NMR spectroscopy, rheology, and molecular modeling, high-resolution data of EpsA-O structure from atom to supramolecular scale was obtained. The repeating unit is composed of the trisaccharide backbone  $[-\rightarrow 3)-\beta\text{-D-QuipNAc}4\text{NAc}-(1\rightarrow 3)-\beta\text{-D-GalpNAc}-(1\rightarrow 3)-\alpha\text{-D-GlcpNAc}-(1)]_n$ , and the side chain  $\beta\text{-D-Galp}(3,4\text{-S-Pyr})-(1\rightarrow 6)-\beta\text{-D-Galp}(3,4\text{-S-Pyr})-(1\rightarrow 6)-\alpha\text{-D-Galp}-(1\rightarrow$  linked to C4 of GalNAc. Close agreement between the primary structure and rheological behavior allowed us to model EpsA-O macromolecular and supramolecular solution structure, which can span the intercellular space forming a gel that leads to a complex 3D biofilm network as corroborated by a mutant strain with impaired ability to produce EpsA-O. This is a comprehensive structure-function investigation of the essential biofilm adhesive exopolysaccharide that will serve as a useful guide for future studies in biofilm architecture formation.

*Bacillus subtilis* is one of the best studied Gram-positive model organisms. It is an important industrial organism, being proficient at secreting proteins and making small fine chemicals, as well as acting as a plant growth promoter<sup>1</sup>. It is ubiquitous, naturally transformable and has an extremely powerful genetic toolbox<sup>2</sup>. Its huge evolutionary and environmental success is supported by its ability to form biofilms on a variety of surfaces<sup>3,4</sup>. It is often used as a model system to dissect the mechanisms controlling biofilm extracellular matrix production and the subsequent transition from planktonic to sessile cells, biofilm development, differentiation, physiological heterogeneity, binary switching, quorum sensing, kin discrimination and dispersion<sup>5-10</sup>.

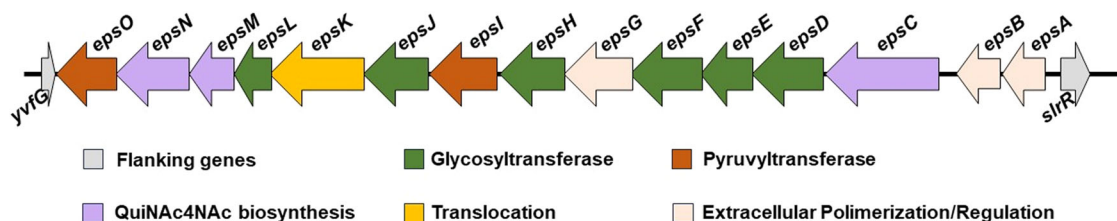
Great advances have been made in recent years in the characterization of *B. subtilis* biofilm matrix components, which include the secreted proteins TasA, TapA and BslA, as well as eDNA. Crystal structures of TasA (PDB ID: 5OF1) and BslA (PDB ID: 4BHU) have been resolved<sup>11,12</sup>. TapA acts as specific chaperone in TasA filament formation<sup>13</sup>, while recently the role of eDNA in early aggregate formation was demonstrated<sup>14</sup>. While previous research<sup>15-21</sup> has revealed that the function of EpsA-O in *B. subtilis* extends beyond providing structural support to the biofilm (i.e. not only supports wrinkled colony morphology, but also plant root colonization, binding of metals, enables the osmotic spreading and protection from

reactive oxygen species), its specific role in the microarchitecture of the biofilm and its chemical structure remains poorly understood.

EpsA-O is the main extracellular polysaccharide in *B. subtilis* and is synthesized by a cluster of 15 gene products of the *epsA-O* operon<sup>15,22</sup> (Fig. 1). EpsA-O is required for complex colony structure and pellicle formation<sup>15,23-25</sup>. The function of only some gene products of the *epsA-O* cluster has already been demonstrated or predicted: EpsA and EpsB are tyrosine kinases that regulate Eps production; EpsC, EpsM and EpsN are responsible for the biosynthesis of *N,N'*-diacetylbaucillosamine (QuiNAc4NAc), a modified monosaccharide synthesized by few bacteria<sup>1</sup>; EpsK is a membrane transporter similar to the Wzx protein of *E. coli*<sup>26</sup>; EpsL binds UDP-QuiNAc4NAc to the lipid acceptor undecaprenol phosphate (UndP) while the glycosyltransferase (GT) EpsD catalyzes the subsequent addition of the *N*-acetylglucosamine<sup>27</sup>.

In spite of several decades of active research, the exact composition of the repeating unit (RU) is still debated with findings from different studies in sharp disagreement<sup>26,28,29</sup>. A recent comprehensive revision<sup>1</sup> concluded that the structural determination of EpsA-O is mandatory in order to move forward in the understanding of *Bacillus subtilis* biofilm architecture and function.

<sup>1</sup>University of Ljubljana, Biotechnical Faculty, Department of Microbiology, Ljubljana, Slovenia. <sup>2</sup>Department of Advanced Translational Diagnostics, Institute for Maternal and Child Health, IRCCS “Burlo Garofolo”, Trieste, Italy. <sup>3</sup>University of Trieste, Department of Life Sciences, Via L. Giorgieri 1, Trieste, Italy. <sup>4</sup>University of Cape Town, Department of Chemistry, Rondebosch, South Africa. ✉e-mail: [pcescutti@units.it](mailto:pcescutti@units.it)



**Fig. 1** | Structure of the *eps* locus of *B. subtilis subsp. subtilis* 6051-HGW. (portion 3514095-3529835 of accession number NC\_020507.1). Arrows indicate the direction, relative length, and function (colored as per legend) of protein-coding genes.

In this work the RU structure of EpsA-O was determined by chemical analysis, and 1D, and 2D NMR spectroscopy. Furthermore, the EpsA-O was characterized rheologically. The data were used to develop a string of beads model of the branched RU to simulate the volume space of the polysaccharide in solution. The results obtained were used to reevaluate the role of EpsA-O in biofilm function and development.

## Results

### EpsA-O has a decisive role in microstructure formation of the biofilm

In the past the CLSM microscopy has been successfully used to reveal the 3D microstructure of *B. subtilis* NDMed strain submerged native biofilms<sup>30,31</sup>. Here we report the 3D microstructure of native mature NCBI3610 *B. subtilis* biofilms on solid-air interface. The pronounced difference in topology in WT against  $\Delta eps$  biofilms was observed in bright-field microscopy (Fig. 2a, b), and in more details using confocal laser scanning microscopy (CLSM) (Fig. 2c, d). Upon 3D reconstruction, significant 3D structures were observed to form only in the WT biofilms. Optical sectioning of these biofilms (Fig. 2e, f) revealed that cells form compact directionally ordered fiber-like structures. When purified EpsA-O was exogenously added to the  $\Delta eps$  mutant cells the roughness and patchiness substantially increased compared to the  $\Delta eps$  mutant (Fig. 3). Compared to the wild type, however, roughness and local ordering were less pronounced. These images (Figs. 2, 3) suggest that EpsA-O acts as adhesive, structuring cells into 3D structures capable of resisting gravity. To resist gravity, the biofilm should be in a gel rather than sol state, which implies that the EpsA-O itself must be in a gel state in the biofilm. In order to elucidate the viscoelastic, physicochemical properties and structure of the EpsA-O, the polysaccharide was first isolated and purified in sufficient amounts.

### The purified EpsA-O has a high molecular mass

The EpsA-O raw isolate, obtained after optimization of the growth medium (Supplementary Fig. 1), was subjected to HPLC-SEC analysis; the elution profile (Supplementary Fig. 2a) showed two main peaks whereas the  $\Delta eps$  mutant profile did not contain the peak at  $t = 18$  min. After performing additional steps of purification, (see Methods) the chromatogram of the purified EpsA-O displayed a single peak at  $t = 18$  min (Supplementary Fig. 2b), containing <2% (w/w) of proteins and nucleic acids (Supplementary Fig. 2d–f). By applying the universal calibration method, the MM (molecular mass) of EpsA-O was estimated to be  $(2.5 \pm 1)$  MDa. For such a large molecule, unusual high values of reducing sugar were obtained with the DNS test, even after extensive dialysis (Supplementary Fig. 2c). This was subsequently attributed to the high content of pyruvate substituent decorating EpsA-O, not the presence of carbonyl reducing ends.

### The EpsA-O forms gels, expands and displays polyelectrolyte properties

Purified EpsA-O solutions showed a pseudoplastic behavior suggesting inter-polymer interactions (Fig. 4a). The critical overlap concentration,  $c^*$ , was at a value of  $(0.010 \pm 0.005)$  g/dL (Fig. 4b). The semi-dilute regime ended with the second critical concentration,  $c^{**}$  at  $(0.15 \pm 0.05)$  g/dL that indicates physical entanglement of polymer chains above this concentration. The intrinsic viscosity,  $[\eta]$ , which is a measure of the volume occupied by the polymer per

unit weight (Fig. 4c), was  $(22.3 \pm 0.1)$  dL/g in MSgg medium indicating that EpsA-O expands well in aqueous environment. The corresponding value in water was higher ( $[\eta] = (42 \pm 2)$  dL/g) implying the polyelectrolyte nature of the polymer. This was further supported by conducting experiments with increasing NaCl concentrations (Fig. 4d). By increasing EpsA-O concentration, the viscoelastic liquid became a gel at concentrations  $\geq 1.0$  g/dL (Fig. 4e). The gel transition point did not change in the absence of salts (Supplementary Fig. 3a), suggesting that the major gelling mechanism does not involve ions. At 1% EpsA-O concentration the gel was weak: it yielded at 25% shear strain deformation and collapsed at 80% shear strain (Fig. 4e). The energy needed to irreversibly deform the structure has to exceed the cohesive energy, which was calculated to be  $3 \times 10^{-28}$  J/nm<sup>3</sup> for EpsA-O gel. The phase transition to the gel state is a second-order continuous transition depending on EpsA-O concentration in solution (Fig. 4f).

When purified EpsA-O was exogenously added to  $\Delta eps$  mutant cells, the rheological behavior changed significantly (Supplementary Fig. 3b). Both storage and loss modulus increased by an order of magnitude. More importantly, the qualitative behavior of the material changed as well. The loose cell material of  $\Delta eps$  mutant that behaved as a weak viscoelastic liquid material changed to viscoelastic solid material. This is indicated by a decrease in loss factor ( $\tan \delta = G''/G'$ ) which changed from 1.3 to 0.3 upon EpsA-O addition. This is a strong indication that *B. subtilis* cells in the presence of EpsA-O form a gel material.

### Composition and linkage analysis of the exopolysaccharide EpsA-O produced by *B. subtilis*

Acid hydrolysis of EpsA-O followed by derivatization to alditol acetates gave the neutral sugars Gal, GlcNAc, GalNAc and Glc in the molar ratios 7.7:1.1:1.0:0.6. Methanolysis followed by derivatization to trimethylsilyl ethers showed Gal, GlcNAc and GalNAc in the molar ratios 2.2:1.5:1.0. Glycosidic linkage analysis identified 6-Galp, 3,4-Galp, 3,4,6-Galp, 3-GlcpNAc, and 3,4-GalpNAc in the molar ratios 1.0:1.0:0.6:1.2:1.1. A very small late eluting peak with key primary fragments at 159 and 172  $m/z$ , and secondary fragments at 117 and 130  $m/z$  was attributed to 3-QuipNAc4NAc. The presence of three branched sugars and the absence of terminal non-reducing residues suggested heavy substitution by non-carbohydrate moieties. The hexoses all have the D absolute configuration.

### <sup>1</sup>H NMR spectroscopy of EpsA-O produced by *B. subtilis*

The <sup>1</sup>H NMR spectrum of the native EpsA-O showed extremely broad lines (Fig. 5a) which did not become sharper upon sonication. Therefore, integration of the methyl region of the sonicated sample gave an approximate estimation of 1 deoxy-sugar (at 1.1 ppm), ~1.8 pyruvyl ketal substituents (at 1.6 ppm) and ~4 N-acetyl groups near 2 ppm. Chemical analysis detected only two N-acetylated sugars (GlcNAc and GalNAc), thus NMR data suggested the presence of a 6-deoxy doubly N-acetylated residue.

### Smith degradation of EpsA-O produced the polysaccharide backbone and part of the side chain

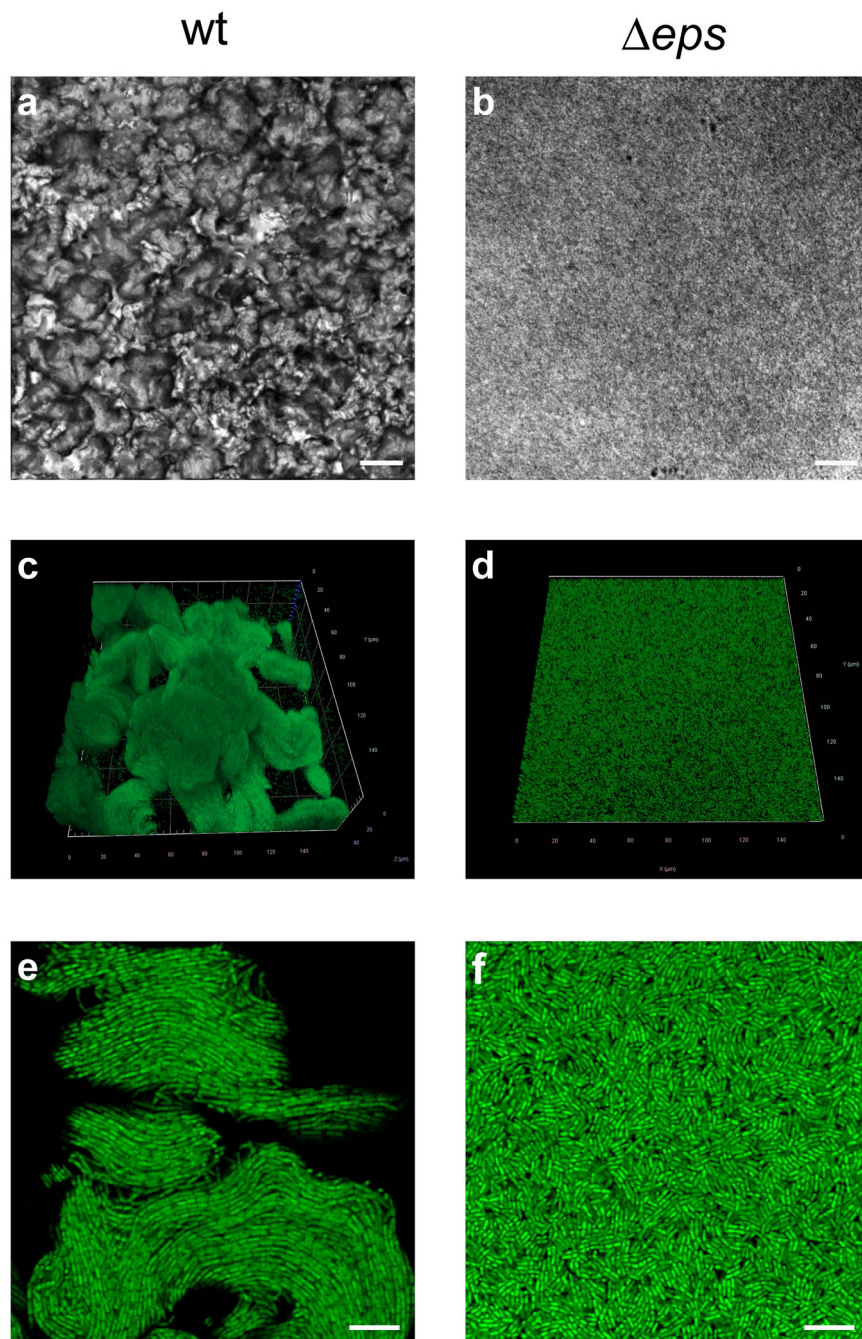
Since the 6-linked Gal is the only residue susceptible to periodate oxidation, the EpsA-O was subjected to complete Smith degradation<sup>32</sup> to yield three fractions, SD1, SD2 and SD3, isolated by size-exclusion chromatography (Supplementary Fig. 4a).

**Fig. 2 | Microscopy images of *B. subtilis* biofilms.**

The mature biofilms were grown on agar-filled 3D-printed chamber slides and observed in situ.

**a, b** Bright-field images taken by air objectives of WT and  $\Delta eps$  *B. subtilis* biofilm surface, respectively.

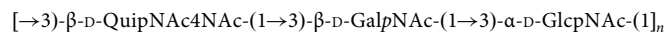
Scale bars represent 50  $\mu\text{m}$ . **c** 3D-reconstructed CLSM image of upper 40  $\mu\text{m}$  of WT biofilm. **d** 3D-reconstructed CLSM image of the upper 10  $\mu\text{m}$  of  $\Delta eps$  biofilm. **e, f** CLSM slice through the biofilm structure of WT and  $\Delta eps$  biofilm, respectively. All CLSM images were taken by silicon oil objective. Scale bars represent 10  $\mu\text{m}$ . To exclude pleiotropic effects as the main cause of the observed biofilm structural differences, further experiments were performed (Fig. 3).



The  $^1\text{H}$ -NMR spectrum of **SD1** (Fig. 5a) showed three anomeric signals of similar intensity at 5.12 (A), 4.64 (B), and 4.51 (C) ppm, indicative of one alpha and two beta anomeric residues. These data together with the absence of reducing end resonances (see inset HSQC plot, Fig. 5b) showed that **SD1** is a polymer composed of a trisaccharide repeating unit (RU). Diagnostic signals in the methyl region were assigned to H6 of a deoxy residue (1.17 ppm, 3 protons), and four N-acetyl substituents (1.95, 1.99, 2.00, and 2.06 ppm, 12 protons). A small signal at 1.59 ppm (0.36 protons) indicated a small amount of pyruvate in comparison with the native EpsA-O (6 protons) and confirmed the success of the Smith degradation.

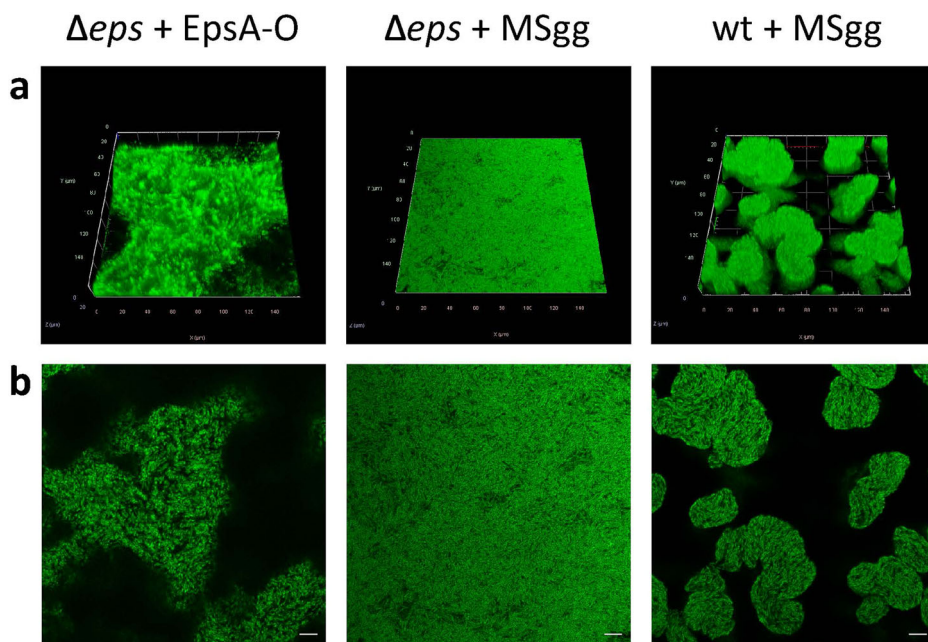
The structure of **SD1** RU was determined by use of an array of  $^1\text{H}$ - $^1\text{H}$  homonuclear experiments (COSY, TOCSY (Supplementary Fig. 5), NOESY, and  $^1\text{H}$ - $^{13}\text{C}$  heteronuclear correlation experiment (HSQC). In the HSQC spectrum (Fig. 5b) the proton/carbon pairs are

labeled and the NMR data are collected in Table 1. Finally, the sequence of sugar residues followed from the NOESY inter-residue correlations, as shown in Supplementary Fig. 6. NMR data agree with those reported for *Bacillus licheniformis* ATCC 9945 exopolysaccharide<sup>33</sup>. The magnitude of the glycosylation shift of  $\beta$ -D-GalpNAc C3 (+7.6 ppm) is identical to the reported<sup>33</sup>, thus establishing the absolute configuration of  $\beta$ -QuipNAc4NAc to be D. Therefore, NMR analysis established the structure of the backbone RU of EpsA-O (structure 1).

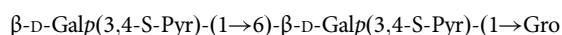


structure 1: backbone repeating unit of EpsA-O

**Fig. 3 | Microscopy images of native *B. subtilis* biofilms following the addition of exogenously purified EpsA-O or MSgg media.** The addition of MSgg media or 0.2% exogenous EpsA-O occurred in three stages: at inoculation, and after 4 and 7 hours of incubation. Mature biofilms were cultivated on agar-filled 3D printed chamber slides and observed in situ. **a** A 3D-reconstructed CLSM image of mature biofilms. **b** CLSM slices through the biofilm structure of  $\Delta eps$  and wt biofilms. All CLSM images were captured using a silicon oil objective, with scale bars indicating 10  $\mu$ m.



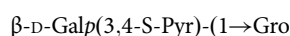
The <sup>1</sup>H-NMR spectrum of **SD2** (Fig. 5a) showed two beta anomeric signals at 4.54 (**D**) and 4.47 (**E**) ppm ( $J_{H1-H2} = 8.2$  Hz), and a methyl signal at 1.60 ppm (6 protons) assigned to pyruvate. As for **SD1**, the use of COSY, TOCSY (Supplementary Fig. 7), and HSQC (Supplementary Fig. 8) led to the assignments of all chemical shifts, reported in Table 2, in agreement with literature values<sup>33,34</sup>, establishing that **SD2** has the following structure (structure 2):



structure 2: oligosaccharide **SD2**

Glycerol (Gro) linked through C1 derives from a 6-linked hexose upon Smith degradation<sup>35</sup>, thus indicating that the pyruvylated disaccharide is linked to C6 of Gal in the native exopolysaccharide, in agreement with the linkage analysis. A 1D NOESY experiment established the S configuration of the pyruvyl substituents<sup>36</sup>.

Moreover, upon elucidation of the <sup>1</sup>H-NMR spectrum (Supplementary Fig. 4b) the structure of **SD3** was established as follows (structure 3):



structure 3: oligosaccharide **SD3**

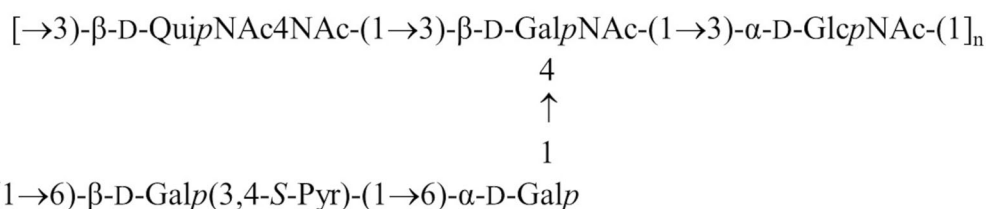
### NMR spectroscopy of fraction **SL1** obtained upon solvolysis of EpsA-O

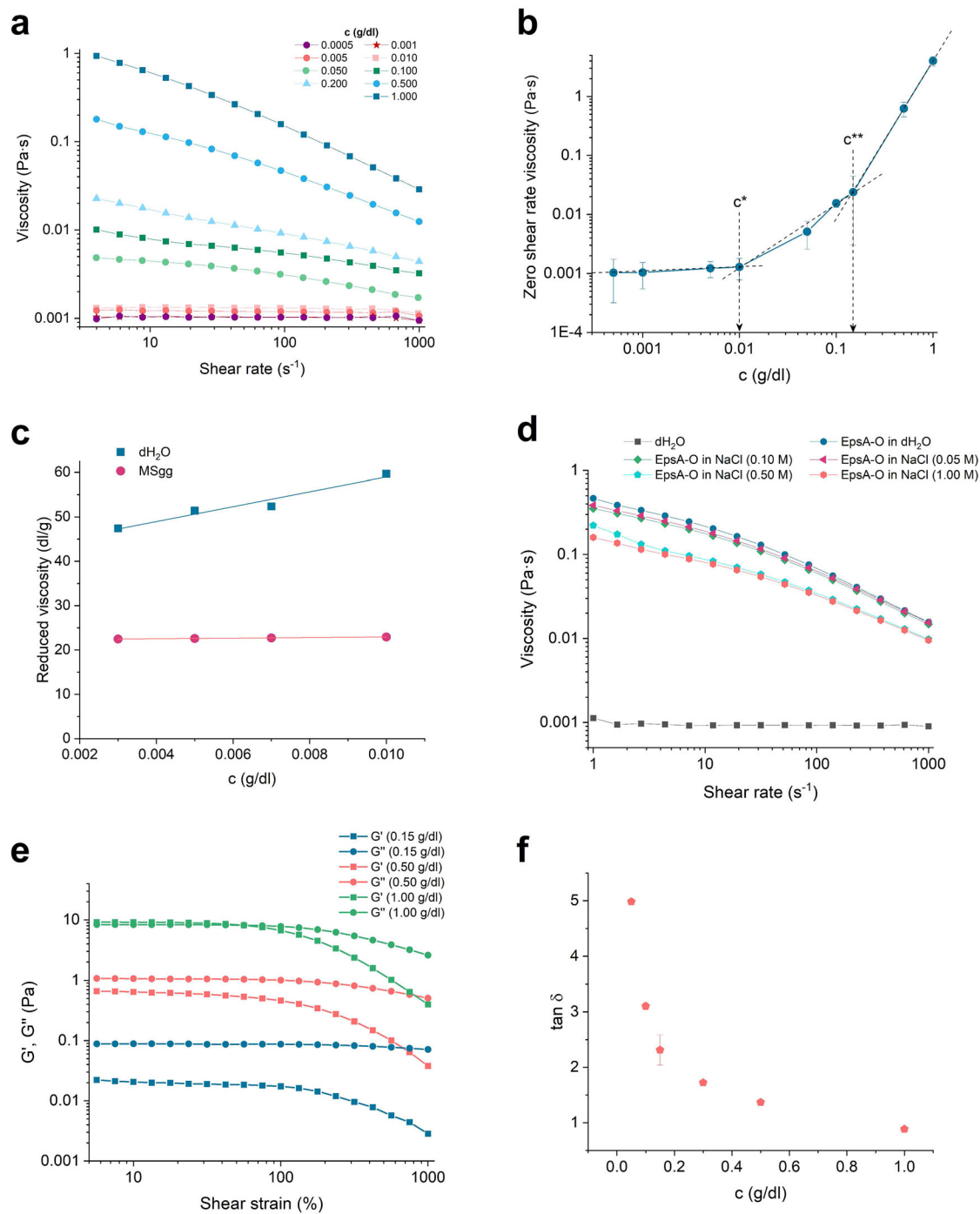
In order to obtain NMR spectra of the EpsA-O containing information on the intact repeating unit, including the 6-Gal residue, the MM of the polysaccharide was decreased by treatment in concentrated TFA

(solvolysis) to yield the fraction named **SL1** (Supplementary Fig. 9). The <sup>1</sup>H NMR spectrum (Fig. 5a) anomeric region showed three separated signals at 5.19, 5.14 and 4.92 ppm ( $\alpha$ -sugars), and two clusters of overlapping resonances at 4.75-4.45 ppm ( $\beta$ -sugars). The methyl region contained the expected signals for H6 of QuiNAc4NAc, N-acetyl groups and pyruvyl substituents with the presence of more than four NAc and two H6 signals indicated some sample heterogeneity arising from solvolysis. Comparison of **SL1** <sup>1</sup>H NMR spectrum with those of **SD1** and **SD2** identified the resonance at 5.14 ppm as H1 of  $\alpha$ -D-GlcNAc and the signal at 4.92 ppm as H1 of 6-linked Gal, present in the NMR of the sonicated EpsA-O (Fig. 5a) but not of the Smith degradation products. The signal at 5.19 ppm was assigned to H1 of a glucose belonging to another polymer, as it is very sharp compared to all EpsA-O resonances, it was not consistently present in all EpsA-O batches and size-exclusion chromatography yielded fractions with variable content. Moreover, the same  $\alpha$ -Glc containing polymer was the only carbohydrate produced by a  $\Delta eps$  mutant strain (Supplementary Fig. 10), indicating the presence of a yet unknown exopolysaccharide in *B. subtilis*.

The structure of **SL1** RU was determined by use of COSY, 1D-TOCSY (Fig. 6a), NOESY, and HSQC (Fig. 6b) experiments. Starting from the proton assignments (Fig. 5a) and by comparison of the **SL1** HSQC with those of **SD1** and **SD2**, the chemical shifts for each spin system were established (Table 3) and the values agree with literature data<sup>33,37</sup>. Finally, 1D NOESY of the 6-Gal H1 showed its binding to GalNAc C4 (Supplementary Fig. 11), thus establishing the following structure of the EpsA-O RU (structure 4, Supplementary Fig. 12):

structure 4: main repeating unit of EpsA-O



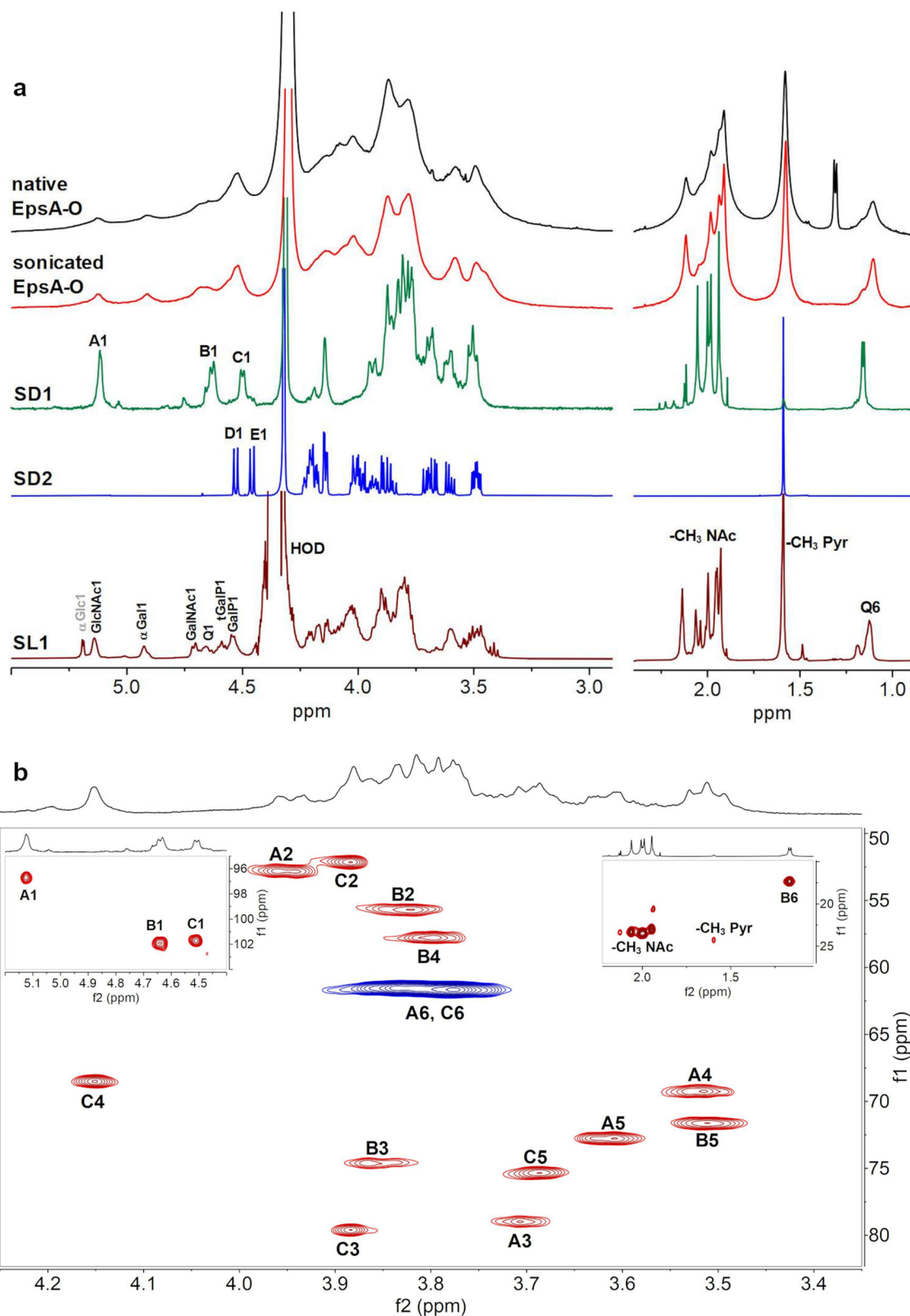


**Fig. 4 | Rheology of EpsA-O.** **a** The flow curves of EpsA-O dissolved in MSgg. The data points are color-coded and shaped according to the concentration of dissolved EpsA-O as follows: blue squares: 1 g/dL; blue circles: 0.5 g/dL; blue triangles: 0.2 g/dL; green squares: 0.1 g/dL; green circles: 0.05 g/dL; orange squares: 0.01 g/dL; orange circles: 0.005 g/dL; red stars: 0.001 g/dL; purple squares: 0.0005 g/dL. **b** Zero-shear viscosity of EpsA-O dissolved in MSgg at different concentrations indicating two critical concentrations,  $c^*$  and  $c^{**}$ . **c** Reduced viscosity ( $\eta_{sp}/c$ ) as a function of EpsA-O dissolved in dH<sub>2</sub>O or MSgg at different concentrations. Blue squares and trendline represent EpsA-O dissolved in dH<sub>2</sub>O, while red circles and trendline represent EpsA-O dissolved in MSgg. **d** Flow curves of EpsA-O aqueous solutions in

the presence of increasing NaCl concentrations. Black squares: dH<sub>2</sub>O; blue circles: EpsA-O in dH<sub>2</sub>O (0.5 g/dL); red triangles: EpsA-O in 0.05 M NaCl (0.5 g/dL); green diamonds: EpsA-O in 0.1 M NaCl (0.5 g/dL); blue pentagons: EpsA-O in 0.5 M NaCl (0.5 g/dL); orange hexagons: EpsA-O in 1 M NaCl (0.5 g/dL). **e** Amplitude sweep test of EpsA-O dissolved in MSgg at different concentrations. The squares represent the storage modulus ( $G'$ ), and the circles represent the loss modulus ( $G''$ ). The color coding for the symbols and lines indicates different concentrations: green for 1 g/dL, orange for 0.5 g/dL, and blue for 0.15 g/dL. **f** Loss factor as a function of EpsA-O concentration in MSgg.

Moreover, the 2:1 ratio of SD2/SD3 obtained from the size exclusion chromatogram together with the 3,4,6-Galp ratio of 0.6 to 1.0 6-Galp suggests side chain variability with up to one-third of the repeating units having a disaccharide with only one pyruvylated Gal instead of

trisaccharide side chain. This is in agreement with a pyruvylation degree of  $0.13 \pm 0.02$  estimated by chemical analysis. Since EpsA-O was purified by repeated treatment with 0.1 M NaOH, the possible presence of alkali labile groups such as acetates could not be established.



**Fig. 5 | 1D and 2D NMR spectroscopy data of EpsA-O, and SD1 and SD2 samples obtained by Smith degradation.** **a**  $^1\text{H}$  NMR spectra of native EpsA-O, sonicated EpsA-O, SD1, SD2, and EpsA-O after solvolysis recorded at 70 °C. Anomeric resonances are labelled with letters and sugar abbreviations, as in Tables 1–3. **b** Ring region of the SD1 HSQC spectrum recorded at 70 °C. Anomeric and methyl regions are reported in the insets. Cross peaks are labeled as in Table 1 (A1 = H1/C1 of residue A).

**Table 1 | <sup>1</sup>H and <sup>13</sup>C chemical shift assignments of the sample SD1 recorded in D<sub>2</sub>O at 500 MHz and 70 °C**

Residue	Nucleus	Chemical shifts (ppm) <sup>a</sup>					
		1	2	3	4	5	6
<b>A</b>	<sup>1</sup> H	5.12	3.95	3.70	3.51	3.61	3.83 - 3.76
→3)-α-D-GlcpNAc-(1→	<sup>13</sup> C	96.7	52.8	79.0	69.3	72.8	61.7
<b>B</b>	<sup>1</sup> H	4.64	3.83	3.85	3.80	3.50	1.17
→3)-β-QuipNAc4NAc-(1→	<sup>13</sup> C	101.9	55.7	74.6	57.8	71.6	17.3
<b>C</b>	<sup>1</sup> H	4.51	3.88	3.88	4.15	3.69	3.83 - 3.76
→3)-β-D-GalpNAc-(1→	<sup>13</sup> C	101.7	52.1	79.6	68.6	75.3	61.7

<sup>a</sup>Chemical shifts are given relative to internal acetone (2.225 ppm for <sup>1</sup>H and 31.07 ppm for <sup>13</sup>C).

**Table 2 | <sup>1</sup>H and <sup>13</sup>C chemical shift assignments of the sample SD2 recorded in D<sub>2</sub>O at 500 MHz and 70 °C**

Residue	Nucleus	Chemical shifts (ppm) <sup>a</sup>					
		1	2	3	4	5	6
<b>D</b>	<sup>1</sup> H	4.53	3.49	4.20	4.15	4.02	3.90
β-D-Galp-3,4 Pyr-(1→	<sup>13</sup> C	103.3	74.2	79.5	75.6	74.0	61.7
<b>E</b>	<sup>1</sup> H	4.46	3.49	4.20	4.15	4.22	4.19 - 4.02
→6)-β-D-Galp 3,4 Pyr(1→	<sup>13</sup> C	103.1	74.2	79.5	75.6	73.0	69.8
Pyr S	<sup>1</sup> H	1.60					
	<sup>13</sup> C	177.9	108.7	24.3			
1-Gro	<sup>1</sup> H	H1'a	H1'b	H2	H3'a	H3'b	
	<sup>13</sup> C	3.99	3.70	3.94	3.68	3.60	
	<sup>13</sup> C	72.0	72.0	71.6	63.4	63.4	

<sup>a</sup>Chemical shifts are given relative to internal acetone (2.225 ppm for <sup>1</sup>H and 31.07 ppm for <sup>13</sup>C).

### Prediction of the reactions catalysed by the EpsA-O gene cluster proteins

Knowledge of the EpsA-O RU structure aided for assignment of the functions of the previously uncharacterized genes *epsE*, *epsF*, *epsG*, *epsH*, *epsI*, *epsJ*, and *epsO* involved in the EpsA-O biosynthesis. This was achieved by a bioinformatic analysis searching for conserved domains (CD) in the proteins encoded by the *epsA-O* cluster. Identified CDs were queried in the CD database (CDD) to find annotated proteins with an enzymatic function consistent with the reactions required for RU biosynthesis. In the case of glycosyltransferases, the CAZY database (<http://www.cazy.org/>) was consulted for a more detailed understanding of the retaining or inverting mechanism of action. The results obtained and a detailed explanation of the bioinformatic work performed are reported in the Supplementary file (Supplementary Table 1, Supplementary Figs. 13-15).

During the biosynthesis of the EpsA-O RU, ketal-pyruvylation of both β-galactose residues occurs before polymerization. Although pyruvylation happens in many glycan structures from bacteria to yeasts, few studies investigated the enzymes governing these reactions. All agree that the reaction occurs in the cytoplasm with the lipid-PP-linked RU as acceptor, but the time at which pyruvylation occurs may vary, depending on the position of the pyruvylated residue in the RU<sup>38</sup>. The annotation of EpsI and EpsO as putative pyruvyltransferases is strongly supported by the presence of three extremely conserved amino acids, suggested as fundamental for *Saccharomyces pombe* pyruvyltransferase activity<sup>39</sup> (Supplementary Fig. 15). The sequence of EpsI and EpsO was compared with that of WcuL<sup>40</sup> which is a pyruvyltransferase catalyzing the addition of a pyruvate group on a terminal β-D-Gal in *Klebsiella pneumoniae* K2 and K30, thus forming 3,4-

Pyr-β-D-Gal. Since BLAST alignment showed that WcuL has higher homology with EpsO than with EpsI (Table S1), pyruvylation of the first β-D-Gal which leads to a terminal 3,4-Pyr-β-D-Gal, a sub-structure identical to that of polysaccharides K2 and K30, is likely catalyzed by EpsO. Then the second β-D-Gal is added to the 3,4-Pyr-β-D-Gal and pyruvylated by EpsI, whose lower similarity to WcuL is probably due to the different acceptor molecule recognized. The biological RU of EpsA-O and the assignments of the gene functions are reported in Fig. 7. The published data on *epsA*, *epsB*, *epsC*, *epsK*, *epsL*, *epsM* and *epsN* functions (Supplementary file) together with our in-silico genes functions predictions, describing six GTs and two pyruvyl transferases among other proteins, perfectly match the structure of EpsA-O.

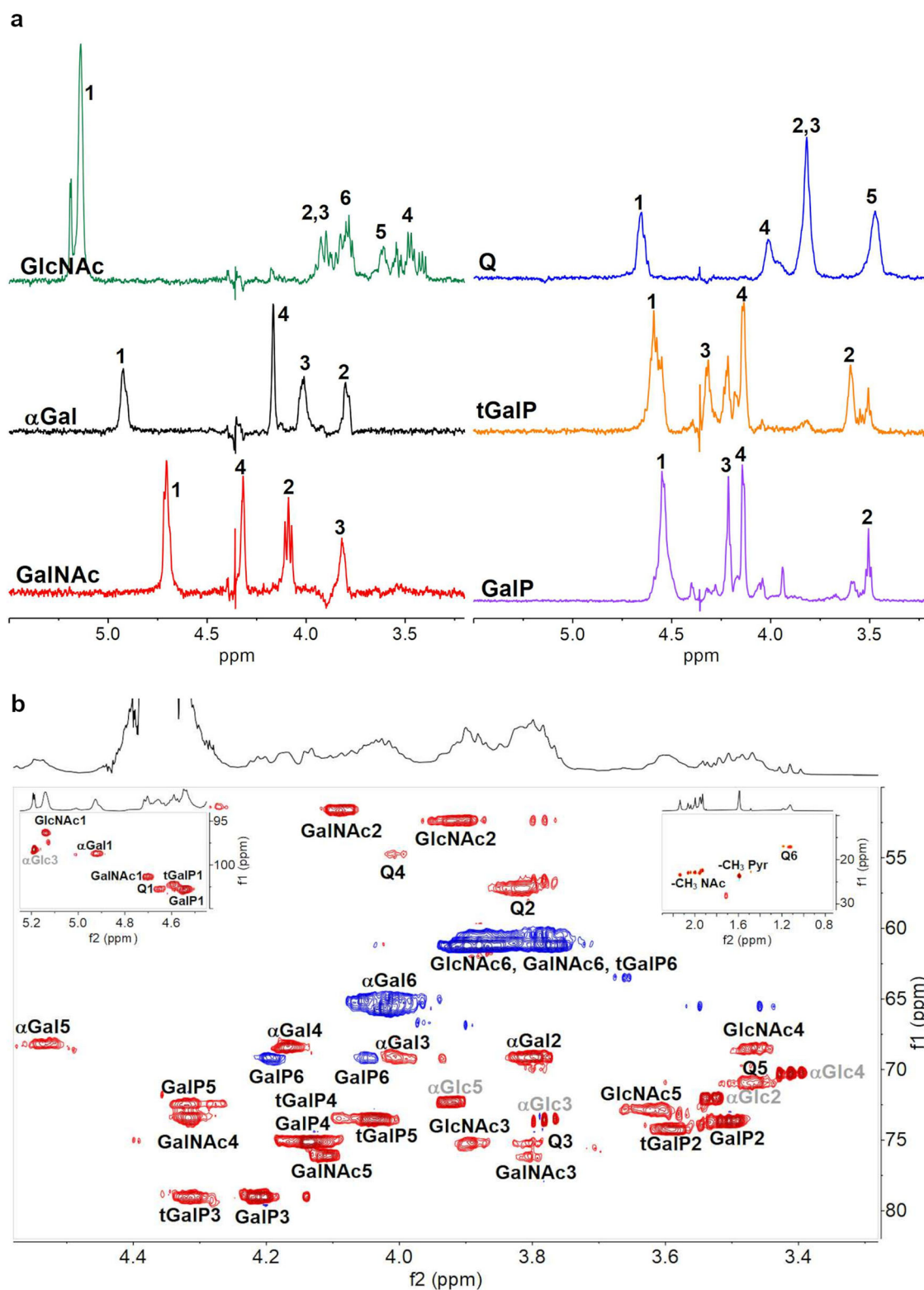
### Modeling EpsA-O spatial structure reveals that EpsA-O is a carefully tailored biofilm adhesive

To visualize EpsA-O by CLSM in the intercellular volume several fluorescent dyes were applied. Concanavalin A (ConA) and wheat germ agglutinin (WGA) did not show specificity against EpsA-O, consistent with their ability to stain teichoic acids in the cell walls of *B. subtilis*<sup>41-44</sup>. As staining by Calcofluor-white produced only weak unspecific fluorescence, we switched to wide-field microscopy techniques. Indian ink stained *B. subtilis* cells, observed by DIC, confirmed the ability of EpsA-O to fill the intercellular space (Fig. 8) enabling cell-cell aggregation.

To visualize EpsA-O in 3D solution space and its ability to act as an adhesive at the molecular level, the EpsA-O in the intercellular volume space was investigated using the String-of-beads model<sup>45-47</sup>. The advantage of this modeling approach is that it can relate the primary structure of EpsA-O to the macromolecular descriptors such as MM and intrinsic viscosity to yield a statistical spatial 3D polymer structure in a given space.

According to its primary structure (Structure 4), the EpsA-O degree of branching, (DB) is 0.5. Therefore, the compression factor for EpsA-O ( $g_{1\eta}$ ) is 0.5, which means that the hypothetical non-branched variant of EpsA-O of identical MM would have twice the intrinsic viscosity of the branched variant (see Material and methods, Modeling spatial structure of EpsA-O). By using the Yamakawa model<sup>48</sup>, the persistence length,  $lp$ , of the non-branched EpsA-O was estimated to be 11 nm. Upon application of the String-of-beads model, the non-branched EpsA-O variant structure with the desired  $lp$  of 11 nm was obtained by setting a random torsion angle and limiting the bond angle to 0.6 rad. To simulate the branched polysaccharide, the same parameters were used after adding side chains on every 3<sup>rd</sup> monosaccharide of the backbone, resulting in a rather large structure (Fig. 9a) that fits into a space of 200 × 200 × 200 nm<sup>3</sup>.

To check how such a structure fits into the *B. subtilis* biofilm space, the supramolecular structure of EpsA-O was simulated at EpsA-O concentrations measured in WT biofilm. The average experimental EpsA-O concentration was (0.61 ± 0.05) mg per 100 mg of wet biofilm which is significantly higher than  $c^{**}$ , suggesting that single polymer



**Fig. 6** | 1D and 2D NMR spectroscopy data of EpsA-O after solvolysis. **a** 1D TOCSY spectra extracted from 2D spectra of sample SL1 obtained after solvolysis of EpsA-O recorded at 70 °C. **b** Ring region of the HSQC spectrum recorded at 70 °C of

the sample SL1. Anomeric and methyl regions are reported in the insets. Peaks and cross peaks are labeled as in Table 3. Signals belonging to a glucose residue not part of the EpsA-O are labeled in grey.

chains may overlap in biofilm. Although this is below the gelling concentration ( $c \geq 1.0$  g/dL  $\approx 1$  mg per 100 mg) required for gelation of the biofilm, the local concentration of EpsA-O in the intercellular space can be significantly higher due to a large cell volume fraction in the biofilm (Fig. 2e). To get a better insight into intercellular volume, the native

biofilm was partially disrupted giving rise to the formation of ordered cell clusters (Fig. 9b). Some of the clusters aligned along the z-axis (Fig. 9c), enabling accurate determination of the intercellular distance at 190 nm. Using 750 nm as the average diameter of *B. subtilis* cells<sup>49</sup>, the volume fraction of the extracellular space was estimated to be

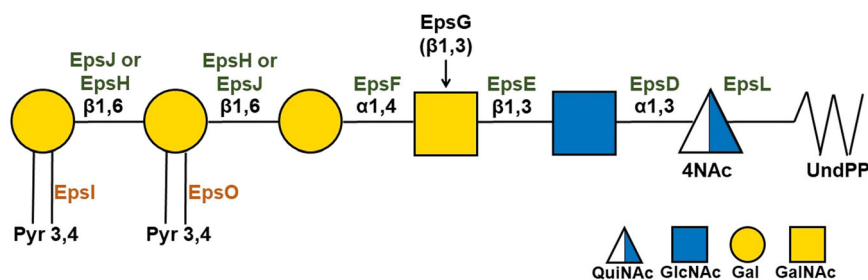
**Table 3** |  $^1\text{H}$  and  $^{13}\text{C}$  chemical shift assignments of the sample SL1 obtained upon solvolysis of EpsA-O recorded in  $\text{D}_2\text{O}$  at 600 MHz and  $70^\circ\text{C}$ 

Residue	Nucleus	Chemical shifts (ppm) <sup>a</sup>					
		1	2	3	4	5	6
—3)— $\alpha$ -D-GlcpNAc-(1→	$^1\text{H}$	5.14	3.92	3.90	3.47	3.61	3.77–390
(GlcNAc)	$^{13}\text{C}$	96.2	52.4	75.3	68.6	73.0	61.2
—6)— $\alpha$ -D-Galp-(1→	$^1\text{H}$	4.92	3.80	4.00	4.17	4.53	4.01
( $\alpha$ Gal)	$^{13}\text{C}$	98.8	69.2	69.2	68.5	68.3	65.2
—3,4)— $\beta$ -D-GalpNAc-(1→	$^1\text{H}$	4.70	4.09	3.82	4.32	4.10	3.77–390
(GalNAc)	$^{13}\text{C}$	101.4	51.7	76.3	73.5 <sup>b</sup> (72.6) <sup>b</sup>	76.1	61.2
—3)— $\beta$ -D-QuipNAc4NAc-(1→	$^1\text{H}$	4.65	3.82	3.81	4.00	3.47	1.13
(Q)	$^{13}\text{C}$	102.8	57.2	75.3	54.8	71.1	17.1
$\beta$ -D-Galp(3,4-S-Pyr)-(1→	$^1\text{H}$	4.59	3.60	4.32	4.15	4.03	3.77–390
(tGalP)	$^{13}\text{C}$	102.3	74.4	79.1	75.3	73.5	61.2
—6)— $\beta$ -D-Galp(3,4-S-Pyr)-(1→	$^1\text{H}$	4.54	3.51	4.21	4.15	4.32	4.19 – 4.05
(GalP)	$^{13}\text{C}$	102.8	73.8	79.1	75.3	72.6 <sup>b</sup> (73.5) <sup>b</sup>	69.2

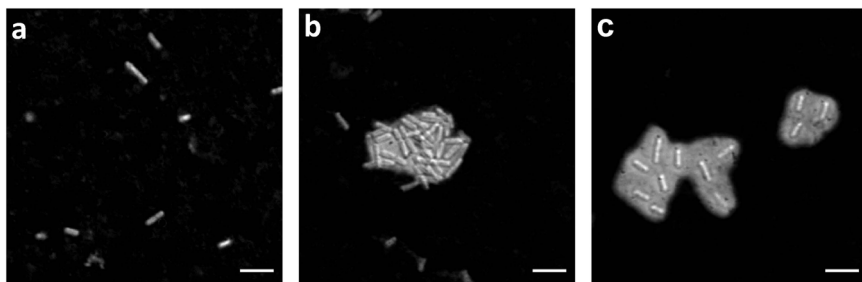
<sup>a</sup>Chemical shifts are given relative to internal acetone (2.225 ppm for  $^1\text{H}$  and 31.07 ppm for  $^{13}\text{C}$ ).

<sup>b</sup>Signals may be exchanged.

**Fig. 7** | *epsA-O* gene functions. Proposed glycosyltransferase, pyruvyltransferase, and polymerase activity of the proteins encoded by the *eps* gene cluster in the biosynthesis of the EpsA-O repeating unit.



**Fig. 8** | Negative staining with Indian ink of the *B. subtilis* wild-type and mutants. DIC microscopic images of *B. subtilis*  $\Delta$ *eps* (a), wt (b), and  $\Delta$ *sinR* strain (c), where polysaccharide capsules were clearly visible around the wild type, and  $\Delta$ *sinR* strain. Scale bars represent 5  $\mu\text{m}$ .

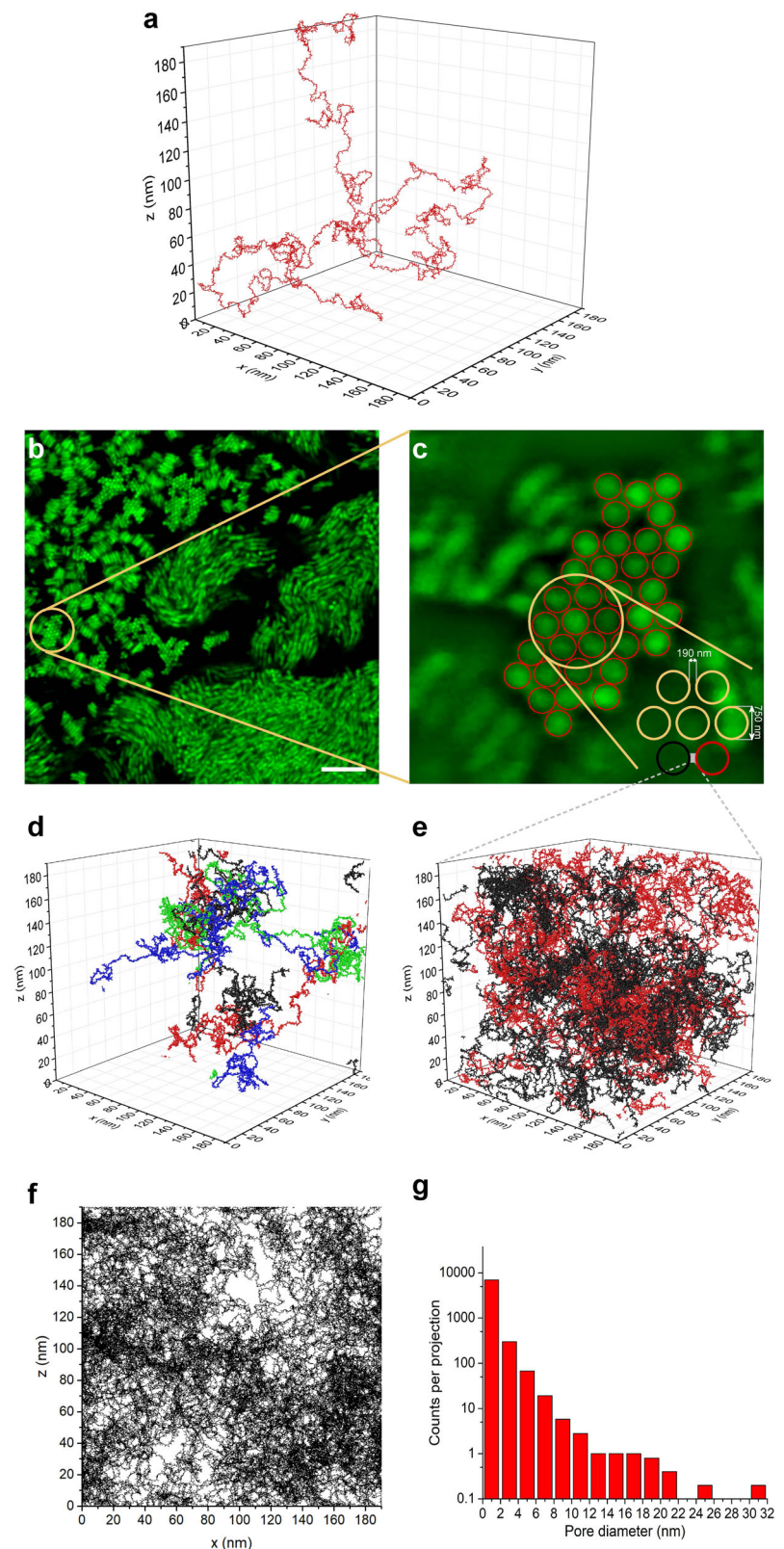


( $50 \pm 10$ ) %, thus increasing the local concentration of EpsA-O to 1.2 g/dL, above its gelling point. By randomly positioning the origins of four polymer chains into the modeling box of  $190 \times 190 \times 190 \text{ nm}^3$ , EpsA-O supramolecular structure at concentration slightly above  $c^{**}$  (at 0.24 g/dL) was simulated (Fig. 9d) and chain entanglements were observed. To get closer to the native state, twenty EpsA-O chains, corresponding to the concentration of 1.2 g/dL, were inserted into the box from opposite sides, simulating EpsA-O production by neighboring cells in the biofilm. The polymer chains entangled across the box (Fig. 9e), thus providing a molecular explanation for the observed gel formation in the biofilm. To estimate the porosity of the resultant entangled network of EpsA-O, 2D projections of x-z plane (Fig. 9f) of the simulated space were constructed and a distribution of different size pores was observed. Although the majority of the pores had diameters  $< 2 \text{ nm}$  (Fig. 9g), there were several of about 10 nm which allow diffusion of polymers within the extracellular space.

## Discussion

The present investigation resolves the structure and function of EpsA-O in *B. subtilis* biofilms. For the first time combined chemical and NMR spectroscopy analyses were used to determine the fine structure of the EpsA-O extracted from pellicles. The primary structure of the main EpsA-O RU consists of a novel branched hexasaccharide made of a trisaccharidic backbone, identical to that of the exopolysaccharide produced by *B. licheniformis*<sup>33</sup>, and a trisaccharide side chain decorated with two pyruvyl substituents (structure 4). The minor contribution of a disaccharide side chain lacking one pyruvylated Gal was also detected. This makes EpsA-O a novel and unique structure among microbial polysaccharides. Attribution of the glycosyltransferases to specific catalyzed reactions was achieved by in silico analysis with a good degree of homology. However, the uniqueness of the side chain structure,  $\beta$ -D-Galp(3,4-S-Pyr)-(1→6)- $\beta$ -D-Galp(3,4-S-Pyr)-(1→6)- $\alpha$ -D-Galp-(1→), impeded to find suitable GTs for sequence homology analysis in data bases, thus leaving the precise identification of the

**Fig. 9 | EpsA-O in the intercellular space of *B. subtilis* biofilm.** **a** The volume space of a single EpsA-O molecule. By combining the data obtained by universal calibration method, intrinsic viscosity and NMR spectroscopy, the EpsA-O molecule was modeled with a String-of-beads model. The representative structure is a symmetrical comb with 0.5 degree of branching and 78 nm radius of gyration. The backbone has a persistence length of 11 nm. The polymer is composed of 12500 impenetrable beads of diameter 0.52 nm corresponding to the size of glucose molecule. **b** The mature WT biofilms were partially disrupted to obtain a better view into cell fibers composing the 3D structures. Scale bar represents 10  $\mu\text{m}$ . **c** The volume fraction of intercellular space and the average distance between two neighboring cells were estimated to be  $(50 \pm 10) \%$  and 190 nm, respectively. **d** In the box of  $190 \times 190 \times 190 \text{ nm}^3$ , EpsA-O was modeled at concentration slightly above  $c^{**}$  (see Fig. 4b); each molecule is in different color. **e** The same box, but the opposing xz sides, representing the cell envelopes of two neighboring cells, were set to be impenetrable and were the starting points of growing EpsA-O molecules, 10 EpsA-O molecules for each bacterium (red for bacterium 1 and black for bacterium 2) are shown, corresponding to the actual EpsA-O concentration. **f** xz projection of the space in **e**, showing the empty space (pores) between bacterium 1 to bacterium 2. **g** Distribution of pore diameters as determined from five independent projections.



sequence of galactosyltransferases action undetermined. The backbone has four N-acetyl groups on three sugar residues, a feature which may drive crucial intra- or inter-chain hydrogen bonds formation and may introduce steric hindrances. On the other hand, ketal-pyruvylation on the side chain adds a ring on both  $\beta$ -galactopyranose residues resulting in bulky substitution and increase of hydrophilicity due to two negative charges.

The bulky and expanded EpsA-O spatial structure in solution is consistent with the measured high intrinsic viscosity, which is comparable only to xanthan among microbial polysaccharides<sup>50</sup>. The high inter-chain connectivity enables gel formation at relatively low polysaccharide concentrations. The primary structure of the EpsA-O dictates its higher-order structure in solution, which in turn drives the supramolecular network

formation that entraps bacteria in the gel-like biofilm. The simulations of supramolecular structure indicate that EpsA-O can span the space between neighboring bacteria in the biofilm acting as an adhesive. This study is unique as the structure of EpsA-O was studied over several orders of magnitude from atomic size to the size of the entire biofilm system. The design of EpsA-O for its adhesive function becomes evident when comparing it to levan, another extracellular polysaccharide produced by *B. subtilis*. Levan, a linear uncharged homopolysaccharide with two-orders of magnitude lower intrinsic viscosity, forms only weak gels with an elastic modulus several orders of magnitude lower than EpsA-O<sup>51</sup>. Although levan can be present in biofilms in much larger quantities<sup>24</sup> than EpsA-O it acts predominantly as a filler<sup>51,52</sup>. On the other hand, EpsA-O is a biological structural polymer that enables the formation of the complex 3D architecture of the biofilm. From rheological data, and assuming that bacteria are connected longitudinally through space by EpsA-O (2000 nm × 190 nm × 190 nm), the total cohesive energy was calculated to be  $2 \times 10^{-20}$  J. To overcome this cohesive energy and irreversibly deform inter-chain connections, the bacterium has to be moved for at least 50 nm, which requires a force of  $5 \times 10^{-13}$  N. This force exceeds a single bacterium weight ( $F_g = 10^{-14}$  N) by 50-fold, indicating that EpsA-O inter-chain forces above the sol-gel transition could easily withstand the gravity force allowing the formation of a complex 3D biofilm structure. The question that remains to be solved is how EpsA-O is connected to the cells. One possibility is through TasA, the extracellular protein that is anchored to the bacteria via TapA<sup>53</sup>. In fact, TasA mutants have a significantly decreased capacity to form floating biofilm even in the presence of EpsA-O<sup>24,25</sup>. Similarly, TasA/EpsA-O double mutant, while initially attempting to form a biofilm, is unable to reinforce it due to the missing structural polymers<sup>25</sup>. The consequence is a biofilm not resilient to shear stress that rapidly disintegrates. Also, the rheology results of biofilms grown on agar<sup>54</sup> are in line with these observations. The  $G'$  drops significantly in  $\Delta eps$  mutant compared to the wt biofilm. In addition, the results of purified EpsA-O gels or cells and EpsA-O mixtures indicate that the interaction of EpsA-O with other components can be only partially recovered in a mixture of EpsA-O and cells.

In conclusion, the structure of the branched RU of EpsA-O was established by composition and linkage analyses, and NMR spectroscopy. The main RU is composed of the backbone  $[\rightarrow 3)\text{-}\beta\text{-D-QuipNAc4NAc-(1}\rightarrow 3)\text{-}\beta\text{-D-GalpNAc-(1}\rightarrow 3)\text{-}\alpha\text{-D-GlcpNAc-(1)}_n$  and the trisaccharide side chain  $\beta\text{-D-Galp(3,4-S-Pyr)-(1}\rightarrow 6)\text{-}\beta\text{-D-Galp(3,4-S-Pyr)-(1}\rightarrow 6)\text{-}\alpha\text{-D-Galp-(1}\rightarrow$  linked to C4 of the GalNAc residue. Up to 1/3 of the RU have a side chain lacking one pyruvylated Gal. This sequence is unique in the microbial polysaccharide world and contains the  $N,N'$ -diacetyl derivative of the prokaryote-specific sugar bacillosamine (QuipNAc4NAc), 4 N-acetyl and  $\sim 1.8$  pyruvyl groups. It is difficult to compare the structure presented with the results obtained by previous studies because only sugar composition or glycosidic linkages were determined, while NMR spectroscopy was not used. The studies also differ for the material investigated: cell lysates to check for intracellular PNAG with ELISA assay<sup>26</sup> and Eps isolated from liquid culture<sup>29</sup>. The presented structure of EpsA-O partially agrees with the composition proposed by Chai et al.<sup>28</sup>, who isolated the Eps from pellicles using mild sonication and found Glc, GalNAc, and Gal.

EpsA-O has very high intrinsic viscosity comparable only to xanthan<sup>50</sup>. Simulation of the EpsA-O structure in solution using the String-of-beads model demonstrated that the polysaccharide can span the space between neighboring bacteria in the biofilm. The supramolecular structure of EpsA-O provides a strong network in the intracellular space that enables gel formation and production of a 3D biofilm architecture. Using different molecular, mechanical and biophysical tools we provide a unique vantage point that spans several orders of magnitude, from atomic to system level in biofilm structure formation that will stimulate further research.

## Methods

### Bacterial strains and growth conditions

*Bacillus subtilis* subs. *subtilis* NCIB 3610 WT (undomesticated prototroph)<sup>55</sup> and its derivative strains were used in this study. The *tasA sinR*

double mutant ( $\Delta tasA::spec \Delta sinR::kan$ ; here referred as  $\Delta sinR$ )<sup>23</sup> and *tasA eps* double mutant ( $\Delta tasA::spec \Delta epsA-O::tet$ ; here referred as  $\Delta eps$ ) were kindly provided by Roberto Kolter<sup>56,57</sup>, while strains TB501 (*comI amyE::Phyperspank-mKate* (Spec) here referred as WT) and TB525 (*comI eps::tet amyE::Phyperspank-mKate* (Spec) here referred as  $\Delta eps$ ) were kindly provided by Ákos T. Kovács<sup>58</sup>. Overnight cultures were grown in LB liquid medium (tryptone 10 g/L; yeast extract 5 g/L; NaCl 5 g/L) containing spectinomycin (100 µg/mL), tetracycline (10 µg/mL) or kanamycin (5 µg/mL) at 37 °C with shaking (200 rpm) for 16 h, when the cultures reached  $OD_{650} \sim 2.0$  AU. For the selection of optimal growth medium, three different liquid growth media LB, TY<sup>29</sup> and MSgg<sup>15</sup> with or without supplements (galactose 0.5% (w/v); glucose 1% (w/v); MSgg with increased glycerol concentration to 12.5 g/L) were inoculated and incubated for 24 h at 37 °C at 200 rpm. The EpsA-O was isolated without the homogenization step (see Isolation of EpsA-O polysaccharide).

For EpsA-O isolation, the biofilms were grown on solid MSgg agar plates with an increased concentration of glycerol (optimal growth medium): 100 mM MOPS (3-(N-morpholino) propane sulfonic acid); 5 mM  $K_3PO_4$ ; 50 mg/L tryptophan; 50 mg/L phenylalanine; 2 mM  $MgCl_2 \cdot 6H_2O$ ; 0.5% (w/v) sodium glutamate; 1.25% (w/v) glycerol; 700 µM  $CaCl_2 \cdot 2H_2O$ ; 50 µM  $FeCl_3 \cdot 6H_2O$ ; 50 µM  $MnCl_2$ ; 1 µM  $ZnCl_2$ ; 1.5% agar; 2 µM thiamine hydrochloride. The pH was adjusted to 7.0. All components of the medium, except thiamine hydrochloride, were autoclaved together at 110 °C. On each MSgg agar plate with a diameter of 90 mm, 250 µL of the overnight culture ( $\Delta sinR$  or  $\Delta eps$  strain) was inoculated in order to obtain continuous biofilm formation. The agar plates were incubated facing up at 37 °C for 24 h in a cardboard box covered with aluminum foil, following the isolation procedure (see Isolation of EpsA-O polysaccharide).

For microscopy, the overnight cultures were grown in LB liquid medium and incubated at 28 °C for 24 h. Biofilms were grown on solid MSgg agar in custom 3D printed 8-well chambered slides (Supplementary Fig. 16). Each well contained 300 µL of MSgg agar with a standard concentration of glycerol (0.5% (w/v)). The overnight cultures (WT,  $\Delta eps$ ) were diluted 1:64 by MSgg liquid medium and 4 µL inoculum was transferred to each agar well. To obtain continuous biofilm each inoculum was spread by a glass bead and incubated in climatic chamber ICH260L (Memmert, Germany) at 37 °C, 80% RH humidity for 15 h. After this time biofilm mass did not increase, indicating the biofilm was in the mature state.

For biofilms with exogenous addition of EpsA-O, the same growth conditions were used, except MSgg media or 0.2% exogenous EpsA-O was added at three stages: at inoculation, and after 4 and 7 hours of incubation. The 0.2% concentration ensured that the EpsA-O polysaccharide remained below the critical level.

### Microscopy of *Bacillus* biofilms

After 15 h of incubation the 8-well slides with grown biofilms were observed under an inverted microscope Axio Observer Z1, LSM 800 (Zeiss, Germany) operated by ZEN 2.6 software. To obtain bright-field images of the biofilm surface, the 8-well slides were observed without cover slip by an air objective LD Plan-Neofluar 20×/0.4 Korr M27. The images were acquired by AxioCam MR Rev3 (Zeiss, Germany), with the resolution of 924×924 pixels. For WT biofilms that showed pronounced surface undulation the images were acquired using extended depth of focus function of ZEN software. To obtain an insight of the biofilm structure, confocal laser scanning module (LSM 800) was used. For this purpose, the 2–6 µL of MSgg (MSgg, where glycerol is replaced by pyruvate) was added on top of the biofilms grown in 8-well slides that were carefully covered by glass cover slip, without touching the biofilm surface. The pinhole size of LSM 800 was set to 1 AU, and the sampling rate in Z-axis was set to by software software-recommended Nyquist rate. The image resolution was set to by ZEN recommended 1610 × 1610 pixels. The images were acquired by LD LCI Plan-Apochromat 40×/1.2 Imm Korr DIC M27 silicon oil objective. The acquired confocal image stacks were deconvolved by ZEN software. The 3D reconstructions were obtained by maximum intensity method in the ZEN software. The same LSM 800 settings and 8-well slide preparations were

made for observation of biofilms with exogenous addition of EpsA-O. At least 4 independent biofilms were observed.

### Visualization of EpsA-O by microscopy

For EpsA-O visualization, *B. subtilis* NCBI 3610 strains  $\Delta eps$ ,  $\Delta sinR$ , and wt were grown in MSgg liquid media at 37 °C with shaking (200 rpm) overnight. 2  $\mu$ L of overnight culture was mixed with an equivalent volume of Indian ink (Royal Talens, Netherlands) on a microscope slide. A cover glass was placed on the slide glass and any excess fluid was pushed out using thumb pressure<sup>59,60</sup>. The negatively stained samples were observed by Plan-Apochromat 100 $\times$ /1.40 Oil DIC M27 oil objective. The presence of EpsA-O was indicated by the exclusion of the Indian ink granules around the bacterial cell or bacterial aggregate. Five independent experiments were performed.

### Isolation of EpsA-O polysaccharide

After 24 hours of incubation, the biofilm was scraped from the MSgg agar plates by microscopy slide and resuspended in a 10-fold volume of saline solution. An Ultra-Turrax T8 homogenizer (set to level 4) was used to achieve an optimally re-suspended sample. Samples were then aliquoted into 1.5 mL cooled centrifuge tubes (7 mL of the sample) and sonicated for 5 seconds per mL with an amplitude of 12  $\mu$ m using an ultrasonic disintegrator with a 3 mm exponential tip (MSE Scientific Instruments). The distribution of samples allowed us to treat all samples under the same conditions and with maximum efficiency. After sonication, samples were stored in ice and NaOH was added to 0.1 M final concentration. The samples were well stirred and incubated for 5 min at room temperature, then stirred again and incubated on ice for an additional 5 min, before adjusting the pH to 7.00  $\pm$  0.02 using cooled HCl. This was followed by centrifugation for 40 min at 12000 rcf at 4 °C. To the supernatant cooled isopropanol was added (1:3 V/V) and incubated at 4 °C overnight. The precipitate was recovered by centrifugation 12,000 rcf at 4 °C for 20 min, and dried for 36 h at 55 °C to yield the raw isolate which was further purified. To obtain purified EpsA-O, the precipitates were further purified without prior drying.

### Purification of EpsA-O polysaccharide

The raw isolate of EpsA-O polysaccharide from  $\Delta sinR$  strain was purified with dialysis membrane and by changing the pH of the solution. Based on the estimated Mw of impurities, determined by high-pressure size exclusion chromatography (HPSEC), appropriate dialysis tubing membranes were used. The purification of the raw isolate was performed in several steps. First, the raw isolate was resuspended in deionized water and isopropanol in 1:1 ratio. In this way, contamination of the samples during the purification procedures as well as contamination by microorganisms that could degrade the polysaccharide were prevented. The sample solutions were then transferred to a dialysis membrane with the cut-off of 300 kDa (Biotech, CE) and dialyzed against dH<sub>2</sub>O for 48 h. Then, NaOH was added to the dialysate, pH was adjusted to 12, and transferred into a fresh dialysis membrane sack with a pore size of 300 kDa. Dialysis was performed in 10 times the volume of NaOH with pH 12 for two hours. The reason for the short exposure to high pH is the instability of the dialysis membrane. The dialysis membrane was then transferred to 300 times the volume of dH<sub>2</sub>O and dialyzed for 4 days with constant mixing. Distilled water was replaced with fresh dH<sub>2</sub>O every 20 hours. After the final step of dialysis, the entire procedure was repeated once more. That is, the dialysate was again exposed to NaOH solution with pH 12 and dialyzed against NaOH (pH 12) for two hours. Afterwards, the dialysis membrane containing the sample was dialyzed against distilled water for 3 more days.

Once the desired purity of the polysaccharide was achieved, the purified isolate was frozen at -80 °C. After 2 h at -80 °C, the samples were transferred to a pre-cooled ScanVac CoolSafe freeze dryer (LaboGene, Denmark) and lyophilized for 36 h at 0.1 mbar and -95 °C.

The purity of the isolate was checked by UV-VIS spectroscopy, HPSEC and by determining the possible protein contaminants by Bradford reagent

as described by Dogsa et al.<sup>24</sup>. In addition, DNA gel and SDS-PAGE electrophoresis of purified EpsA-O samples were performed. For DNA gel electrophoresis, 0.8% (w/v) agarose gel in 1 $\times$  TAE buffer was prepared. Samples of EpsA-O polysaccharide (0.1% and 0.2%, diluted in dH<sub>2</sub>O) and isolated genomic DNA from *B. subtilis* PS-216 were mixed with loading dye (6 $\times$  TriTrack - Thermo Scientific) at a 3:1 (v/v) ratio and loaded onto the gel. A standard DNA ladder (GeneRuler DNA Ladder Mix - Thermo Scientific) was added to the first well. Electrophoresis was conducted for 60 minutes at 75 V. The gel was then stained with GelRed dye for 25 minutes and rinsed with dH<sub>2</sub>O. Gels images were captured using a G-box (SynGene, UK). For SDS-PAGE electrophoresis, precast gels (Mini-PROTEAN TGX Gels - Biorad) were used. BSA at different concentrations (0.05%, 0.10%, 0.15%, 0.20% w/v) was used as standard. Samples of EpsA-O polysaccharide were prepared at concentrations of 0.1% and 0.2%, diluted in dH<sub>2</sub>O. To 25  $\mu$ L of each sample, 23.75  $\mu$ L of Laemmli buffer and 1.25  $\mu$ L of 2-mercaptoethanol were added. The mixtures were then vortexed and incubated in a thermoblock for 8 minutes at 107 °C. After incubation, 10  $\mu$ L of the cooled samples were carefully pipetted into the gel wells. Two wells were loaded with the standard (PageRuler Prestained Protein Ladder, Thermo Scientific) to determine the molecular weight (Mw) of the proteins. The gel was then placed in a buffer tank, filled with 1 $\times$  SDS-PAGE running buffer, and run for 30 minutes at 200 V. After electrophoresis, the gels were stained in 40 mL of staining solution (12 mL distilled water, 20 mL methanol, 4 mL acetic acid, and 4 mL of 0.1% Coomassie Brilliant Blue dye) for 30 minutes. Destaining was carried out for 30 minutes in 40 mL of destaining solution (20 mL distilled water, 16 mL methanol, and 4 mL acetic acid) and repeated three times. Finally, the gels were imaged using a G-box (SynGene, UK).

### Reducing sugars quantification, pyruvylation degree estimation and EpsA-O concentration determination

The quantification of reducing sugars was attempted by DNS assay<sup>61</sup>. For calibration curve, glucose solutions in concentrations from 0 to 5 mM were used. DNS reagent and standards with known concentrations were mixed in 1:1 volume ratio. The samples (raw isolates from  $\Delta eps$  and  $\Delta sinR$  strains, purified EpsA-O polysaccharide isolate from  $\Delta sinR$  strain) were dissolved in dH<sub>2</sub>O at a concentration of 0.5% (w/v) and then prepared in the same way as standard solutions for measuring the amount of reducing sugars. The prepared samples were incubated in a thermoblock set at 112 °C for 20 min. The temperature in the test tubes was 100 °C  $\pm$  2 °C. 300  $\mu$ L of each sample was transferred into 96-well microtiter plate. This was followed by absorbance measurement with Multiscan reader (Thermo Scientific, USA) at a wavelength of 575 nm.

The degree of pyruvylation was determined as described by Pinto et al.<sup>62</sup>. Briefly, for the calibration curve various dilutions of sodium pyruvate (0%, 0.005%, 0.01%, 0.02%, 0.04%, 0.06%, 0.08%) were prepared. An aliquot (1 mL) of each sample was transferred to a 15 mL centrifuge tube, 1 mL of 2,4-dinitrophenylhydrazine (0.5% in 2 M HCl), and 1 mL of dH<sub>2</sub>O added and the mixture incubated for 10 minutes in a 37 °C water bath. After incubation, 5 mL of 0.6 M ammonium hydroxide was added. The hydrolysis of the samples was performed using 0.5% EpsA-O solution in 2 M HCl for 5 hours at 100 °C. Then, 2 mL of the hydrolyzed material was transferred to a 15 mL centrifuge tube and 9 mL of distilled water was added. After that, 1 mL of different samples was mixed with 1 mL of 2,4-dinitrophenylhydrazine (0.5% in 2 M HCl) and 1 mL of dH<sub>2</sub>O, and incubated the mixture for 10 minutes at 37 °C, followed by the addition of 5 mL of 0.6 M ammonium hydroxide. The absorbance was measured at 420 nm.

The EpsA-O content of dry raw isolate was determined using a modified phenol-sulfuric acid method<sup>63</sup>, with samples incubated for 20 min at 100 °C. For the calibration curve, the purified EpsA-O isolate was dissolved at 0.5 mg/mL, 1 mg/mL, and at 3 mg/mL in dH<sub>2</sub>O. For these analyses at least three replicates were made.

### High-performance size exclusion chromatography

The pullulan standards (PSS Polymer Standards Service, GMBH, Germany), raw isolates, and purified isolate were dissolved in 0.05 M NaOH to

obtain a final concentration of 0.2% (w/v). All samples were heated in a thermoblock for 30 min at 55 °C in order to reduce the concentration of dissolved gases and improve the solubility of the polymers. A PSS SUPREMA GPC/SEC column was used in combination with a PSS SUPREMA pre-column (PSS Suprema Analytical, Germany). On-line detection was carried out by UV absorbance using a K-2501 detector (Knauer, Germany) at 260 nm and by Smartline RI 2300/2400 detector (Knauer, Germany). The mobile phase consisted of 0.05 M NaOH at pH 12.5. All measurements were obtained at 35 °C using a mobile phase flow of 1.00 mL/min and a sample loop of 112 µL.

### Viscoelasticity of EpsA-O polysaccharide and EpsA-O polysaccharide - cells mixtures

To determine the viscoelasticity, two dynamic moduli depending on the strain were measured with a modular oscillating rheometer Anton Paar Physica MCR 302 with a cone plate-plate (CP50-1,  $d = 50$  mm) measuring system. The measurements temperature was kept at 20.0 °C. The elastic modulus ( $G'$ ) measures the elastic response of the material, which measures stored energy, while the viscous modulus ( $G''$ ) measures the viscous response of the material, where energy is dissipated as heat. The viscoelasticity of the polysaccharide solutions was measured at a constant frequency of 10 rad/s and by increasing the strain from 5 to 1000%. During the measurement, 17-19 logarithmically spaced measurement points were acquired.

To prepare the mixtures of EpsA-O polysaccharide and cells, biofilms of the  $\Delta eps$  strain on MSgg agar plates were grown for 24 hours. The biofilm (100 mg) was scraped into a microcentrifuge tube and 300 µL of 2 mg/mL EpsA-O polysaccharide solution were added. For the cell sample, we mixed 100 mg of biofilm (cells) with 300 µL of dH<sub>2</sub>O and for the EpsA-O sample, we mixed 100 mg of dH<sub>2</sub>O with 300 µL of 2 mg/mL EpsA-O. The EpsA-O was added at this concentration to ensure it remained below the critical level. The samples were placed in a desiccator to concentrate the polysaccharide until only 100 µL remained in the tube. Finally, the rheological properties were examined as described above. For all viscoelastic tests at least three independent measurements were performed.

### Molecular mass determination of EpsA-O

The MM of EpsA-O was determined by universal calibration method<sup>64</sup>. To construct the calibration curve, the intrinsic viscosity,  $[\eta]$ , and elution volume,  $V_e$ , on HPSEC of pullulan standards with known MM dissolved in 0.05 M NaOH were determined. The MM of EpsA-O was determined by comparing the calibration curve to the intrinsic viscosity and elution volume of purified EpsA-O in 0.05 M NaOH.

### Modeling the spatial structure of EpsA-O

To obtain the spatial structure of EpsA-O the String-of-beads model<sup>45-47</sup> was upgraded in order to simulate branched, comb-like polymers.

Briefly, in this model, the polymer molecular conformation is represented as a string of  $N$  monomer units, which are approximated by homogeneous impenetrable hard spheres called beads, of diameter,  $d$ . The branching is controlled by setting the length of the main chain and side chain segments composed of beads. A specific structural situation is defined with an individual set of four shape parameters  $\Theta_p$ ,  $\Phi_p$ ,  $\Theta_{plim}$ , and  $p^*$ , where  $\Theta_p$  and  $\Phi_p$  represent the bond and torsion angles between the adjacent beads, respectively;  $p^*$  is the probability for the random variation of the two angles, and  $\Theta_{plim}$  is the upper limit of the bond angle in the case of its random variation. If  $p^* = 0$  all the beads in the string have constant values of  $\Theta_p$  and  $\Phi_p$ , leading to a helical string, otherwise the higher the value of the parameter  $p^*$ , the more beads are assigned with random values of these angles. The value of the parameter  $p^*$  therefore indicates the degree of randomness (dynamics) in the molecular structure, while the value of the parameter  $\Theta_{plim}$  reflects the stiffness of the polymer chain. Due to the random steps in the growth of the modeled molecular configuration, an individual set of four shape parameters represents an infinite set of molecular configurations that are structurally alike and mimic the case of real

polymers that are not static and rigid structures. Therefore, individual structures shown in the figures can be regarded as snapshots of polymer spatial structures.

By combining the data on MM of EpsA-O, as obtained by the universal calibration method, and its primary structure, it was possible to calculate  $N$  and the size of the main and side chain segments. The size of the monomer unit was assumed to be the size of anhydrous glucose (i.e.  $d = 0.52$  nm) and having MM = 200 Da. In the absence of detailed spatial information, the EpsA-O was modeled as a freely rotating chain ( $p^* = 1$ ), where the chain stiffness was controlled by setting the  $\Theta_{plim}$  to a constant value. The choice of the value of  $\Theta_{plim}$  is, however, not straightforward. The  $\Theta_{plim}$  controls the chain stiffness and therefore it is strongly related to the persistence length,  $lp$  that our model can calculate for each simulated polymer. Therefore, if the  $lp$  of the polymer is known, one can then obtain the value of  $\Theta_{plim}$  by stepwise increasing  $\Theta_{plim}$  until the desired  $lp$  of the polymer is obtained.

According to the model by Yamakawa et al.<sup>48</sup>, the intrinsic viscosity can be predicted by knowing the persistence length,  $lp$ , of the linear polymer, contour length, chain thickness, and MM. The application of this model to infer  $lp$  of the EpsA-O, is however not straightforward, as the model is valid for linear polymers and microbial polysaccharides are often branched. The ratio of branched to linear dilute solution properties of polymers is often expressed in contraction factors,  $g_{Rg^2}$ , for the ratio of radii of gyration and  $g_{[\eta]}$ , for the ratio of intrinsic viscosities of branched and respective linear versions of the same polymer. For the random coil polymer that is branched as a symmetrical comb, the analytical solution for  $g_{Rg^2}$  was established by Casassa & Berry<sup>65</sup>. Defining the number of main chain segments to be proportional to the difference between total chain mass and side chain mass, and assuming there are many polymer branches, the expression for  $g_{Rg^2}$ <sup>65</sup> can be simplified to:

$$g_{Rg^2} = 1 - DB \quad (1)$$

However, for our modeling  $g_{[\eta]}$  was needed which is related to  $g_{Rg^2}$ <sup>66</sup>:

$$g_{[\eta]} = (g_{Rg^2})^\epsilon \quad (2)$$

Although it is still unclear what is the exact value of  $\epsilon$  in the case of comb-like polymers, there is growing evidence that  $\epsilon \approx 1$ <sup>67-69</sup>, which was assumed to be the case in EpsA-O.

The DB resulted from the EpsA-O primary structure and  $g_{[\eta]}$  was then calculated. Therefore, it was possible to determine the theoretical intrinsic viscosity of the linear variant of the otherwise branched native EpsA-O for which the intrinsic viscosity was measured. By using the Yamakawa model equation, the value of  $lp$  of EpsA-O linear variant was retrieved, and subsequently also the value of  $\Theta_{plim}$  of simulated EpsA-O.

To determine the volume fraction of extracellular space in the WT *B. subtilis* biofilms the CLSM images of biofilms were analyzed by Fiji-ImageJ<sup>70,71</sup>. As the lateral resolution (i.e. x-y plane) is larger than the axial resolution in CLSM, parts of the biofilm were examined, where the fiber-like bacterial structures were oriented perpendicular to the x-y image plane in CLSM. The identified regions were enveloped by the Freehand selection tool, followed by area calculation. The number of *B. subtilis* cells that appeared as circles of 750 nm in diameter<sup>49</sup> inside the region was manually counted. The volume fraction of extracellular space was then obtained as a ratio of the area of all cells divided by the total area of the enveloped region. To obtain an average and SD, six regions were analyzed.

The 3D structures and their x-z projections were plotted in OriginPro (OriginLab, Northampton, USA). For pore diameter distributions, the images were exported to Fiji-ImageJ, where the pore areas were analyzed. The diameters were calculated as the square roots of the pore areas and the distributions were obtained in OriginPro. The pore distribution shown is an average of the distribution of five projections of five independent simulations.

### Cohesive energy and forces calculations

The estimate of cohesive energy density,  $E_c$ , of the EpsA-O in the biofilm was calculated<sup>72</sup>:

$$E_c = \frac{1}{2} G' \times \gamma_c^2 \quad (3)$$

where  $G'$  is the value of elastic modulus of EpsA-O 1% (w/v) in Mmsg medium in the linear viscoelastic range,  $\gamma_c$  is the critical strain, where, beyond this point, the structure is irreversibly broken down. Based on microscopy data, the intercellular volume, occupied by EpsA-O, of two neighboring bacillus cells in a single layer of biofilm extending along a longer axis of bacillus was calculated to be 2000 nm × 190 nm × 190 nm. The cohesive energy of EpsA-O in this volume is  $2 \times 10^{-20}$  J (at  $\gamma_c = 25\%$ , Fig. 4e). The distance needed to irreversibly move two cells apart is about 50 nm (25% from 190 nm). The force required to perform this work, which is equal to the cohesive energy, is therefore  $5 \times 10^{-13}$  N.

### Composition and linkage analysis of EpsA-O

Before chemical derivatizations, EpsA-O was dissolved in water, sonicated in an ice bath with 20 bursts of 1 min each, separated by 1 min intervals using a Branson sonifier equipped with a microtip at 2.8 A, and lyophilized. Composition analysis was carried out by chemical derivatization of the sonicated EpsA-O both as alditol acetates and trimethyl glycosides. Alditol acetates were obtained as described<sup>73</sup> after hydrolysis with 2 M trifluoroacetic acid (TFA) for 1 h at 125 °C. Trimethylsilyl methylglycosides were obtained by derivatization with Sylon™ HTP (Merck) after methanolysis of the EpsA-O with 3 M HCl in methanol at 85 °C for 16 h<sup>74</sup>. Linkage analysis, through derivatization to partially methylated alditol acetates (PMAA), was performed following the protocol developed by Harris<sup>75</sup>; for the reduction step, sodium borodeuteride (NaBD<sub>4</sub>) was used. The absolute configuration (except for QuiNac4Nac) was established by GLC analysis of the trimethylsilylated (+)-2-butyl glycosides derivatives<sup>76,77</sup>. The derivatized samples were analyzed by GLC using an Agilent Technologies 6850 gas chromatograph equipped with a flame ionization detector, an HP-1 capillary column (Agilent Technologies, 30 m), and using He as the carrier gas. The following temperature programs were used: for alditol acetates 3 min at 150 °C, 150–270 °C at 3 °C/min, 2 min at 270 °C, for trimethylsilyl methylglycosides 150–280 °C at 3 °C/min, for trimethylsilyl (TMS) (+)-2-butyl glycosides 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, and for PMAA 1 min at 125 °C, 125–240 °C at 4 °C/min, 2 min at 240 °C. GLC-MS analyses were carried out on an Agilent Technologies 7890 A gas chromatograph coupled to an Agilent Technologies 5975 C VL MSD using the same column and the temperature programs of the GLC analysis. Values of the integrated area of the PMAA were corrected by the effective carbon response factors<sup>78</sup>.

### Smith degradation of EpsA-O

Smith degradation was performed as previously described<sup>32</sup>. Twelve mg of sonicated EpsA-O were dissolved in 9.75 mL H<sub>2</sub>O; 60 μL of 0.76 M NaIO<sub>4</sub> and 190 μL of 2.1 M MgCl<sub>2</sub> were added and the solution was incubated for 7 days at 10 °C in the dark with stirring. Glycerol was added to stop the reaction and NaBH<sub>4</sub> was added to pH 10 for reduction of the aldehyde groups (16 h at 25 °C). The reaction was stopped with a dropwise addition of 50% acetic acid, the solution was dialyzed, lyophilized, and hydrolyzed with 5 mL 0.1 M TFA at 24 °C for 7 days. The acid was removed by evaporation under reduced pressure, isopropanol was added and the sample was taken to dryness; the latter procedure was repeated three times. The sample, named **EpsA-O SD**, was dissolved in 1 mL of H<sub>2</sub>O, taken to pH = 7.3, and recovered by lyophilization. **EpsA-O SD** was dissolved in 1.9 mL of 0.05 M NaNO<sub>3</sub> and separated by size exclusion chromatography on a Bio-Gel P2 column (1.6 cm i.d. × 90 cm) at a flow rate of 6.8 mL/h. Fractions were collected at 12 min intervals. Elution was monitored using a refractive index detector (Knauer, RI detector K-2301, Lab-Service Analytica) which was interfaced with a computer via PicoLog software. Three main peaks were obtained (**SD1**, **SD2**, **SD3**) and they were desalted by medium-pressure

chromatography using a Bioline chromatography system equipped with a Superdex G30 glass column (1 cm i. d. × 82 cm) connected to a Smartline 1050 pump (Knauer). Elution with water at 1 mL/min flow rate was monitored with a refractive index detector (Smartline 2300, Knauer, Lab-Service Analytica) which was interfaced with a computer via Clarity software. Fractions were collected at 30 s intervals and those belonging to the same peak were pooled, recovered by lyophilization, and subjected to 1D and 2D NMR spectroscopy investigation.

### Solvolysis of EpsA-O

Solvolysis was performed by adding TFA (1.2 mL) to 16.4 mg of EpsA-O previously sonicated as reported above. The solution was stirred at 40 °C for 1 h<sup>79</sup>. The sample was dried under a nitrogen flux and washed three times with anhydrous methanol to remove TFA. The newly formed aldehydic functions were reduced with aqueous sodium borohydride for 16 h at RT. The reduction was then stopped with acetic acid and the sample was taken to dryness by evaporation under reduced pressure. The procedure was repeated three times after adding 10% acetic acid in methanol and three times with methanol. The dried sample was dissolved in 0.15 M NaCl and subjected to middle-pressure size-exclusion chromatography on the Bioline system previously described, using a Sephacryl S-300 HR column (1 cm id × 45 cm). Elution was carried out with 0.15 M NaCl at 1 mL/min flow rate. The selection of fractions to be pooled together was obtained by applying a fitting procedure using a Gaussian function and hypothesizing two main peaks, the degraded polysaccharide, and salt. According to the fitting curves obtained, (Supplementary Fig. 9a), test tubes were pooled in three fractions, named **SL1**, **SL2**, and **SL3**, the latter being mainly salt. Fractions **SL1** and **SL2** were then desalted using a Superdex G30 prep grade column (1 cm i.d. × 82 cm) with water as eluent at 1.5 mL/min flow rate, recovered by lyophilization, and dissolved in 10 mM, pH = 7 sodium phosphate buffer, blown with N<sub>2</sub> to prevent degradation of QuiNac4Nac. <sup>1</sup>H NMR analysis showed that fraction **SL1** was less heterogeneous than **SL2** and thus it was chosen for thorough NMR investigation (Supplementary Fig. 9b).

### NMR spectroscopy

Samples subjected to NMR were exchanged twice with 99.9% D<sub>2</sub>O by lyophilization, dissolved in 0.6 mL of 99.96% D<sub>2</sub>O, and introduced into a 5 mm NMR tube for data acquisition. Spectra of all samples except **SL1** were recorded on a 500 MHz VARIAN spectrometer operating at 343 K after setting the proper pw90° pulse. 2D experiments were performed using standard VARIAN pulse sequences and pulsed field gradients for coherence selection when appropriate. HSQC spectra were recorded using 140 Hz (for directly attached <sup>1</sup>H–<sup>13</sup>C correlations). HMBC experiments were recorded using a coupling constant of 8 Hz (for long-range <sup>1</sup>H–<sup>13</sup>C correlations) and a relaxation time 1.2 s. TOCSY spectra were acquired using 100 ms spin-lock time and 1.0 s relaxation time. NOESY experiments were recorded with 200 ms mixing time and 1.0 s relaxation time. Chemical shifts are expressed in ppm using acetone as an internal reference (2.225 ppm for <sup>1</sup>H and 31.07 ppm for <sup>13</sup>C). NMR spectra were processed using MestreNova software. **SL1** spectra (1D <sup>1</sup>H and 2D, COSY, TOCSY, and HSQC) were obtained using a Bruker Avance 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. The 2D TOCSY experiment was recorded using a mixing time of 160 ms and the 1D variants using a mixing time of 200 ms. The 1D NOESY experiments were performed using a mixing time of 300 ms. The HSQC (with multiplicity editing) experiment was optimized for J = 145 Hz (for directly attached <sup>1</sup>H–<sup>13</sup>C correlations). 2D experiments were recorded using non-uniform sampling: 40% for homonuclear and 20% for heteronuclear experiments.

### Prediction of proteins function

Prediction of the function of enzymes encoded by the epsA-O cluster was performed using protein analysis tools available on the NCBI website. Sequence similarity was assessed using the BLASTP suite<sup>80</sup> and conserved domains were identified using the CD Search Service<sup>81</sup>. Multiple sequence

alignments were performed using the Constraint-based Multiple Alignment Tool (COBALT)<sup>82</sup>.

### Data presentation and statistical analysis

All the experiments were done in 3–6 independent biological replicates. CLSM micrographs and data of the most representative series are shown. The statistical analysis and data presentation were performed in OriginPro (OriginLab, USA) program. Experimental errors indicated in text or represented as error bars are given as standard deviations. For testing statistical significance, a two-tailed unpaired t-test was used. Samples showing  $p$ -value < 0.05 were considered statistically significantly different.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### Data availability

The authors declare that the data supporting the findings of this study are available within the paper and its Supplementary Material. Raw microscopic image data CLSM sets are available at FigShare database at <https://doi.org/10.6084/m9.figshare.26878876>. Additional data are available from the corresponding author upon a reasonable request.

### Code availability

The String-of-beads model is available from the corresponding authors upon a reasonable request.

Received: 24 January 2024; Accepted: 20 August 2024;

Published online: 02 October 2024

### References

- Arnaouteli, S., Bamford, N. C., Stanley-Wall, N. R. & Kovács, Á. T. *Bacillus subtilis* biofilm formation and social interactions. *Nat. Rev. Microbiol.* **19**, 600–614 (2021).
- Errington, J. & Aart, L. T. V. Microbe Profile: *Bacillus subtilis*: model organism for cellular development, and industrial workhorse. *Microbiology* **166**, 425–427 (2020).
- Earl, A. M., Losick, R. & Kolter, R. Ecology and genomics of *Bacillus subtilis*. *Trends Microbiol.* **16**, 269–275 (2008).
- Vlamakis, H., Chai, Y., Beaugregard, P., Losick, R. & Kolter, R. Sticking together: building a biofilm the *Bacillus subtilis* way. *Nat. Rev. Microbiol.* **11**, 157–168 (2013).
- Sretenovic, S. et al. An early mechanical coupling of planktonic bacteria in dilute suspensions. *Nat. Commun.* **8**, 213 (2017).
- Lopez, D., Fischbach, M. A., Chu, F., Losick, R. & Kolter, R. Structurally diverse natural products that cause potassium leakage trigger multicellularity in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA* **106**, 280–285 (2009).
- Chai, Y., Chu, F., Kolter, R. & Losick, R. Bistability and biofilm formation in *Bacillus subtilis*. *Mol. Microbiol.* **67**, 254–263 (2008).
- Špacapan, M. et al. The ComX Quorum sensing peptide of *Bacillus subtilis* affects biofilm formation negatively and sporulation positively. *Microorganisms* **8**, 1131 (2020).
- Bolješić, M., Kraigher, B., Dogsa, I., Jerič Kokelj, B. & Mandić-Mulec, I. Kin discrimination modifies strain distribution, spatial segregation, and incorporation of extracellular matrix polysaccharide mutants of *Bacillus subtilis* strains into mixed floating biofilms. *Appl. Environ. Microbiol.* **22**, 88 (2022).
- Dergham, Y. et al. Direct comparison of spatial transcriptional heterogeneity across diverse *Bacillus subtilis* biofilm communities. *Nat. Commun.* **14**, 7546 (2023).
- Diehl, A. et al. Structural changes of TasA in biofilm formation of *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA* **115**, 3237–3242 (2018).
- Hobley, L. et al. BslA is a self-assembling bacterial hydrophobin that coats the *Bacillus subtilis* biofilm. *Proc. Natl. Acad. Sci. USA* **110**, 13600–13605 (2013).
- Roske, Y. et al. TapA acts as specific chaperone in TasA filament formation by strand complementation. *Proc. Natl. Acad. Sci. USA* **120**, e2217070120 (2023).
- Dogsa, I., Kostanjšek, R. & Stopar, D. eDNA provides a scaffold for autoaggregation of *B. subtilis* in Bacterioplankton suspension. *Microorganisms* **11**, 332 (2023).
- Branda, S. S., Gonzalez-Pastor, J. E., Ben-Yehuda, S., Losick, R. & Kolter, R. Fruiting body formation by *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA* **98**, 11621–11626 (2001).
- Kearns, D. B., Chu, F., Branda, S. S., Kolter, R. & Losick, R. A master regulator for biofilm formation by *Bacillus subtilis*. *Mol. Microbiol.* **5**, 739–749 (2005).
- Beaugregard, P. B., Chai, Y., Vlamakis, H., Losick, R. & Kolter, R. *Bacillus subtilis* biofilm induction by plant polysaccharides. *Proc. Natl. Acad. Sci. USA* **110**, E1621–E1630 (2013).
- Azulay, D. N. et al. Multiscale X-ray study of *Bacillus subtilis* biofilms reveals interlinked structural hierarchy and elemental heterogeneity. *Proc. Natl. Acad. Sci. USA* **119**, e2118107119 (2022).
- Shemesh, M., Kolter, R. & Losick, R. The biocide chlorine dioxide stimulates biofilm formation in *Bacillus subtilis* by activation of the histidine kinase KinC. *J. Bacteriol.* **192**, 6352–6356 (2010).
- Seminara, A. et al. Osmotic spreading of *Bacillus subtilis* biofilms driven by an extracellular matrix. *Proc. Natl. Acad. Sci. USA* **109**, 1116–1121 (2012).
- Oppenheimer-Shaanan, Y. et al. Spatio-temporal assembly of functional mineral scaffolds within microbial biofilms. *npj Biofilms Microbiome* **2**, 15031 (2016).
- Branda, S. S. et al. Genes involved in formation of structured multicellular communities by *Bacillus subtilis*. *J. Bacteriol.* **186**, 3970–3979 (2004).
- Guttenplan, S. B., Blair, K. M. & Kearns, D. B. The EpsE flagellar clutch is bifunctional and synergizes with EPS biosynthesis to promote *Bacillus subtilis* biofilm formation. *PLoS Genet.* **6**, e1001243 (2010).
- Dogsa, I., Brložnik, M., Stopar, D. & Mandić-Mulec, I. Exopolymer diversity and the role of levan in *Bacillus subtilis* biofilms. *PLoS ONE* **8**, e62044 (2013).
- Krajnc, M. et al. Systems view of *Bacillus subtilis* pellicle development. *npj Biofilms Microbiome* **8**, 25 (2022).
- Roux, D. et al. Identification of Poly-N-acetylglucosamine as a Major Polysaccharide Component of the *Bacillus subtilis* Biofilm Matrix. *J. Biol. Chem.* **290**, 19261–19272 (2015).
- Arbour, C. A. et al. Defining early steps in *Bacillus subtilis* biofilm biosynthesis. *mBio* **14**, e0094823 (2023).
- Yunrong, C. et al. Galactose metabolism plays a crucial role in biofilm formation by *Bacillus subtilis*. *mBio* **3**, 4 (2012).
- Jones, S. E., Paynich, M. L., Kearns, D. B. & Knight, K. L. Protection from intestinal inflammation by bacterial exopolysaccharides. *J. Immunol.* **192**, 4813–4820 (2014).
- Bridier, A., Meylheuc, T. & Briandet, R. Realistic representation of *Bacillus subtilis* biofilms architecture using combined microscopy (CLSM, ESEM and FESEM). *Micron*. **48**, 65–69 (2013).
- Dergham, Y. et al. *Bacillus subtilis* NDmed, a model strain for biofilm genetic studies. *Biofilm* **6**, 100152 (2023).
- Hay, G. W., Lewis, B. A. & Smith, F. Periodate Oxidation of Polysaccharides: General Procedures in Methods in Carbohydrate Chemistry Vol. V (ed. Whistler, R. L., BeMiller, J. N. & Wolfrom, M. L.) 357–361 (Academic Press, New York, 1965).
- Schäffer, C. et al. Purification and structure elucidation of the N-acetylglucosamine-containing polysaccharide from *Bacillus licheniformis* ATCC 9945. *Eur. J. Biochem.* **268**, 857–864 (2001).

34. James, C. & Leitch, R. Determination of the structure and absolute configuration of the glycerolphosphate containing capsular polysaccharide of *Pasteurella haemolytica* serotype T3 by high-resolution nuclear magnetic resonance spectroscopy. *Can. J. Chem.* **68**, 1574–1584 (1990).
35. Lindberg, B., Lönngrén, J. & Svensson, S. Specific degradation of polysaccharides. *Adv. Carbohydr. Chem. Biochem.* **31**, 185–240 (1975).
36. Bilan, M. I., Vinogradova, E. V., Shashkov, A. S. & Usov, A. I. Structure of a highly pyruvylated galactan sulfate from the Pacific green alga *Codium yezoense* (Bryopsidales, Chlorophyta). *Carbohydr. Res.* **342**, 586–596 (2007).
37. Cahill, S. M. et al. Elucidation of the K32 capsular polysaccharide structure and characterization of the KL32 gene cluster of *Acinetobacter baumannii* LUH5549. *Biochemicals* **85**, 241–247 (2020).
38. Hager, F. F., Sützl, L., Stefanovic, C., Blaukopf, M. & Schäffer, C. Pyruvate substitutions on glycoconjugates. *Int. J. Mol. Sci.* **20**, 4929 (2019).
39. Higuchi, Y. et al. A rationally engineered yeast pyruvyltransferase Pvg1p introduces sialylation-like properties in neo-human-type complex oligosaccharide. *Sci. Rep.* **6**, 26349 (2016).
40. Pan, Y. J. et al. Genetic analysis of capsular polysaccharide synthesis gene clusters in 79 capsular types of *Klebsiella* spp. *Sci. Rep.* **5**, 15573 (2015).
41. Doyle, R. J. & Birdsell, D. C. Interaction of concanavalin A with the cell wall of *Bacillus subtilis*. *J. Bacteriol.* **109**, 652–658 (1972).
42. Koyano, Y., Okajima, K., Mihara, M. & Yamamoto, H. Visualization of Wall Teichoic acid decoration in *Bacillus subtilis*. *J. Bacteriol.* **205**, e0006623 (2023).
43. Mirouze, N., Ferret, C., Cornilleau, C. & Carballido-López, R. Antibiotic sensitivity reveals that wall teichoic acids mediate DNA binding during competence in *Bacillus subtilis*. *Nat. Commun.* **9**, 5072 (2018).
44. Hayhurst, E. J., Kailas, L., Hobbs, J. K. & Foster, S. J. Cell wall peptidoglycan architecture in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA* **105**, 14603–14608 (2008).
45. Dogsa, I., Štrancar, J., Laggner, P. & Stopar, D. Efficient modeling of polysaccharide conformations based on Small-Angle X-ray Scattering experimental data. *Polymer* **49**, 1398–1406 (2008).
46. Dogsa, I. et al. Amorphous supramolecular structure of carboxymethyl cellulose in aqueous solution at different pH values as determined by rheology, small angle X-ray and light scattering. *Carbohydr. Polym.* **111**, 492–504 (2014).
47. Dogsa, I., Cerar, J., Jamnik, A. & Tomšič, M. Supramolecular structure of methyl cellulose and lambda- and kappa-carrageenan in water: SAXS study using the string-of-beads model. *Carbohydr. Polym.* **172**, 184–196 (2017).
48. Yamakawa, H. & Fujii, M. Intrinsic viscosity of wormlike chains. determination of the shift factor. *Macromolecules* **7**, 128–135 (1974).
49. Dion, M. F. et al. *Bacillus subtilis* cell diameter is determined by the opposing actions of two distinct cell wall synthetic systems. *Nat. Microbiol.* **4**, 1294–1305 (2019).
50. Milas, M., Rinaudo, M. & Tinland, B. The viscosity dependence on concentration, molecular weight and shear rate of xanthan solutions. *Polym. Bull.* **14**, 157–164 (1985).
51. Benigar, E. et al. Structure and dynamics of a polysaccharide matrix: aqueous solutions of bacterial levan. *Langmuir.* **30**, 4172–4182 (2014).
52. Stojković, B., Sretenovic, S., Dogsa, I., Poberaj, I. & Stopar, D. Viscoelastic properties of Levan-DNA mixtures important in microbial biofilm formation as determined by micro- and macrorheology. *Biophys. J.* **108**, 758–765 (2015).
53. Romero, D., Vlamakis, H., Losick, R. & Kolter, R. Functional analysis of the accessory protein TapA in *Bacillus subtilis* amyloid fiber assembly. *J. Bacteriol.* **196**, 1505–1513 (2014).
54. Liu, S. et al. Probing the growth and mechanical properties of *Bacillus subtilis* biofilms through genetic mutation strategies. *Synth. Syst. Biotechnol.* **7**, 965–971 (2022).
55. Conn, H. J. The identity of *Bacillus subtilis*. *J. Infect. Dis.* **44**, 341–350 (1930).
56. Branda, S. S., Chu, F., Kearns, D. B., Losick, R. & Kolter, R. A major protein component of the *Bacillus subtilis* biofilm matrix. *Mol. Microbiol.* **59**, 1229–1238 (2006).
57. Chu, F., Kearns, D. B., Branda, S. S., Kolter, R. & Losick, R. Targets of the master regulator of biofilm formation in *Bacillus subtilis*. *Mol. Microbiol.* **59**, 1216–1228 (2006).
58. Dragoš, A. et al. Division of labor during biofilm matrix production. *Curr. Biol.* **28**, 1903–1913 (2018).
59. Aucken, H. M., Wilkinson, S. G. & Pitt, T. L. Identification of capsular antigens in *Serratia marcescens*. *J. Clin. Microbiol.* **35**, 59–63 (1997).
60. Abe, K. et al. Developmentally-regulated excision of the SPb prophage reconstitutes a gene required for spore envelope maturation in *Bacillus subtilis*. *PLoS Genet.* **10**, e1004636 (2014).
61. Miller, G. L. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal. Chem.* **31**, 426–428 (1959).
62. Pinto, E. P., Furlan, L. & Vendruscolo, C. T. Chemical deacetylation natural xanthan (Jungbunzlauer). *Polimeros* **21**, 47e52 (2011).
63. Cuesta, G., Suarez, N., Bessio, M. I., Ferreira, F. & Massaldi, H. Quantitative determination of pneumococcal capsular polysaccharide serotype 14 using a modification of phenol-sulfuric acid method. *J. Microbiol. Methods* **52**, 69–73 (2003).
64. Grubisic, Z., Rempp, P. & Benoit, H. A universal calibration for gel permeation chromatography. *Polym. Lett.* **5**, 753–759 (1967).
65. Casassa, E. F. & Berry, G. C. Angular distribution of intensity of rayleigh scattering from comblike branched molecules. *J. Polym. Sci. Part A-2: Polym. Phys.* **4**, 881–897 (1966).
66. Kurata, M. & Fukatsu, M. Unperturbed dimension and translational friction constant of branched polymers. *J. Chem. Phys.* **41**, 2934–2944 (1964).
67. Lee, S. & Chang, T. Branching analysis of comb-shaped polystyrene with long chain branches. *Macromol. Chem. Phys.* **218**, 1700087 (2017).
68. Ivancic, R. J. S., Orski, S. V. & Audus, D. J. Structure–dilute solution property relationships of comblike macromolecules in a good solvent. *Macromolecules* **55**, 766–775 (2022).
69. Radke, W. & Müller, A. H. E. Synthesis and characterization of comb-shaped polymers by SEC with on-line light scattering and viscometry detection. *Macromolecules* **38**, 3949–3960 (2005).
70. Rueden, C. T. et al. ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinforma.* **18**, 529 (2017).
71. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
72. Tadros, T. F. *Rheology of Dispersions: Principles and Applications* (ed. Tadros, T. F.) (Wiley-VCH: Weinheim, Germany, 2010).
73. Albersheim, P., Nevins, D. J., English, P. D. & Karr, A. A method for the analysis of sugars in plant cell-wall polysaccharides by gas-liquid chromatography. *Carbohydr. Res.* **5**, 340–345 (1967).
74. Kakehi, K. & Honda, S. Silyl ethers of carbohydrates in Analysis of Carbohydrates by GLC and MS. 1st ed. Volume 1. (ed. Biermann, C. J. & McGinnis, G. D.) 43–85 (CRC Press; Boca Raton, FL, USA 1988).
75. Harris, P. J., Henry, R. J., Blakeney, A. B. & Stone, B. A. An improved procedure for the methylation analysis of oligosaccharides and polysaccharides. *Carbohydr. Res.* **127**, 59–73 (1984).
76. Gerwig, G. J., Kamerling, J. P. & Vliegthart, J. F. G. Determination of the d and l configuration of neutral monosaccharides by high-resolution capillary g.l.c. *Carbohydr. Res.* **62**, 349–357 (1978).
77. Gerwig, G. J., Kamerling, J. P. & Vliegthart, J. F. G. Determination of the absolute configuration of monosaccharides in complex carbohydrates by capillary g.l.c. *Carbohydr. Res.* **77**, 1–7 (1979).

78. Sweet, D. P., Shapiro, R. H. & Albersheim, P. Quantitative analysis by various g.l.c. response-factor theories for partially methylated and partially ethylated alditol acetates. *Carbohydr. Res.* **40**, 217–225 (1975).
79. Vyacheslav, L. et al. Solvolysis with trifluoroacetic acid: an efficient method for selective cleavage of polysaccharides. *Mendeleev Commun.* **26**, 279–281 (2016).
80. Johnson, M. et al. NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, W5–W9 (2008).
81. Marchler-Bauer, A. et al. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* **39**, D225–D229 (2011).
82. Papadopoulos, J. S. & Agarwala, R. *Bioinformatics* **23**, 1073–1079, (2007).
- and designed the experiments, performed Smith degradation of EpsA-O, and interpreted NMR spectra. All authors contributed to writing the manuscript. All authors read and approved the final manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41522-024-00555-z>.

**Correspondence** and requests for materials should be addressed to Paola Cescutti.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

### Acknowledgements

The work was supported by the Slovenian Research and Innovation Agency (ARIS) J1-3021 research grant and National program grant P4-0116. The microscopy experiments were supported by funding from ARIS to Infrastructural Centre Microscopy of Biological Samples (MRIC UL, I0-0022-0481-08), at Biotechnical Faculty, University of Ljubljana, Slovenia. When the experiments were performed Barbara Bellich was working at the University of Trieste, Department of Life Sciences, Via L. Giorgieri 1, Trieste, Italy.

### Author Contributions

I.D. conceived and designed the experiments, performed modeling, rheology, and microscopy data analysis, is responsible for funds and project management. B.B. determined the composition and linkages of EpsA-O, performed solvolysis of EpsA-O, and recorded NMR spectra. M.B. performed microscopy and rheology experiments and isolated EpsA-O under I.D. and D.S. supervision. C.L. assigned the enzymatic functions within the *epsA-O* gene cluster. N.R. recorded and interpreted NMR spectra. R.R. conceived and designed the experiments, and interpreted NMR spectra. D.S.: Conceived and designed the experiments. P.C. conceived