

SUPPLEMENTARY INFORMATION

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

Iztok Dogsa¹, Barbara Bellich^{2,#}, Mojca Blaznik¹, Cristina Lagatolla³, Neil Ravenscroft⁴, Roberto Rizzo³, David Stopar¹, Paola Cescutti^{3,*}

¹ University of Ljubljana, Biotechnical Faculty, Department of Microbiology, Večna pot 111, Ljubljana, Slovenia

² Department of Advanced Translational Diagnostics, Institute for Maternal and Child Health, IRCCS "Burlo Garofolo", Via dell'Istria 65, 34137 Trieste, Italy

³ University of Trieste, Department of Life Sciences, Via L. Giorgieri 1, Trieste, Italy

⁴ University of Cape Town, Department of Chemistry, Rondebosch, 7701, South Africa

* Corresponding author: Department of Life Sciences, University of Trieste, via L. Giorgieri 1, Bdg C11, 34127 Trieste, Italy. E-mail address: pcescutti@units.it (P. Cescutti)

When the experiments were performed Barbara Bellich was working at Department of Life Sciences, University of Trieste, Via Licio Giorgieri 1, Bdg. C11, 34127 Trieste, Italy.

Content of Supplementary Information:

Supplementary Figure 1 | The amount of EpsA-O raw isolate in different growth media.

Supplementary Figure 2 | HPLC-SEC, DNS analysis and purity test of isolated EpsA-O.

Supplementary Figure 3 | Amplitude sweep test of EpsA-O samples.

Supplementary Figure 4 | Analysis of the Smith degradation products of EpsA-O.

Supplementary Figure 5 | 1D NMR spectra of SD1 recorded at 70 °C and 500 MHz.

Supplementary Figure 6 | NOESY plot of sample SD1 recorded at 70 °C.

Supplementary Figure 7 | TOCSY plots of SD2 sample recorded at 70 °C.

Supplementary Figure 8 | Ring region of the SD2 HSQC spectrum recorded at 70 °C.

Supplementary Figure 9 | Analysis of EpsA-O sample after solvolysis.

Supplementary Figure 10 | Evidence that EpsA-O contains a minor glucan impurity.

Supplementary Figure 11 | 1D NOESY spectrum obtained with selective excitation of H1 of the 6-linked α -Gal at 4.92 ppm present in the SL1 fraction.

Supplementary Figure 12 | Repeating unit structure of EpsA-O.

Supplementary Data 1: Functions assignments of the genes *epsE*, *epsH*, *epsJ* and *epsF* in the *epsA-epsO* gene cluster of *B. subtilis*.

Supplementary Table 1. Similarity level between Eps enzymes and proteins with known function obtained by BLAST alignment.

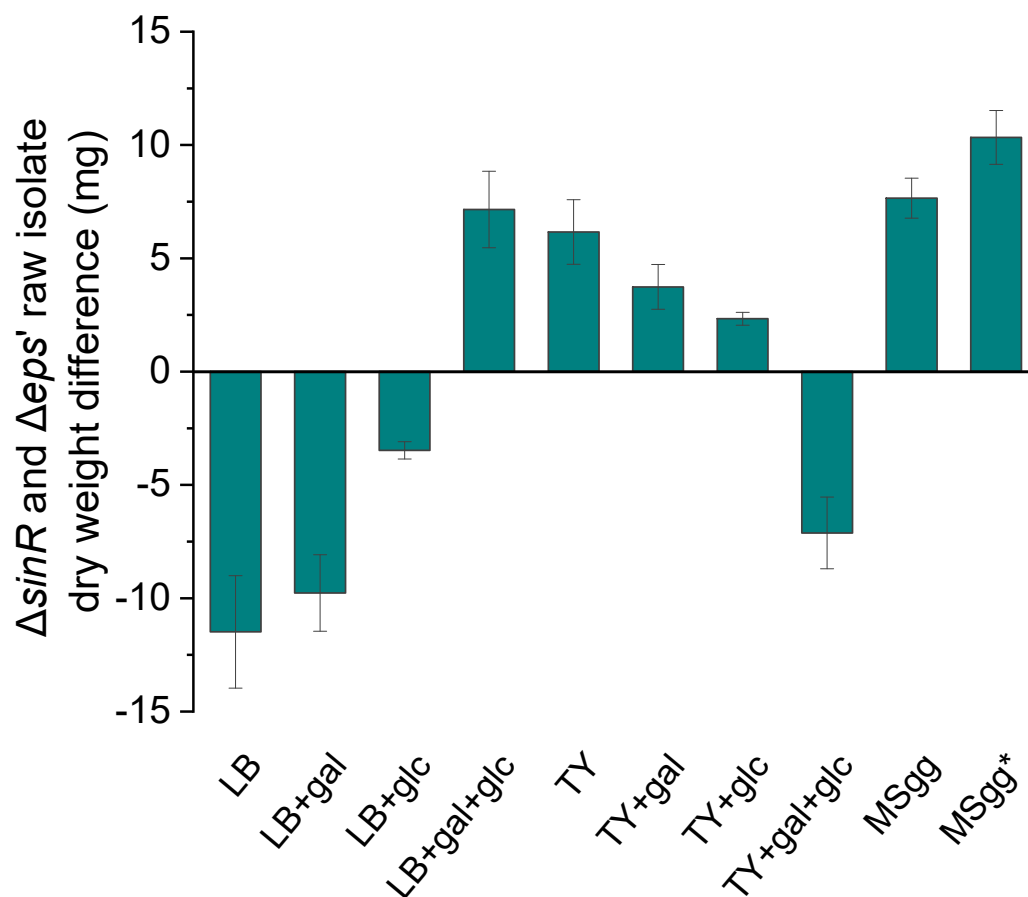
Supplementary Figure 13 | COBALT Multiple alignment of EpsF of *B. subtilis* and its homologous wbcN of *Yersinia enterocolitica* with some representatives of the GT4-CapH-like Family Protein.

Supplementary Figure 14 | Graphic Summary of BLAST alignments of EpsH with EpsJ and EpsE.

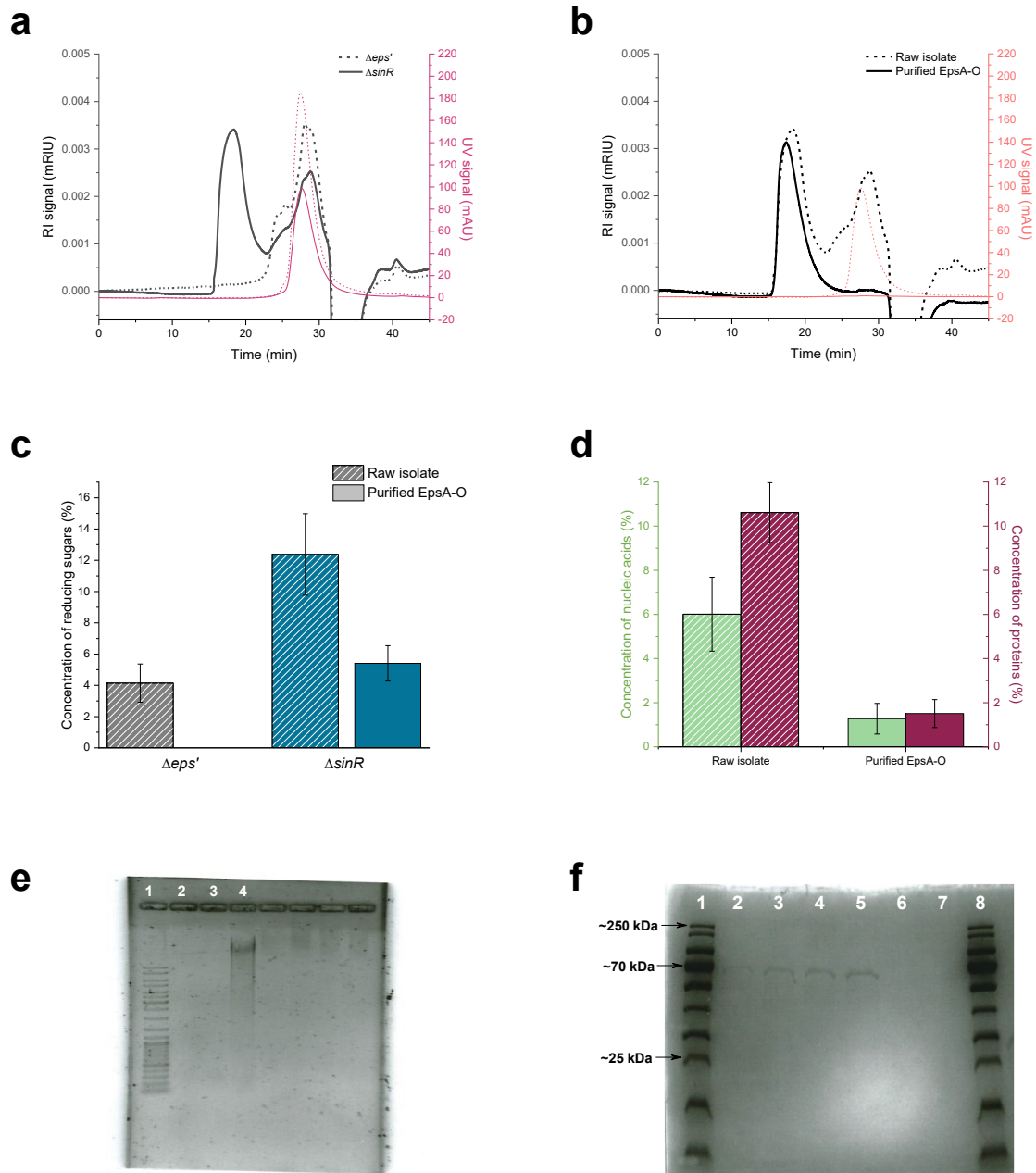
Supplementary Figure 15 | COBALT Multiple alignment of EpsI and EpsO pyruvyltransferases of *B. subtilis* with the enzymes WclU of *Klebsiella pneumoniae* and Pvg1 of *Saccharomyces pombe*.

Supplementary Figure 16 | The workflow for preparing slides for observing biofilms under a microscope.

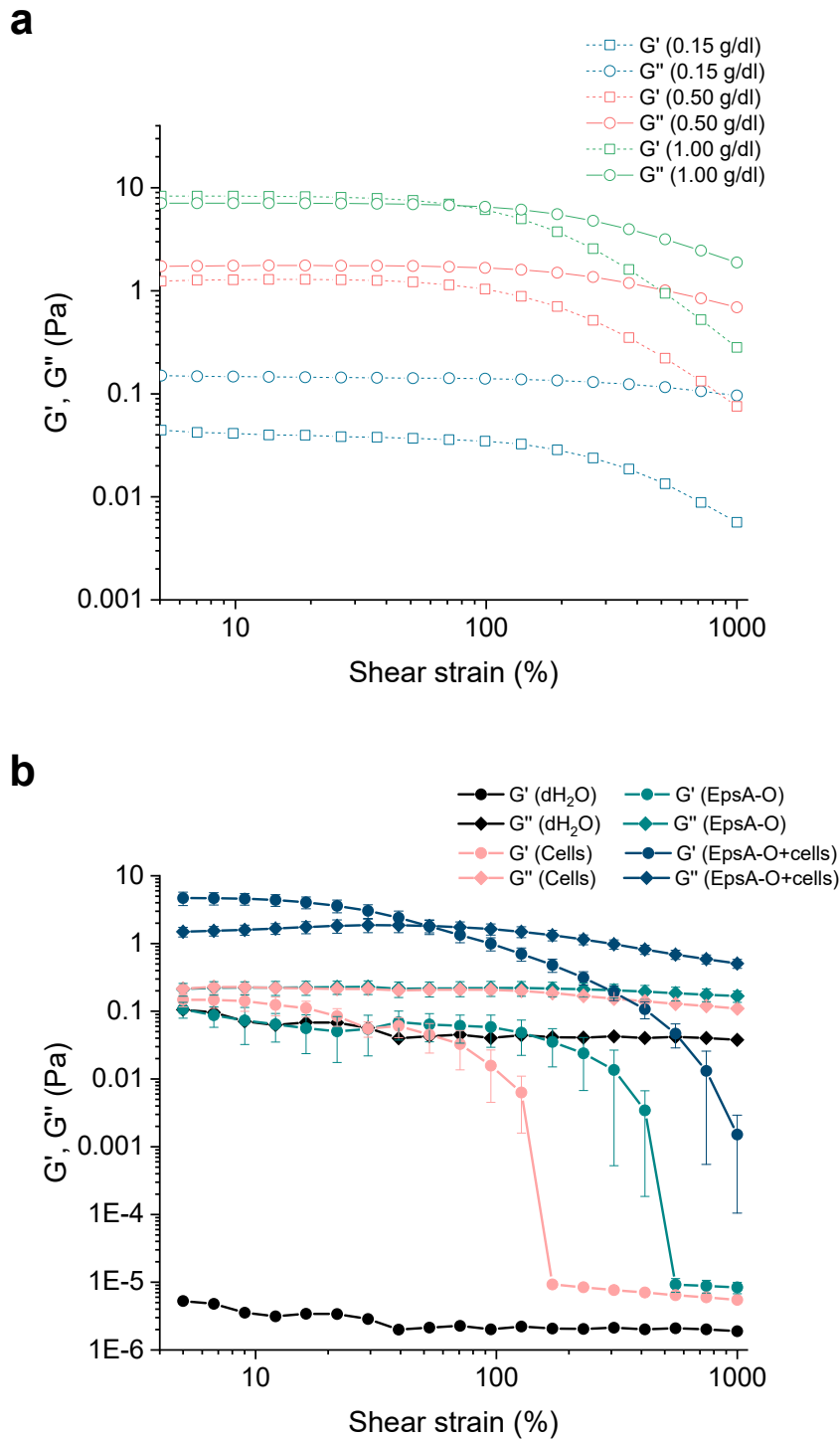
Supplementary References



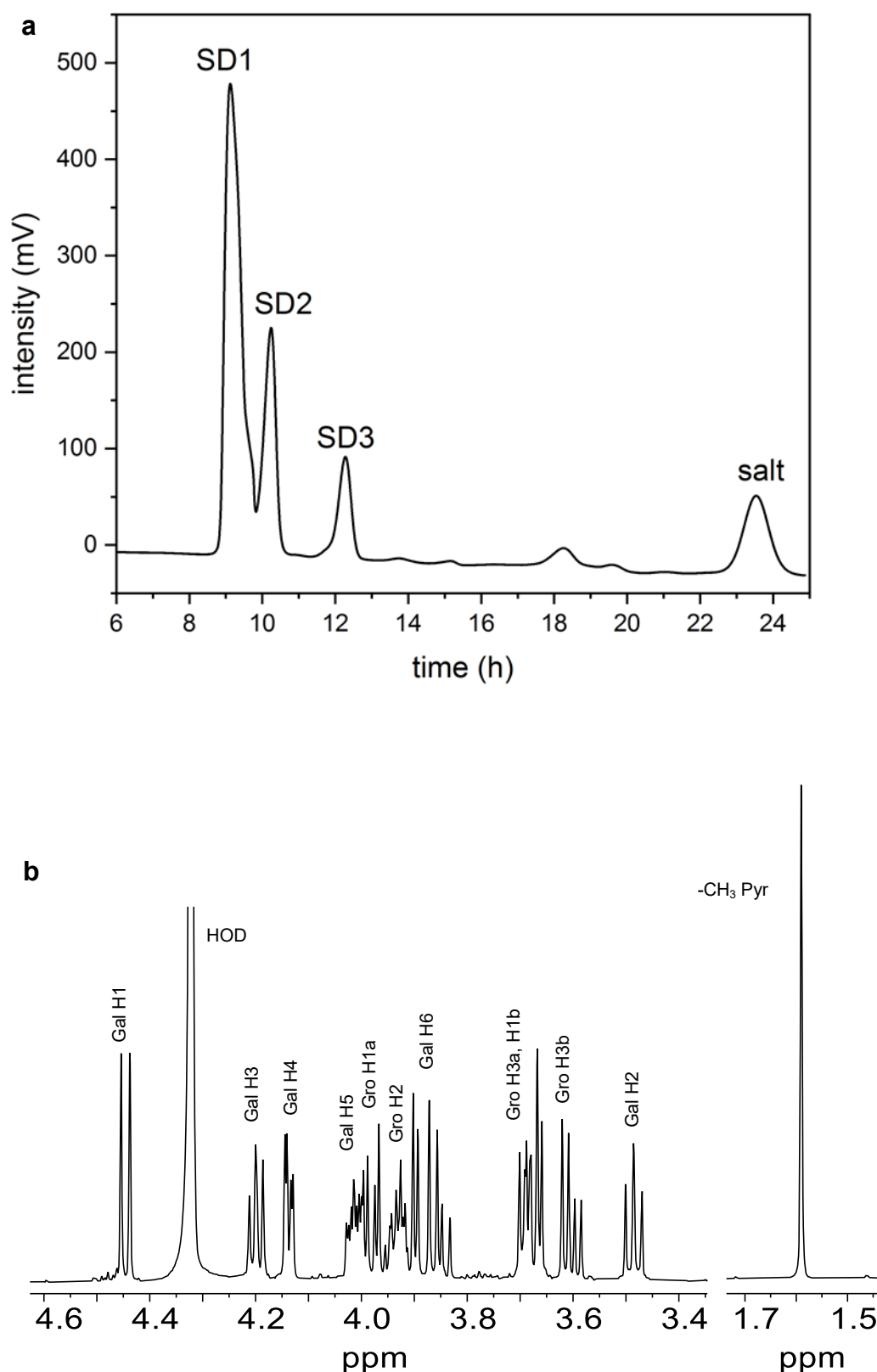
Supplementary Figure 1 | The amount of EpsA-O raw isolate in different growth media. The mass of EpsA-O raw isolate was calculated as the difference of raw isolate obtained from 12 mL of spent liquid growth-media of *B. subtilis* $\Delta sinR$ (EpsA-O hyper-producer) and *B. subtilis* $\Delta eps'$ (deficient in EpsA-O production-negative control). Three different growth media (LB, TY, MSgg) with supplements (gal = galactose 0.5 % (w/v); glc = glucose 1 % (w/v); MSgg* = MSgg with increased concentration of glycerol, from 5 g/L to 12.5 g/L) were inoculated and incubated for 16 h at 37 °C at 200 rpm, following isolation procedure (see Materials and Methods, Isolation of EpsA-O polysaccharide).



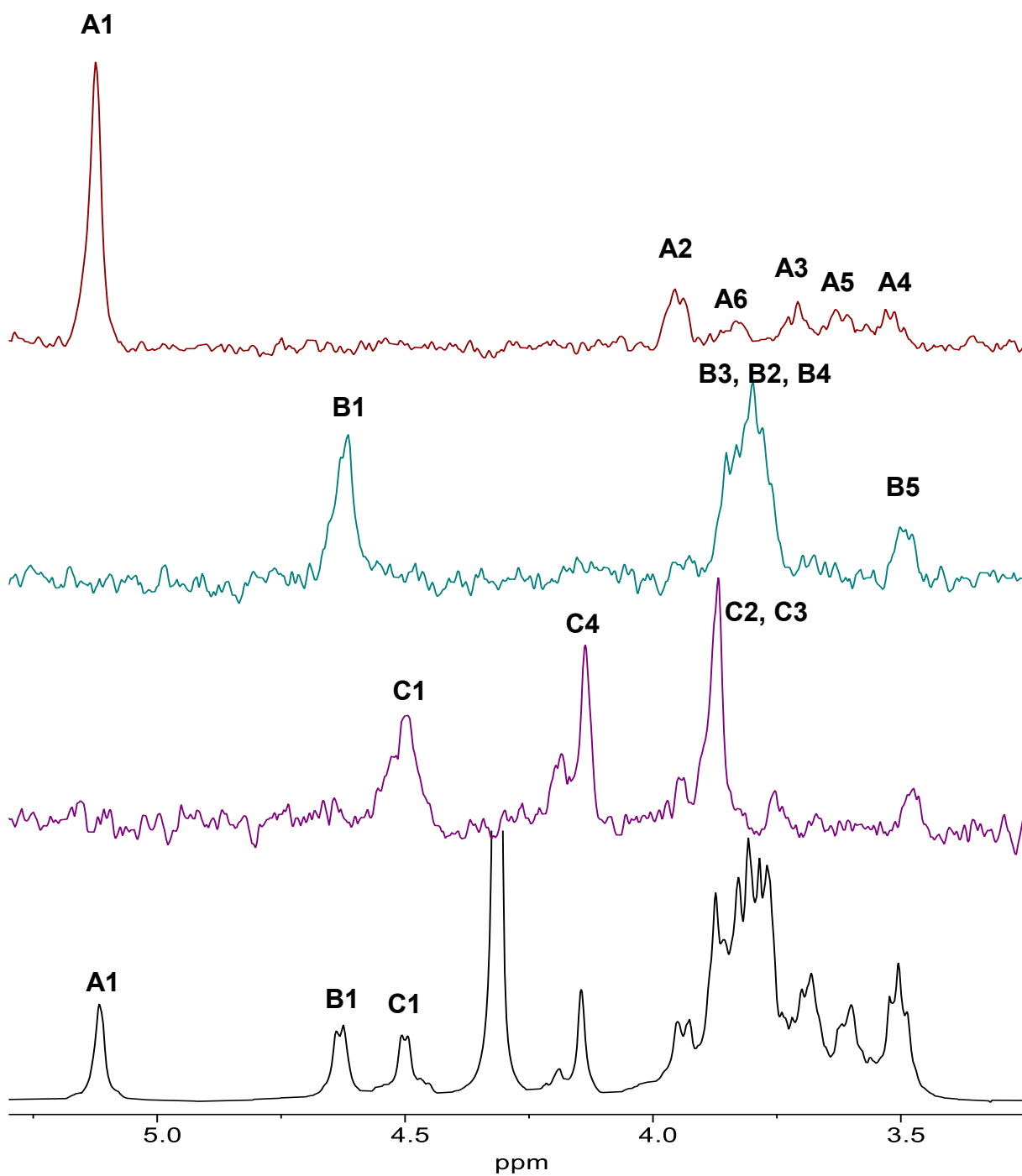
Supplementary Figure 2 | HPLC-SEC, DNS analysis and purity test of isolated EpsA-O. **a**, HPLC-SEC chromatogram recorded on RI and UV detectors of raw isolated material from $\Delta sinR$, (EpsA-O hyperproducer) and $\Delta eps'$ (deficient in EpsA-O and TasA production). Black curves: RI detector signal (mRIU); red curves: UV detector signal (mAU). The dashed lines correspond to the *B. subtilis* $\Delta eps'$ strain, and the solid curves correspond to the *B. subtilis* $\Delta sinR$. **b**, $\Delta sinR$ was further purified to obtain purified EpsA-O. Black curves: RI detector signal (mRIU); orange curves: UV detector signal (mAU). The dashed lines correspond to raw isolate, and the solid curves correspond to the purified EpsA-O. **c**, DNS analysis. Patterned bar represents raw isolates, while unpatterned bar represents purified EpsA-O sample. **d**, Concentration of nucleic acid and proteins contaminants of EpsA-O isolates in % (w/v) as determined by UV-VIS and Bradford assay, respectively. **e**, DNA gel electrophoresis of EpsA-O sample (1, DNA ladder mix; 2, EpsA-O 0.1 %; 3, EpsA-O 0.2 %; 4, *B. subtilis* PS-216 genomic DNA). **f**, SDS-page gel electrophoresis of EpsA-O sample (1, Protein ladder mix; 2, BSA 0.05 %; 3, BSA 0.10 %; 4, BSA 0.15 %; 5, BSA 0.20 %; 6, EpsA-O 0.1 %; 7, EpsA-O 0.2 %; 8, Protein ladder mix).



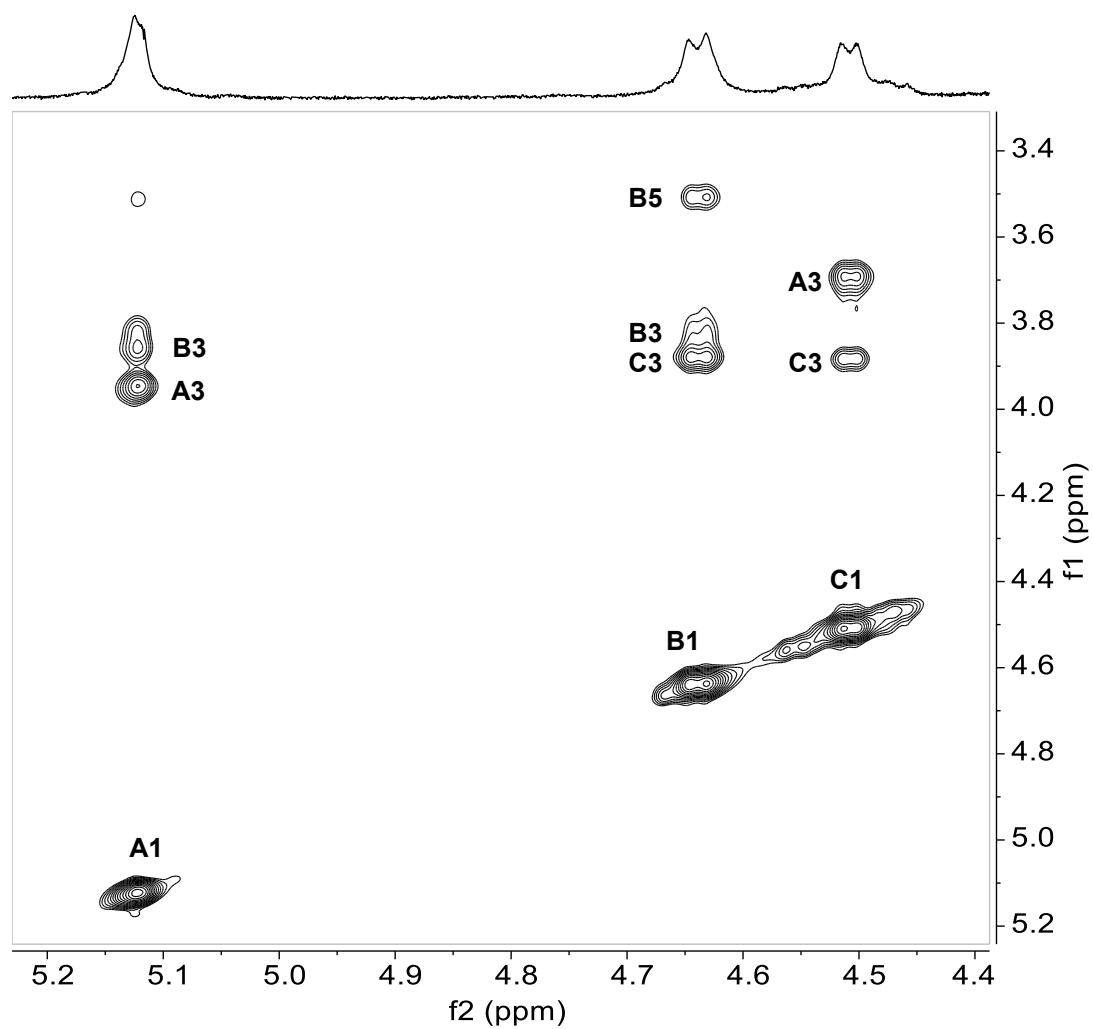
Supplementary Figure 3 | Amplitude sweep test of EpsA-O samples. a, Amplitude sweep test of EpsA-O dissolved in dH₂O at different EpsA-O concentrations. Squares = storage modulus (G'); circles = loss modulus (G''). The color coding for the symbols and lines indicates different concentrations: green 1 g/dL, orange 0.5 g/dL, and blue 0.15 g/dL. **b,** The formation of a gel structure in *B. subtilis* Δeps cells upon the addition of exogenous EpsA-O. An amplitude sweep test was conducted on Δeps cells, Δeps cells supplemented with exogenous EpsA-O, and a purified sample of EpsA-O. The final concentration of exogenous EpsA-O was 0.6 %. Circles = storage modulus (G'); diamonds = loss modulus (G''). The color coding for the symbols and lines indicates different samples: black for dH₂O, orange for cells only, green for 0.6 % EpsA-O polysaccharide solution, and blue for the mixture of bacterial cells and EpsA-O in a final concentration of 0.6 %.



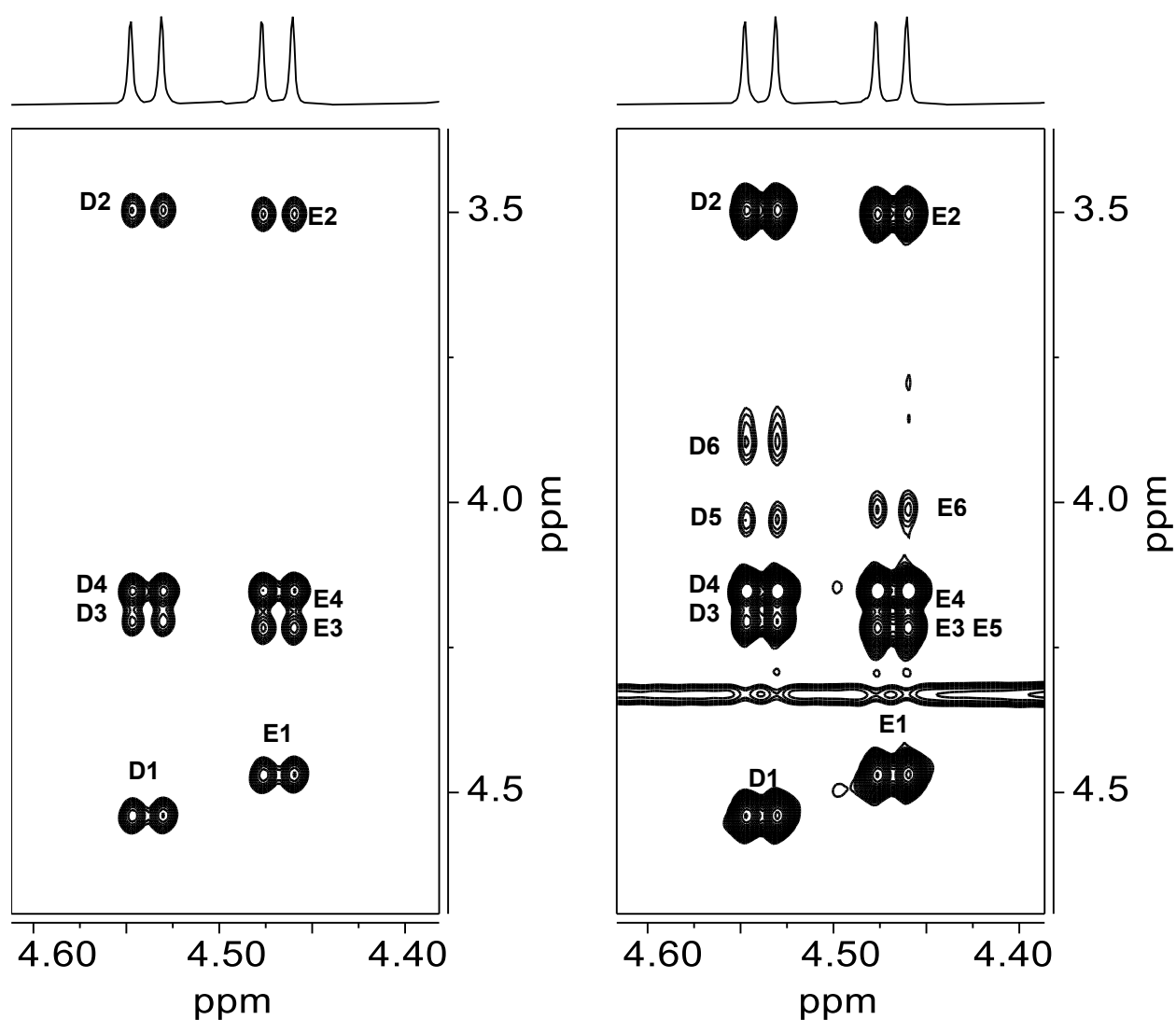
Supplementary Figure 4 | Analysis of the Smith degradation products of EpsA-O. **a**, Size-exclusion chromatogram of the Smith degradation products obtained using a Bio Gel P2 column: peaks are labelled **SD1** (eluted at V_0), **SD2** and **SD3**. Integration of the **SD2** and **SD3** peak areas gave a 2:1 ratio. **b**, ^1H NMR spectrum of fraction **SD3** showed that it is constituted by $\beta\text{-D-Galp}(3,4\text{-}S\text{-Pyr})\text{-(1}\rightarrow\text{Gro)}$ Assignments are indicated. (Gal = galactose; Pyr = pyruvyl substituent; Gro = glycerol)



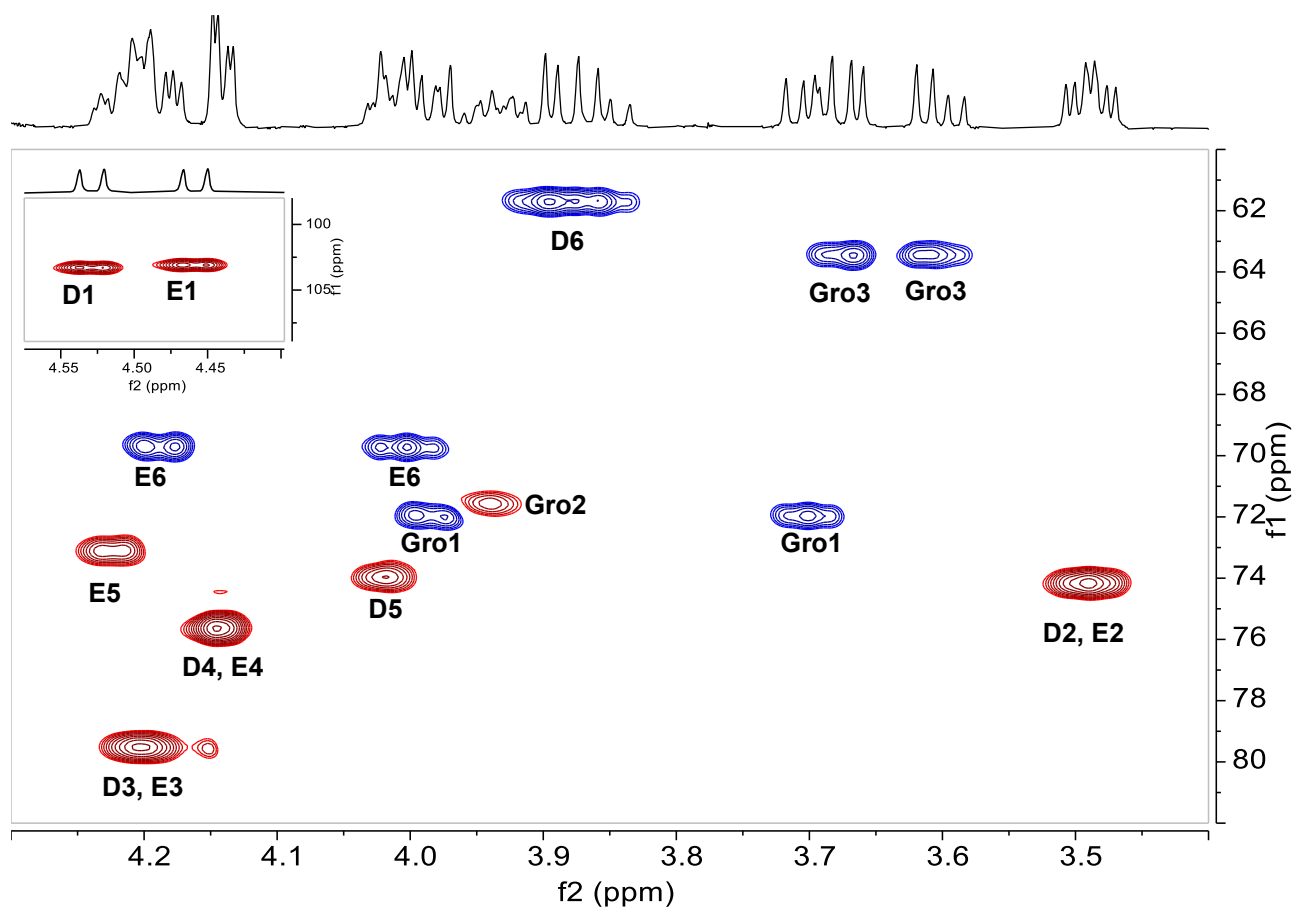
Supplementary Figure 5 | 1D NMR spectra of SD1 recorded at 70 °C and 500 MHz. 1D TOCSY (three top) and ¹H NMR (bottom) spectra of sample **SD1**. Peaks are labelled according to Table 1 (**A1** = H1 of residue **A**). For **B1** transfer of magnetization occurred up to **B6** (not shown).



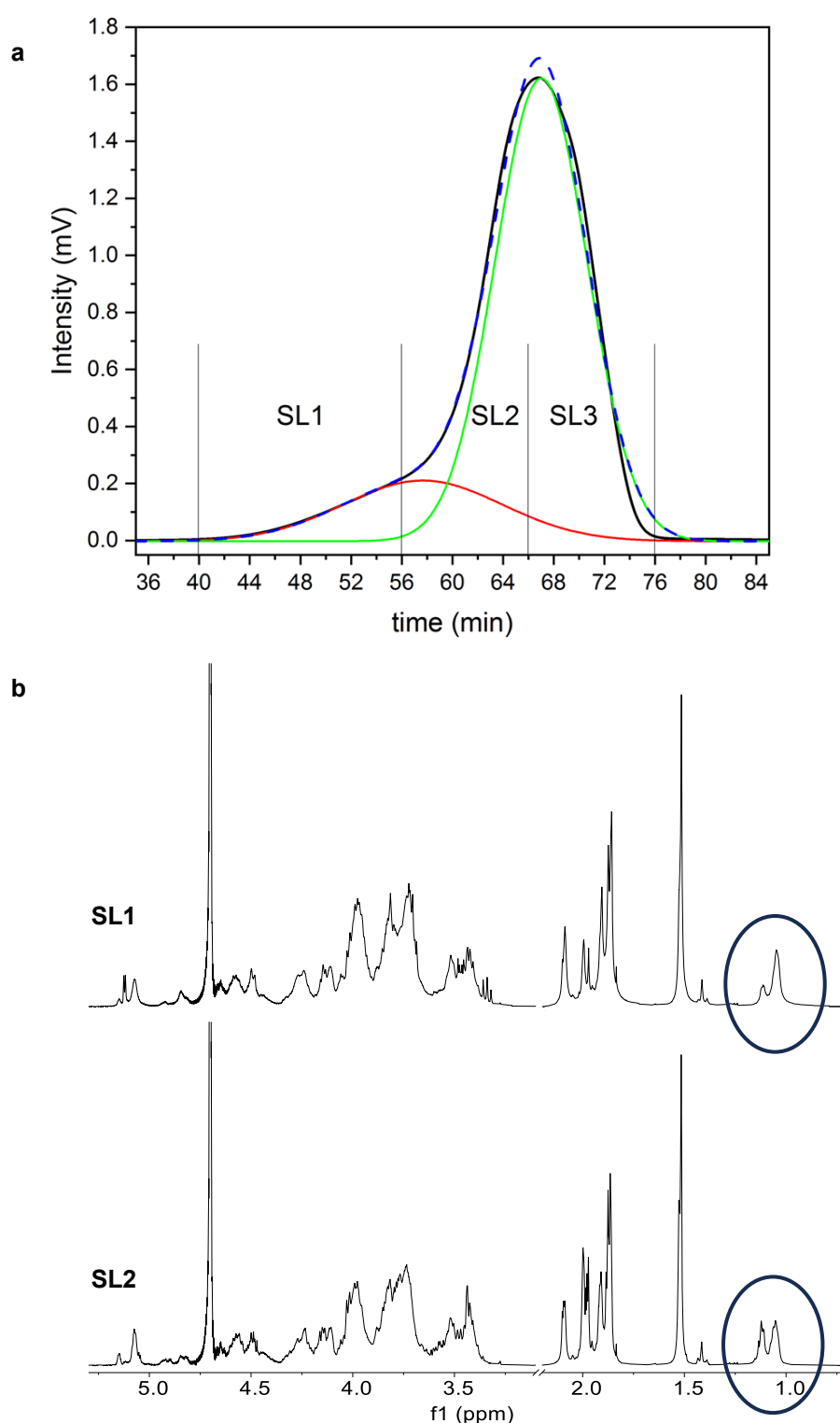
Supplementary Figure 6 | NOESY plot of sample SD1 recorded at 70 °C. Relevant inter-residues cross peaks are labelled (A1 = H1 of residue A).



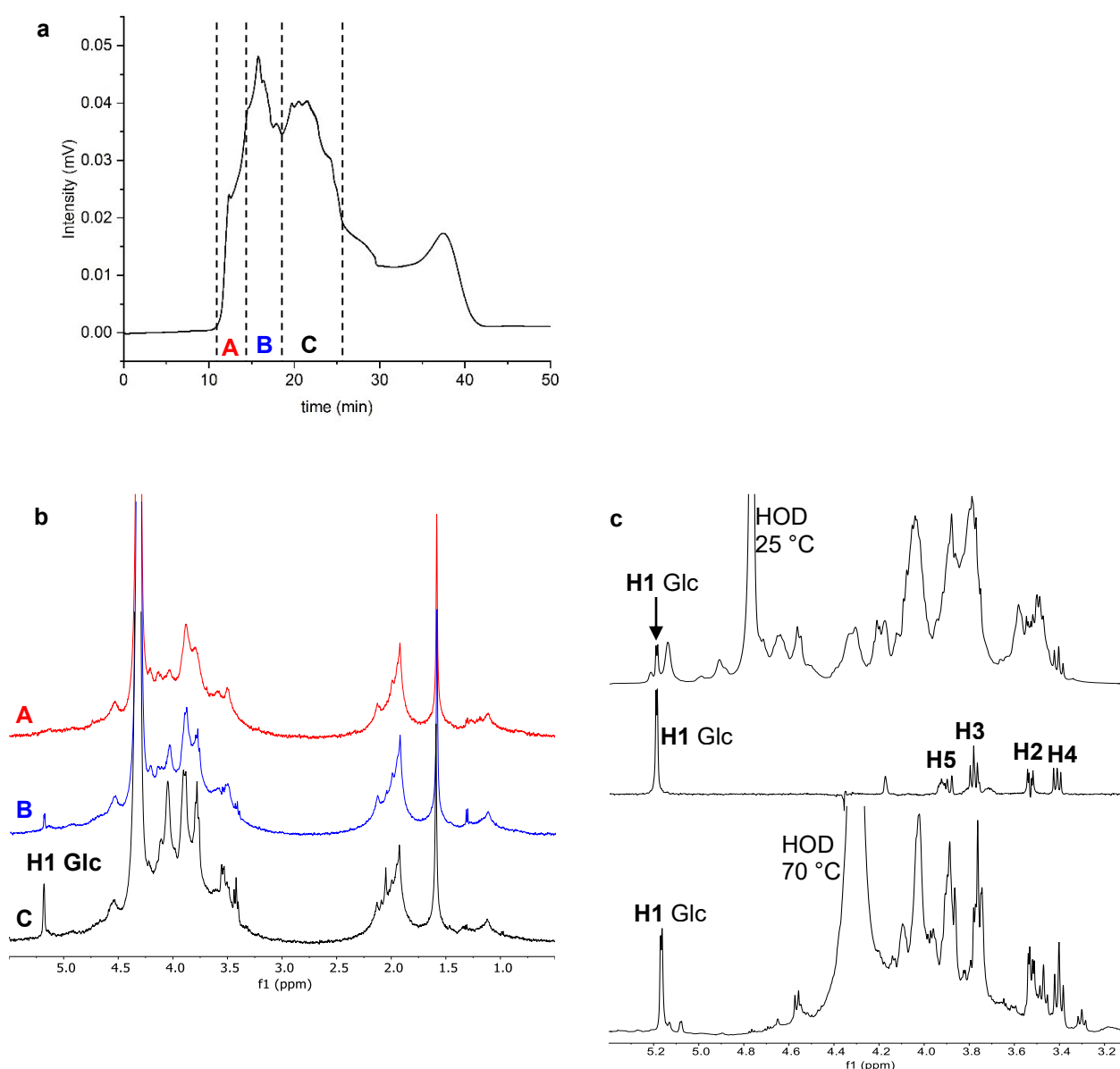
Supplementary Figure 7 | TOCSY plots of SD2 sample recorded at 70 °C. Cross peaks are labelled as in Table 2 (D1 = H1 of residue D). Left, low intensity plot; right, high intensity plot showing also H5 and H6 of residues D and E.



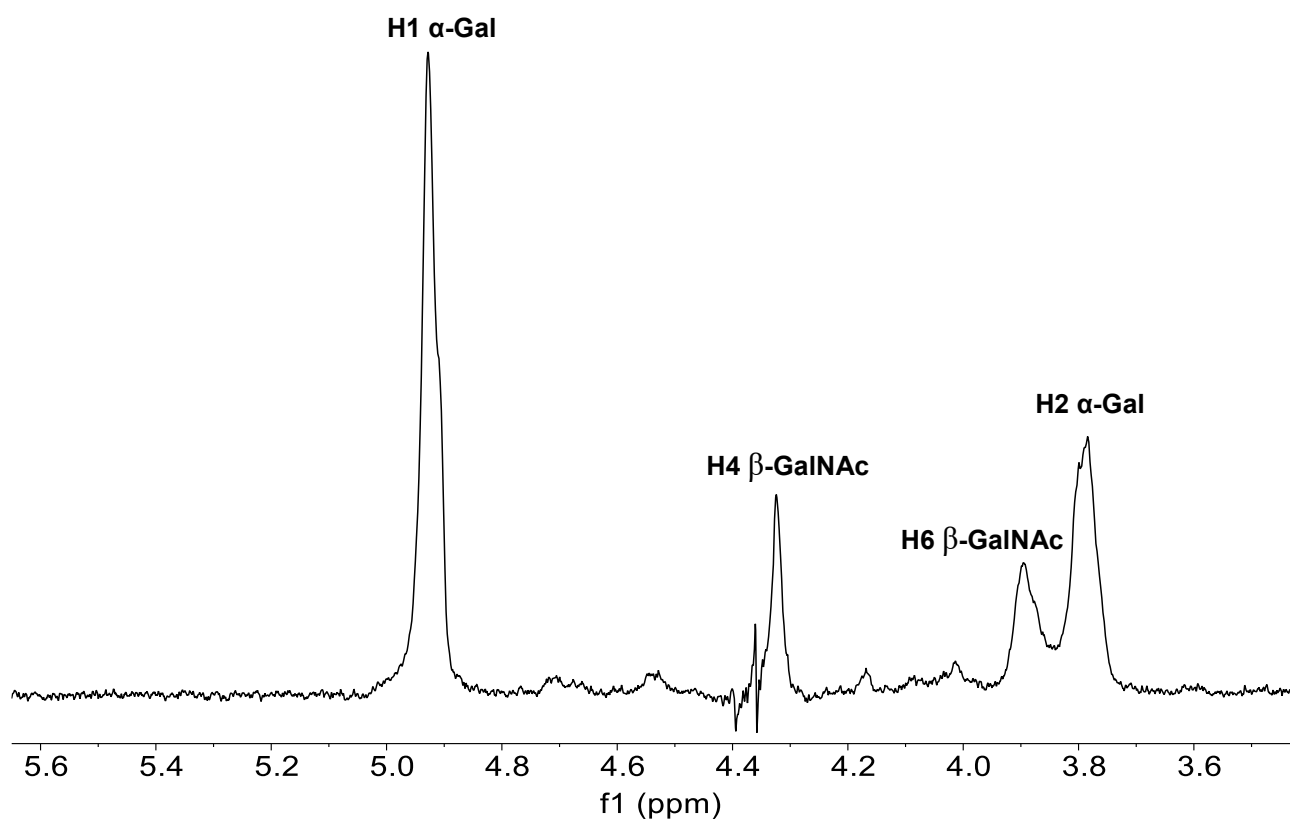
Supplementary Figure 8 | Ring region of the SD2 HSQC spectrum recorded at 70 °C. The anomeric region is reported in the inset. Cross peaks are labelled as in Table 2 (**D1** = H1/C1 of residue **D**).



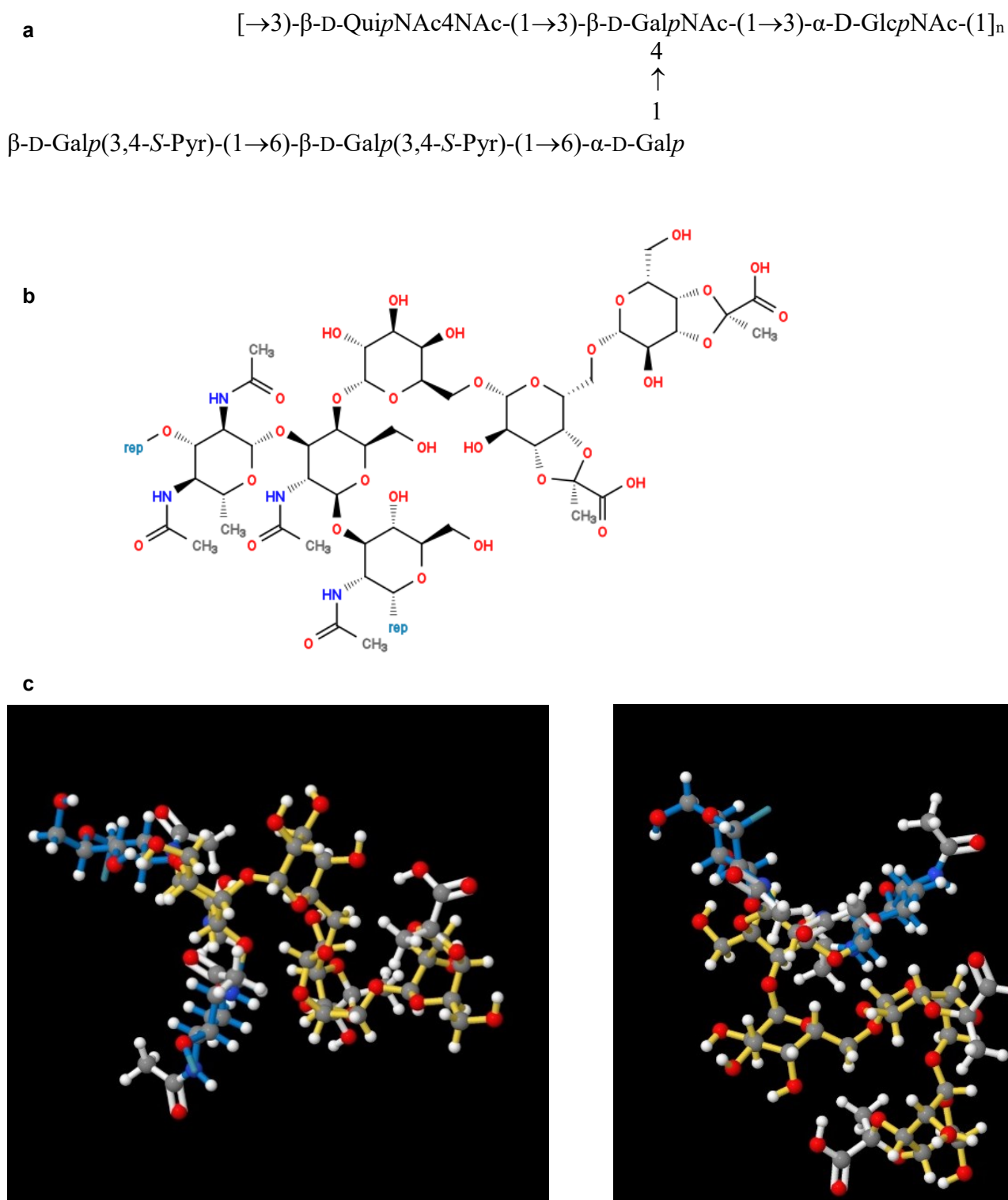
Supplementary Figure 9 | Analysis of EpsA-O sample after solvolysis. **a**, Chromatographic profile of EpsA-O after solvolysis in TFA (black) obtained on a S-300 HR column (1 cm i. d. X 45 cm) using a Bioline chromatographic system (see Methods). Elution was performed with 0.15 M NaCl at 1 mL/min. Green and red curves were obtained by applying a fitting procedure using a Gaussian function and hypothesizing two main peaks containing the degraded polysaccharide and salt; the dashed blue curve represents the cumulative fit; bars define which fractions were pooled together according to the fitting. **b**, ^1H NMR spectra of samples **SL1** and **SL2** recorded at 500 MHz and 25 °C. Circles indicate the methyl group (H6) resonance of QuiNAc4NAc which appears as two distinct peaks of equal intensity in **SL2** indicating a higher degree of structural heterogeneity of this sample.



Supplementary Figure 10 | Evidence that EpsA-O contains a minor glucan impurity. **a**, Size exclusion chromatography profile of the native EpsA-O separated on a Sephacryl S-300 HR column (1 cm i. d. X 45 cm) using a Bioline chromatographic system (see Methods). Elution was performed with 0.15 M NaCl at 1 mL/min and fractions were collected at 30 s interval. Fractions between bars were pooled (**A**, **B**, **C**), desalted, exchanged with D₂O and subjected to ¹H NMR spectroscopy. **b**, ¹H NMR spectra of the fractions **A**, **B** and **C** showing that the intensity of signal H1 of glucose (H1 Glc) increases as the fractions MM decreases, thus proving that it is not part of the EpsA-O polymer. **c**, ¹H NMR spectra of **SL1** sample (top), the polysaccharide produced by a $\Delta eps'$ mutant (bottom) where the resonances related to the EpsA-O are absent, while the signal of H1 of glucose is very intense, and the 1D TOCSY spectrum (center) obtained upon excitation of H1 of glucose at 5.19 ppm present in the **SL1** fraction.



Supplementary Figure 11 | 1D NOESY spectrum obtained with selective excitation of H1 of the 6-linked α -Gal at 4.92 ppm present in the SL1 fraction. Inter-residues n.o.e. effects were detected between H1 of α -Gal and H4 and H6 of β -GalNAc, thus confirming the sequence α -Gal(1-4) β -GalNAc. Assignments are indicated.



Supplementary Figure 12 | Repeating unit structure of EpsA-O. **a** structure of the EpsA-O repeating unit; **b** Haworth projection of the EpsA-O repeating unit; **c** two different views of the molecular structure of the repeating unit: blue bonds indicate Glc type rings; yellow bonds indicate Gal type rings. Structures b) and c) were obtained using the software available at <http://csdb.glycoscience.ru/database/> (Scherbinina et al., 2021).

Supplementary Data 1

Functions assignments of the genes *epsE*, *epsH*, *epsJ* and *epsF* in the *epsA-epsO* gene cluster of *B. subtilis*

The *epsA-epsO* gene cluster of *B. subtilis* 6051-HGW encodes 15 enzymes involved in the biosynthesis of the EpsA-O. The function of some of them has already been determined or predicted as follows:

- EpsA and EpsB are tyrosine kinases that regulate Eps production (Gerwig et al., 2014);
- EpsC, EpsM and EpsN are responsible for the biosynthesis of *N,N'*-diacetyl bacillosamine (QuiNAc4NAc) (Kaundinya et al., 2018);
- EpsK is a putative membrane transporter similar to the Wzx protein of *E. coli* (Roux et al., 2015);
- EpsL binds UDP-QuiNAc4NAc to the lipid acceptor undecaprenol phosphate (UndP) (Arnaouteli et al., 2021);
- EpsD adds the N-acetylglucosamine residue (Arbour et al., 2023) to the QuiNAc4NAc-UndPP.

Among the remaining genes, *epsE*, *epsH* and *epsJ* encode for CAZy GT-2 enzymes with an inverting mechanism of action, whereas *epsF* encodes for a GT-4 enzyme, with a retaining mechanism. Therefore, EpsF is the only GT which can bind the α -D-Galactose. This hypothesis was corroborated by the homology between EpsF and *Yersinia enterocolitica* galactosyltransferase WbcN (Supplementary Table 1), which catalyzes exactly the same reaction (Pinta et al., 2010). Moreover, *in silico* analysis showed that EpsF and WbcN belong to the GT4_CapH-like family protein, characterized by common motives thought to be involved in the formation of an “ATP-binding pocket” important for catalytic activity (Supplementary Figure 13). Among the remaining GTs, EpsE, EpsH, EpsJ, one must be an N-acetylgalactosaminyltransferase, and two galactosyltransferases. All these enzymes belong to the CAZy (<http://www.cazy.org/>) GT-2 family, characterized by two distinct domains: the donor and the acceptor binding sites at the -NH₂ and -COOH domains, respectively (Ünlügil and Rini, 2000). Comparison of their sequences evidenced high similarity between EpsH and EpsJ (mainly in the -NH₂ domain, whereas their -COOH domains were quite different), while EpsE showed low similarity with both GTs (Supplementary Figure 14). Moreover, EpsH and EpsJ -NH₂ domains showed homology with the same region of *B. anthracis* GtsE galactosyltransferase (Supplementary Table 1), thus suggesting that these enzymes are galactosyltransferases. Unfortunately, it was not possible to distinguish which enzyme catalyzes the addition of the first or the second galactose, because neither homologous -COOH domain sequences nor a known GT forming the linkage β -D-Galp-(1 $\rightarrow$ 6)- β -D-Galp or β -D-Galp-(1 $\rightarrow$ 6)- α -D-Galp was found in

available structural databases. By exclusion, EpsE was assigned to the N-acetylgalactosaminyltransferase that binds β -D-GalNAcp to C3 of α -D-GlcNAcp. Such hypothesis was also supported by its homology with the GalNAcp binding domain of *E. coli* K4 chondroitin polymerase (Supplementary Table 1) (Osawa et al., 2009). Finally, EpsG belongs to the GT-C superfamily, which are transmembrane proteins with 7 to 13 transmembrane domains exhibiting inverting glycosyltransferase activity starting from a polyisoprenol-linked donor substrate (Albuquerque et al., 2019). This finding, and the homology with the Wzy protein from *A. baumannii* (Kenyon et al., 2017) (Supplementary Table 1) suggest that EpsG is the polymerase that catalyzes the addition of RUs on the outer side of the membrane.

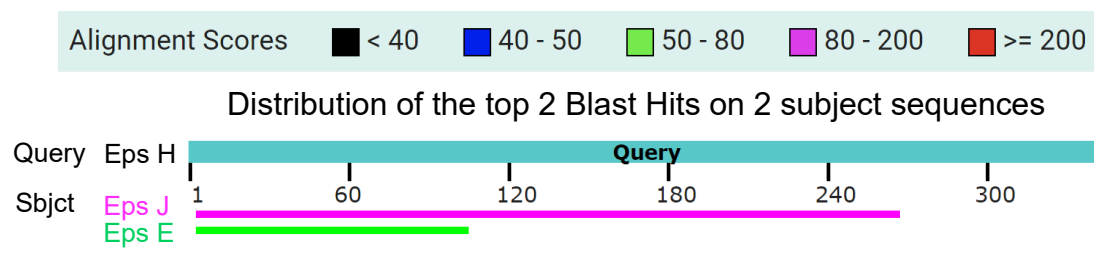
In the *epsA-epsO* cluster the genes responsible for the biosynthesis and activation of the common sugars Gal, GlcNAc and GalNAc are absent and they are probably located elsewhere, as it happens in other bacterial genomes.

Supplementary Table 1. Similarity level between Eps enzymes and proteins with known function obtained by BLAST alignment. Eps proteins: length and Accession numbers are indicated; Homologous proteins: name of the protein, bacterial species where it was described, their Accession Number and function are indicated.							
	Homologous protein		Similarity				
	Description	Enzymatic activity	E value	% Coverage	% Identities	% Positives	% Gaps
EpsE 278 aa WP_003244557.1	Condroitin synthase <i>E. coli</i> 2Z87_A	GT-2 GalNAcp transferase (Osawa et al., 2009)	3x10 ⁻¹⁶	54	28	45	4
EpsF 384 aa WP_003228257.1	WbcN <i>Y. enterocolitica</i> CAA87702	GT-4 Galp transferase (Pinta et al., 2010)	2x10 ⁻¹⁷	79	27	42	14
EpsH 344 aa WP_003243644.1	GtsE <i>B. anthracis</i> WP_000672693	GT-2 Galp transferase (Chateau)	1x10 ⁻¹²	45	33	48	5
EpsJ 344 aa WP_003228264.1			5x10 ⁻⁸	43	32	51	4
EpsI 358 aa WP_003228262.1	WcuL <i>K. pneumoniae</i> BAT23553.1	Pyruvyl transferase (Pan et al., 2015)	7x10 ⁻²²	93	25	44	12
EpsO 322 aa WP_003228270.1			4x10 ⁻⁷⁷	98	39	59	7
EpsG 367 aa WP_003228259.1	Wzy <i>A. baumannii</i>	Sugar polymerase (Kenyon et al., 2017)	7x10 ⁻¹⁸	56	31	49	15

CDD 340840: capsular polysaccharide biosynthesis glycosyltransferase CapH and similar proteins; cd03812
Feature 1: putative ADP-binding pocket (chemical binding site)

Feature 1		#			
WP_003228257.1	1	M	NSSQKRVLVHLS-GMNRGSAETMVMNLY-RKMDKSKVQDFLTyrNDPCAYD--EEILSLGGRLFYVPSIGQSNP	72	<i>B. subtilis</i> [EpsF]
CAA87702.1	1	-	-----MYEACHNVMIISLTQETLVR-----PND-----GIQLNELK-LDKAPFS-----L	39	<i>Y. enterocolitica</i> [wbcN]
AAA64647.1	1	M	-----IKVMHIFS-RMNRGSAETLMDTM-KLLNR---EFEFHVCAATSGKRKGELDDELESMGITIHYLDIKKFSFP	67	<i>S. aureus</i>
NP_391312.1	1	M	NSSQKRVLVHLS-GMNRGSAETMVMNLY-RKMDKSKVQDFLTyrNDPCAYD--EEILSLGGRLFYVPSIGQSNP	72	<i>B. subtilis</i> str. 168
EAN09128.1	1	-	---MKRILHFQG-RMGLGSAESFMNLY-RKIDRTNYQDFLIYEDYADVQDYHSEIERLGGRIFFVVPNPKK-NI	69	<i>E. faecium</i>
CAI34583.1	1	M	[11]GVDMIKILYVNGGMDRGVSTFMNLY-EKMHSEKIQIDFLVHTLSEGVDR--EDILNLGGKIFRVVARGK-NP	83	<i>S. pneumoniae</i>
Feature 1					
WP_003228257.1	73		LTfVRNVRNAIKENG-PFSAVHAHTDFQTGFIALAARLAGVPVRVCHSHNTS-WKTGFNWKDRLQLLVFRRLILANATAL	150	<i>B. subtilis</i> [EpsF]
CAA87702.1	40		FRGLFEVKKIIRFK--EDIVHSHMFHANLFAIRILRVFTKIPAHICIAHNTNeGSSSLMLAVKYTDKL-----ASLS	109	<i>Y. enterocolitica</i> [wbcN]
AAA64647.1	68		FKFIKLLK---KKNI---DVVHSHILEMSGLIQLLSFSANVRNRIHFRTSKdSKEQYNKIRKARNKVLKAIIEIFSTKI	141	<i>S. aureus</i>
NP_391312.1	73		LTfVRNVRNAIKENG-PFSAVHAHTDFQTGFIALAARLAGVPVRVCHSHNTS-WKTGFNWKDRLQLLVFRRLILANATAL	150	<i>B. subtilis</i> str. 168
EAN09128.1	70		FKYLIENVNLLK-KE-SFSIVHNQVYFGGGINLWAKNGIRQRIASHATE-DGKSONIVMNVLRKFLTKLLQONATDY	146	<i>E. faecium</i>
CAI34583.1	84		LKNYRQIKQIML-NG-NYDVVHAHADAGNRITLSIAKECDIPIRISHCHNT--NYTNKSILKKFLNEQFKKQIPRYATHL	159	<i>S. pneumoniae</i>
Feature 1					
WP_003228257.1	151		CACGEDAGRFLFGQSNMERERVHLLPNGIDLELFAPNGQAADDEKAARGIAADRLIIGHVAFHEVKNHAFLLKLAHLK	230	<i>B. subtilis</i> [EpsF]
CAA87702.1	110		TNVSQDAVDSFIHKGA SSTGRMIAVSNIGIDASQDFDSMDERKVKRSELGIFNDTPILLSVGLTEAKDYPNLLTAFSLLI	189	<i>Y. enterocolitica</i> [wbcN]
AAA64647.1	142		LYVSNIANRNLSMKLPKKHKTI-----YNG--FEISNINKNFKKEENSFIYVSPFIHTKNQLFLLDVIEILK	208	<i>S. aureus</i>
NP_391312.1	151		CACGEDAGRFLFGQSNMERERVHLLPNGIDLELFAPNGQAADDEKAARGIAADRLIIGHVAFHEVKNHAFLLKLAHLK	230	<i>B. subtilis</i> str. 168
EAN09128.1	147		LAVSQAGESLFOQNHFE-----IVHNGIDLELYSKNSEAKVNRKRELDISMSTFVVGNGIGLEKQKNQSYLLEIFLEIL	221	<i>E. faecium</i>
CAI34583.1	160		WACSEKAGEWLYGNHSEFE-----VIPNAIDVQKFYISPOLSKDLRKLNLNLE-NKFVIGHVGRFDYQKNHDFLKLVFTEFI	233	<i>S. pneumoniae</i>
Feature 1					
WP_003228257.1	231		ER---GIRFQVLVLAG--DGPLCGEIEEARQONLLSDVFLGTEER#HELMRTFDVFVMPSLYEGLPV#LVEAQASGLPC	305	<i>B. subtilis</i> [EpsF]
CAA87702.1	190		KDnslQSEFQFLFIYG--TGHLDGYLKNSKEFGIDKYVTLFGQRDDLDQMCADIDFVLSSEWEGFPLVITEAMACKKI I	267	<i>Y. enterocolitica</i> [wbcN]
AAA64647.1	209		KE---FNTNIEITFVGniQTDYGGKFLSIANERGLNKNIKVIGEVNNELDYLTSEYFLFPSELEGLPGALIEAHHNCIV	286	<i>S. aureus</i>
NP_391312.1	231		ER---GIRFQVLVLAG--DGPLCGEIEEARQONLLSDVFLGTEER#HELMRTFDVFVMPSLYEGLPV#LVEAQASGLPC	305	<i>B. subtilis</i> str. 168
EAN09128.1	222		KM---NSDSYLILVG--EGSLRNKLKAKAVELDIKRVKFLGERNDPELLSVMDVFVMPSLYEGLPV#GLEAQAKLKL	296	<i>E. faecium</i>
CAI34583.1	234		NE---REDAHLVLIG--KGLEEVIKKQANHLGILDKISFLGESSN#NELIN#VFDVGVPFSLFEGFSI#MVEMQVNGLPL	308	<i>S. pneumoniae</i>
WP_003228257.1	306		IISDSITEKVDAGLGLVTRLISLSEPISVWAETIARAAAAGRPKREFIKETLAQLGYDAQQNVGALLNVYNISTEKDHN#	384	<i>B. subtilis</i> [EpsF]
CAA87702.1	268		VATDAGGITFALGDC--GSIVPIKDPNSLSQAINKMIKLSDNEKEILGNKARERIOTN-SIEKILFELGCLFIINLKNNG	344	<i>Y. enterocolitica</i> [wbcN]
AAA64647.1	287		IISNINKENSEVNYQFKDSFELELIPKTWASTIKKLIS--RKKHISFNDSNV---FDINMTQELKEIYMSKTL---	355	<i>S. aureus</i>
NP_391312.1	306		IISDSITEKVDAGLGLVTRLISLSEPISVWAETIARAAAAGRPKREFIKETLAQLGYDAQQNVGALLNVYNISTEKDHN#	384	<i>B. subtilis</i> str. 168
EAN09128.1	297		VLADITISKDTKLT--NNVDFVSLDETPOEWAKKVIKEP-----YANKITQELKSYDVSYTTEQMESIY---SRDH---	361	<i>E. faecium</i>
CAI34583.1	309		VVSDNVPSINLTL--DNIRFLSLDETVKYWKCTILTEK--GRDITGAVDKIIA--KGYNLSDMVHKLTKTY---ERVVD#	379	<i>S. pneumoniae</i>

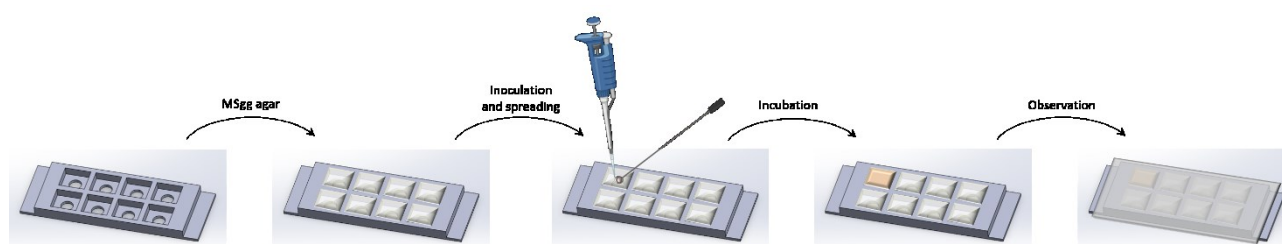
Supplementary Figure 13 | COBALT Multiple alignment of EpsF of *B. subtilis* and its homologous wbcN of *Yersinia enterocolitica* with some representatives of the GT4-CapH-like Family Protein. The conserved feature residue patterns forming the putative ADP binding pocket, necessary for the catalytic activity of these enzymes, are highlighted (yellow and #).



Supplementary Figure 14 | Graphic Summary of BLAST alignments of EpsH with EpsJ and EpsE. Alignment of EpsH with EpsJ gave an E-value of 2×10^{-42} , while alignment of EpsH with EpsE gave an E-value of 7×10^{-15} .

NP_593959.1	1	M[30]DLQTLK[5]TSPSSSTSVDKKKPLFTKSPRNSASCESTITLQSNLLFTYYKH[4]KKVALIGFPDHPNKGSSAI	109	S. pombe [Pvg1]
BAT23553.1	1	-----MHQNNLKNKLNELPFIKDK	39	K. pneumoniae K30 [WclU]
WP_003228270.1	1	-----MDSKHSMSLQKLSGLLD-VIPKQS-----	43	B. subtilis [EpsI]
WP_003228262.1	1	M SLQSLK INFAEWLLKVKYP-----SQYWLGAADQPVKAAHQ-----	58	B. subtilis [EpsQ]
NP_593959.1	110	YVAEKKLLDALNIE--VYITAQADYASSELKSII[4]RDEFALAFHGGGNFGDLYPDHQLRELVRDFFSFTTISFP	188	S. pombe [Pvg1]
BAT23553.1	40	YHGTFNFFKNNIN--VKLRNSNARYSVKEIEKYL NPNTVILMQGGGNFGDLYPEFQKFRQDVVSSFPDTRIIILP	113	K. pneumoniae K30 [WclU]
WP_003228270.1	44	MKGTEAFFKEHGIR--VRKRWNPDNFIPIGRKLDPNL ----IIVCQGGGNFGDLYPYQGFREKIVQTYPNHKIVILP	114	B. subtilis [EpsI]
WP_003228262.1	59	AYASKAFLEQEQYDFdIVEVDMKDIYKSAKSLIRSR HPEDVMVFIIGGNMGDLYR-YEWEWTRRFIITKTFHDYRVVQL	134	B. subtilis [EpsQ]
NP_593959.1	189	QSVWYNEQQL LEQASILYAENPNITLVTDRQSYGFAVDAFGKhNEVLLTPDIVFFMGPIPEIREATPIThdvLILA	265	S. pombe [Pvg1]
BAT23553.1	114	QTAFFENEHE LDKSVKAFGAHKDLIIFS DTRLSLSIFKKFTSQ---TFLMPDMAHSLWGTLPKSQKSGT---LYLI	184	K. pneumoniae K30 [WclU]
WP_003228270.1	115	QSIYFQNKDN LKRTAEIFS KANHLHIMTEKASYATAQAYFTT-NHIQLLPDMAHQLFPVPTQQPSNQK---LRFI	187	B. subtilis [EpsI]
WP_003228262.1	135	PATAHFSDTK[5]LKRAQKIYNAHPGLLLMADETTYQFMKQHFQE-KTILKQPDMLVLYLDRSKAPAEREGVY-----MC	210	B. subtilis [EpsQ]
NP_593959.1	266	RLDHEGGQHQHGAEDYYRD TLNAAALTYSVEDWLLWDPPV-----AQNPDSFDDRGQARYEAG---AEFLASARVV	333	S. pombe [Pvg1]
BAT23553.1	185	RNDKEINKVORRLEHKK[8]EDILTSKDLLMRKLCRKLDGIGVTNLVSLKNISNVIWNYTYKMVQRYALYFTSHEKV	269	K. pneumoniae K30 [WclU]
WP_003228270.1	188	RTDHEANQALQEHAAEAS YDWRTVLSASDRRTIAFLQTLNVLNKKAGNPLPIAYIWEKYSYDIYQKAIRFFSRYESV	264	B. subtilis [EpsI]
WP_003228262.1	211	LREDQESVLQEQQRNRVK AALCEEFGEEKSTFTTIGRRVSRDTR-----EHELEAL---WSKLQSAEAV	271	B. subtilis [EpsQ]
NP_593959.1	334	ITD LHAHILSTLMGIPHIVVENSQmGKITNYHN TWLHGCTLDGVSVVVDSVDKALSLLEWNEAGYF	401	S. pombe [Pvg1]
BAT23553.1	270	VTS MHGHIFSCLLSLPNDVIDNAY-GKNSGYFK ---EWTYDIDGTALLEE-----	316	K. pneumoniae K30 [WclU]
WP_003228270.1	265	ETS LHGHILSSLLQKENTVIDNSY-GKNANYFH TWMEGVPS---TRLIHASKKENLPAHM-----	322	B. subtilis [EpsI]
WP_003228262.1	272	VTD LHGMIFCALTGTPCVVIRSFD-HKVMGEYQ[8]KLIEHPPEPERVTAAVNELLTKETSRAGFPDVFYF[12]	358	B. subtilis [EpsQ]

Supplementary Figure 15 | COBALT Multiple alignment of EpsI and EpsO pyruvyltransferases of *B. subtilis* with the enzymes WclU of *Klebsiella pneumoniae* and Pvg1 of *Saccharomyces pombe*. Both enzymes catalyze the pyruvylation of a galactose residue. In Pvg1 the binding sites for galactose (balck circles) and for pyruvate (black diamond) have been identified by crystallography (Higuchi et al., 2016)



Supplementary Figure 16 | The workflow for preparing slides for observing biofilms under a microscope. The biofilm growth on solid MSgg agar was obtained in custom 3D printed 8-well chambered slides. The cover glass rested on the slide without touching the biofilm surface.

Supplementary References

- Albuquerque-Wendt, A., Hütte, H. J., Buettner, F. F. R., Routier, F. H. & Bakker H. Membrane Topological Model of Glycosyltransferases of the GT-C Superfamily. *Int. J. Mol. Sci.* **20**, 4842; 10.3390/ijms20194842 (2019).
- Arbour, C. A. et al. Defining Early Steps in *B. subtilis* Biofilm Biosynthesis. *mBio* **14**:e00948-23. <https://doi.org/10.1128/mbio.00948-23>.
- Arnauteli, 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). <https://doi.org/10.1038/s41579-021-00540-9>.
- Chateau, A. et al. Distinct pathways carry out α and β galactosylation of secondary cell wall polysaccharide in *Bacillus anthracis*. *J. Bacteriol.* **202**:e00191-20. (2020). <https://doi.org/10.1128/JB.00191-20>.
- Gerwig, J., Kiley, T. B., Gunka, K., Stanley-Wall, N. & Stulke, J. The protein tyrosine kinases EpsB and PtkA differentially affect biofilm formation in *Bacillus subtilis*. *Microbiology* **160**, 682-691 (2014).
- 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). <https://doi.org/10.1038/srep26349>
- Kaundinya, C. R., Savithri, H. S., Rao, K. K. & Balaji, P. V. EpsM from *Bacillus subtilis* 168 has UDP-2,4,6-trideoxy-2-acetamido-4-amino glucose acetyltransferase activity in vitro. *Biochem. Biophys. Res. Commun.* **505**, 1057-1062 (2018).
- Kenyon, J. J. et al. The KL24 gene cluster and a genomic island encoding a Wzy polymerase contribute genes needed for synthesis of the K24 capsular polysaccharide by the multiply antibiotic resistant *Acinetobacter baumannii* isolate RCH51. *Microbiology* **163**, 355-363 (2017) <https://doi.org/10.1099/mic.0.000430>.
- Osawa, T. et al. Crystal structure of chondroitin polymerase from *Escherichia coli* K4. *Biochem. Biophys. Res. Comm.* **378**, 10-14 (2009). <https://doi.org/10.1016/j.bbrc.2008.08.121>
- Pan, Y. J. et al. Genetic analysis of capsular polysaccharide synthesis gene clusters in 79 capsular types of *Klebsiella* spp. *Sci. Rep.* **5**, 15573; [10.1038/srep15573](https://doi.org/10.1038/srep15573) (2015).
- Pinta, E., et al. Characterization of the Six Glycosyltransferases Involved in the Biosynthesis of *Yersinia enterocolitica* Serotype O:3 Lipopolysaccharide Outer Core. *J. Biol. Chem.* **285**, 28333-28342 (2010) <https://doi.org/10.1074/jbc.M110.111336>.
- 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). <https://doi.org/10.1074/jbc.M115.648709>.
- Scherbinina, S. I., Frank, M. & Toukach, P. V. Carbohydrate Structure Database oligosaccharide conformation tool. *Glycobiology*, **32**, 460-468 (2022). <https://doi.org/10.1093/glycob/cwac011>.
- Ünlügil, U. M., & Rini, J. M. Glycosyltransferase structure and mechanism. *Curr. Opin. Struct. Biol.* **10**, 510-517 (2000). [https://doi.org/10.1016/s0959-440x\(00\)00124-x](https://doi.org/10.1016/s0959-440x(00)00124-x).