

SUPPLEMENTARY INFORMATION

Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic *Bacteroides*

Ana S. Luis^{1¶}, Jonathon Briggs^{1¶}, Xiaoyang Zhang^{1¶}, Benjamin Farnell^{2,3}, Didier Ndeh¹, Aurore Labourel¹, Arnaud Baslé¹, Alan Cartmell¹, Nicolas Terrapon⁴, Katherine Stott⁵, Elisabeth C. Lowe¹, Richard McLean², Kaitlyn Shearer², Julia Schüchel⁶, Immacolata Venditto¹, Marie-Christine Ralet⁷, Bernard Henrissat^{4,8,9}, Eric C. Martens¹⁰, Steven C. Mosimann³, D. Wade Abbott^{2,3*} and Harry J. Gilbert^{1,*}

¹*Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne NE2 4HH, UK.* ²*Lethbridge Research and Development Centre, Agriculture and Agri-Food Canada, Lethbridge, Alberta, Canada, T1J 4B1.* ³*Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, Alberta, Canada, T1K 6T5.*

⁴*Architecture et Fonction des Macromolécules Biologiques, Centre National de la Recherche Scientifique (CNRS), Aix-Marseille University, F-13288 Marseille, France.*

⁵*Department of Biochemistry, University of Cambridge, Cambridge, U.K.*

⁶*Department of Plant and Environmental Sciences, Faculty of Science, University of Copenhagen, Copenhagen, Denmark.* ⁷*INRA, UR1268 Biopolymères Interactions Assemblages, 44300 Nantes, France.* ⁸*INRA, USC 1408 AFMB, F-13288 Marseille, France.* ⁹*Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia.* ¹⁰*Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI, USA.* ¶Authors contributed equally. *e.mail:

wade.abbott@agr.gc.ca; harry.gilbert@ncl.ac.uk,

Supplementary Discussion

Proteins encoded by RGI-PUL

GH2 and GH35 enzymes: RGI-PUL encodes one GH35 (BT4160) and three GH2 (BT4151, BT4156 and BT4181) proteins. Each protein is predicted to be a β 1,4-galactosidase that removes remnants of galactan side chains appended to the RGI backbone via glycosidic linkage to rhamnose. The data (**Supplementary Table 1** and **Supplementary Fig. 9**) showed that BT4160 displayed significant activity against oligosaccharides with a degree of polymerization >3 displaying a 10-fold preference for galactotriose over galactoheptaose, and was the only enzyme that released galactose from galactan. BT4151 and BT4156 were, essentially, only active on RGI. Thus, for long galactan remnants degradation is initiated by BT4160 with BT4156 or BT4151 removing the final galactose attached to the RGI backbone. The importance of BT4156 and BT4151 is illustrated by the observation that when the enzyme is omitted from RGI degrading enzyme cocktails there was a significant reduction in the release of rhamnose and GalA from potato RGI (RGI-P), while BT4160 had little impact on the depolymerizing process. This indicates that the RGI-P used here contains predominantly short galactan stubs, likely reflecting the extensive enzyme treatment of potato galactan to generate predominantly the RGI backbone. In an *in vivo* context, however, the remaining galactan chains are likely to have a more variable degree of polymerization explaining the requirement for BT4160. The third GH2 enzyme encoded by RGI-PUL, BT4181, appeared to function as a β -D-glucuronidase against sugar beet arabinan in which GalA in the RGI backbone is known to be decorated at O3 with β -D-GlcA¹. This is surprising as the enzyme

contains a lysine in place of the invariant glutamate that functions as the catalytic acid-base. However, a number of *Bacteroides* GH2 enzymes also contain the Glu to Lys substitution suggesting that in a cohort of glycoside hydrolases this amino acid replacement does not influence activity. Indeed, in GH127 the catalytic acid-base glutamate is replaced by glutamine in an aceric acidase. The substrate contains a carboxylate and it was suggested that this group may contribute to substrate assisted catalysis².

BT4158 was shown to hydrolyse 4-nitrophenyl-acetate and is thus an acetyl esterase (**Supplementary Table 1**). The enzyme appeared to perform a similar role to BT4156 and BT4151. By removing acetyl groups from the backbone of RGI, likely appended to O3 of GalA residues, BT4158 relieved the steric block imposed by these decorations enabling enzymatic depolymerisation of this pectic glycan (**Supplementary Figure 9**). BT4158 was shown to be a general acting carbohydrate esterase as it was active on esterified birchwood xylan and glucogalactomannan in addition to RGI.

To explore the mechanism of catalysis and substrate recognition by BT4170, a key component of the RGI degrading apparatus, the crystal structure of the enzyme was determined in complex with reaction products. The data, **Supplementary Fig. 3**, showed that the enzyme displays a right-handed parallel β -helix fold with three short parallel β -sheets (PB1, PB2 and PB3) and 10 turns. When the structure of BT4170 was overlaid with Pel9A, **Supplementary Fig. 3³**, the two key components of the active site of the two enzymes were conserved. These comprise a calcium binding site on coils 5 and 6 (Asp215, Asp246, Asp247 and Asp250; designated Ca-1), and a lysine on coil 7, Lys285, which functions as the catalytic base abstracting a proton from C5³. In BT4170 Asp280, Gly212 and Asp246 comprise a second calcium binding site (Ca-2). Electron density for two oligosaccharides were present in the substrate binding site. The density in the negative subsites (-3 to -1) was built as α -L-Rha- α 1,4-D-GalA- α 1,2-L-Rha and in the positive subsites (+1 to +3) as Δ 4,5-GalA- α 1,2-L-Rha- α 1,4-D-GalA (**Supplementary Fig. 3**).

Substrate recognition at the negative subsites was mediated by hydrogen bonds between Gln155 and O3 and O4 of L-Rha (-1 subsite) and Arg125 and His153 with the carboxyl group of the D-GalA at (-2 subsite), **Supplementary Fig. 3**. Three residues, Tyr184, Asn213 and His220, make apolar contacts with the trisaccharide. Mutational analysis showed that Arg125 and His220 dominated substrate binding in the negative subsites, **Supplementary Table 6**. Substrate recognition at the distal positive subsites was through hydrogen bonds between Asn324 and O4 of L-Rha (+2 subsite), Arg319 and O5 of D-GalA; and Tyr322 and His325 with O6 of D-GalA in the +3 subsite. Mutation of these polar residues showed that they had limited impact on activity and thus make a minor contribution to substrate binding, **Supplementary Table 6**.

A central feature of substrate recognition in the active site (+1 subsite) are two calcium ions, Ca-1 and Ca-2, that make salt bridges with O δ 2 of the carboxylate of the bound Δ 4,5-GalA, **Supplementary Fig. 3**. Only the Ca-1 binding site is conserved in PL9 lyases. Ile252 makes apolar contacts with Δ 4,5-GalA at +1 and L-Rha at +2. I252A substitution decreased activity by ~500-fold, **Supplementary Table 6**, suggesting that the isoleucine contributes to substrate binding likely through

its interaction with the uronic acid. Substitution of any of the carboxylates that form the Ca-1 binding site completely inactivated BT4170, indicating that this calcium plays an essential role in substrate binding and/or catalysis. The 100-fold reduction in activity displayed by the D280A mutant suggests that this residue contributes to Ca-binding and that the Ca-2, in turn, plays a role in catalysis.

All PLs including those in PL9 are predicted to cleave their target glycosidic bond through a β -elimination reaction. A catalytic base extracts a proton from C5 with stabilization of the resulting anion by charge delocalization at the C-6 carbonyl group through polar interactions with calcium⁴. This leads to lytic cleavage of the O4:C4 bond that is facilitated by proton donation from a catalytic acid to the glycosidic oxygen (O4). In BT4170 the only candidate general base that can extract the H5 proton is Lys285, consistent with a pH optimum of 9 and the complete loss of activity by alanine substitution of the lysine **Supplementary Table 6**. The identity of the catalytic acid that protonates O4 to encourage leaving group departure is unclear.

Inspection of the overlay of BT4170 and Pel9A, **Supplementary Fig. 3**, provides insight into the specificity determinants of these conserved, but functionally distinct PLs. Apart from the catalytic apparatus, there is little conservation of substrate binding residues. Of particular note is the substitution of the critical substrate binding residues Arg125 and His220 with a glutamine (Gln130) and lysine (Lys214), respectively. Indeed the lysine in Pel9A is likely to make substantial steric clashes with C6 of the L-Rhap in the -1 subsite, in addition to creating a highly unfavorable polar environment.

The GH106 α -L-rhamnosidase, BT4145, releases L-Rha from the non-reducing end of AM-RG-I. The activity of the enzyme did not increase against oligosaccharides >2 indicating that this enzyme has only a -1 and +1 subsite that bind Rha and GalA, respectively. In the presence of EDTA the enzyme is inactive indicating that it is metal dependent, similar to the recently characterized GH106 α -L-rhamnosidase BT0986². Site-directed mutagenesis was used to evaluate the catalytic importance of putative active site residues. The selection of amino acids was based on sequence conservation within the GH106 family, and BT0986 in particular as a crystal structure of this enzyme is available and the active site has been mapped. Alanine substitution of the proposed general acid/base pair (Glu253/Glu395) and the calcium binding residues (Asp250, Gln328 and Glu394), based on conservation of the catalytic apparatus of BT0986, inactivated the enzyme, **Supplementary Table 6**. These data show that the predicted catalytic apparatus is conserved between GH106 enzymes and suggest that the calcium ion has a major role in catalytic activity.

The GH106 enzymes BT0986 and BT4145 target L-Rha- α 1,2-L-Arap in Chain B of RGII² and L-Rha- α 1,4-D-GalA linkages in the backbone of RGI, respectively. It is expected, therefore, that these enzymes will have different +1 subsites. Alignment of the two sequences showed that BT4145 does not contain a sequence equivalent to Pro179 to Ser364 of BT0986. In the crystal structure this sequence corresponds to the C domain, which contributes to the leaving group (positive) subsites of the enzyme. The lack of this domain in BT4145 is consistent with the different specificities for substrate bound in the positive subsites.

Protein encoded by Ara-PUL

Although BT0349 displays distant sequence identity to GH127 enzymes, it is too divergent to be included in this family. Thus, BT0349 comprises a GH family that was previously unknown. To explore the specificity of the enzyme sugar beet arabinan was digested to completion with all the α -L-arabinanases encoded by Ara-PUL. Mass spectrometry showed that the limit product was an arabinotetraose. This oligosaccharide was analysed by NMR spectroscopy in order to deduce the anomeric configuration and linkages, **Supplementary Fig. 7**. Chemical-shift assignments were obtained using 2D ^1H - ^1H TOCSY, ROESY and DQFCOSY alongside 2D ^{13}C HSQC and HSQC-TOCSY experiments. The assignment was complete and is shown in **Supplementary Table 7**. Four arabinofuranose residues were detected. ^{13}C -1 chemical shifts revealed the internal Araf to be in the α anomeric configuration (δ c. 108 ppm) while the non-reducing-end residue was in the β anomeric configuration (δ c. 102 ppm). All the linkages were clear from intense NOE crosspeaks in the ROESY spectrum and downfield ^{13}C shifts of the relevant linked carbons. The information taken together revealed a sequence of Araf- β 1,3-Araf- α 1,3-Araf- α 1,5-Araf, **Supplementary Fig. 7**. BT0349 released the β 1,3-linked arabinose from the tetrasacchride, and the resulting arabinotriose was now accessible to exo-hydrolysis by the α 1,3-L-arabinofuranosidase BT0368. This indicates that the arabinotetraose contains a β -arabinose at the non-reducing terminus, which blocks the action of exo-acting α -L-arabinofuranosidases. BT0349 was also shown to hydrolyse β 1,2-L-arabinobiose demonstrating that the enzyme is a non-specific β -L-arabinofuranosidase.

The crystal structure of BT0349 was solved using SAD methods to a resolution of 2.1 Å. The 804 residue protein consists of four domains (**Supplementary Fig. 8**). The catalytic domain, located at the N-terminus, comprises an $(\alpha/\alpha)_6$ barrel (residues His25 to Ile432), and is followed by three β -sandwich domains defined as 1 (Tyr433 to His535), 2 (Val540 to Ala647) and 3 (Val680 to Lys804). Although the linkers between the domains are generally short, β -sandwich domains 2 and 3 are connected via a 23 residue α -helix. A Dali search for functionally relevant structural homologs of BT0349 revealed that the GH127 β -L-arabinofuranosidase BLLJ_0211 (PDB code [3WKX](#)) was the closest homolog with a Z-score of 31, root mean square deviation of 3.3 Å, and sequence identity of 16% over 526 aligned residues. An overlay of the enzymes revealed that the respective active site pockets, which both contain an arabinofuranose residue, are located in the centre of the $(\alpha/\alpha)_6$ barrel catalytic domain. In close proximity to the arabinose in BT0349 is a zinc atom that is co-ordinated with three cysteines and a glutamate. The arabinose makes numerous polar contacts with amino acids that line the active site pocket and hydrophobic interactions with a tyrosine at the base of the pocket. The zinc binding site and the proposed catalytic apparatus of GH127 enzymes are completely conserved in BT0349. There are, however, some structural differences. While the hydrophobic platform is invariant and O1 and O2 of the arabinose make identical polar contacts with the respective active sites, the amino acids that form hydrogen bonds with O3 and O5 of the furanose sugar are not conserved between the GH127 enzymes and BT0349. Furthermore, the GH127 enzymes lack β -sandwich domain 3 evident in BT0349. It is likely that the location of this C-terminal domain in crystal is not biologically relevant, as it is positioned over the active site burying the bound arabinose and thus preventing catalysis. Thus, it is likely that the location of this domain is highly flexible in solution. Although the *B. thetaiotaomicron*

arabinofuranosidase displays very low sequence identity with GH127 enzymes, the conservation of the catalytic apparatus (which infers a conserved mechanism), the fold of the catalytic domain and the stereochemistry of the target glycosidic bond, indicates an evolutionary relatedness between the two enzymes. We propose, therefore, that the family founded by BT0349 should form a clan with GH127.

Protein encoded by *B. ovatus* Gal-PUL

The BACOVA_05493 gene is only present in some *Bacteroides* species. As described in the main paper the enzyme was shown to be an β 1,4-galactosidase that targets long β 1,4-galactooligosacchrides displaying significant activity against galactan. Sequence alignments indicate that BACOVA_05493 establishes a previously undiscovered GH family that is a member of the GH-A clan, which enabled the catalytic residues (Glu203 and Glu300) to be predicted. The proposed role of Glu203 and Glu300 as the catalytic acid/base and nucleophile, respectively, is consistent with the observation that the mutants E300A and E203A displayed no catalytic activity, **Supplementary Table 6**. Further support for the location of BACOVA_05493 in clan GH-A is provided by analysis of the reaction by NMR. The data showed that the β -anomer of galactose was released from β 1,4-galactotriose within two min but after one hour both the β - and α -anomer of galactose was evident indicating that mutarotation had occurred (**Supplementary Fig. 16**). These data show that BACOVA_05493 cleaves galactosidic bonds through a double displacement mechanism leading to retention of anomeric configuration, as observed in all members of the GH-A clan.

Bacteroides Mutant Growths

To assess the contribution to of each PUL component to the utilisation of the target glycan knockout or inactivation mutations were generated in *B. thetaiotaomicron*. These variants were then exposed to the relevant glycan in minimal media and growth monitored.

In the arabinan PUL *bt0362*, *bt0364* and *bt0365* were deleted from the *B. thetaiotaomicron* genome generating the mutants $\Delta bt0362$, $\Delta bt0364$ and $\Delta bt0365$. Despite an initial growth rate analogous to the wild type *B. thetaiotaomicron*, $\Delta bt0365$ was unable to achieve as high a cell density (**Supplementary Fig. 5**). Deletion of the SGBP, BT_0365, is not catastrophic for arabinan utilisation, indicating the pair of GH43 enzymes at the cell surface are capable of adequate activity without the substrate capturing function of the BT0365.

The Gal-PUL SusD homologue was deleted in $\Delta bt4670$. When grown on galactan the mutant was capable of utilising galactan at a similar rate to the wild type bacterium and achieved similar final cell density (**Supplementary Fig. 5**). This indicates the galactan SusD-homologue is not integral to galactooligosaccharide transport, implying the associated SusC-homologue has a greater role in substrate binding prior to transport into the periplasm than in previously published glycan utilisation systems⁵.

Of the β -galactosidases encoded by RGI-PUL, BT4156 was shown, through biochemical analysis, to have the greatest impact on the degradation of the RGI backbone. Deletion of *bt4156*, however, had a relatively modest impact on growth on RGI. The $\Delta bt4156$ mutant displayed an increased lag phase and slightly reduced

exponential growth rate on RGI when compared to the wild type *B. thetaiotaomicron* (**Supplementary Fig. 5**). The steric block created by the galactose residues on the RGI backbone are either removed by other enzymes at a much slower rate or full deconstruction of the RGI oligosaccharides is not required for growth. The RGI-PUL encodes two SusC_H/SusD_H pairs (BT4164/BT4165 and BT4168/BT4169). To explore which of these two protein pairs import RGI-derived oligosaccharides the respective gene pairs were deleted. Surprisingly the data showed that the mutants $\Delta bt4164/\Delta bt4165$ and $\Delta bt4168/\Delta bt4169$ were unable to grow on RGI-AM suggested that the two importers acted in synergy to transport the glycans. A recent study has shown that SusC_H/SusD_H pairs form dimers⁶. It is possible, that in the RGI degrading system, rather than forming homodimers, the two SusC_H/SusD_H interact and thus generate heterodimers.

Supplementary Table 1 Kinetics of enzymes encoded by RGI-PUL

PL9 and PL11					
Enzyme	Activity	Substrate	k_{cat} (min ⁻¹)	K_M (mg/ml)	k_{cat}/K_M (min ⁻¹ mg ⁻¹ ml)
BT4170 (PL9)	RGI lyase	RGI-AM	1.7(± 0.62) x 10 ³	2.3(± 0.3) x 10 ⁻²	7.1(± 0.1) x 10 ⁴
		RGI-P	Trace activity		
BT4175 (PL11)	RGI lyase	RGI-AM	6.6(± 0.3) x 10 ²	3.5(± 0.4) x 10 ⁻²	1.2(± 0.3) x 10 ⁴
		RGI-P	2.2(± 0.09) x 10 ²	1.91(± 0.3) x 10 ⁻¹	1.2 (± 0.2) x 10 ³
BT4183 (PL9)	RGI lyase	RGI-AM	1.7 (± 0.05) x 10 ²	1.5 (± 0.2) x 10 ⁻²	1.1(± 0.2) x 10 ⁴
		RGI-P	Inactive		
GH106					
Enzyme	Activity	Substrate	k_{cat}/K_M (min ⁻¹ M ⁻¹)		
BT4145	RGI-specific α-L-rhamnosidase	α-L-Rha-α1,4-D-GalA	7.1(± 0.03) x 10 ⁷		
		(α1,2-L-Rha-α1,4-D-GalA) ₂	8.7 x 10 ⁶		
		(α1,2-L-Rha-α1,4-D-GalA) ₃	1.5 x 10 ⁷		
GH28					
Enzyme	Activity	Substrate	k_{cat}/K_M (min ⁻¹ M ⁻¹)		
BT4146	RGI disaccharide specific α-D-galacturonidase	α-D-GalA-α1,2-L-Rha	2.6(± 0.2) x 10 ⁵		
		(α1,4-D-GalA-α1,2-L-Rha) ₂	Inactive		
		(α1,4-D-GalA) ₂			
BT4153	RGI specific α-D-galacturonidase	α-D-GalA-α1,2-L-Rha	2.0(± 0.1) x 10 ⁶		
		(α1,4-D-GalA-α1,2-L-Rha) ₂	4.5(± 0.03) x 10 ⁶		
		(α1,4-D-GalA) ₂	Inactive		
BT4155	HG specific α-D-galacturonidase	(α1,4-D-GalA) ₂	3.2(± 0.2) x 10 ⁵		
		(α1,4-D-GalA) ₃	2.5(± 0.1) x 10 ⁵		
		α-D-GalA-α1,2-L-Rha	9.2(± 0.07) x 10 ⁵		
		(α1,4-D-GalA-α1,2-L-Rha) ₂	Inactive		
		Pectin DE 30%	9.4 ± 0.32*		
		RGI-P	2.7 ± 0.17*		
		RGI-AM	Inactive		

GH28 (cont)						
Enzyme	Activity	Substrate	k_{cat}/K_M ($\text{min}^{-1} \text{M}^{-1}$)			
BT4149	RGI specific α -D- galacturonidase	α -D-GalA- α 1,2-L-Rha	2.8(\pm 0.1) $\times 10^4$			
		(α 1,4-D-GalA- α 1,2-L-Rha) ₂	2.0(\pm 0.3) $\times 10^8$			
		(α 1,4-D-GalA- α 1,2-L-Rha) ₃	5.2(\pm 0.7) $\times 10^7$			
		(α 1,4-D-GalA- α 1,2-L-Rha) ₄	4.8(\pm 0.8) $\times 10^7$			
		(α 1,4-D-GalA-) ₂	Inactive			
		(α 1,4-D-GalA-) ₃				
		RGI-P				
		RGI-AM				
		Branched Arabinan				
		PGA orange				
		Pectin DE 30%				
GH105						
Enzyme	Activity	Substrate	k_{cat} (min^{-1})	K_M (M)	k_{cat}/K_M ($\text{min}^{-1} \text{M}^{-1}$)	% Activity**
BT4174	RGI-specific unsaturated- galacturonidase	Δ 4,5-GalA- α 1,2-L-Rha- α 1,4-D-GalA- α 1,2-L-Rha	27.1	3.3×10^{-4}	8.3×10^4	100
		Δ 4,5-GalA- α 1,2-L-Rha				1.1
BT4176	RGI-disaccharide specific unsaturated- galacturonidase	Δ 4,5-GalA- α 1,2-L-Rha	73.7	3.5×10^{-4}	2.1×10^5	100
		Δ 4,5-GalA- α 1,2-L-Rha- α 1,4-D-GalA- α 1,2-L-Rha				2.5
GH2						
Enzyme	Activity	Substrate	k_{cat} (min^{-1})	K_M (mg/ml)	k_{cat}/K_M ($\text{min}^{-1} \text{mg}^{-1} \text{ml}$)	
BT4151	β 1,4-D- galactosidase	RGI-P	22.5 \pm 0.6	0.2 \pm 0.02	95.7 \pm 1.8	
		(β 1,4-Galp-) ₂	Inactive			
		(β 1,4-Galp-) ₃				
		(β 1,4-Galp-) ₄				
		(β 1,4-Galp-) ₅	Trace activity			
		(β 1,4-Galp-) ₆				
		Galactan				

GH2 (cont)							
Enzyme	Activity	Substrate	k_{cat} (min ⁻¹)	K_M (mg/ml)	k_{cat}/K_M (min ⁻¹ mg ⁻¹ ml)		
BT4156	β 1,4-D-galactosidase	RGI-P	1.2(\pm 0.09) x 10 ²	2.4 \pm 0.5	49.7 \pm 5.1		
		(β 1,4-Galp) ₂	Inactive				
		(β 1,4-Galp) ₃					
		(β 1,4-Galp) ₄					
		BT4181	β -glucuronidase	(β 1,4-Galp) ₅	Trace activity		
				(β 1,4-Galp) ₆			
				Galactan			
Galactan							
BT4160	β 1,4-D-galactosidase	RGI-P	2.1(\pm 0.3) x 10 ^{3*}				
		(β 1,4-Galp) ₂					
		(β 1,4-Galp) ₃					
		(β 1,4-Galp) ₄					
		(β 1,4-Galp) ₅					
		(β 1,4-Galp) ₆					
		Galactan					
GH35							
Enzyme	Activity	Substrate	k_{cat}/K_M (min ⁻¹ M ⁻¹)				
BT4160	β 1,4-D-galactosidase	RGI-P	2.1(\pm 0.3) x 10 ^{3*}				
		(β 1,4-Galp) ₂	Trace activity				
		(β 1,4-Galp) ₃	2.0 x 10 ⁵				
		(β 1,4-Galp) ₄	7.0 x 10 ⁴				
		(β 1,4-Galp) ₅	3.5 x 10 ⁴				
		(β 1,4-Galp) ₆	2.0 x 10 ⁴				
		Galactan	Trace activity				
GH27							
Enzyme	Activity	Substrate	k_{cat} (min ⁻¹)	K_M (M)	k_{cat}/K_M (min ⁻¹ M ⁻¹)		
BT4157	α -D-galactosidase	4NP- α -D-Galp	7.6 \pm 0.2	1.3(\pm 0.01) x 10 ⁻⁴	5.8(\pm 0.7) x 10 ⁴		
Carbohydrate esterase							
Enzyme	Activity	Substrate	k_{cat} (min ⁻¹)	K_M (M)	k_{cat}/K_M (min ⁻¹ M)		
BT4158	Acetyl esterase	4NP-Acetate	6.9 \pm 0.4	1.9(\pm 0.4) x 10 ⁻³	3.6(\pm 0.1) x 10 ³		
		Acetylated Birchwood Xylan	Active				
		Acetylated Glucomannan					

*min⁻¹mg⁻¹ ml; ** Activity using a single substrate concentration

Supplementary Table 2 Kinetics of enzymes encoded by HG-PUL

PL1					
Enzyme	Activity	Substrate	k_{cat} (min ⁻¹)	K_M (mg/ml)	k_{cat}/K_M (min ⁻¹ mg ⁻¹ ml)
BT4115	HG-lyase	Pectin DE 30%	$8.1(\pm 0.4) \times 10^2$	$5(\pm 0.1) \times 10^{-2}$	1.8×10^4
		Pectin DE 60%	$1.2(\pm 0.7) \times 10^3$	0.6 ± 0.1	2.3×10^3
		PGA orange	$2.3(\pm 0.9) \times 10^3$	$5(\pm 0.1) \times 10^{-2}$	4.5×10^4
BT4116	HG-lyase	Pectin DE 30%	$1.4(\pm 0.5) \times 10^3$	$2(\pm 0.2) \times 10^{-1}$	7.2×10^3
		Pectin DE 60%	$1.4(\pm 0.4) \times 10^3$	1.1 ± 0.1	1.3×10^3
		PGA orange	$6.6(\pm 0.3) \times 10^2$	$4(\pm 0.1) \times 10^{-2}$	1.6×10^3
BT4119	HG-Lyase	PGA orange	Generates long oligosaccharides		
GH28					
Enzyme	Activity	Substrate	k_{cat}/K_M (min ⁻¹ M ⁻¹)		
BT4123	HG-specific α -D- galacturonidase	(α 1,4-D-GalA) ₂	$1.6(\pm 0.07) \times 10^5$		
		(α 1,4-D-GalA) ₃	$2.8(\pm 0.09) \times 10^5$		
GH105					
Enzyme	Activity	Substrate	k_{cat} (min ⁻¹)	K_M (M)	k_{cat}/K_M (min ⁻¹ M ⁻¹)
BT4108	HG-specific unsaturated- galacturonidase	Δ 4,5-GalA- α 1,4- D-GalA	$1.0(\pm 0.04) \times 10^2$	$8(\pm 0.1) \times 10^{-2}$	$1.3(\pm 0.3) \times 10^3$

Supplementary Table 3 Kinetics of enzymes encoded by Gal-PUL

GH2 and GH53 enzymes					
Enzyme	Activity	Substrate	k_{cat} (min ⁻¹)	K_M (M)	k_{cat}/K_M (min ⁻¹ M ⁻¹)
BT4667	β1,4-galactosidase	β-D-Glc-β1,4-Gal	15.7 ± 0.6	5.8(± 0.1) × 10 ⁻³	2.7(± 0.3) × 10 ³
		(β1,4-Galp) ₂	1.5(± 0.2) × 10 ³	1.5(± 0.2) × 10 ⁻³	1.0(± 0.1) × 10 ⁶
		(β1,4-Galp) ₃	1.7(± 0.1) × 10 ³	1.2(± 0.3) 10 ⁻³	1.4(± 0.2) × 10 ⁶
		(β1,4-Galp) ₄	1.3(± 0.1) × 10 ³	7.6(± 0.1) × 10 ⁻⁴	1.7(± 0.2) × 10 ⁶
		(β1,4-Galp) ₅	1.0(± 0.05) × 10 ³	1.7 (± 0.3) × 10 ⁻³	5.8(± 0.9) × 10 ⁵
		RGI-P	70.4 ± 0.3	2.9 ± 0.2*	24.3 ± 0.5*
GH53					
Enzyme	Activity	Substrate	k_{cat} (min ⁻¹)	K_M (M)	k_{cat}/K_M (min ⁻¹ M ⁻¹)
BT4668 (GH53)	Endp-β1,4-D-galactanase	(β1,4-Galp) ₂			Inactive
		(β1,4-Galp) ₃			2.8(± 0.2) × 10 ²
		(β1,4-Galp) ₄			4.2(± 0.5) × 10 ³
		(β1,4-Galp) ₅			2.3(± 0.1) 10 ⁴
		(β1,4-Galp) ₆			1.9(± 0.07) × 10 ⁴
		Galactan	1.2(± 0.03) × 10 ³	0.3 ± 0.03*	4.2(± 0.3) × 10 ^{3*}
GH146					
Enzyme	Activity	Substrate	k_{cat} (min ⁻¹)	K_M (M)	k_{cat}/K_M (min ⁻¹ M ⁻¹)
BACOVA 05493	β1,4-D-galactosidase	(β1,4-Galp) ₂			Low activity
		(β1,4-Galp) ₃	1.7(± 0.5) × 10 ²	7.5(± 0.1) × 10 ⁻⁴	2.3 (± 0.5) × 10 ⁵
		(β1,4-Galp) ₄	2.1(± 0.4) × 10 ²	5.2(± 0.9) × 10 ⁻⁴	4.0(± 0.6) × 10 ⁵
		(β1,4-Galp) ₅	2.7(± 0.3) × 10 ²	7.3(± 0.7) × 10 ⁻⁴	3.7(± 0.2) × 10 ⁵
		(β1,4-Galp) ₆	2.8(± 0.3) × 10 ²	1.8(± 0.5) × 10 ⁻⁴	1.5(± 0.05) × 10 ⁶
		RGI-P	74.8 ± 6.1	8.3 ± 1.8*	9.0 ± 0.9*
		Galactan	1.2(± 0.03) × 10 ³	0.3 ± 0.03*	4.2(± 0.3) × 10 ^{3*}

* K_M (mg/ml), k_{cat}/K_M (min⁻¹mg⁻¹ ml)

Supplementary Table 4 Kinetics of enzymes encoded by Ara-PUL

GH43					
Enzyme	Activity	Substrate	k_{cat} (min ⁻¹)	K_M (M)	k_{cat}/K_M (min ⁻¹ M ⁻¹)
BT0360 [†]	Endo- α 1,5-L-arabinanase	Branched arabinan			5.1(\pm 0.6) x 10 ^{2*}
		Unbranched arabinan			57.8 \pm 10.2*
BT0367 [†]	Endo- α 1,5-L-arabinanase	Branched arabinan			52.4 \pm 3.2*
		Unbranched arabinan			3.2(\pm 0.2) x 10 ^{2*}
BT0369 [†]	β 1,2-L-arabinofuranosidase	4NP- α -L- Araf	4.0(\pm 0.7) x 10 ⁴	8.0(\pm 0.2) x 10 ⁻³	5.0(\pm 0.2) x 10 ⁹
		(α 1,5-Araf) ₃	16.0 \pm 2.4	1.8(\pm 0.6) x 10 ⁻⁴	9.0(\pm 0.4) x 10 ⁵
		(α 1,5-Araf) ₆			3.6(\pm 0.3) x 10 ³
		Branched arabinan	2.7(\pm 0.3) x 10 ³	2(\pm 0.1) x 10 ^{-2*}	1.3(\pm 0.2) x 10 ^{7*}
GH51					
Enzyme	Activity	Substrate	k_{cat} (min ⁻¹)	K_M (M)	k_{cat}/K_M (min ⁻¹ M ⁻¹)
BT0348	α 1,3-L-arabinofuranosidase	(α 1,5-Araf) ₂	26.6 \pm 1.95	5.6(\pm 0.5) x 10 ⁻³	4.8(\pm 0.4) x 10 ³
		(α 1,5-Araf) ₃	36.7 \pm 1.8	1.6(\pm 0.3) x 10 ⁻³	2.4(\pm 0.3) x 10 ⁴
		(α 1,5-Araf) ₄	65.6 \pm 4.11	4.3(\pm 0.5) x 10 ⁻³	1.5(\pm 0.04) x 10 ⁴
		(α 1,5-Araf) ₅	45.1 \pm 3.22	2.4(\pm 0.3) x 10 ⁻³	1.9(\pm 0.2) x 10 ⁴
		Branched arabinan	3.1(\pm 0.3) x 10 ²	1.7 \pm 0.2*	1.8(\pm 0.2) x 10 ^{2*}
		Unbranched arabinan	19.8 \pm 0.19	1.9 \pm 0.4*	10.4 \pm 0.8*
BT0368	Endo- α 1,5-L-arabinanase	(α 1,5-Araf) ₂	1.7(\pm 0.2) x 10 ⁴	2.8(\pm 0.4) x 10 ⁻³	6.2(\pm 0.6) x 10 ⁶
		(α 1,5-Araf) ₃	2.7(\pm 0.06) x 10 ⁴	2.6(\pm 0.3) x 10 ⁻³	1.0(\pm 0.1) x 10 ⁷
		(α 1,5-Araf) ₄	2.7(\pm 0.4) x 10 ⁴	3.0(\pm 0.4) x 10 ⁻³	9.0(\pm 0.9) x 10 ⁶
		(α 1,5-Araf) ₅	2.8 (\pm 0.2) x 10 ⁴	3.0(\pm 0.3) x 10 ⁻³	9.4(\pm 0.2) x 10 ⁶
		(α 1,5-Araf) ₆	2.8(\pm 0.2) x 10 ⁴	2.3(\pm 0.3) x 10 ⁻³	1.2(\pm 0.1) x 10 ⁶
		Branched arabinan			14.7 \pm 0.4*
		Unbranched arabinan			53.9 \pm 5.6*
GH127					
Enzyme	Activity	Substrate	k_{cat} (min ⁻¹)	K_M (M)	k_{cat}/K_M (min ⁻¹ M ⁻¹)
BT0349	β 1,2-L-arabinofuranosidase	4NP- α -L- Araf			Inactive
		Branched arabinan	25.0 \pm 2.1	3.9 \pm 0.8*	6.4 \pm 0.4
		Unbranched arabinan	22.7 \pm 2.0	1.7 \pm 0.4*	13.4 \pm 0.9

* K_M (mg/ml), k_{cat}/K_M (min⁻¹ mg⁻¹ ml); [†] published in⁷.

Supplementary Table 5 Binding of surface proteins and HTCS to pectin glycans

RGI-PUL						
Protein	Ligand	K_A ($\times 10^3 M^{-1}$)	ΔG (kcal mole ⁻¹)	ΔH (kcal mole ⁻¹)	$T\Delta S$ (kcal mole ⁻¹)	N
BT4167 (SGBP)	RGI-P	57 ± 0.03	-6.5	-11.4 ± 0.4	-4.9	1.02 ± 0.031
	RGI-AM	43 ± 0.04	-6.3	-13.5 ± 0.5	-7.2	1.00 ± 0.0
BT4165 (SusD-like)	RGI-P	No binding				
	RGI-AM					
	$\Delta 4,5\text{-GalA-(}\alpha 1,2\text{-L-Rha-}\alpha 1,4\text{-D-GalA)}_n$					
	$(\alpha 1,2\text{-L-Rha-}\alpha 1,4\text{-D-GalA)}_n$					
BT4178 (HTCS)	RGI-AM	130 ± 0.004	-7.0	-1.1 ± 0.1	5.9	1.00 ± 0.0
	RGI-AM + BT4174	No binding				
	$\Delta 4,5\text{-GalA-}\alpha 1,2\text{-L-Rha}$	No binding				
	$\Delta 4,5\text{-GalA-}\alpha 1,2\text{-L-Rha-}\alpha 1,4\text{-D-GalA-}\alpha 1,2\text{-L-Rha}$	8.5 ± 1.3	-5.4	-3.1 ± 0.1	2.3	1.00 ± 0.0
	$\alpha\text{-L-Rha-}\alpha 1,4\text{-D-GalA-}\alpha 1,2\text{-L-Rha}$	No binding				
	Unsaturated hexa and octasaccharide of RGI-AM	Binding too weak to quantify				
HG-PUL						
Protein	Ligand	K_A ($\times 10^3 M^{-1}$)	ΔG (kcal mole ⁻¹)	ΔH (kcal mole ⁻¹)	$T\Delta S$ (kcal mole ⁻¹)	N
BT4112 (SGBP)	Pectin DE 30%	42 ± 1.1	-6.3	-10.5 ± 0.2	-4.17	1.01 ± 0.01
BT4113 (SusD-like)	$(\alpha 1,4\text{-D-GalA})_4/(\alpha 1,4\text{-D-GalA})_5$	2.2 ± 0.2	-4.6	-3.2 ± 0.2	1.4	1
	$\Delta 4,5\text{-GalA-(}\alpha 1,4\text{-D-GalA)}_3/(\alpha 1,4\text{-D-GalA)}_4$	5.3 ± 1.3	-5.0	-2.4 ± 0.3	2.6	1
BT4122 (SusD-like)	$(\alpha 1,4\text{-D-GalA})_4/(\alpha 1,4\text{-D-GalA})_5$	No binding				
	$\Delta 4,5\text{-GalA-(}\alpha 1,4\text{-D-GalA)}_3/(\alpha 1,4\text{-D-GalA)}_4$					
BT4111 (HTCS)	$(\alpha 1,4\text{-D-GalA})_2$	7.4 ± 0.4	-5.3	-7.0 ± 0.1	-1.73	1.00 ± 0.0
	$(\alpha 1,4\text{-D-GalA})_3$	5.5 ± 0.4	-5.1	-10.0 ± 0.2	-4.9	1.00 ± 0.0
	$(\alpha 1,4\text{-D-GalA})_n$	3.2 ± 0.6	-4.8	-18.4 ± 2.5	-13.7	1.00 ± 1.3
	$\Delta 4,5\text{-GalA-(}\alpha 1,4\text{-D-GalA)}_n$	No binbing				

Gal-PUL						
Protein	Ligand	K_A ($\times 10^3 M^{-1}$)	ΔG (kcal mole $^{-1}$)	ΔH (kcal mole $^{-1}$)	$T\Delta S$ (kcal mole $^{-1}$)	N
BT4669 (SGBP)	($\beta 1,4$ -Galp) $_2$	No binding				
	($\beta 1,4$ -Galp) $_3$	0.4 \pm 0.02	-3.1	-21.7 \pm 4.6	-18.2	0.99
	($\beta 1,4$ -Galp) $_4$	2.1 \pm 0.18	-4.6	-28.1 \pm 1.9	-23.6	1.06
	($\beta 1,4$ -Galp) $_5$	10.4 \pm 1.5	-5.5	-24.4 \pm 0.7	-18.9	1.02
	($\beta 1,4$ -Galp) $_6$	124 \pm 42	-7.0	-21.9 \pm 0.2	-14.9	0.95
	($\beta 1,4$ -Galp) $_7$	138 \pm 40	-7.0	-22.6 \pm 0.2	-15.6	1.00
	Galactan	304 \pm 14	-7.6	-68.4 \pm 0.05	-60.8	1.03
BT4670 (SusD-like)	($\beta 1,4$ -Galp) $_2$	No binding				
	($\beta 1,4$ -Galp) $_3$	No binding				
	($\beta 1,4$ -Galp) $_4$	8.1 \pm 1.1	-5.3	-8.1 \pm 1.1	-2.8	1.05
	($\beta 1,4$ -Galp) $_5$	11.1 \pm 1.2	-5.5	-7.6 \pm 1.7	-2.1	1.07
	($\beta 1,4$ -Galp) $_6$	16.8 \pm 1.4	-5.8	-7.7 \pm 0.8	-2.0	1.04
	($\beta 1,4$ -Galp) $_7$	22.0 \pm 2.5	-5.9	-7.2 \pm 0.8	-1.3	1.09
	Galactan	84.8 \pm 2.2	-5.7	-6.2 \pm 0.6	-0.5	1.02
BT4673 (HTCS)	Galp	No binding				
	($\beta 1,4$ -Galp) $_2$	38.0 \pm 2.4	-6.2	-17.9 \pm 0.9	-11.7	1.05
	($\beta 1,4$ -Galp) $_3$	48.5 \pm 9.2	-6.4	-16.1 \pm 2.3	-9.7	1.01
	($\beta 1,4$ -Galp) $_4$	98.5 \pm 9.2	-6.8	-12.9 \pm 0.4	-6.2	1.04
	($\beta 1,4$ -Galp) $_5$	86.9 \pm 1.4	-6.7	-6.1 \pm 0.3	-0.7	1.08
Ara-PUL						
Protein	Ligand	K_A ($\times 10^3 M^{-1}$)	ΔG (kcal mole $^{-1}$)	ΔH (kcal mole $^{-1}$)	$T\Delta S$ (kcal mole $^{-1}$)	N
BT0365 (SGBP)	($\alpha 1,5$ -Araf) $_4$	0.6 \pm 0.1	-3.8	-23.4 \pm 1.2	-19.6	0.99
	($\alpha 1,5$ -Araf) $_5$	7.3 \pm 1.2	-5.3	-26.2 \pm 1.2	-21.0	0.98
	($\alpha 1,5$ -Araf) $_6$	14.4 \pm 0.7	-4.3	-25.5 \pm 7.2	-20.7	1.02
	($\alpha 1,5$ -Araf) $_7$	68.4 \pm 8.2	-6.6	-32.1 \pm 1.7	-25.5	1.01
	($\alpha 1,5$ -Araf) $_8$	112 \pm 15	-6.8	-31.6 \pm 1.3	-24.8	1.08
	Branched arabinan	170 \pm 15	-7.2	-33.8 \pm 0.6	-26.6	1.02
	Unbranched arabinan	246 \pm 25	-7.0	-30.3 \pm 0.5	-23.0	1.02

Ara-PUL (cont.)						
Protein	Ligand	K_A ($\times 10^3 M^{-1}$)	ΔG (kcal mole ⁻¹)	ΔH (kcal mole ⁻¹)	$T\Delta S$ (kcal mole ⁻¹)	N
BT0361 (SusD-like)	($\alpha 1,5$ -Araf) ₄	No binding				
	($\alpha 1,5$ -Araf) ₇					
	Branched arabino oligosaccharides					
	Branched arabinan					
	Unbranched arabino oligosaccharides					
	Unbranched arabinan					
BT0363 (SusD-like)	($\alpha 1,5$ -Araf) ₄					
	($\alpha 1,5$ -Araf) ₇					
	Branched arabino oligosaccharides					
	Branched arabinan					
	Unbranched arabino oligosaccharides					
	Unbranched arabinan					
BT0366	($\alpha 1,5$ -Araf) ₅ [†]	No binding				
	($\alpha 1,5$ -Araf) ₆ [†]	3.9 ± 0.9	-4.9	7.6 ± 0.7	12.5	1.2
	($\alpha 1,5$ -Araf) ₇ [†]	35 ± 6.0	-6.2	0.6 ± 0.3	6.8	1.3
	($\alpha 1,5$ -Araf) ₈ [†]	48 ± 1.0	-6.4	1.6 ± 0.7	8.0	0.7
	Unbranched arabinan [†]	100 ± 0.0	-6.8	-26.8 ± 0.7	-20.0	----
	Unbranched arabino oligosaccharides	52 ± 0.5	-6.4	-21.6 ± 1.0	-15.2	
	Branched arabinan	No binding				
	Branched arabino oligosaccharides					

The thermodynamics of binding was quantified by isothermal titration calorimetry;

[†] published in⁸.

Supplementary Table 6. Enzyme Mutant Activities

RGI-PUL				
Enzyme	Activity	Substrate	Mutation	k_{cat}/K_M ($\text{min}^{-1} \text{M}^{-1}$)
BT4145	RGI-specific α -L-rhamnosidase	α -L-Rha- α 1,4-D-GalA	WT	$7.1 (\pm 0.03) \times 10^7$
			D250A	Inactive
			E253A	Inactive
			E349A	Inactive
			E395A	Inactive
			Q328E	Inactive
BT4155	HG specific α -D-galacturonidase	$(\alpha$ 1,4-D-GalA) ₂	WT	$3.2 (\pm 0.2) \times 10^5$
			D249A	Inactive
			D270A	Inactive
BT4156	β 1,4-galactosidase	RGI-P	WT	49.7 ± 5.1
			E422A	Inactive
BT4170	RGI lyase	RGI-AM	WT	$3.63 (\pm 0.05) \times 10^{4*}$
			R125A	Inactive
			H153A	$7.26 (\pm 0.2) \times 10^{3*}$
			Q155A	$1.93 (\pm 0.04) \times 10^{4*}$
			Y184A	$8.72 (\pm 0.1) \times 10^{3*}$
			G212A	$1.77 (\pm 0.03) \times 10^{4*}$
			N213A	$1.66 (\pm 0.06) \times 10^{3*}$
			D215A	Inactive
			H220A	Inactive
			D246A	Inactive
			D247A	Inactive
			D250A	Inactive
			I252A	$70.5 \pm 2.8^*$
			N253A	$1.85 (\pm 0.04) \times 10^{4*}$
			D280A	$376.0 \pm 26.0^*$
			K285A	Inactive
			L318A	$1.92 (\pm 0.02) \times 10^{4*}$
			R319A	$1.92 (\pm 0.1) \times 10^{3*}$
Y322A	$3.46 (\pm 0.09) \times 10^{3*}$			
N324A	$3.68 (\pm 0.2) \times 10^{3*}$			
H325A	$1.22 (\pm 0.03) \times 10^{3*}$			

Gal-PUL				
Enzyme	Activity	Substrate	Mutation	k_{cat}/K_M ($\text{min}^{-1}\text{mg}^{-1}\text{ml}$)
BT4668	Endo- β 1,4-galactanase	Galactan	WT	$4.2 (\pm 0.3) \times 10^3$
			E180A	Inactive
			E292A	Inactive
BACOVA 05493	β 1,4-galactosidase	Galactan	WT	$2.3 (\pm 0.5) \times 10^5$
			E203A	Inactive
			E300A	Inactive
Ara-PUL				
Enzyme	Activity	Substrate	Mutation	k_{cat}/K_M ($\text{min}^{-1}\text{mg}^{-1}\text{ml}$)
BT0349	β 1,2-L-arabinofuranosidase	Branched arabinan	WT	6.4 ± 0.4
			E318A	Inactive
			C414A	Inactive
BT0360	Endo- α 1,5-arabinanase	Branched arabinan	WT	$5.1 (\pm 0.6) \times 10^{2\dagger}$
			D154A	Inactive
BT0367	Endo- α 1,5-arabinanase	Branched arabinan	WT	$52.4 \pm 3.2^\dagger$
			D63A	Inactive

[†] published in'

Supplementary Table 7. ^1H and ^{13}C NMR assignments of the arabinotetraose at 25 °C in D_2O .

	H-1	H-2	H-3	H-4	H-5	C-1	C-2	C-3	C-4	C-5
$\beta\text{-Araf}_1$	5.082	4.135	4.044	3.904	3.808, 3.728	102.25	77.07	75.00	82.81	63.76
$\alpha\text{-Araf}_2$	5.193	4.372	3.953	4.151	3.733, 3.831	107.80	80.43	84.54	83.44	61.93
$\alpha\text{-Araf}_3$	5.105	4.290	4.052	4.171	3.762, 3.862	108.28	80.06	82.59	83.65	61.75
$\alpha\text{-Araf}_4$	5.253	4.035	4.037	4.232	3.780, 3.849	101.92	82.14	76.62	82.12	67.35
$\beta\text{-Araf}_4$	5.301	4.095	3.945	-	-	96.12	-	-	-	-

Supplementary Table 8. Growth profiles of *Bacteroides* species on pectins

Species	Galactan	Galactan oligos	SBA	SBA oligos	dBA	dBA oligos	HG	HG oligos	RGI-AM	RGI-AM oligos
<i>B. caccae</i>	0.68	0.63	0.05	0.02	0	-0.02	0.58	0.56	0	-0.04
<i>B. cellulosilyticus</i>	0.57	0.45	0.69	0.59	0.68	0.56	0.49	0.59	-0.1	0.52
<i>B. clarus</i>	0.7	0.61	0.13	0.05	0.05	0.04	0.6	0.48	0.01	0.55
<i>B. dorei</i>	0.7	0.57	0.8	0.55	0.66	0.61	0.59	0.63	0.15	0.64
<i>B. eggerthii</i>	0.03	0.38	0.61	0.48	0.45	0.41	0.09	0.05	-0.1	0
<i>B. fingoldii</i>	0.7	0.88	0.7	0.62	0.59	0.68	0.44	0.42	1.04	0.58
<i>B. fluxus</i>	-0.04	0.46	-0.01	0.06	-0.07	-0.07	0.39	0.37	0.01	-0.02
<i>B. fragilis</i>	0.61	0.36	0.59	0.48	0.53	0.59	0.02	0.38	-0.14	-0.01
<i>B. intestinalis</i>	0.71	0.74	0.76	0.64	0.61	0.63	0.46	0.54	0	0.58
<i>B. massiliensis</i>	0.69	0.57	0.65	0.54	0.58	0.54	-0.07	0.58	-0.09	0.56
<i>B. nordii</i>	0.06	0.32	0.01	0	-0.01	-0.03	0.54	0.48	-0.05	0.06
<i>B. oleiciplenus</i>	0.74	0.56	0.69	0.59	0.61	0.59	0.53	0.54	0.03	0.53
<i>B. ovatus</i>	0.81	0.8	0.15	0.43	0.2	0.33	0.49	0.52	0.95	0.48
<i>B. plebeius</i>	-0.01	0.33	0.01	0	-0.05	-0.07	-0.12	0.07	-0.06	-0.02
<i>B. salyersiae</i>	0.77	0.66	0.05	0.06	-0.02	0.03	0.04	0.52	-0.06	0.48
<i>B. stercoris</i>	0.75	0.53	0.58	0.57	0.56	0.57	0.57	0.53	0.01	0.52
<i>B. thetaiotaomicron</i>	0.67	0.58	0.79	0.58	0.71	0.44	0.57	0.54	1.01	0.57
<i>B. uniformis</i>	0.11	0.48	0.04	0.08	0.05	0.06	0.04	0.08	-0.13	-0.02
<i>B. vulgatus</i>	0.78	0.53	0.62	0.63	1	0.61	-0.04	0.64	-0.07	0.62
<i>B. xylanisolvens</i>	0.04	0.67	0.75	0.46	0.54	0.57	0.42	0.47	0.97	0.57

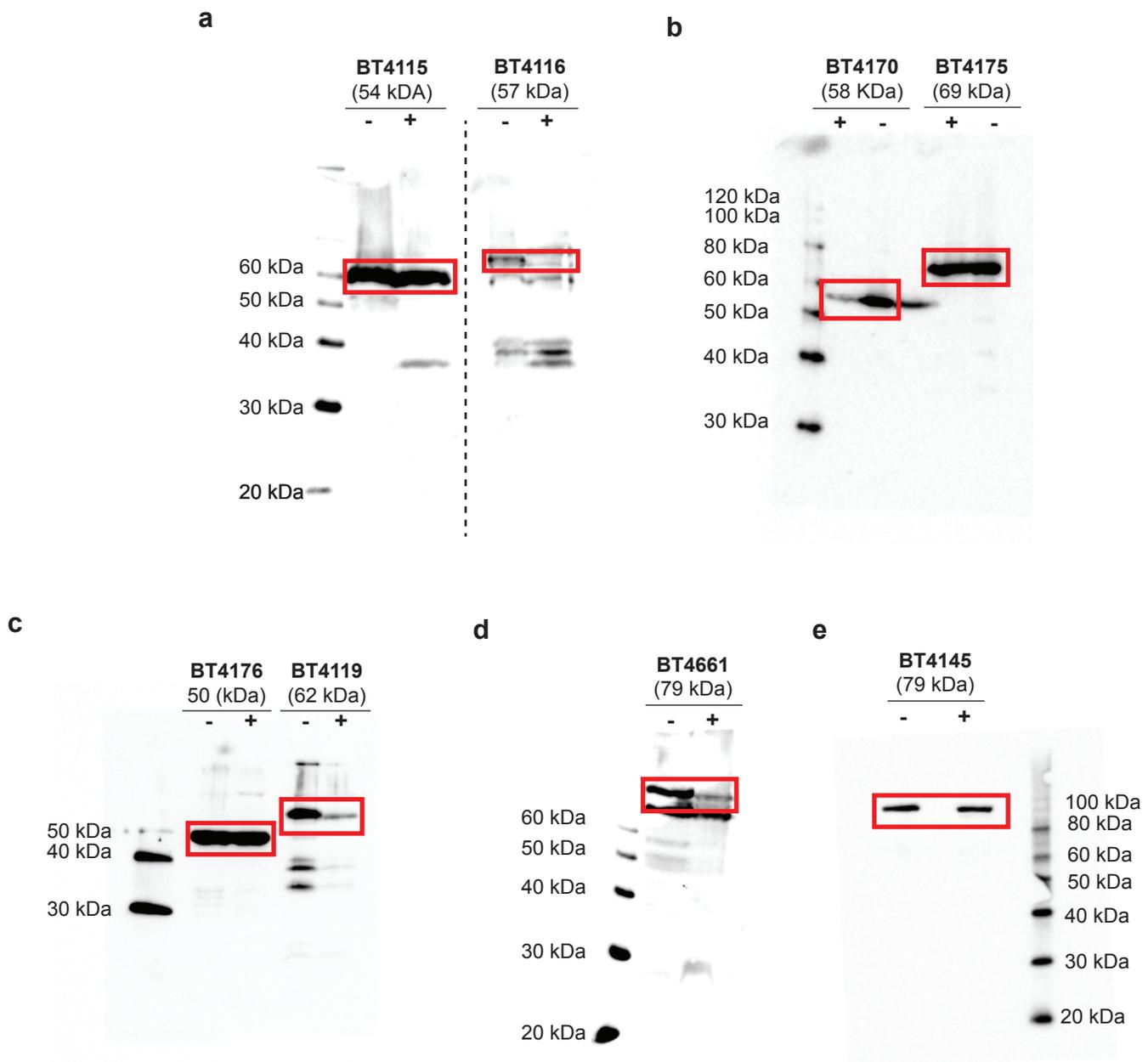
The numerical values represent the optical density of stationary phase cultures. Panels that are coloured are indicative of growth. SBA, sugar beet arabinan; dBA, debranched sugar beet arabinan; RGI-AM, undecorated RGI from Arabidopsis seed mucilage; HG, homogalacturonan; oligos, oligosaccharides generated by partial enzyme digestion of the cognate polysaccharide.

Supplementary Table 9. Data collection and refinement statistics.

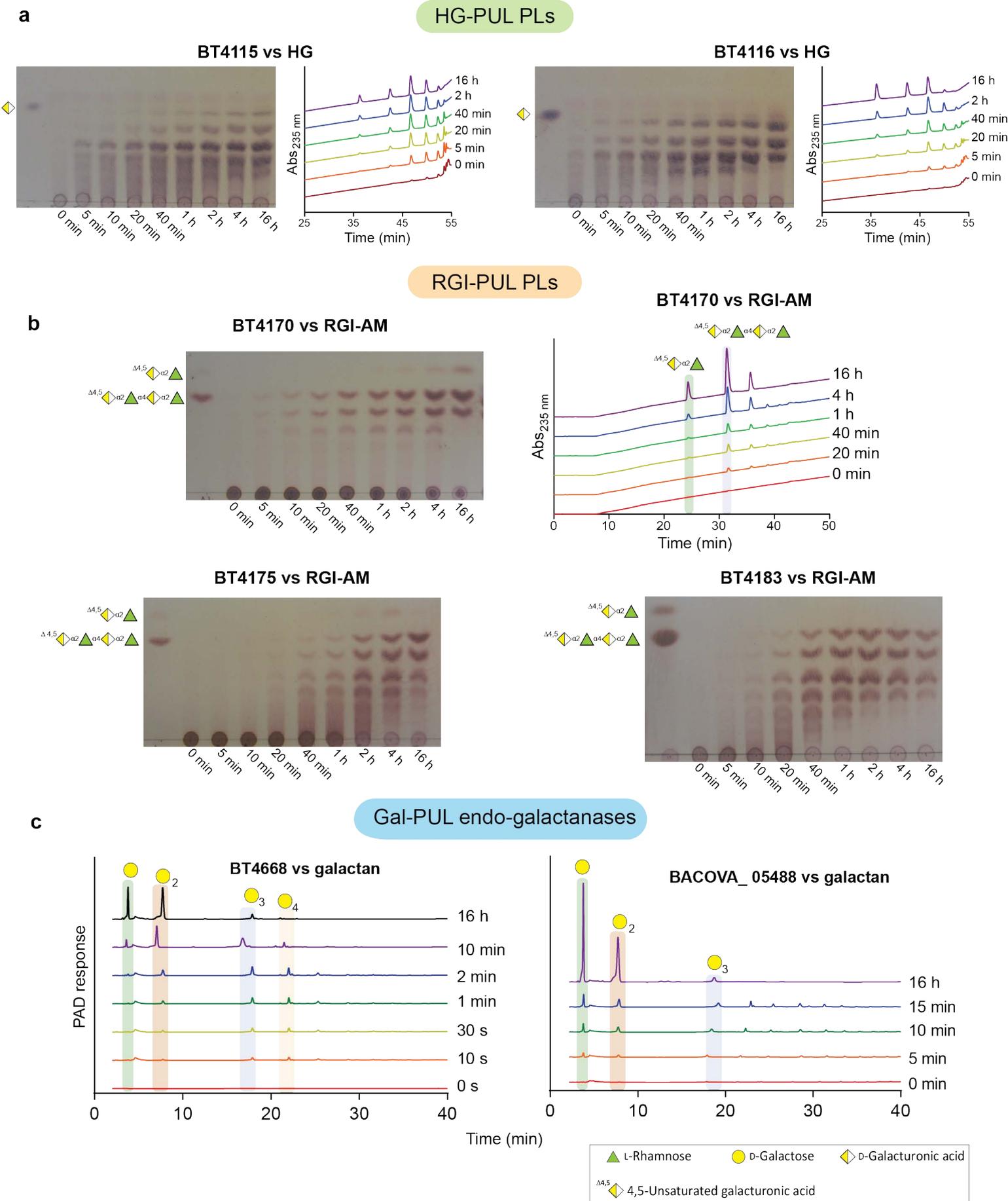
Data collection	BT4170 apo	BT4170 K285A-ligand	WT BT4170-ligand	BT4155	BT0349-arabinose
Date	02/07/14	16/10/15	12/12/14	21/06/16	05/07/17
Source	I04-1	I04	I04	I04-1	I04
Wavelength (Å)	0.9200	0.9791	0.9794	0.9282	0.9796
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	I422	P4 ₃ 2 ₁ 2
Cell dimensions					
<i>a</i> , <i>b</i> , <i>c</i> (Å)	77.5 123.8 138.4	45.4 76.6 122.8	77.5 123.9 137.6	127.5 127.5 263.6	96.21 96.21 187.84
α , β , γ (°)	90	90	90	90	90
No. of measured reflections	1547996 (62292)	155355 (11830)	1712463 (15540)	1969121 (125488)	4058474 (206181)
No. of independent reflections	221195 (10868)	22486 (1650)	536956 (9262)	73408 (4477)	56151 (2753)
Resolution (Å)	39.65 – 1.48 (1.51 – 1.48)	61.7 – 2.20 (2.26 – 2.20)	48.4 – 1.07 (1.09 – 1.07)	48.73 – 2.00 (2.04 – 2.00)	28.54-2.05 (2.09-2.05)
CC _{1/2}	0.999 (0.736)	0.998 (0.553)	0.998 (0.378)	0.998 (0.620)	0.999(0.831)
<i>I</i> / σ <i>I</i>	13.4 (2.6)	10.7 (1.6)	11.3 (0.9)	12.8 (1.6)	17.33 (2.11)
Completeness (%)	99.9 (99.9)	99.8 (99.5)	94.3 (33.1)	100.0 (100.0)	100 (100)
Redundancy	7.0 (5.7)	6.9 (7.2)	3.2 (1.7)	26.8 (28)	72.37 (74.92)
Anomalous completeness (%)				100.0 (100.0)	99.97(99.88)
Anomalous multiplicity				14.0 (14.4)	38.42(39.02)
Refinement					
<i>R</i> _{work} / <i>R</i> _{free}	12.68 / 16.89	23.48 / 27.48	14.47/16.64	18.19 / 21.50	16.82 / 20.71
No. atoms					
Protein	9891	3257	10037	6859	6145
Ligand/Ions	18	99	126	4	11
Water	1519	3	1757	450	355
B-factors					
Protein	15.6	70.3	10.7	29.2	35.1
Ligand/Ions	18.9	47.04	11.6	30.2	30.5
Water	28.1	38.1	25.7	36.5	39.6
R.m.s deviations					
Bond lengths (Å)	0.012	0.009	0.011	0.011	0.01
Bond angles (°)	1.54	1.32	1.54	1.51	1.55
PDB code	5OLQ	5OLR	5OLS	5OLP	5OPJ

Supplementary References

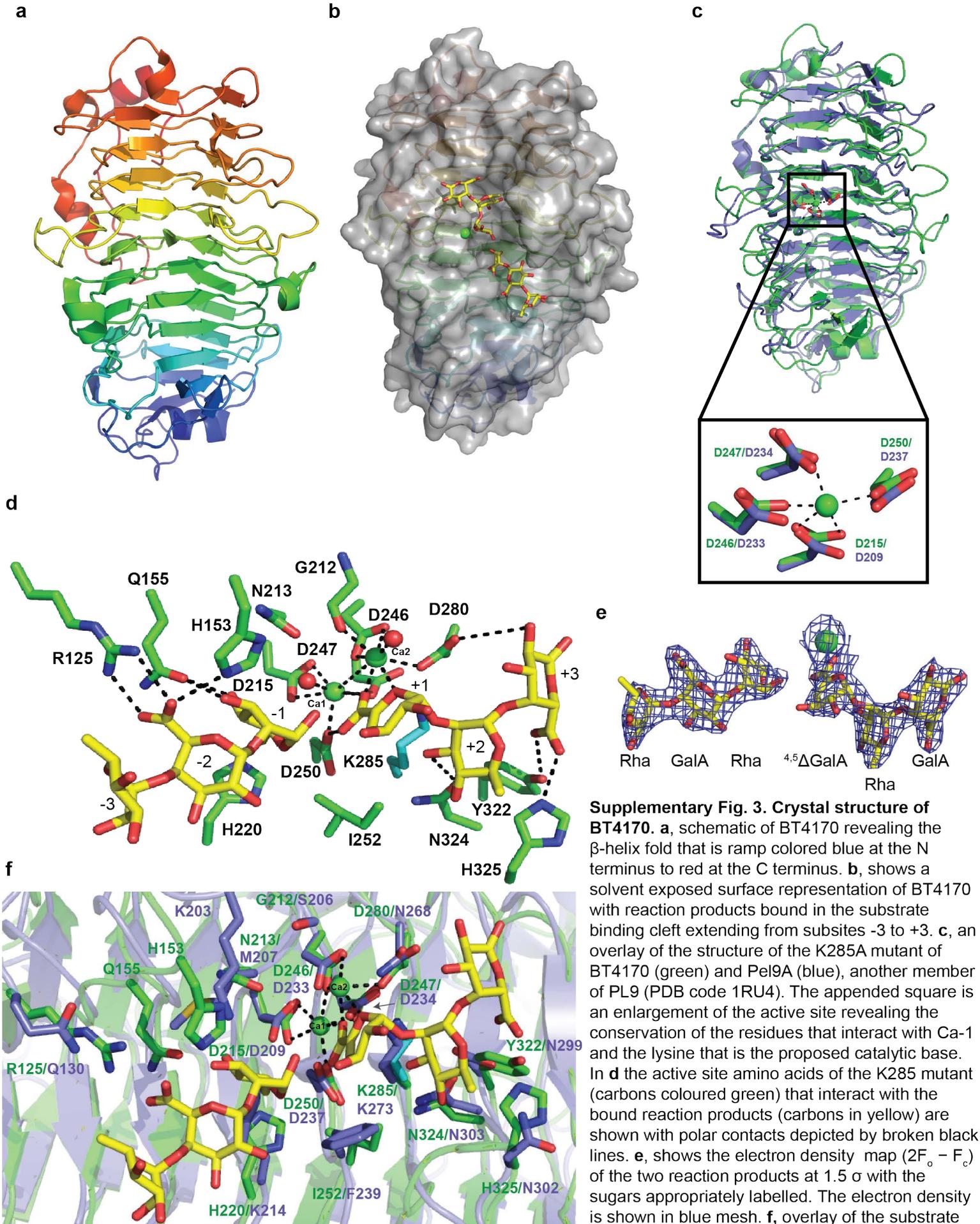
- 1 Renard, C. M., Crepeau, M. J. & Thibault, J. F. Glucuronic acid directly linked to galacturonic acid in the rhamnogalacturonan backbone of beet pectins. *Eur J Biochem* **266**, 566-574 (1999).
- 2 Ndeh, D. *et al.* Complex pectin metabolism by gut bacteria reveals novel catalytic functions *Nature* **544**, 65-70 (2017).
- 3 Jenkins, J., Shevchik, V. E., Hugouvieux-Cotte-Pattat, N. & Pickersgill, R. W. The crystal structure of pectate lyase Pel9A from *Erwinia chrysanthemi*. *J Biol Chem* **279**, 9139-9145 (2004).
- 4 Garron, M. L. & Cygler, M. Uronic polysaccharide degrading enzymes. *Curr Opin Struct Biol* **28**, 87-95 (2014).
- 5 Cameron, E. A. *et al.* Multifunctional nutrient-binding proteins adapt human symbiotic bacteria for glycan competition in the gut by separately promoting enhanced sensing and catalysis. *MBio* **5**, e01441-01414 (2014).
- 6 Glenwright, A. J. *et al.* Structural basis for nutrient acquisition by dominant members of the human gut microbiota. *Nature* **541**, 407-411 (2017).
- 7 Cartmell, A. *et al.* The structure and function of an arabinan-specific alpha-1,2-arabinofuranosidase identified from screening the activities of bacterial GH43 glycoside hydrolases. *J Biol Chem* **286**, 15483-15495 (2011).
- 8 Martens, E. C. *et al.* Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol* **9**, e1001221 (2011).



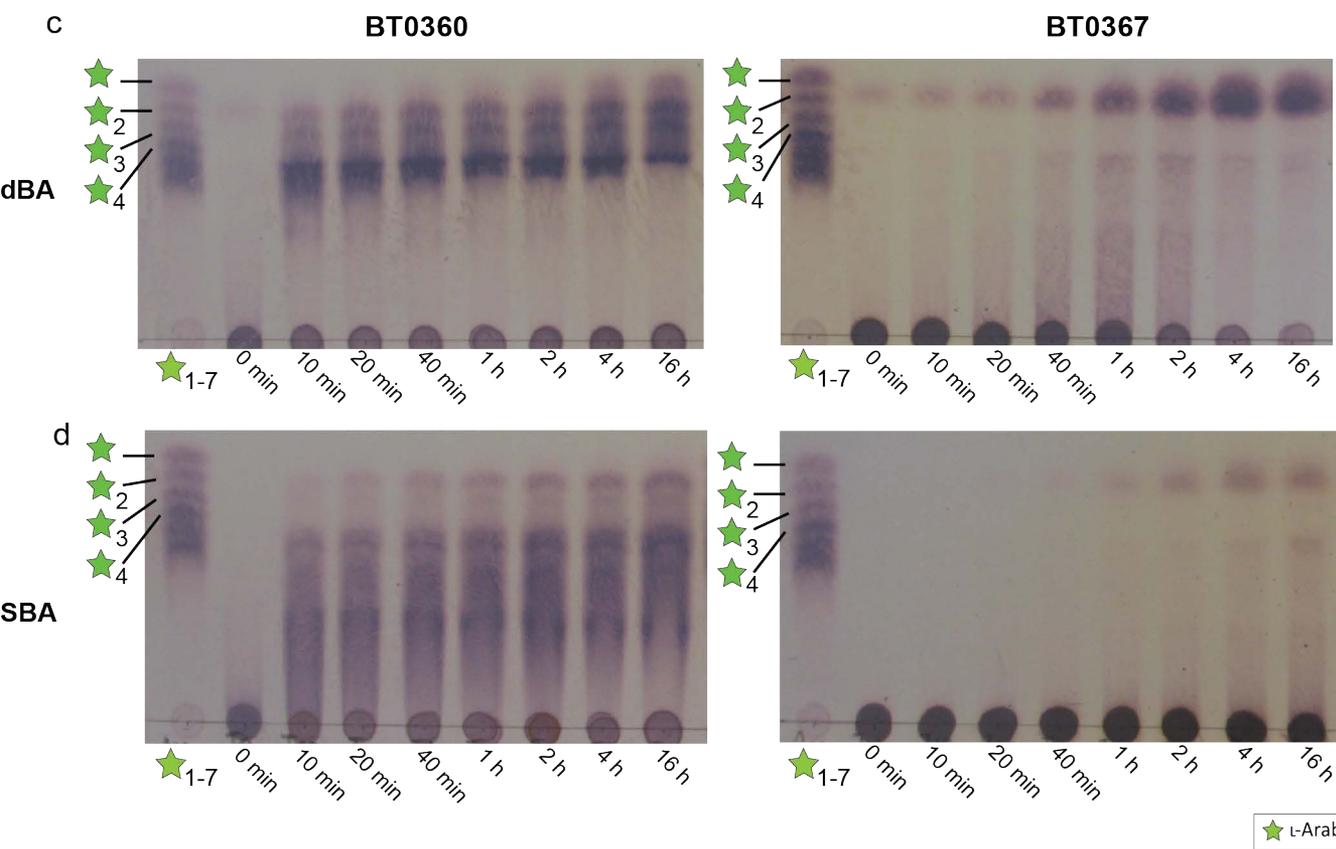
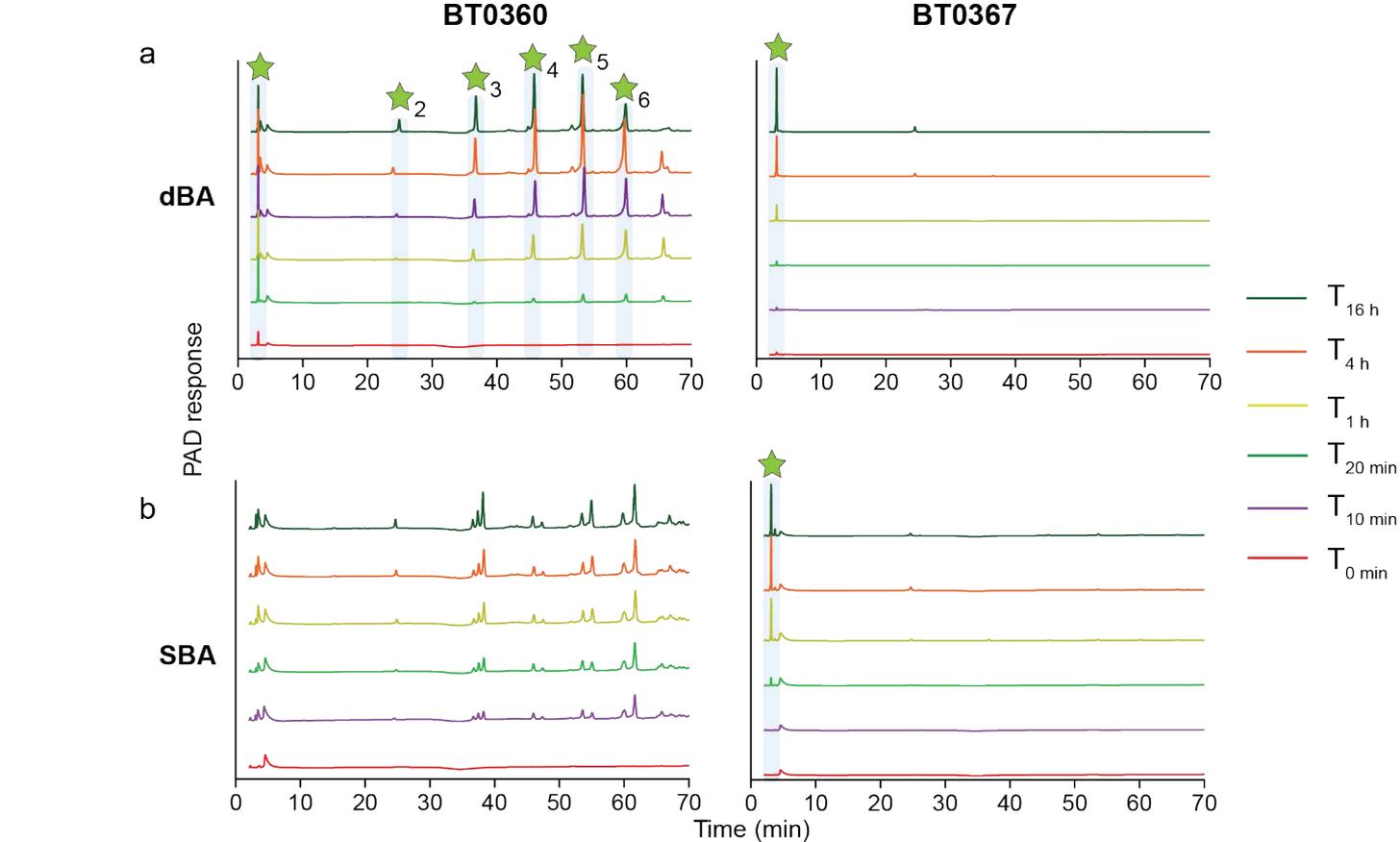
Supplementary Figure 1. Immunodetection of pectin degrading enzymes. Western blot detection of selected *B. thetaiotaomicron* enzymes encoded by the HG-PUL and RGI-PUL after treatment of whole cells grown on HG or RG-I with proteinase K (+) or untreated (-). Blots **a** and **d** were probed with polyclonal primary antibodies generated against target proteins, blots **b**, **c** and **e** were probed with anti-flag tag antibody, then all blots used appropriate HRP labelled secondary antibody. In **a**, a single gel was run and blotted, the membrane was cut up approximately as indicated by dotted lines for probing with different primary antibodies, and subsequently imaged as a reassembled membrane to enable comparison with the molecular weight marker. The sections cropped and shown in Fig 2c are highlighted with red rectangles. In the case of BT4170 and BT4175, the lanes were loaded with Proteinase K treated on the left, and untreated on the right. In Figure 2c these have been flipped to show untreated on the left, and Proteinase K treated on the right, to retain consistency with the other blots depicted. The examples shown are from biological replicates n = 3.



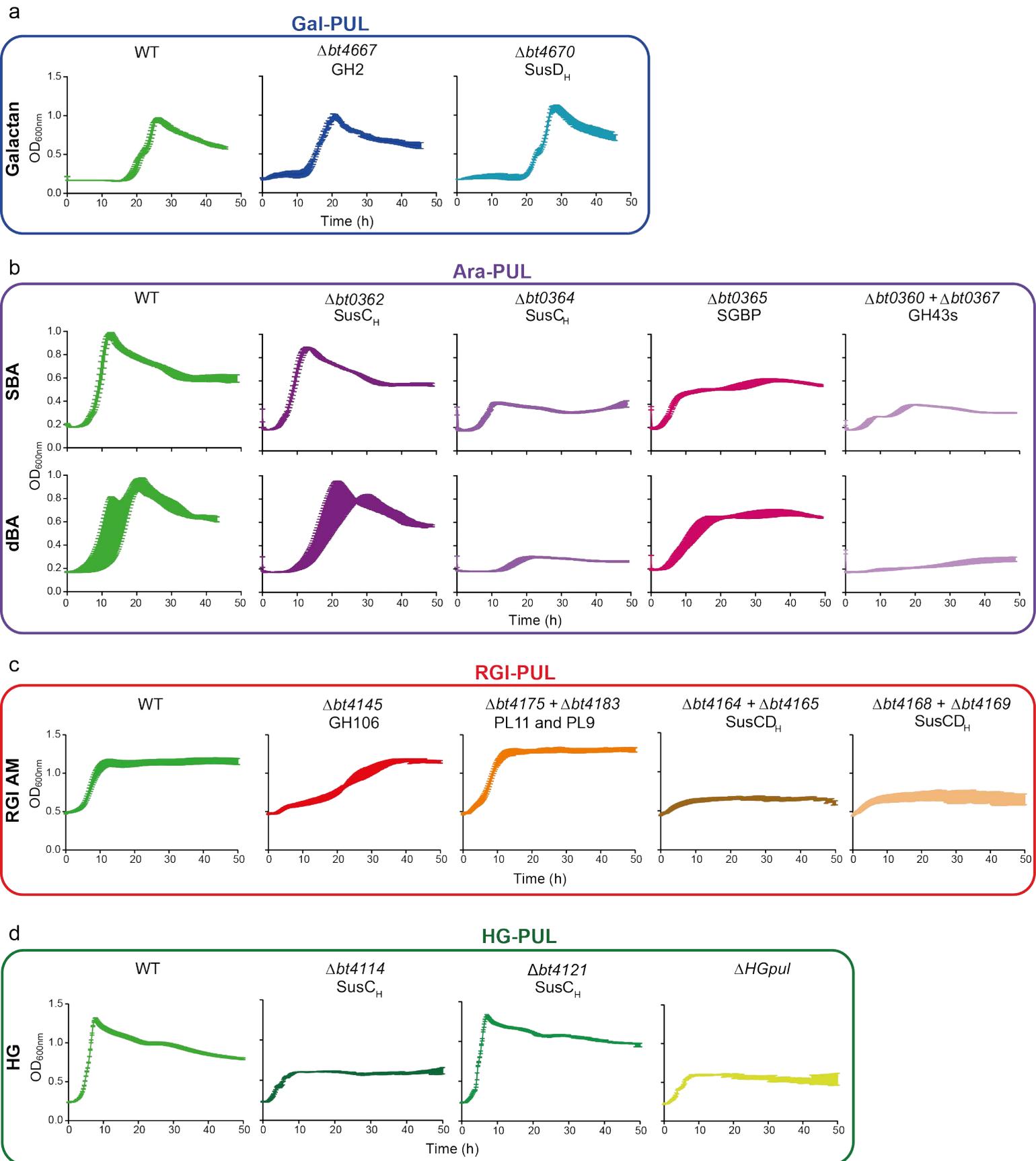
Supplementary Fig. 2. Reaction profile of endo-acting enzymes encoded by HG-PUL, RGI-PUL and Gal-PUL. recombinant enzymes encoded by a, HG-PUL, b, RGI-PUL and c, Gal-PUL were incubated with 5 mg/ml of the cognate polysaccharide in reaction conditions described in Methods. At regular intervals aliquots were removed and analysed by HPAEC, in which carbohydrate detection was by pulsed amperometric detection (PAD) response or through an increase in absorbance at 235 nm ($Abs_{235\text{ nm}}$). Samples were also analysed by TLC. Sugar standards on TLC are labelled on the left hand lane. The concentration of enzyme was 250 nM for BT4115 and BT4116, 30 nM for BT4170, 400 nM for BT4175, 300 nM for BT4183, 500 nM for BT4668 and 1 μ M for BACOVA_05488. The enzymes were either glycoside hydrolases (GHs) or polysaccharide lyases (PLs). The data presented are a representative from biological replicates $n = 3$.

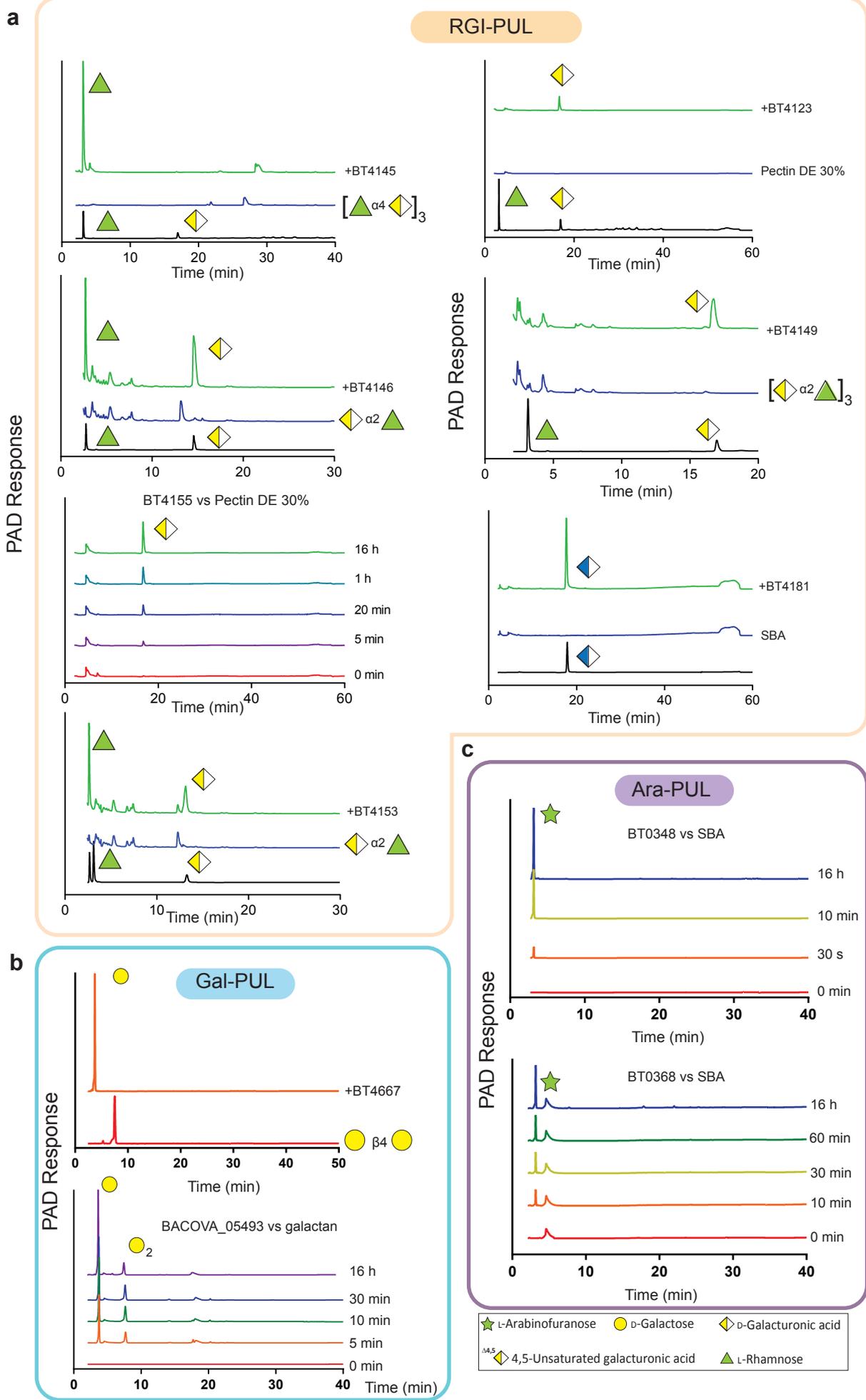


Supplementary Fig. 3. Crystal structure of BT4170. **a**, schematic of BT4170 revealing the β -helix fold that is ramp colored blue at the N terminus to red at the C terminus. **b**, shows a solvent exposed surface representation of BT4170 with reaction products bound in the substrate binding cleft extending from subsites -3 to +3. **c**, an overlay of the structure of the K285A mutant of BT4170 (green) and Pel9A (blue), another member of PL9 (PDB code 1RU4). The appended square is an enlargement of the active site revealing the conservation of the residues that interact with Ca-1 and the lysine that is the proposed catalytic base. In **d** the active site amino acids of the K285 mutant (carbons coloured green) that interact with the bound reaction products (carbons in yellow) are shown with polar contacts depicted by broken black lines. **e**, shows the electron density map ($2F_o - F_c$) of the two reaction products at 1.5σ with the sugars appropriately labelled. The electron density is shown in blue mesh. **f**, overlay of the substrate binding cleft of BT4170 (K285 mutant) and Pel9A revealing amino acids in the equivalent positions in the proteins. In all panels amino acids are shown in stick format with BT4170 coloured green, Pel9A blue and ligand yellow. In **f** the schematic of BT4170 is green and Pel9A blue. Lys285 in **e** and **f** is shown in cyan and is derived from an overlay of the apo structure of BT4170 with the K285A mutant of the enzyme.

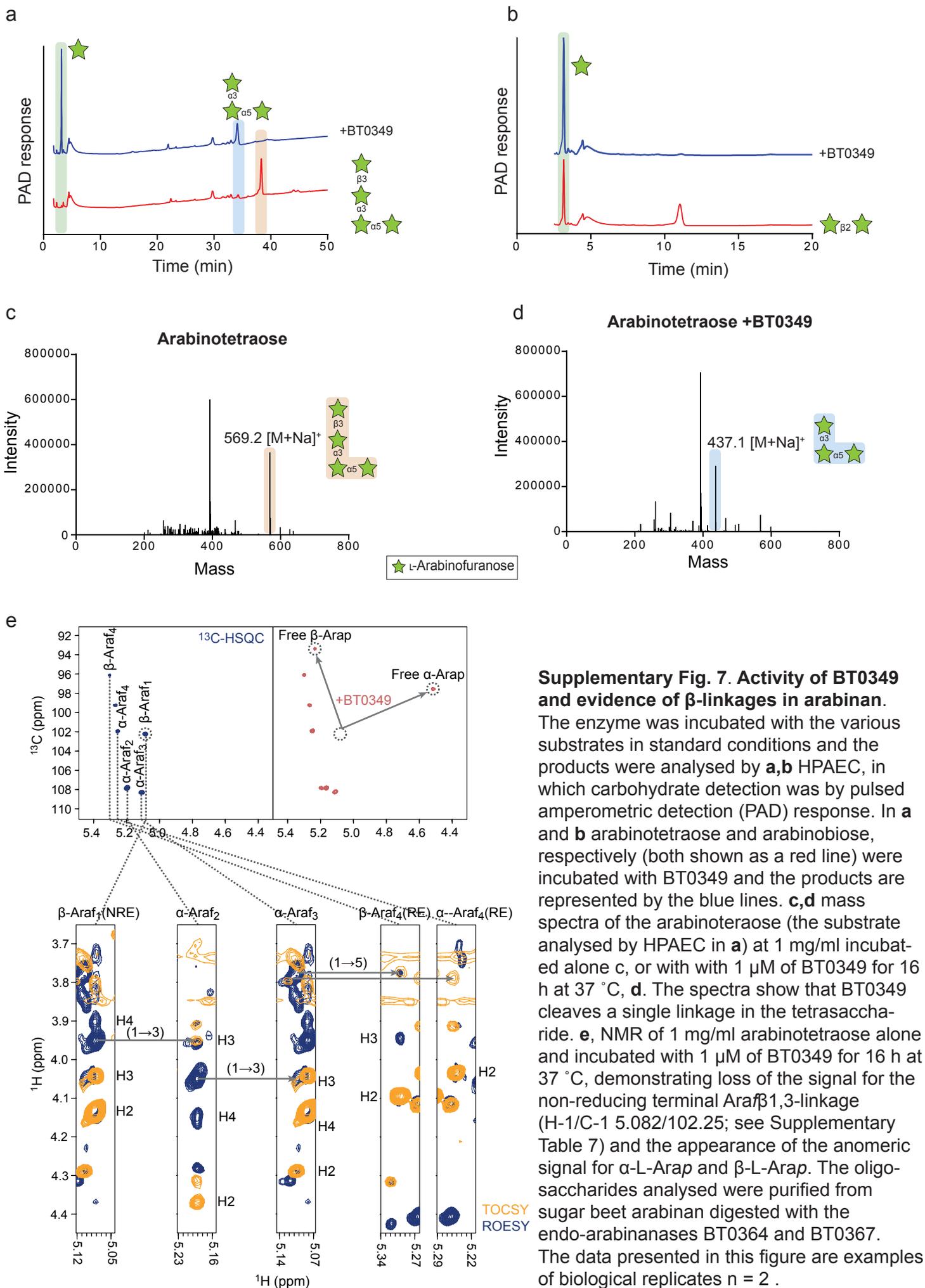


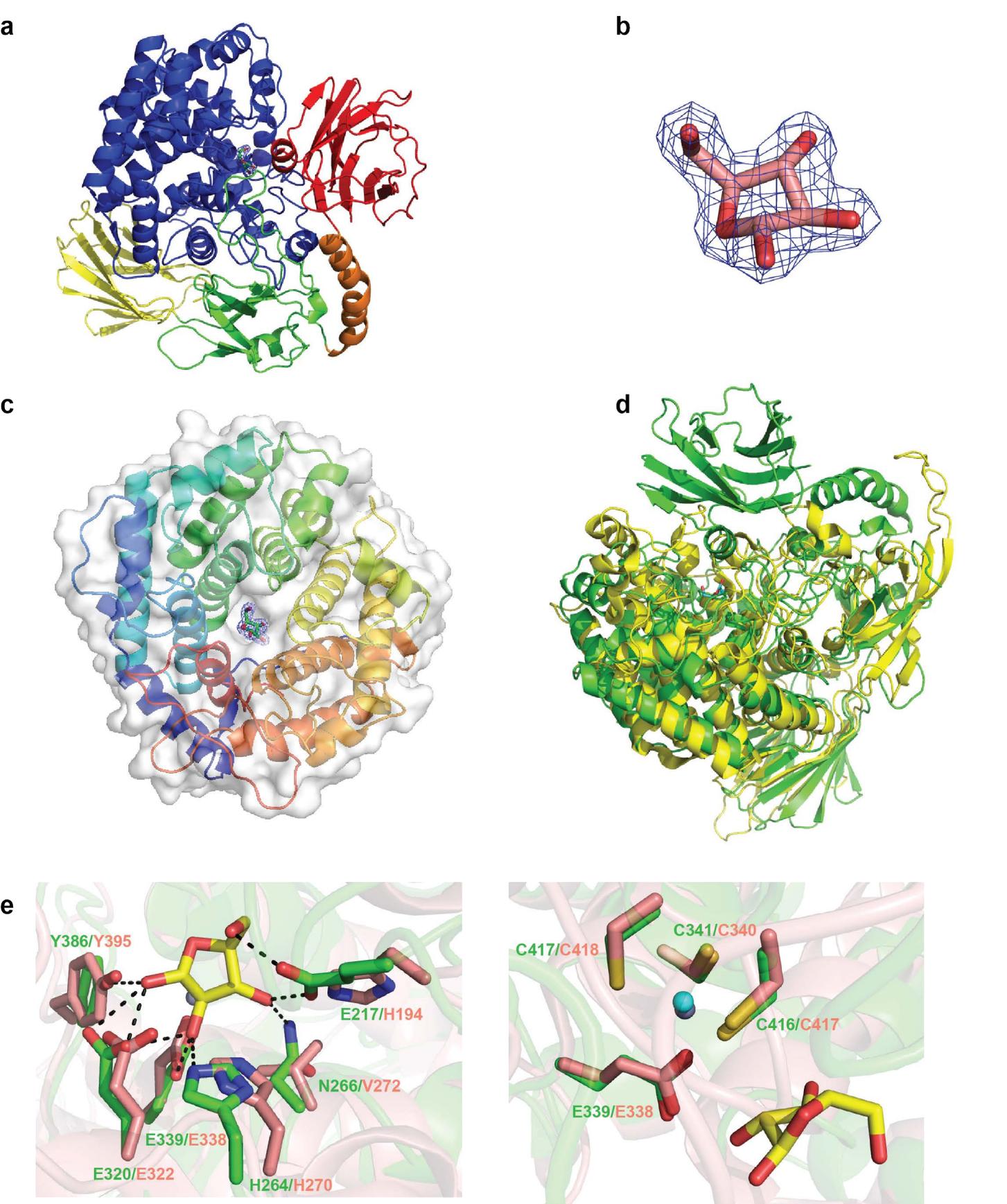
Supplementary Figure 4. Reaction products generated by surface arabinanases. The two surface arabinanases BT0360 and BT0367 were incubated with (a and c) debranched arabinan (dBA) or (b and d) sugar beet arabinan (SBA) at 5 mg/ml in PBS and at time intervals aliquots were removed and analysed by HPAEC in which carbohydrate detection was by pulsed amperometric detection (PAD) response (a and b), and TLC (c and d). BT0360 was at 500 nM for both reactions, while BT0367 was at 4 μM and 2 μM when incubated with SBA and dBA, respectively. The data presented are representative of biological replicates $n = 3$.



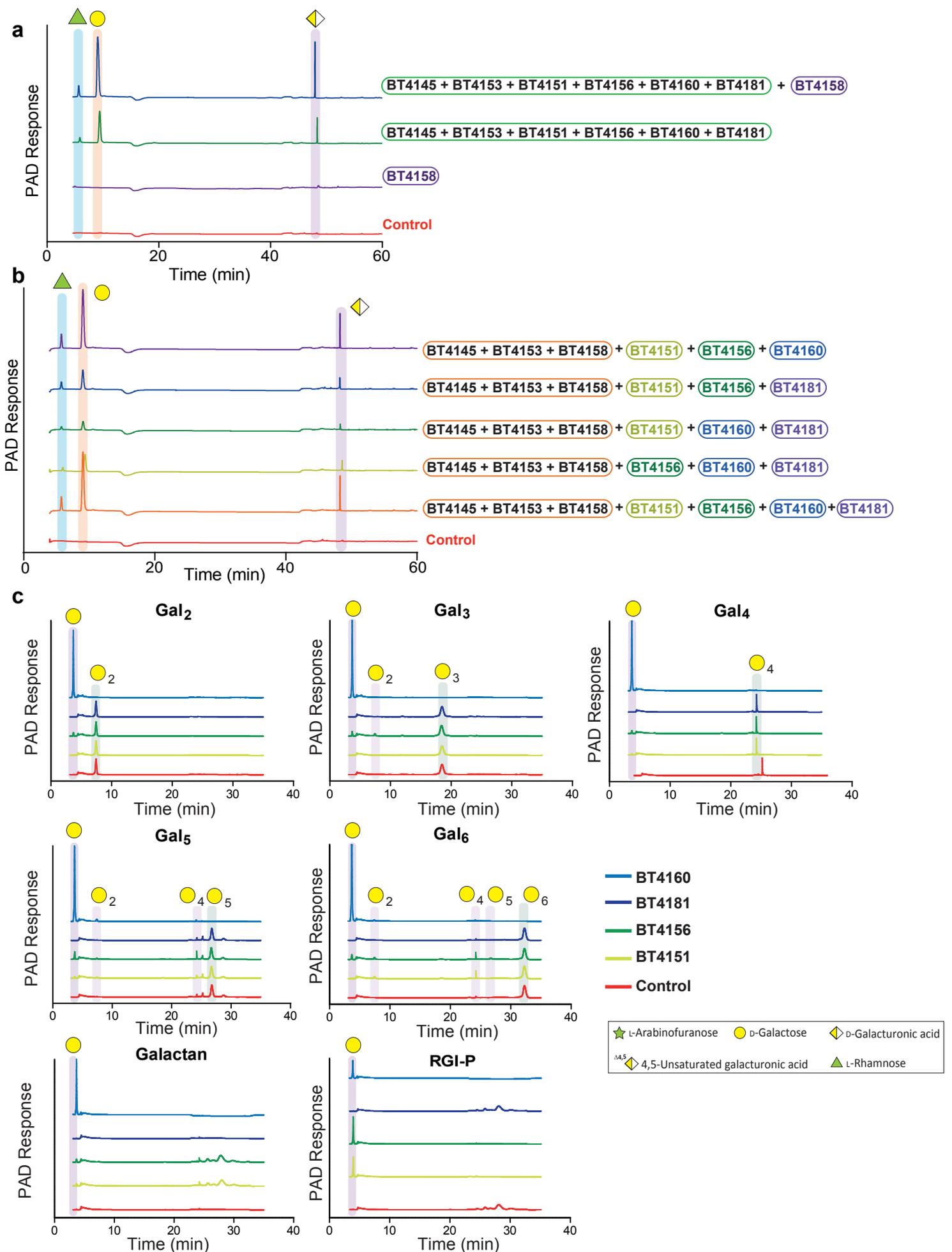


Supplementary Fig. 6. Reaction profile of exo-acting enzymes encoded by the pectic PULs. The recombinant enzymes at 1 μ M, derived from **a**, RGI-PUL, **b**, Gal-PUL and **c**, Ara-PUL were incubated with 5 mg/ml of the cognate polysaccharide or 2 mM of oligosaccharide in reaction conditions described in Methods. The reactions were incubated for 16 h and analysed by HPAEC, detected by pulsed amperometric detection (PAD). In the chromatographs labelled +BTXXXX (where X are numbers) the substrates were incubated with the cognate enzyme. For the *B. thetaiotaomicron* enzyme BT4123 and *B. ovatus* enzyme BACOVA_05493 various time points were analysed. The data presented are representative of biological replicates $n = 2$.

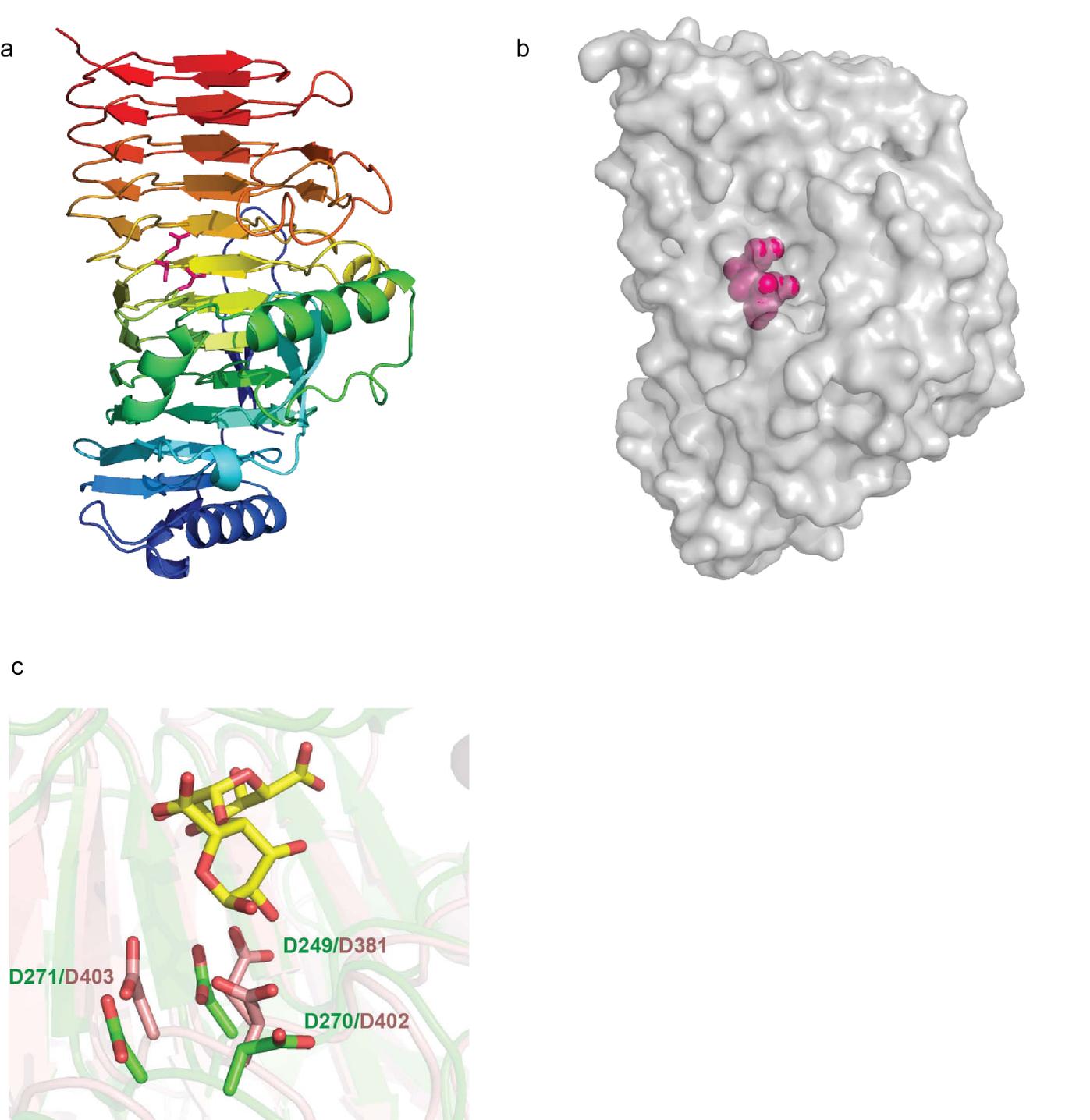




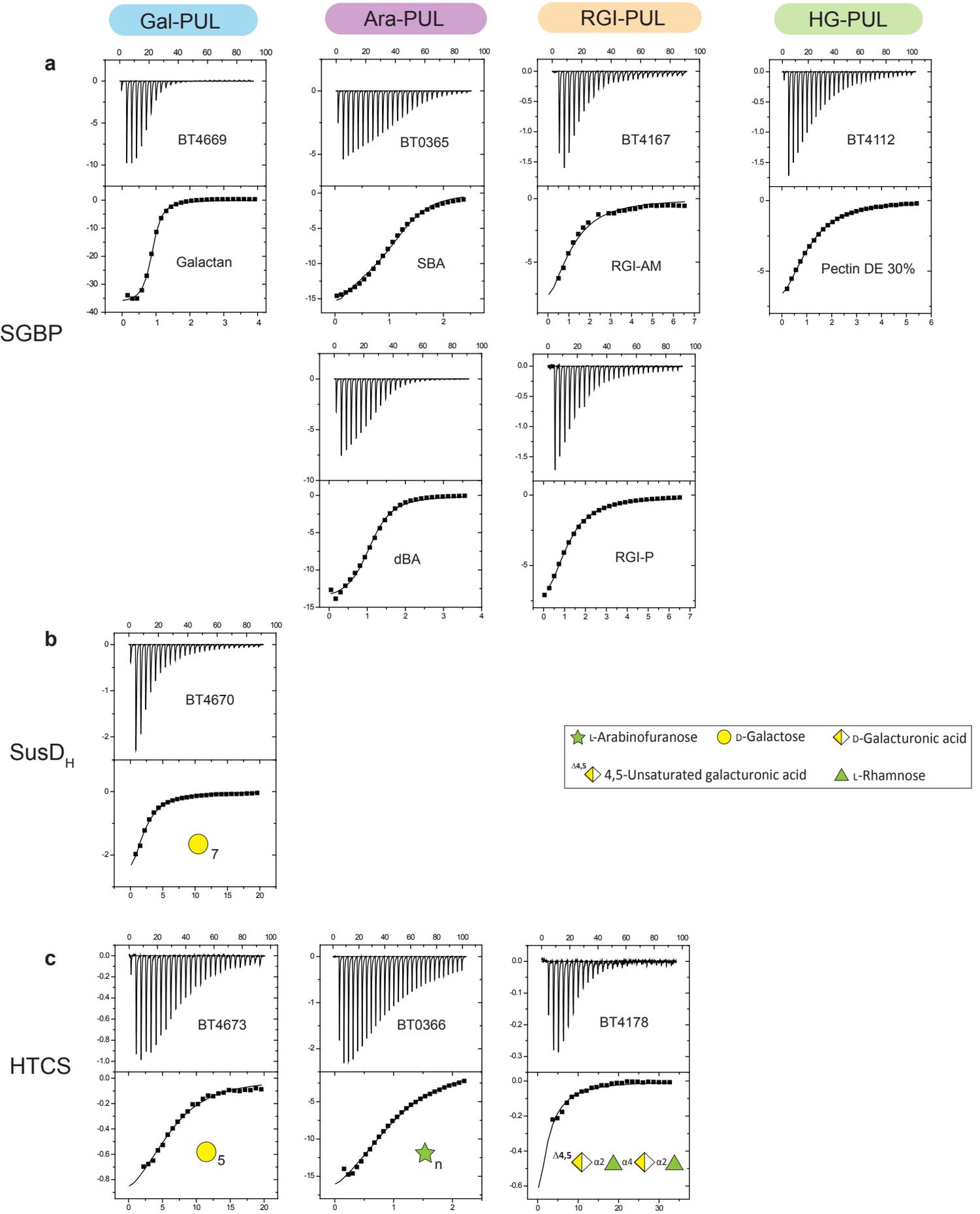
Supplementary Fig. 8. Crystal structure of BT0349. **a**, schematic of BT0349 that is coloured to show in order of sequence in the protein the $(\alpha/\alpha)_6$ barrel catalytic domain (blue) and β -sandwich domains 1 (yellow), 2 (green) and 3 (red). The α -helix connecting β -sandwich domains 2 and 3 is coloured orange. **b**, shows a solvent exposed surface representation of the catalytic domain of BT0349 with arabinose bound in the active site pocket at the centre of the $(\alpha/\alpha)_6$ barrel. **c**, shows the electron density map ($2F_o - F_c$) of the furanose sugar at 1.5σ with the sugars appropriately labelled. The electron density is shown in blue mesh. **d**, an overlay of the structure of BT0349 (green) and BLLJ_0211 (yellow), a member of family GH127 (PDB code 3WKX). In **e** the active site amino acids (carbons coloured green and cyan in BT0339 and BLLJ_0211, respectively) that interact with the bound arabinofuranose products (carbons in salmon pink) are shown with polar contacts depicted by broken black.



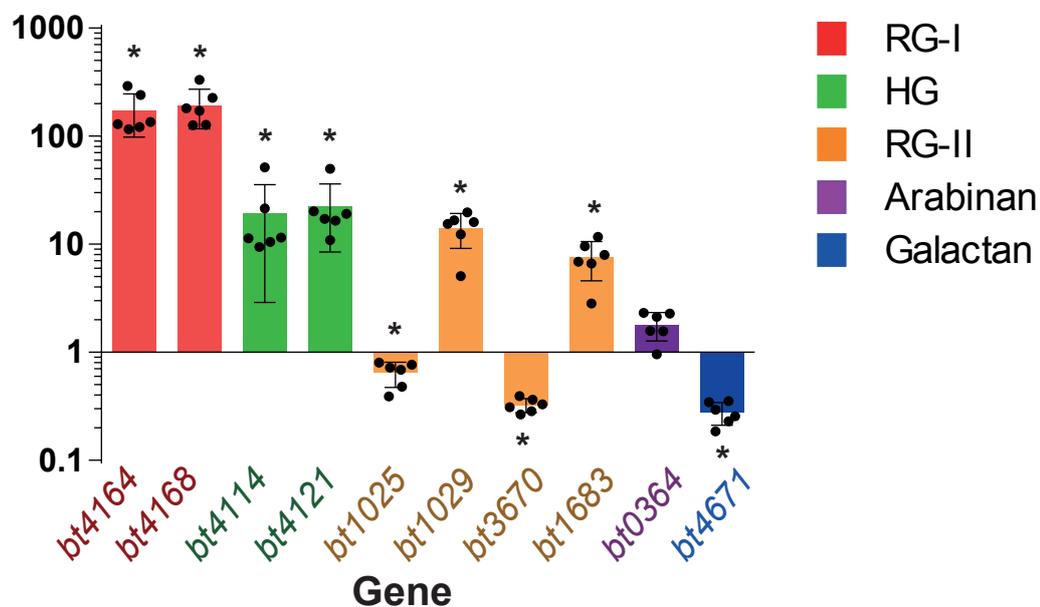
Supplementary Fig. 9. Analysis of the influence of galactosidase and esterase activity on RGI degradation. **a**, assesses the importance of the esterase BT4158 in the degradation of potato RGI (RGI-P). The role of the GH2 and GH35 β -galactosidases in RGI-P deconstruction, **b**, and in the cleavage of β -1,4-galactooligosaccharides **c**, were evaluated. RGI-P and the galactooligosaccharides at 1 mg/ml and 1 mM, respectively, were incubated with 1 μ M of enzyme for 16 h using standard conditions and the reactions were analysed by HPAEC. The data presented are representative of biological replicates $n = 4$



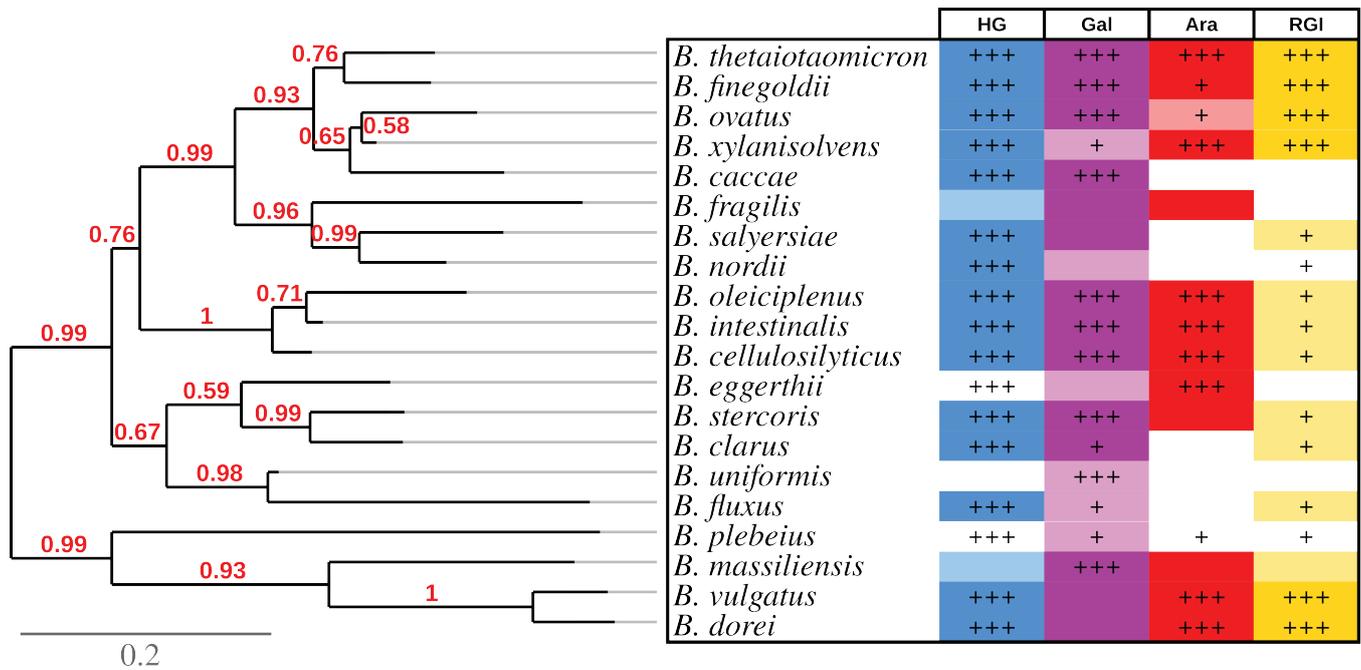
Supplementary Fig. 10. Crystal structure of BT4155. **a**, schematic of BT4155 revealing the β -helix fold that is ramp colored blue at the N terminus to red at the C terminus. **b**, shows a solvent exposed surface representation of BT4170 with putative catalytic residues highlighted in red. **c**, an overlay of the structure of BT4155 (green) with the GH28 exo-polygalacturonidase YeGH28 (salmon pink) bound to di-galacturonic acid shown in yellow (PDB code 2UVF).



Supplementary Fig. 12. Evaluation of ligand binding of HTCS proteins, SGBPs and a SusD_H by isothermal titration calorimetry. Representative isothermal titration calorimetry (ITC) data of SGBP (a), SusD_H (b) and HTCS (c) proteins titrated with carbohydrates as indicated. The relevant PUL is indicated at the top of the figure. The soluble ligand (4 mg/ml for polysaccharide and ~10 mM for oligosaccharide in the syringe) was titrated into 50 μM of the appropriate protein in the cell. The top half of each titration shows the raw injection heats; the bottom half, the integrated peak areas fitted using a single-site model (MicroCal Origin v7.0). ITC was carried out in 50 mM Na-HEPES pH 7.5 at 25 °C. The data presented are from biological replicates n = 2



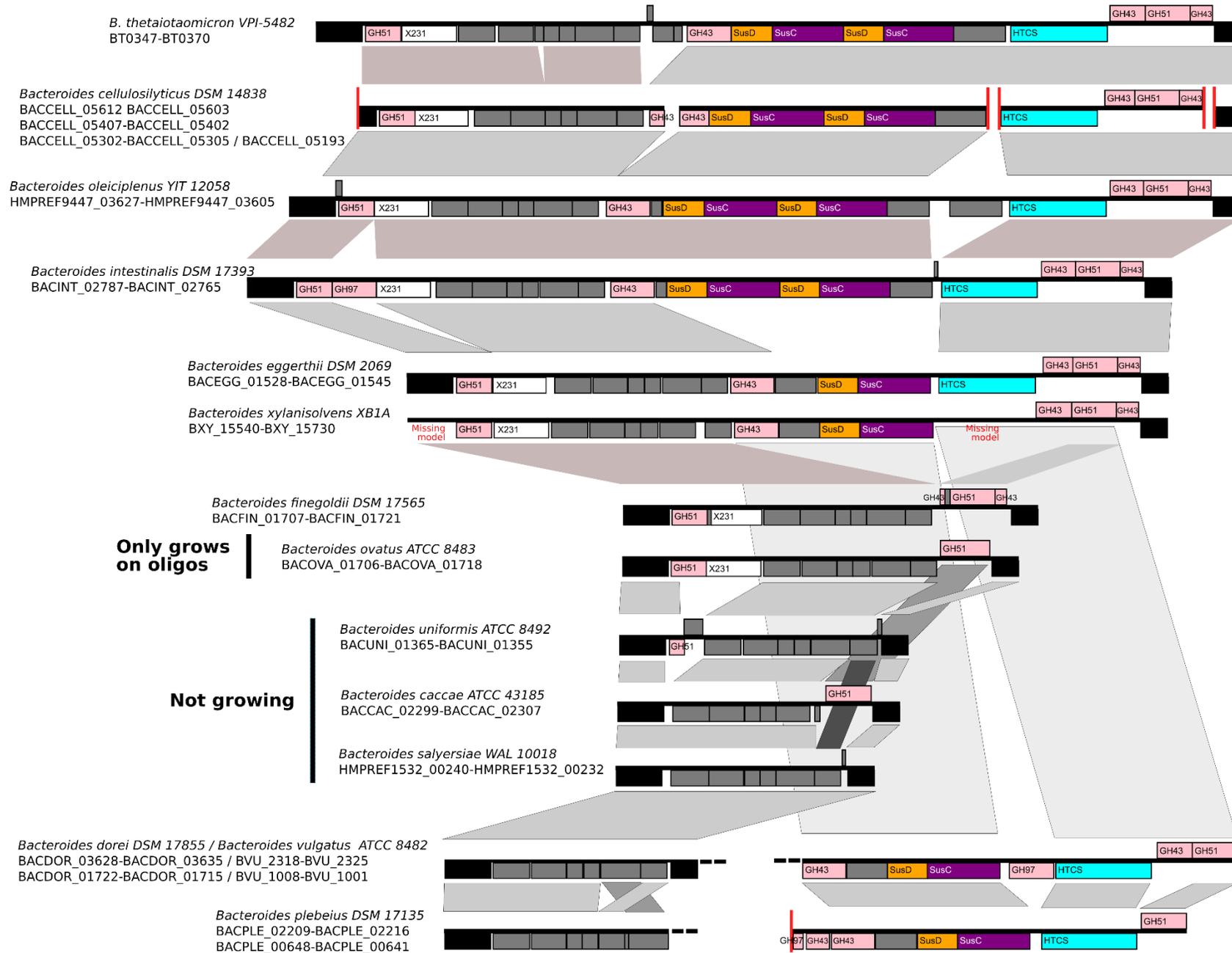
Supplementary Fig. 13. Analysis of the expression of susC homologues of the pectic PULs when *B. thetaiotaomicron* was cultured on RGI-AM. The bacterium was grown to mid-exponential phase on RGI-AM of glucose and RT-qPCR was used to determine the ratio of transcripts of the susCH genes in the two cultures. The colour of the genes corresponds to RGI-PUL (red), HG-PUL (green), RGII-PUL (orange), Ara-PUL (purple) and Gal-PUL (blue). * indicates that the transcripts of the susC_H genes from the pectin culture were significantly up- or down-regulated compared to the glucose culture with p values from t-test: bt4164 - 0.0002, bt4168 - 0.0001, bt4114 - 0.021, bt4121 - 0.0038, bt1025 - 0.041, bt1029 - 0.0008, bt3670 - 0.029, bt1682 - 0.0004, bt0364 - 0.079, bt4671 - 0.00003. The error bars in the bar chart are standard deviations



Supplementary Fig. 14. Relationship between phylogeny, PUL composition and the growth of HGM *Bacteroides* species on pectins. Growth is represented by colours (dark colour: growth on polysaccharides; light colour: growth only on oligosaccharides; white: no growth). The presence of corresponding PULs in each genome is indicated in each cell: (+++) highly similar to *B. thetaiotaomicron*, (+) distantly related. A void cell indicates that no PUL corresponding to *B. thetaiotaomicron* could be identified in the genome. The phylogenetic tree was reconstructed at <http://phylogeny.fr34> with branch length represented in black, bootstrap values for 100 replicates labelled in red and a bottom scale indicating the mean number of nucleotide substitutions per site in grey. The growth data presented are from biological replicates n = 4.

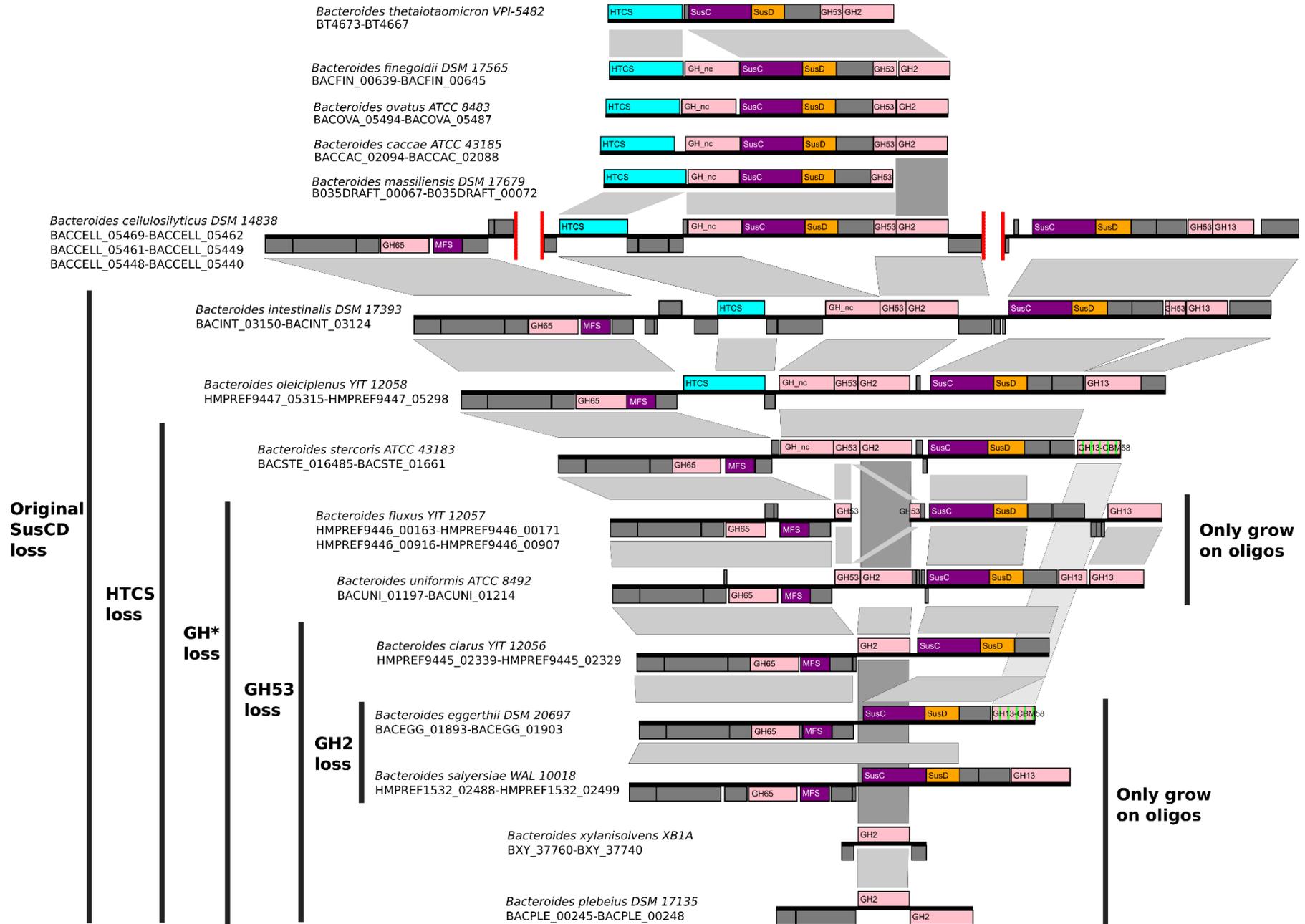
a

Arabinan PULs



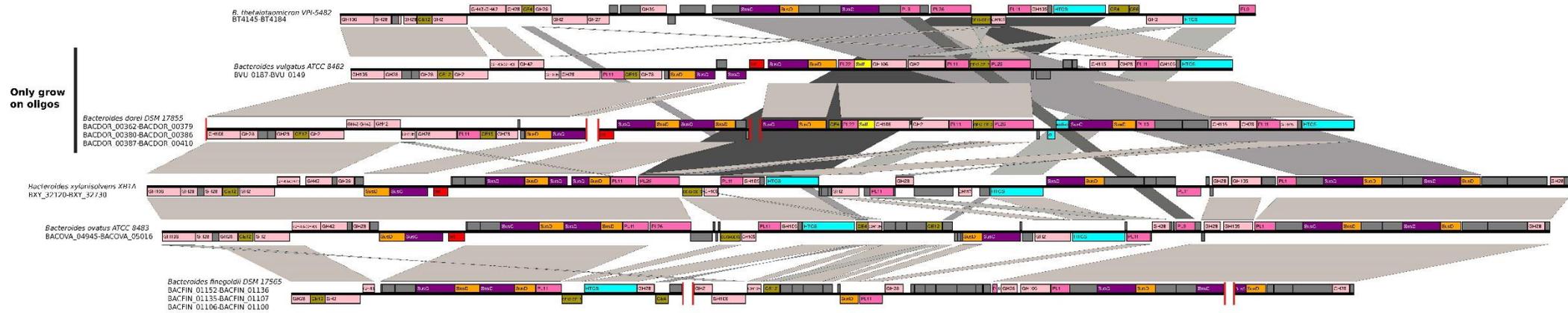
b

Galactan PULs

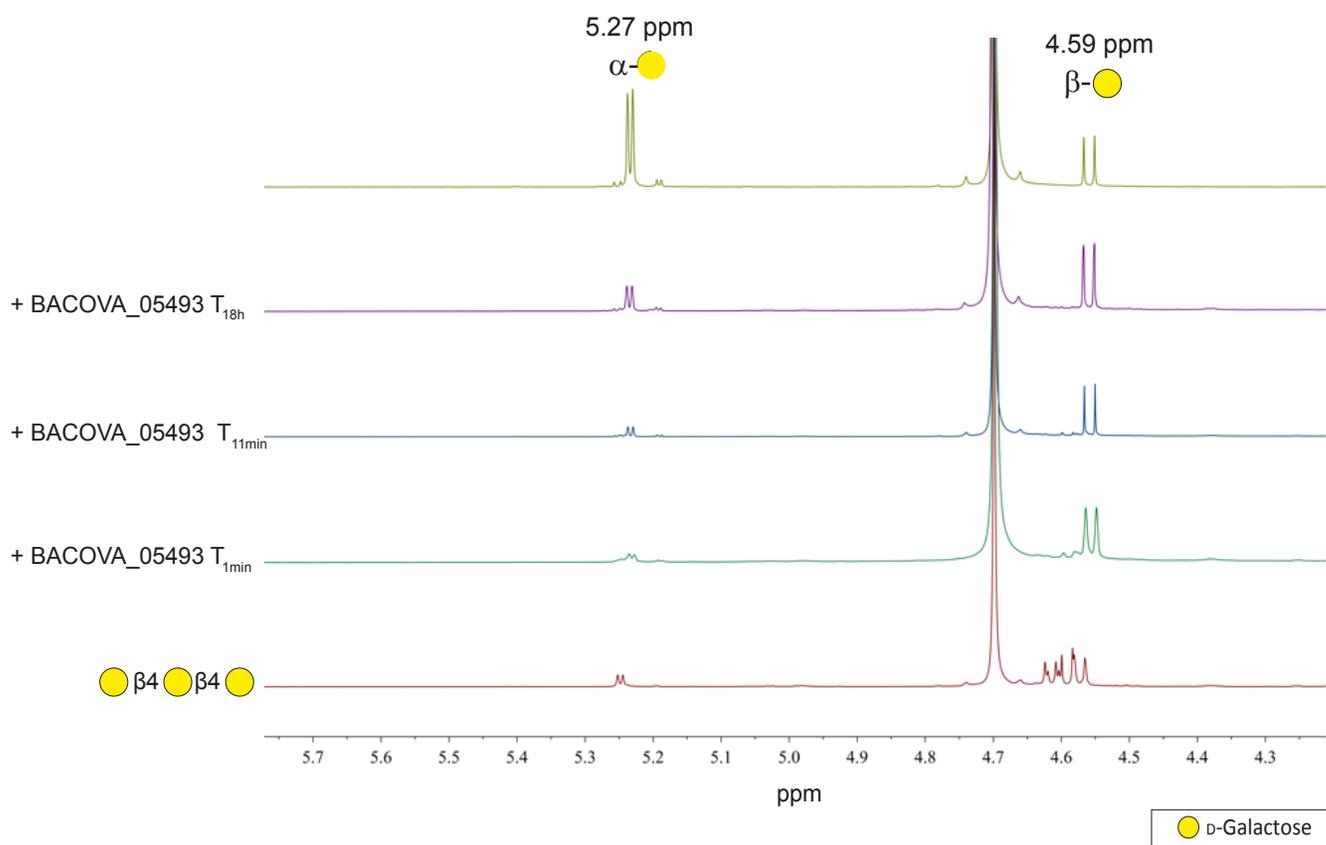


d

RGI PULs



Supplementary Fig. 15: Genomic view of *B. thetaiotaomicron* pectin PULs and syntenic regions in the species analyzed in this work. PUL protein-coding genes are represented by rectangles colored according to their functional modules: GH in light pink, PL in dark pink, CBM in green, CE in brown, HTCF and ECF- σ /anti- σ regulators in light blue, MFS and SusC-like transporters in purple, SusD-like outer membrane proteins in orange, ROK (regulators, ORFs, kinases) in blue, sulfatases in yellow, peptidases in gold, and integrases in red, and grey otherwise. Conserved proteins flanking the PUL, but not part of the PUL, are depicted by black rectangles. Genes are represented either above or below the central black line according to the coding strand. Every represented species grew on the corresponding polysaccharide, unless otherwise stated by 'growth only on oligosaccharide' or 'no growth'. Grey polygons outline the most conserved homologues and likely genomic rearrangements between species.



Supplementary Fig. 16. NMR analysis of the activity of BACOVA_05493. The enzyme at 81 μM was incubated with 16 mM β 1,4-galactotriose in standard conditions in a solvent of D_2O . Spectra were recorded at the times indicated. The data presented are representative from $n = 2$ biological replicates.