



Metabolism of a hybrid algal galactan by members of the human gut microbiome

Craig S. Robb^{1,2,8}, Joanne K. Hobbs^{1,8}, Benjamin Pluvinage¹, Greta Reintjes^{3,4}, Leeann Klassen³, Stephanie Monteith³, Greta Giljan⁴, Carolyn Amundsen³, Chelsea Vickers^{1,7}, Andrew G. Hettle¹, Rory Hills¹, Nitin⁵, Xiaohui Xing³, Tony Montana⁶, Wesley F. Zandberg⁵, D. Wade Abbott³ and Alisdair B. Boraston¹✉

Native porphyran is a hybrid of porphyran and agarose. As a common element of edible seaweed, this algal galactan is a frequent component of the human diet. Bacterial members of the human gut microbiota have acquired polysaccharide utilization loci (PULs) that enable the metabolism of porphyran or agarose. However, the molecular mechanisms that underlie the deconstruction and use of native porphyran remains incompletely defined. Here, we have studied two human gut bacteria, porphyranolytic *Bacteroides plebeius* and agarolytic *Bacteroides uniformis*, that target native porphyran. This reveals an *exo*-based cycle of porphyran depolymerization that incorporates a keystone sulfatase. In both PULs this cycle also works together with a PUL-encoded agarose depolymerizing machinery to synergistically reduce native porphyran to monosaccharides. This provides a framework for understanding the deconstruction of a hybrid algal galactan, and insight into the competitive and/or syntrophic relationship of gut microbiota members that target rare nutrients.

Our understanding of carbohydrate biomass metabolism by bacteria in the environment and animal gut microbiota is seeing unprecedented advancement (see, for example, refs. ^{1,2}). Despite these advances, the processes that microbes use to break down and metabolize sulfated polysaccharides, particularly those of marine origin, remain an area of microbial glycobiology that is underexplored. Sulfated polysaccharides are particularly abundant in marine algae where they represent an abundant carbon sink that is recycled by the action of microbes^{3,4}. Indeed, carrageenan and ulvan have provided models for the microbial turnover of sulfated algal polysaccharides by marine bacteria^{5–7}. Some of these chemically distinct groups of polysaccharides are normal dietary components of humans⁸. As such, the metabolic adaptability and diversity of the human gut microbiome is further demonstrated by its bacterial members that have acquired the capacity to use the unusual polysaccharides from marine algae. This is enabled by polysaccharide utilization loci (PUL) that are proposed to have been laterally acquired from marine bacteria^{9–13}. Such rare nutrients, like porphyran from marine algae of the *Porphyra* genus, are sometimes referred to as ‘privileged’ nutrients and can be used for selection of specific gut bacteria that possess the ability to catabolize them^{13–15}.

Porphyran and agarose are closely related red algal galactans that both comprise a backbone of modified L-galactose residues linked α -1,3 to D-galactose, with this repeating disaccharide unit joined by β -1,4-glycosidic linkages (Extended Data Fig. 1a,b)¹⁶. The defining difference between the two polysaccharides is that the L-galactose in agarose is 3,6-anhydro-L-galactose (LA) but it is L-galactose-6-sulfate (L6S) in porphyran. This reflects the common biosynthetic route of the two polysaccharides whereby porphyran

is synthesized first then converted to agarose by the cyclization of L6S to LA via a ‘sulfate eliminase’ catalyzed reaction^{17,18}. Complete conversion of porphyran would result in a homogeneous polysaccharide of agarose. However, naturally occurring polysaccharides containing porphyran are typically hybrids, comprising blocks of porphyran and blocks of agarose (Extended Data Fig. 1c), presumably by incomplete conversion. We refer to this naturally occurring hybrid polysaccharide as native porphyran. The relative content of porphyran and agarose in native porphyran varies by the source of the polysaccharide^{3,19}. Although native porphyran is an ecologically relevant polysaccharide and a common component of the human diet (for example, nori), many details regarding the mechanism by which microbes completely deconstruct this sulfated hybrid polysaccharide are unresolved.

Given the evidence of gene acquisition from marine bacteria, human gut bacteria have made good models for understanding the overall features of algal polysaccharide degradation. These models have also provided unique insights into the evolution of the diverse metabolic capabilities of the gut microbiome. In this respect, two relevant bacteria are *B. uniformis* (strain NP1) and *B. plebeius* (strain DSM 17135) whose genomes encode an agarose PUL (AgaPUL) and a porphyran PUL (PorPUL), respectively (Extended Data Fig. 1d and Supplementary Table 1)^{10,20}. The proteins encoded by these PULs endow both strains with the capacity to grow on native porphyran; *B. uniformis* can grow on agarose as well but *B. plebeius* cannot²⁰. Biochemical analysis of some of the PUL-encoded gene products, however, indicate that both PULs have unexplained activities that are consistent with overlapping agarose and porphyran metabolism (that is, both PULs encode agarases

¹Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia, Canada. ²Department of Biochemistry and Molecular Biology and Centre for Blood Research, The University of British Columbia, Vancouver, British Columbia, Canada. ³Lethbridge Research and Development Centre, Agriculture and Agri-Food Canada, Lethbridge, Alberta, Canada. ⁴Max Planck Institute for Marine Microbiology, Bremen, Germany. ⁵Irving K Barber Faculty of Science, The University of British Columbia, Department of Chemistry, Kelowna, British Columbia, Canada. ⁶Southern Alberta Genome Sciences Centre, Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, Alberta, Canada. ⁷Present address: School of Biological Sciences, Victoria University, Wellington, New Zealand. ⁸These authors contributed equally: Craig S. Robb, Joanne K. Hobbs. ✉e-mail: boraston@uvic.ca

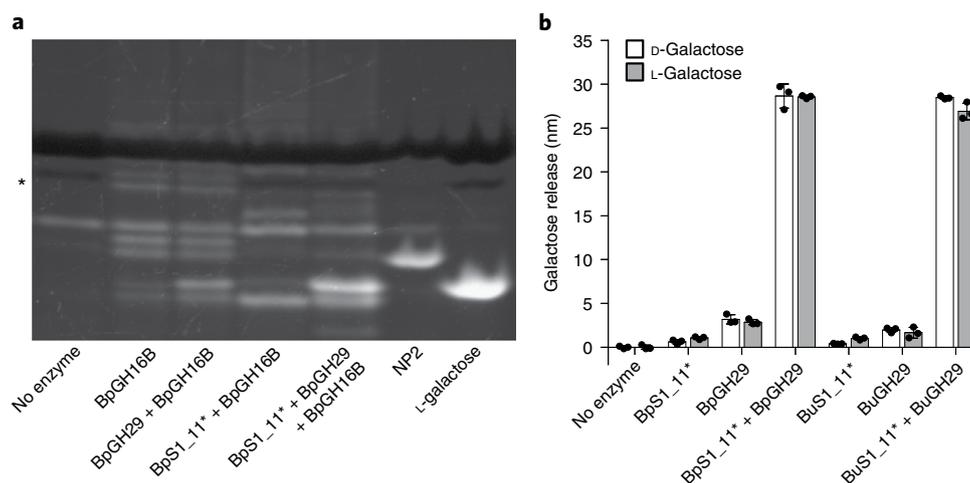


Fig. 1 | Identification of initiating exo-porphyrinase enzymes. **a**, Fluorophore-assisted carbohydrate electrophoresis analysis of porphyrin digestion by the *endo*- β -porphyrinase BpGH16B in combination with BpS1_11* and BpGH29. This gel is representative of independent experiments performed three times ($n=3$) with reproducible results. **b**, Hydrolysis of 30 nM of neoporphyrinose (NP2) by the S1_11* and GH29 from both *B. plebeius* and *B. uniformis* as detected by L- and D-galactose release. $n=3$ independent reactions were performed with each replicate shown along with the mean \pm standard deviation (s.d.).

and porphyrinases) (Supplementary Table 1)^{10,20,21}. We therefore hypothesized that both PULs are equipped to target the complete depolymerization of agarose and porphyrin in the context of more complex native porphyrin. Through more detailed analysis of the AgaPUL and PorPUL components, we determined the biochemical pathway for native porphyrin depolymerization. Further, the results reveal the potentially synergistic actions of agarose and porphyrin active enzymes that are present in both PULs to depolymerize native porphyrin. Thus, the PULs have a shared dual capacity to maximize the yield of monosaccharides from native porphyrin. However, *B. uniformis* and *B. plebeius* are unable to use homogeneous versions of porphyrin and agarose, respectively, due to deficiencies in binding, sensing and/or uptake of these specific polysaccharides, which has implications toward their selective maintenance in a microbial population.

Results

Identification of an L-galactose dehydrogenase (LGDH).

The PorPUL possesses a gene having 83% amino acid sequence identity to a characterized *Bacteroides vulgatus* LGDH, which performs the NADP⁺ dependent oxidation of L-galactose to L-galactono-1,5-lactone (ref. 22). We assayed the PorPUL ortholog, called BpLGDH, which like the *B. vulgatus* enzyme displayed specificity for L-galactose (Extended Data Fig. 2a) and NADP⁺ as a cofactor (Extended Data Fig. 2b–d). Thus, the PorPUL can process L-galactose that would be liberated from porphyrin while providing BpLGDH as a specific tool to quantitatively detect free L-galactose.

The BpLGDH crystallized as a dimer with twofold symmetry having a buried interface of 2,153 Å² that is predicted to be stable in solution (Extended Data Fig. 3a). Each monomer adopts the (α/β)₈-barrel fold of the aldo-keto reductase superfamily. The active site was identified by a conserved catalytic histidine at the center of the β -barrel (Extended Data Fig. 3b). The most structurally similar protein is AKR11C1 from *Bacillus halodurans*, which reduces the aldehyde of the substrate 4-hydroxy-2,3-*trans*-nonenal in an NADPH dependent manner (root mean square deviation of 2.37 Å over 241 aligned residues, Extended Data Fig. 3c)²³. Several of the active site residues are conserved between the two proteins, suggesting shared features in their mechanisms of aldehyde substrate recognition and oxidation (Extended Data Fig. 3c).

GH29 and S1_11 activity. The PorPUL contains a single gene encoding a protein (BpS1_11) that is classified into the S1 sulfatase family, subfamily 11, of the SulfAtlas classification²⁴. An adjacent gene encodes a predicted family 29 glycoside hydrolase (BpGH29) of the carbohydrate-active enzyme (CAZy) classification²⁵. We examined the activities of recombinant BpGH29 and an S83C mutant of BpS1_11 (denoted as BpS1_11*, generated to allow maturation of the cysteine to a formylglycine catalytic residue on coexpression of a formylglycine-generating enzyme (FGE)) on porphyrin in combination with the β -porphyrinase BpGH16B (ref. 20) (Fig. 1a). BpGH16B and BpGH29 together did not affect the product profile compared with BpGH16B alone. Combining BpGH16B and BpS1_11* changed the product profile consistent with the disappearance of neoporphyrinose (NP2, L6S-G where G refers to D-galactose). Treatment with all three enzymes increased the amount of a product with migration properties the same as those of L-galactose. The sequential action of BpGH29 and BpS1_11*, as well as recombinant versions of the BuGH29 and BuS1_11* (S80C catalytic residue mutation of BuS1_11) homologs from the AgaPUL, were further investigated by hydrolysis of purified NP2 (Fig. 1b, see Methods for the structural characterization of NP2). Only the cotreatment with enzyme pairs (S1_11* with GH29), not the enzymes alone, resulted in the quantitative production of L- and D-galactose: roughly 28 nM of each compared with the expected 30 nM of input material. These results are consistent with S1_11* performing desulfation of the L-galactose-6-sulfate in NP2 before hydrolysis of the α -1,3-glycosidic bond by GH29.

Structure of an exo-6-sulfo-L-galactose sulfatase. Neither BpS1_11* nor BuS1_11* was able to hydrolyze *p*-nitrophenyl sulfate, indicating the necessity of the carbohydrate portion of the substrate for substrate recognition and hydrolysis. The kinetic parameters for the desulfation of NP2 by BpS1_11* yielded K_m and k_{cat} values of 598 (± 50) μ M and 3.8 (± 0.1) s⁻¹, respectively (Extended Data Fig. 4a). We note, however, that the k_{cat} value presumes the presence of completely matured and active sulfatase.

The structure of native BuS1_11 (that is, nonmutated and therefore not catalytically matured) revealed the typical S1 sulfatase two-domain organization that comprises a conserved $\alpha/\beta/\alpha$ alkaline phosphatase fold and a C-terminal accessory domain of mixed

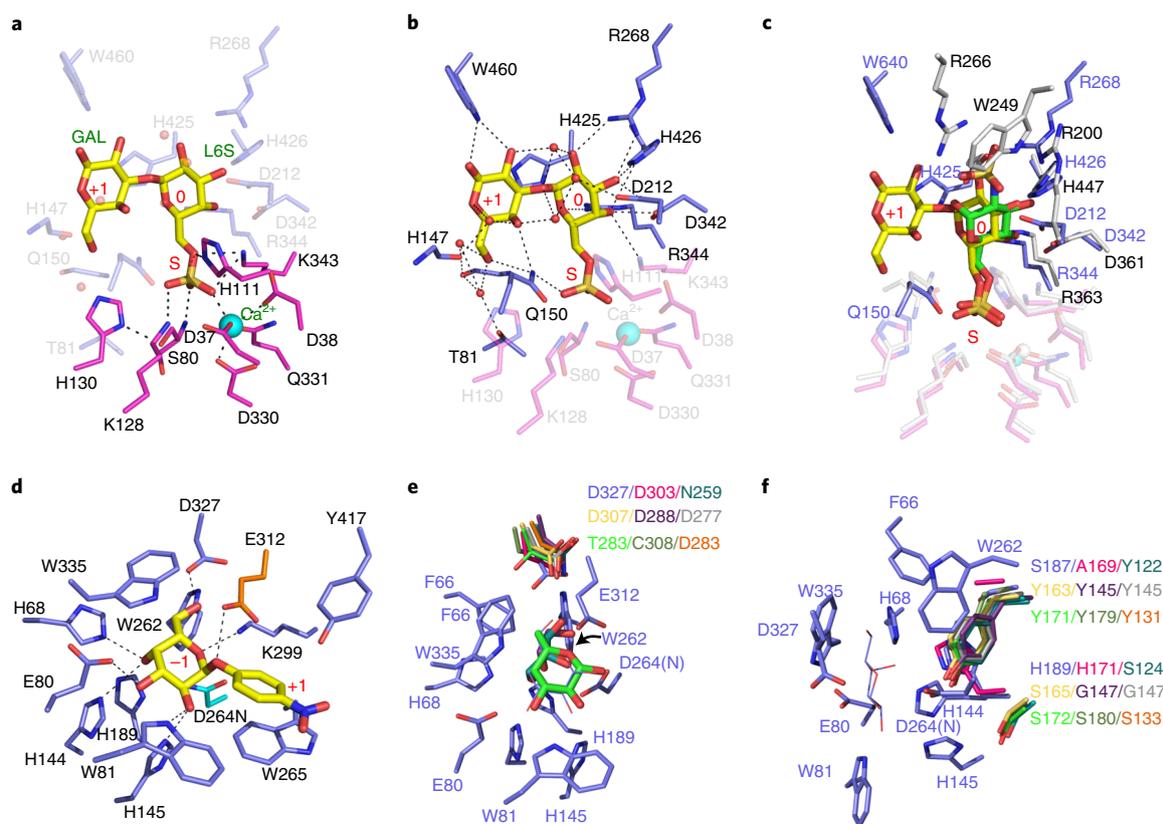


Fig. 2 | Structural analysis of initiating exo-porphyrinase enzymes. **a**, View of the BpS1₁₁* active site focusing on the sulfate coordinating residues (magenta), including the catalytic residues that are conserved in the S1 family. **b**, View of the BpS1₁₁* active site focusing on the carbohydrate coordinating residues (blue) that define the unique specificity of this sulfatase. **c**, A structural overlap of the BpS1₁₁* (magenta, blue and yellow) active site with BT4656 (gray and green, PDB 5G2V)²⁸. **d**, Two views of the BpGH29 active site focusing on substrate coordinating residues. The putative acid-base is shown in orange and the putative nucleophile, which would otherwise be an aspartic acid, in cyan. **e**, the BpGH29 active site showing highly conserved residues involved in the recognition of the C6 group (shown as colors). **f**, the BpGH29 active site showing variable residues in the active site that are potentially involved in L-fucose and L-galactose discrimination (shown as colors). The color coding is as follows: blue, BpGH29; magenta, FgFCO1 from *Fusarium graminearum* (PDB 4PSR); teal, GH29 from unknown rumen bacterium (PDB 5K9H); yellow, BACOVA_04357 from *Bacteroides ovatus* (PDB 4ZRX); purple, BT3798 from *Bacteroides thetaiotaomicron* (PDB 3GZA); gray, BT2192 from *B. thetaiotaomicron* (PDB 4OUE); light green, GH29 from *Thermotoga maritima* (PDB 1ODU); dark green, BT2970 from *B. thetaiotaomicron* (PDB 2WVU) and orange, GH29 from *Bifidobacterium longum* (PDB 3UET).

α/β topology²⁶ (Extended Data Fig. 4b). A complex of BuS1₁₁ with NP2 (Extended Data Fig. 4c) showed the L6S residue occupied the bottom of the active site with the O4 hydroxyl buried and pointing to the interior of the protein, and the sulfate group completely sequestered in an adjoining pocket (Fig. 2a and Extended Data Fig. 4d). The catalytic center contained the conserved calcium ion and S1 sulfatase catalytic machinery, including the position of the proto-catalytic nucleophile Ser⁸⁰ and catalytic acid His¹¹¹ (Fig. 2b). Recognition of the L- and D-galactose portions of the substrate in the 0 and +1 subsites, respectively (using the subsite nomenclature of Hettle et al.²⁷), occurs through a specific network of direct and water-mediated hydrogen bonds (Fig. 2b). The active site architecture and the mode by which L6S occupies it together indicate that BuS1₁₁ uses an *exo*-mode of sulfatase activity, whereby the enzyme can act on the terminal sulfate of the nonreducing end of porphyrin chains regardless of chain length.

BuS1₁₁ displays 41% amino acid sequence identity and a root mean square deviation of 1.231 Å (over 420 matched Ca atoms) with BT4656 (Protein Data Bank (PDB) ID 5G2V), an S1₁₁ sulfatase with *exo*-2-N,6-O-disulfo-D-glucosamine-6-sulfate 6-O-sulfohydrolase activity²⁸. S-subsites of both enzymes are highly conserved (Extended Data Fig. 4e). The zero subsites, however, which accommodate the sugar moieties bearing the sulfate group,

display differences consistent with recognizing structurally distinct carbohydrates (for example, L- versus D-). A set of sidechains that are structurally conserved between the two sulfatasases recognize O3 and O4 of the sugars in the 0 subsite, despite the monosaccharides having axial and equatorial O4 groups, but accommodation of the remaining portions of the substrates is different. The recontouring of each enzyme active site via the presence of inserted structural elements that contribute amino acid sidechains to the active site influences the variability of the carbohydrate-specific subsites (Extended Data Fig. 4e).

Molecular basis of *exo*- α -L-galactosidase activity. Neither BpGH29 nor BuGH29 showed activity on *p*-nitrophenyl- α -L-fucopyranoside (pNP- α -L-Fuc) but both were active on *p*-nitrophenyl- α -L-galactopyranoside (pNP- α -L-Gal). Kinetic analysis of BpGH29 on pNP- α -L-Gal yielded a pH optimum between 5 and 6, and K_m and k_{cat} values of 0.22 (\pm 0.02) mM and 0.79 (\pm 0.02) s⁻¹, respectively, at pH 5.6 (Supplementary Fig. 3).

The structure of BpGH29 revealed the (α/β)₈ catalytic domain and a two- β -sheet accessory domain arrangement that is typical of family 29 glycoside hydrolases (Extended Data Fig. 5a). The loop containing the catalytic acid/base residue, Glu³¹², was retracted from the active site of BpGH29 (Extended Data Fig. 5a). The

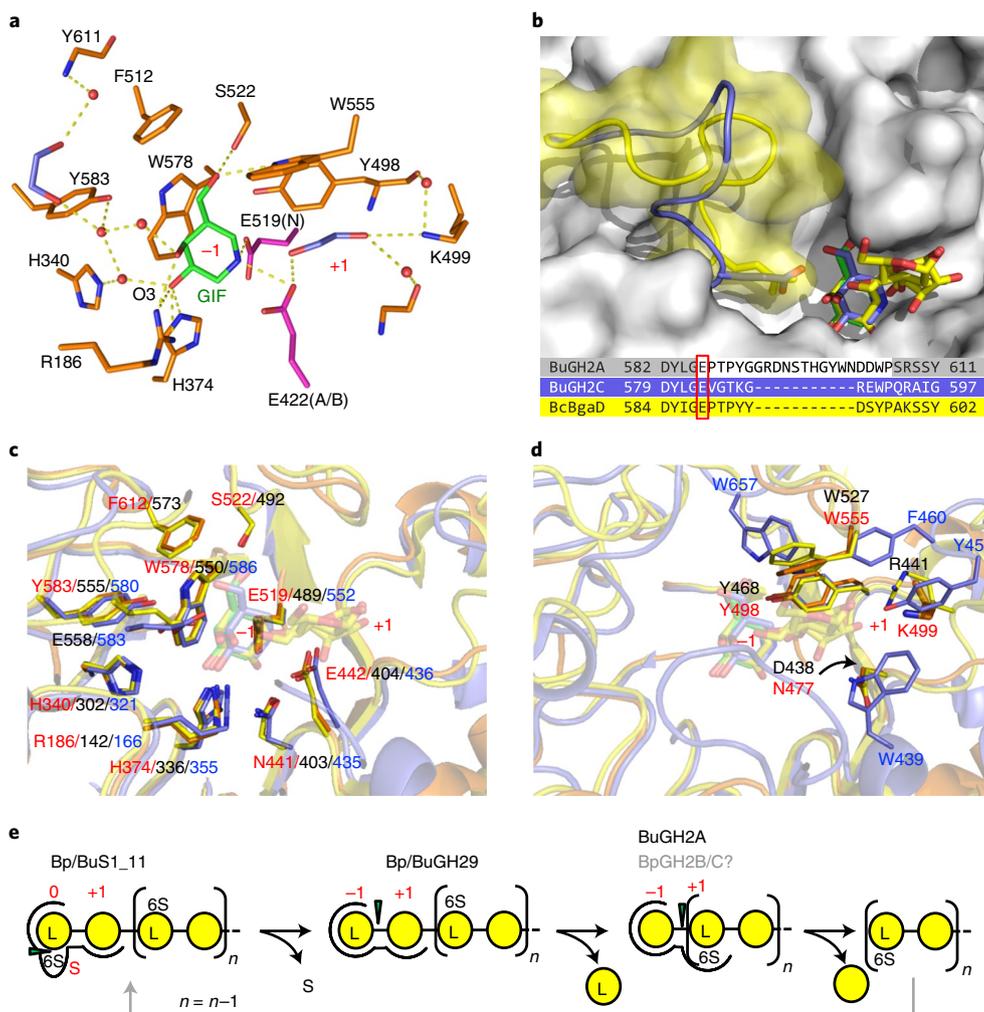


Fig. 3 | An exo- β -D-porphyraoligosaccharide hydrolase leads to the exo-based model of neoporphyraoligosaccharide hydrolysis. **a, Galactoisofagomine (GIF) bound to the BuGH2A active site focusing on inhibitor coordinating residues. The putative catalytic residues are shown in magenta with nucleophile (N) and the acid-base (A/B) indicated. **b**, BuGH2A overlapped with BuGH2C (5T9G) and BcBgaD (7CWD), focusing on the disordered active site loop in BuGH2A. The loop sequence alignment is shown at the bottom of the panel with the disordered residues in BuGH2A shown without highlighting (white background) and the key glutamate residue (shown as sticks) boxed in red. The surface of BuGH2A is shown in gray and the bound GIF as green sticks. The loops in BuGH2C and BcBgaD are shown in blue and yellow, respectively. The surface of the BcBgaD loop is shown in transparent yellow. The GIF bound to BuGH2C is shown as blue sticks and galactose and/or glucose monosaccharides bound to BcBgaD as yellow sticks. **c,d**, the -1 (**c**) and $+1$ (**d**) subsites, respectively, of BuGH2A (orange, red labels), BuGH2C (blue, dark blue labels) and BcBgaD (yellow, black labels). **e**, schematic of the exo-based model of neoporphyraoligosaccharide hydrolysis. Enzyme subsites are labeled in red and the point of bond cleavage by a green triangle.**

binding of the product L-galactose (Extended Data Fig. 5b,c) was accompanied by the rearrangement of two loops: the one harboring the catalytic acid or base and one donated by the first of the two C-terminal β -sandwich domains (Extended Data Fig. 5d). This structural change results in engagement of the Glu³¹² with the oxygen of the glycosidic bond, positioning it for its catalytic role (Extended Data Fig. 5c,d).

The structure of an inactive Asp²⁶⁴ to Asn mutant of BpGH29 was determined in complex with intact pNP- α -L-Gal (Extended Data Fig. 5e). The dimensions of the active site, and poise of the substrate in it, were consistent with a purely exo-mode of nonreducing end L-galactose hydrolysis from disaccharides, oligosaccharides or extended chains. The pocket-like nature of the -1 subsite encloses the L-galactose residue and would prevent recognition of a 6-sulfated L-galactose (Extended Data Fig. 5f). Asp³²⁷, which is in position to hydrogen bond with O6 of the L-galactose, blocks the portion of the active site that would accommodate the 6-sulfate

modification (Fig. 2d), thus rationalizing the need for previous desulfation of neoporphyraoligosaccharides by the S1_11 sulfatase. The *p*-nitrophenyl group of pNP- α -L-Gal occupies the $+1$ subsite, which would typically accommodate the D-galactose residue of a desulfated NP2 substrate. The placement of the Trp²⁶⁵ sidechain in this subsite provides an aromatic platform for the nitrophenyl group and, most likely, D-galactose in the native substrate.

The BpGH29 active site is well conserved among structurally characterized α -fucosidase members of this GH family, even though it is selective for α -L-galactose over α -L-fucose. The region of the active site near C6 of the substrate does not display the variability expected of a potential role in selectivity for α -L-galactose (Fig. 2e). However, notable differences between GH29 fucosidases and BpGH29 are a serine or glycine in place of His¹⁸⁹ and a tyrosine residue at the position of Ser¹⁸⁷ (Fig. 2f). Despite retaining its thermal stability, a S187Y/H189S double mutant of BpGH29 failed to display significant activity on either pNP- α -L-Fuc or pNP- α -L-Gal,

pointing to unknown subtleties in the active site architectures of these enzymes that impart their selectivity.

Identification of a β -porphyraoligosaccharide hydrolase.

To generate a suitable substrate to test the possible porphyran-specific *exo*- β -D-galactosidase activity (that is, an *exo*- β -D-porphyraoligosaccharide hydrolase) of BuGH2A, we purified a tetrasaccharide from enzyme hydrolyzed native porphyran and characterized the oligosaccharide by high-performance liquid chromatography–mass spectrometry (HPLC–MS) (see Methods for structural characterization of oligosaccharides). The tetrasaccharide preparation, referred to as N4, comprised mainly neoporphyratetraose (NP4, L6S-G-L6S-G at roughly 40% of total, or roughly 50% if partially sulfated variants are included) with the remainder being primarily a hybrid porphyran or agarose tetrasaccharide (NAP4, L6S-G-LA-G). Treatment of 13 nM of input N4 (measured by total sugar assay) with BuS1_11* and BuGH29 released an amount of L-galactose consistent with all oligosaccharides terminating in nonreducing end L-galactose-6-sulfate (Extended Data Fig. 6). This enzyme combination would produce P3 (G-L6S-G) and A3 (G-LA-G). Inclusion of BuGH2C, which was previously shown to be active on A3, with BuS1_11* and BuGH29 resulted in the release of D-galactose but no additional L-galactose, consistent with the known activity of BuGH2C¹⁰. The inclusion of BuGH2A, which we previously demonstrated is not active on A3 (ref. ¹⁰), in reactions with both BuS1_11* and BuGH29 resulted in the release of substantial amounts of D-galactose and additional L-galactose (Extended Data Fig. 6), providing evidence of activity on the nonreducing terminal β -linked galactose of P3. This supports the assignment of BuGH2A as a β -galactosidase with *exo*- β -D-porphyraoligosaccharide hydrolase activity. Full enzymatic hydrolysis data of N4 with combinations of the *B. uniformis* enzymes support both the proposed specificity of BuGH2A and the structural assignment of N4 (Extended Data Fig. 6).

Structure an *exo*- β -D-porphyraoligosaccharide hydrolase.

The structure of BuGH2A has the typical five-domain GH2 fold (Extended Data Fig. 7a). The complex of BuGH2A with the galactosidase inhibitor galactoisofagomine (GIF, a kind gift from K. Stubbs) (Extended Data Fig. 7b) showed occupation of the –1 subsite (Fig. 3a). A loop adjacent to the active site comprising residues 586–606 was disordered and could not be modeled. A comparison of BuGH2A with the agarose specific BuGH2C (PDB 5T9G, 37% amino acid identity) and *Bacillus circulans* β -galactosidase BcBgaD (PDB 7CWD, 43% amino acid identity) indicates that this loop contains a conserved glutamate residue that is in position to hydrogen bond with the O4 of a galacto-configured sugar in the –1 subsite and generally contours the active site for *exo*-galactosidase activity (Fig. 3b). The disordered loop in BuGH2A likely performs the same function, while sequestration of the C3 hydroxyl group prevents *endo*-type recognition of internal D-galactose residues in a porphyran chain (Fig. 3a,b). The –1 subsites of BuGH2A, BuGH2C and BcBgaD are highly conserved (Fig. 3c), consistent with *exo*- β -galactosidase activity. BuGH2A and BcBgaD have similar +1 subsites, including a conserved tryptophan residue, suggesting the architecture of the +1 subsite in BuGH2A may accommodate L6S in a comparable manner to galactose or glucose recognition by BcBgaD (Fig. 3d). BuGH2C, however, has a different predicted +1 subsite, likely reflecting the need to accommodate 3,6-anhydro-L-galactose in this subsite. The structural and functional analysis of BuGH2A taken together with the activities of the S1_11 and GH29 enzymes points to an ‘*exo*-based’ mode of porphyran oligosaccharide depolymerization (Fig. 3e).

Saccharification of native porphyran and enzyme synergy. We estimated native porphyran extracted from our commercially

obtained nori to contain a roughly 50:50 ratio of porphyran to agarose regions (Methods and Supplementary Fig. 4). We treated this with *B. uniformis* and *B. plebeius* enzymes in combination with *endo*-acting porphyranases (using BuGH2A as a surrogate for the unidentified *B. plebeius* *exo*- β -D-porphyraoligosaccharide hydrolase). Single enzymes, double enzyme combinations and ternary mixtures lacking either the S1_11* or GH29 produced low to negligible amounts of D- and L-galactose (Supplementary Fig. 5). The ternary mixture of S1_11*, GH29 and BuGH2A produced substantial amounts of D- and L-galactose, which was increased roughly 60% by the addition of *endo*-porphyranase (BuGH16C or BpGH86A) to the mixtures (Fig. 4a,b). The ternary mixture lacking BuGH2A was unexpectedly effective at releasing monosaccharides from native porphyran (only roughly 10–15% less than the quaternary mix), which we attribute to the release of substantial NP2 from native porphyran by the *endo*- β -porphyranases BpGH86A (ref. ²⁰) and BuGH16C (Supplementary Fig. 6). In all cases, the ratio of D-galactose released to L-galactose was roughly one, which would be expected if only porphyran regions of the native polysaccharide were being targeted by these combinations of enzymes.

Attempts at complete saccharification of 95 nM of native porphyran (determined as equivalents of NP2) by combining all porphyranases or all agarases yielded roughly 18 and 11 nM of D-galactose from the porphyranase mix and agarase mix, respectively (Fig. 4c). This equates to overall conversions of roughly 19 and 12%. The combination of all seven enzymes resulted in the release of roughly 43 nM of D-galactose (Fig. 4c), giving an increase in overall conversion to roughly 45%, which is larger than the additive conversions of the agarases and porphyranases alone (approximately 31%). Although the agarases alone were incapable of releasing L-galactose, the inclusion of these enzymes with the porphyranases gave about a 30% increase in L-galactose release relative to the porphyranases alone (increase from roughly 15 to 20 nM), indicating that the agarases are effective in revealing porphyranase substrate in a highly polymerized hybrid polysaccharide; presumably the reverse is also true. Overall, this reveals the synergy between the porphyranase and agarase enzyme mixes to more efficiently reduce the hybrid native porphyran polysaccharide into monosaccharides.

Growth of *Bacteroides* sp. on native porphyran. BuAgaSGBP-B is the cell-surface agarose binding protein from *B. uniformis* that we previously noted also bound to native porphyran¹⁰. BACPLE_01696, or BpPorSGBP-B, is the putative analogous porphyran binding protein in the PorPUL. Qualitative affinity electrophoresis confirmed binding of BuAgaSGBP-B to agarose and native porphyran and showed that BpPorSGBP-B bound native porphyran but not agarose (Supplementary Fig. 7). Estimates of binding affinity based on quantitative affinity electrophoresis showed that BuAgaSGBP-B bound to agarose roughly eightfold more tightly than to native porphyran (K_d of 0.06 mM, compared to 0.5 mM; note that a previous determination of the latter K_d gave a value of 0.4 mM, ref. ¹⁰). This difference in affinity is not explained by the stoichiometric influence of the 50:50 ratio of porphyran to agarose in native porphyran (Supplementary Fig. 8). BpPorSGBP-B displayed an intermediate affinity for native porphyran (K_d of 0.2 mM). Taken together, the results indicate a preference of BuAgaSGBP-B for agarose and suggest a preference of BpPorSGBP-B for porphyran regions. Native porphyran enzymatically depleted of agarose (PENP, roughly 75% porphyran) or native porphyran enzymatically depleted of porphyran (AENP, roughly 80% agarose) were too depolymerized to be effectively used in affinity electrophoresis. Thus, we used fluorescently labeled polysaccharides (FLA-PS^{29,30}) to examine native porphyran and PENP binding and uptake by *B. plebeius*. Both preparations of labeled sugar were taken up by *B. plebeius* cells (Fig. 5a and Supplementary Fig. 9a) but not *B. thetaiotaomicron* cells (Supplementary Fig. 9b,c). The selectivity of the carbohydrate receptors revealed in these results

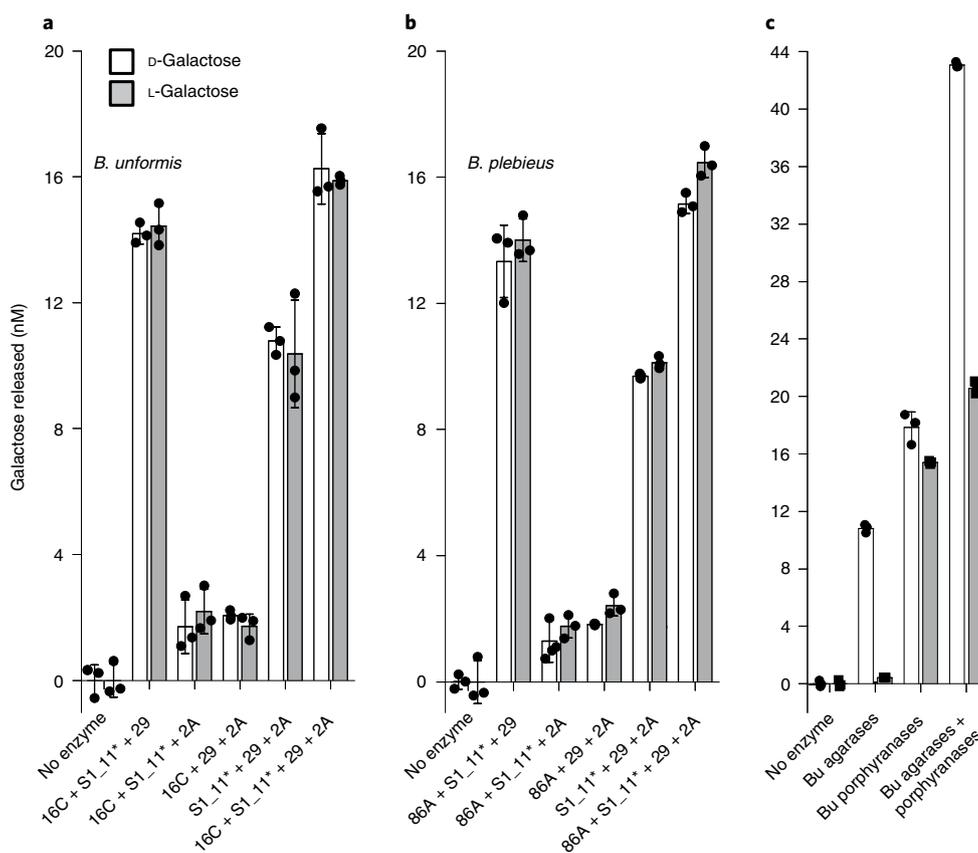


Fig. 4 | Depolymerization of native porphyrin. a, b, Monosaccharide release from native porphyrin using combinations of *B. uniformis* porphyrin hydrolyzing enzymes (**a**) and *B. plebeius* enzymes (**b**). Note that in **b** BuGH2A is used in place of the unidentified *B. plebeius* *exo*- β -D-porphyrano oligosaccharide hydrolase. **c**, Monosaccharide release from native porphyrin using combinations of *B. uniformis* porphyrin (BuS1₁₁*, BuGH16C, BuGH29 and BuGH2A) and/or agarose (BuGH117B, BuGH2C and BuGH16B) hydrolyzing enzymes. In all reactions, the equivalent amount of polysaccharide to 95 nM of NP2 was used. $n = 3$ independent reactions were performed with each replicate shown along with the mean \pm s.d. See Supplementary Fig. 5 for complete hydrolysis data.

were consistent with, and supported by, growth experiments that showed growth of *B. uniformis* on native porphyrin and AENP, but poor growth on PENP. *B. plebeius* grew well on native porphyrin and PENP, but poorly on AENP (Fig. 5b,c). *B. thetaiotaomicron* only displayed growth on galactose (Supplementary Fig. 9d). Thus, native porphyrin is a shared substrate for these two species of bacteria that otherwise have differential abilities to use porphyrin or agarose. To provide additional insight into this selectivity we performed coculture experiments with the two microbes. In general, *B. plebeius* showed a much weaker growth phenotype, even on a readily used substrate such as galactose (Fig. 5b,c), so the experiments were initiated with a 10:1 ratio of *B. plebeius* to *B. uniformis* cells. Under these conditions, *B. plebeius* was completely unable to compete with *B. uniformis* when provided with galactose or agarose as a growth substrate (Fig. 5d). On native porphyrin and PENP, *B. uniformis* showed dominance of the cultures at initial time points. This was unanticipated for PENP; however, *B. uniformis* did display some growth on this substrate in monoculture and, overall, displayed more rapid growth characteristics than *B. plebeius*, which likely accounts for its initial proliferation in the mixed culture. However, at later time points *B. plebeius* did have the potential to compete with *B. uniformis* for growth, particularly on PENP, as it began to comprise an increasing fraction of the culture (Fig. 5d). This suggests that the two microbes can be co-maintained on native porphyrin but may be individually selected for using porphyrin or agarose.

Discussion

B. uniformis and *B. plebeius*, model organisms for dietary algal galactan metabolism in the human gut, display similar mechanisms of initial polysaccharide breakdown and import (Fig. 6a,b). Although slightly different *endo*-acting enzymes are deployed by the two organisms at the cell surface, enzymes sometimes referred to as ‘vanguard’ enzymes³¹, similar pools of oligosaccharides composed of neoporphyrano oligosaccharides, neoagarooligosaccharides and hybrid oligosaccharides are predicted to be generated in the periplasmic space (Fig. 6a,b). We previously demonstrated that the AgaPUL deploys agarolytic enzymes that are consistent with an *exo*-based model of agarose-derived oligosaccharide depolymerization. This relies on sequential hydrolysis of nonreducing terminal LA and galactose from neoagarooligosaccharides by a GH117 *exo*- α -(1,3)-neoagarooligosaccharide hydrolase (BuGH117B) followed an *exo*- β -(1,4)-D-agarooligosaccharide hydrolase from GH2 (BuGH2C)^{10,32} (Fig. 6c). At present we have been unable to produce stable forms of the *B. plebeius* GH2 enzymes and thus we have not identified the functional homolog of BuGH2C from the PorPUL. However, BpGH2A shares 64% amino acid sequence identity with BuGH2C and is, therefore, most likely an *exo*- β -(1,4)-D-agarooligosaccharide hydrolase encoded by the PorPUL. As shown previously²¹, and here, BpGH117 from the PorPUL functions as an *exo*- α -(1,3)-neoagarooligosaccharide hydrolase. Thus, like the AgaPUL, the PorPUL appears to encode the complete machinery for the *exo*-based model of neoagarooligosaccharide depolymerization (Fig. 6c).

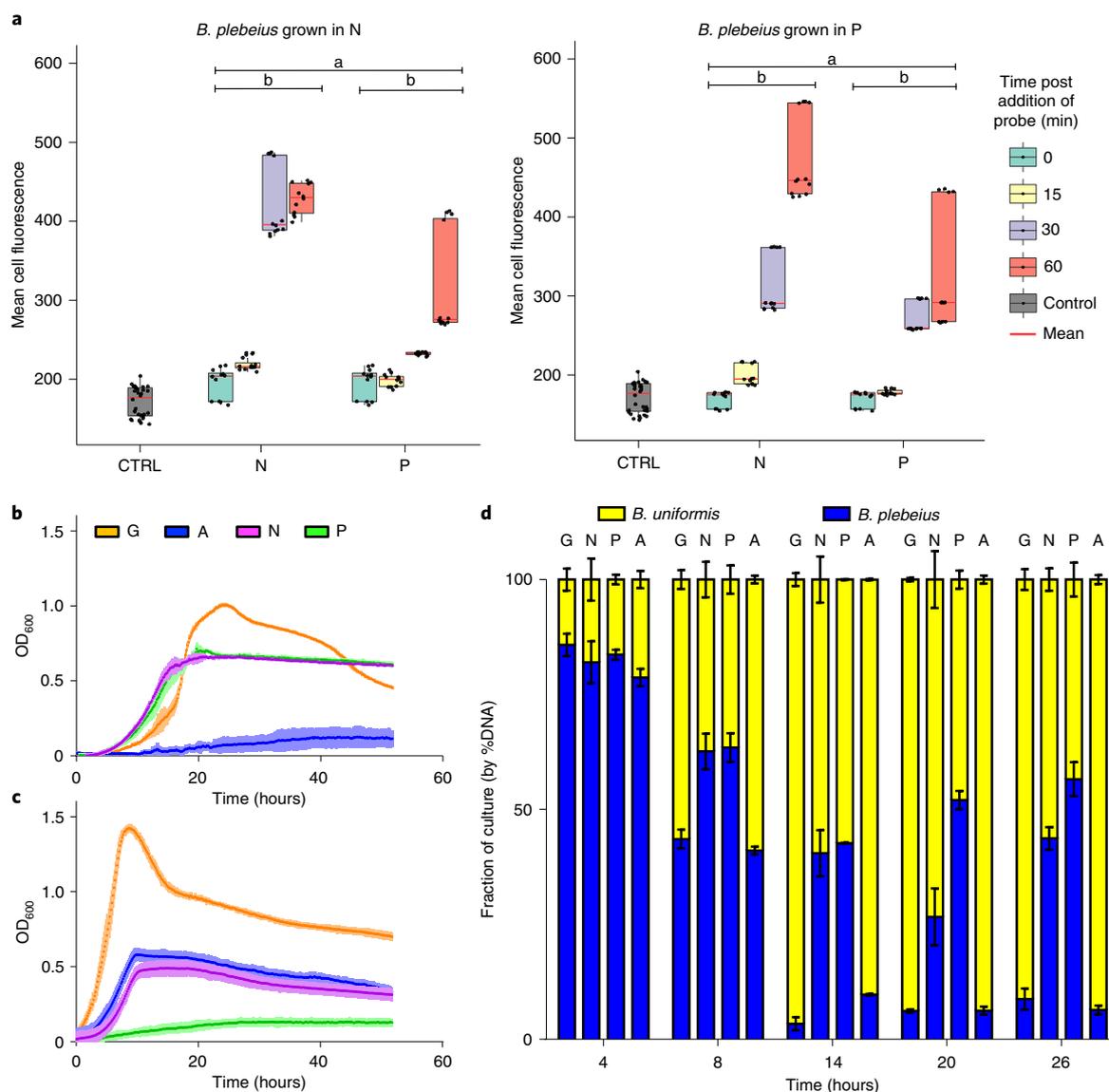


Fig. 5 | Native porphyrin as a growth substrate. **a**, Box plots showing the mean fluorescence intensity of *B. plebeius* cells incubated with fluorescently labeled native porphyrin (FLA-nPOR, indicated with N) and porphyrin enriched native porphyrin (FLA-POR, indicated with P), and a control (*B. plebeius* incubated with unlabeled native porphyrin, indicated as CTRL), over time (0, 15, 30, 60 min). The number of samples for the controls was 32 ($n=32$) and 48 for the samples ($n=48$) with 10,000 cells counted for each sample. The minimum and maximum were 143 and 546, respectively, with the center at 200. The bounds of the boxes are the percentiles P1, 175 and P3, 286. The error bars show standard deviations. The statistical difference within and between treatments were calculated by one-sided Welch's and paired *t*-tests (**a,b**). Difference in the mean fluorescence of each FLA-PS (**a**, $P=2 \times 10^{-16}$). Difference in the change in mean fluorescence within each incubation over time (**b**, $P=2 \times 10^{-16}$). A Tukey test shows that in all FLA-PS incubations the change in mean fluorescence over time was significant (all $P \leq 2.8 \times 10^{-4}$). **b,c**, Growth profiles of *B. plebeius* (**b**) and *B. uniformis* (**c**) on algal galactans. Growth substrates are indicated by the color in the legend. Data points and vertical lines indicate the mean \pm s.d., respectively, of $n=4$ independent cultures. **d**, Coculture of *B. plebeius* and *B. uniformis* on the four growth substrates. Cultures were inoculated with a 10:1 ratio of *B. plebeius* to *B. uniformis* cells. The abundance of individual bacterial species was monitored by qPCR. In all panels, the growth substrates were G, galactose; N, native porphyrin; P, porphyrin enriched native porphyrin and A, agarose enriched native porphyrin. The mean of $n=3$ biological replicates \pm s.d. is shown.

The mode of porphyrin saccharification by both the AgaPUL and PorPUL also displays the features of an *exo*-model (Figs. 3d and 6d). Depolymerization of neoporphyrin oligosaccharides in the periplasm is initiated by, and dependent on, desulfation of the non-reducing terminal L6S residue by the S1_11 sulfatase followed by hydrolysis of the resulting terminal L-galactose by the GH29. In *B. uniformis*, removal of the exposed nonreducing end β -linked galactose is catalyzed by BuGH2A. Given that BpGH2A is most likely an *exo*- β -(1,4)-D-agarooligosaccharide hydrolase, BpGH2B

or BpGH2C are candidates for the *exo*- β -(1,4)-D-porphyrin oligosaccharide hydrolase in the PorPUL; however, BpGH2C is predicted to be intracellular, leaving BpGH2B as the most likely option. The sequential action of enzymes with these specificities would thereby be able to reduce neoporphyrin oligosaccharides to their composite monosaccharides (Figs. 3d and 6d).

Although galactans such as native porphyrin often contain 6-O-methyl-D-galactose, our model does not account for accommodation of this modification. Some marine microbes are known

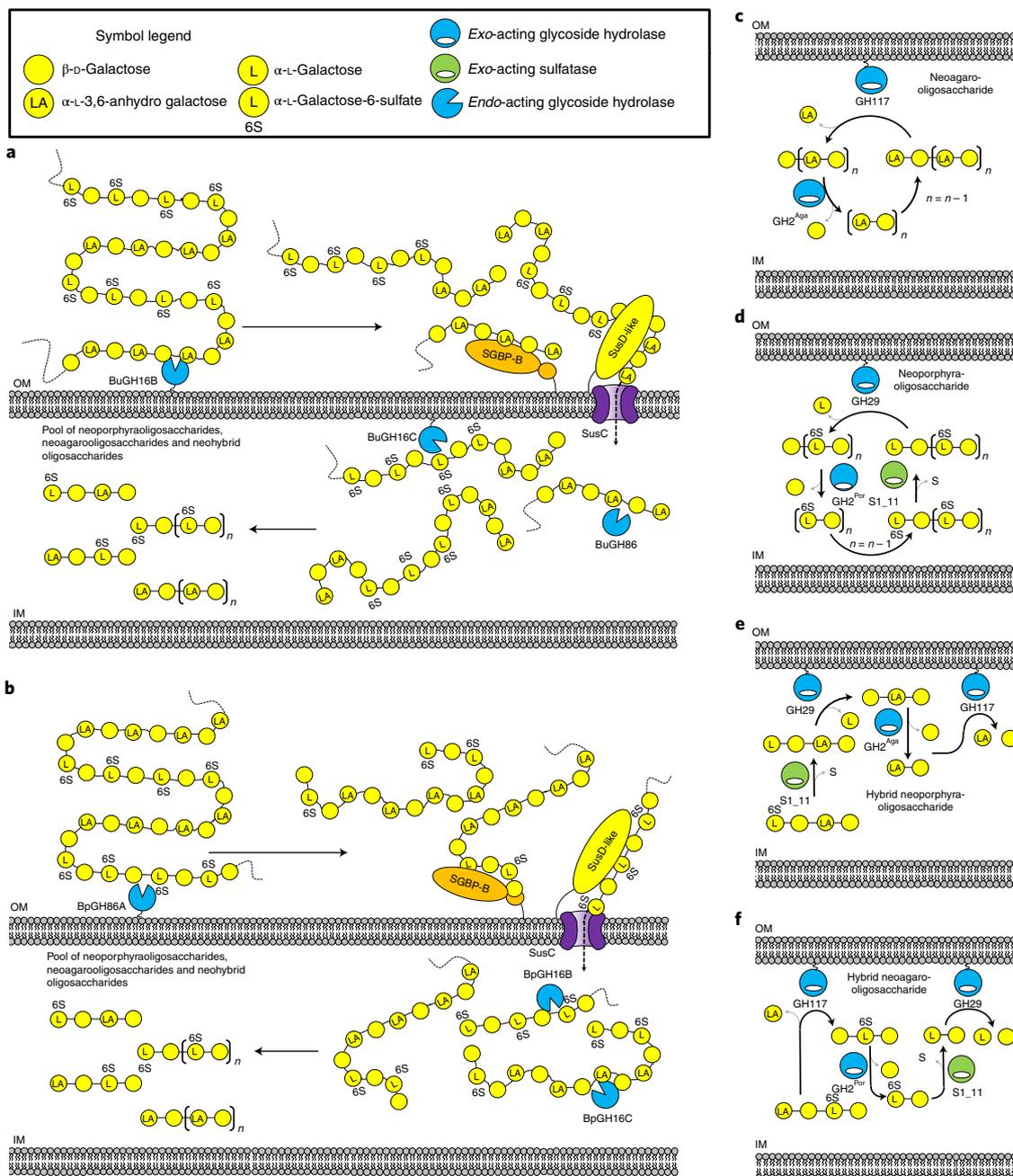


Fig. 6 | Schematics of the depolymerization of native porphyran and oligosaccharides. a, b, The pathways of initial native porphyran import and depolymerization by *B. uniformis* (**a**) and *B. plebeius* (**b**). The specific pathways for depolymerization of neoagaro-oligosaccharides (**c**), neoporphyra-oligosaccharides (**d**), hybrid neoporphyra-oligosaccharides (**e**), and hybrid neoagaro-oligosaccharides (**f**). GH2^{Aga} and GH2^{Por} are used to indicate the *exo*- β -galactosidases that are specific to agarose and porphyran, respectively. The GH2^{Aga} in *B. plebeius* is presumed based on high amino acid sequence identity (>60%) with the homologous *B. uniformis* enzyme. GH2^{Por} remains unidentified in *B. plebeius* but is most likely one of the two additional GH2 enzymes encoded by the PorPUL (Supplementary Table 1). OM indicates the outer membrane and IM the inner membrane.

to accomplish demethylation of galactose using cytochrome P450 monooxygenases^{33,34}; however, there is no evidence for the presence of genes encoding similar proteins in the genomes of either bacterium. Consistent with this, NP2 and N4 derived from enzymatic hydrolysis of native porphyran by two different porphyranases lacked significant methylation (Supplementary Figs. 10–16 and Supplementary Table 8), suggesting that the *endo*-hydrolysis of native porphyran by the enzymes of these microbes may preferentially produce oligosaccharides lacking methylation. Indeed, BuGH2A also displayed poor capacity to accommodate

6-O-methyl-D-galactose (Methods). This may explain why a complete cocktail of enzymes did not achieve full conversion of native porphyran. Thus, we suggest the enzyme systems encoded by the PorPUL and AgaPUL may be unable to cope with methylation of the D-galactose residues in native porphyran. However, it is formally possible that an *endo*-active demethylase that would act before the *endo*- β -porphyranases remains unidentified.

Both the AgaPUL and the PorPUL, therefore, have the biochemical capacity to reduce both agarose and porphyran to monosaccharides. However, based on the dichotomous growth profiles

(Fig. 5b–d), the specificities of the sensing and uptake systems of these PULs appear to dictate the capacity of the microbes to deal with pure polysaccharides of agarose or porphyran (that is, not hybrid native porphyran). The question of why these PULs overlap in their biochemical capacity is answered by the observation that native galactans are typically heterogeneous, as demonstrated by the hybrid nature of native porphyran¹⁹ (Extended Data Fig. 1a–c). With both saccharolytic capacities deployed simultaneously, the machineries encoded by these PULs can synergistically depolymerize naturally occurring native porphyran and the resulting hybrid oligosaccharides, thus optimizing the release of monosaccharides (Fig. 6). This is particularly important with respect to D-galactose, which is a readily used carbon source that comprises roughly 50% of galactans, regardless of their heterogeneity. The remaining monosaccharides, however, which are L-galactose and/or 3,6-anhydro-L-galactose, appear to be differentially used by the bacteria as dictated by the specific monosaccharide processing enzymes encoded by the PULs (Extended Data Fig. 1d and Supplementary Table 1)¹⁰.

The PorPUL encodes an enzyme with LGDH activity that is orthologous to the initiating enzyme of the L-galactose processing pathway in *B. vulgatus*²². Indeed, the PorPUL encodes genes (locus tags BACPLE_01674 to BACPLE_01679) with high amino acid sequence identity (roughly 50–75%) to the remaining enzymes in the L-galactose processing pathway⁹. There is no evidence that the genome of *B. plebeius* encodes enzymes tailored to the metabolism of 3,6-anhydro-L-galactose. In contrast, *B. uniformis* lacks an obvious L-galactose processing pathway, but the AgaPUL encodes proteins whose properties are most consistent with 3,6-anhydro-L-galactose metabolism¹⁰. Thus, despite the overlap of the polysaccharide depolymerization systems, the AgaPUL and the PorPUL not only appear to display selectivity based on specific polysaccharide recognition and import, but they may also have some selectivity via the capacity to metabolize certain proportions of the released monosaccharides in the form of L-galactose or 3,6-anhydro-L-galactose.

The evolution of two related PULs that are highly specialized for a mixed algal galactan, yet which are responsive to specific and chemically distinct components of these galactans, suggests two possibilities. These two strains of bacteria either compete with each other using a fitness advantage dictated by the composition of ingested red seaweeds or they have coevolved to coexist in a syntrophic relationship when provided with mixed agars. Coculture experiments conducted here suggest that although there is a clear advantage for *B. uniformis* on agarose, the populations reach a stoichiometric equilibrium when provided with native porphyran (Fig. 5d). Most likely, this reflects the chemical heterogeneity of native porphyran, which provides blocks of agarose and porphyran for foraging by both strains. Indeed, our model of galactan use, which relies on import of large polysaccharide fragments that are ultimately depolymerized in the periplasmic space, is consistent with a largely ‘selfish’ mechanism of polysaccharide use^{2,35} (Fig. 6). However, ‘pure’ polysaccharide fragments resulting from initial extracellular native porphyran breakdown that are likely not recognized by the import system of the bacterium (that is, agarose fragments for *B. plebeius*, or porphyran fragments for *B. uniformis*) would remain extracellular and therefore could be scavenged by other microbes, such as in this case where the cocultured microbes have complementary PULs. This represents a unique example of cooperative selfish metabolism, in which complementary pathways have been acquired in two different species for the metabolism of a hybrid substrate. Supporting the potential in vivo relevance of this observation, a previous analysis of the AgaPUL and PorPUL frequency in human gut metagenomes indicated that the co-occurrence of the PULs in individual metagenomes is consistent with random distribution: 0.06% of individual metagenomes contained the AgaPUL, 0.14% for the PorPUL and 0.01% for co-occurrence of both PULs, which is the predicted probability of

co-occurrence based on the individual PUL frequencies¹⁰. Further investigation into the competitive or synergistic feeding behaviors displayed by gut microbes containing algal polysaccharide-specific PULs, and other PULs that target complex polysaccharide matrices³⁶, will continue to unlock the evolutionary history and foraging mechanisms of these fluid biochemical pathways.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41589-022-00983-y>.

Received: 28 May 2021; Accepted: 27 January 2022;

Published online: 14 March 2022

References

- Ndeh, D. et al. Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature* **544**, 65–70 (2017).
- Cuskin, F. et al. Human gut *Bacteroidetes* can utilize yeast mannan through a selfish mechanism. *Nature* **517**, 165–169 (2015).
- Kloareg, B. & Quatrano, R. S. Structure of the cell walls of marine algae and ecophysiological functions of the matrix polysaccharides. *Annu. Rev.* **26**, 259–315 (1988).
- Hehemann, J.-H., Boraston, A. B. & Czjzek, M. A sweet new wave: structures and mechanisms of enzymes that digest polysaccharides from marine algae. *Curr. Opin. Struct. Biol.* **28**, 77–86 (2014).
- Ficko-Blean, E. et al. Carrageenan catabolism is encoded by a complex regulon in marine heterotrophic bacteria. *Nat. Commun.* **8**, 1685 (2017).
- Hettle, A. G. et al. Insights into the κ /I-carrageenan metabolism pathway of some marine *Pseudoalteromonas* species. *Commun. Biol.* **2**, 474 (2019).
- Reisky, L. et al. A marine bacterial enzymatic cascade degrades the algal polysaccharide ulvan. *Nat. Chem. Biol.* **15**, 803–812 (2019).
- Wells, M. L. et al. Algae as nutritional and functional food sources: revisiting our understanding. *J. Appl. Phycol.* **29**, 949–982 (2017).
- Hehemann, J.-H. et al. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464**, 908–912 (2010).
- Pluvinage, B. et al. Molecular basis of an agarose metabolic pathway acquired by a human intestinal symbiont. *Nat. Commun.* **9**, 1043 (2018).
- Mathieu, S. et al. Ancient acquisition of ‘alginate utilization loci’ by human gut microbiota. *Sci. Rep.* **8**, 8075 (2018).
- Thomas, F., Hehemann, J.-H., Rebuffet, E., Czjzek, M. & Michel, G. Environmental and gut bacteroidetes: the food connection. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2011.00093> (2011).
- Pudlo, N. A. et al. Extensive transfer of genes for edible seaweed digestion from marine to human gut bacteria. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.06.09.142968> (2020).
- Shepherd, E. S., DeLoache, W. C., Pruss, K. M., Whitaker, W. R. & Sonnenburg, J. L. An exclusive metabolic niche enables strain engraftment in the gut microbiota. *Nature* **557**, 434–438 (2018).
- Kearney, S. M., Gibbons, S. M., Erdman, S. E. & Alm, E. J. Orthogonal dietary niche enables reversible engraftment of a gut bacterial commensal. *Cell Rep.* **24**, 1842–1851 (2018).
- Rinaudo, M. in *Comprehensive Glycoscience* 691–735 (Elsevier, 2007).
- Rees, D. Enzymic synthesis of 3:6-anhydro-L-galactose within porphyran from L-galactose 6-sulphate units. *Biochem. J.* **81**, 347–352 (1961).
- Rees, D. Enzymic desulphation of porphyran. *Biochem. J.* **80**, 449–453 (1961).
- Correc, G., Hehemann, J.-H., Czjzek, M. & Helbert, W. Structural analysis of the degradation products of porphyran digested by *Zobellia galactanivorans* β -porphyranase A. *Carbohydr. Polym.* **83**, 277–283 (2011).
- Hehemann, J.-H., Kelly, A. G., Pudlo, N. A., Martens, E. C. & Boraston, A. B. Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. *Proc. Natl Acad. Sci. USA* **109**, 19786–19791 (2012).
- Hehemann, J. H., Smyth, L., Yadav, A., Voadlo, D. J. & Boraston, A. B. Analysis of a keystone enzyme in agar hydrolysis provides insight into the degradation of a polysaccharide from red seaweeds. *J. Biol. Chem.* **287**, 13985–13995 (2012).
- Hobbs, M. E., Williams, H. J., Hillerich, B., Almo, S. C. & Raushel, F. M. L-Galactose metabolism in *Bacteroides vulgatus* from the human gut microbiota. *Biochemistry* **53**, 4661–4670 (2014).
- Marquardt, T. et al. High-resolution crystal structure of AKR11C1 from *Bacillus halodurans*: an NADPH-dependent 4-hydroxy-2,3-trans-nonanal reductase. *J. Mol. Biol.* **354**, 304–316 (2005).

24. Barbeyron, T. et al. Matching the diversity of sulfated biomolecules: creation of a classification database for sulfatases reflecting their substrate specificity. *PLoS ONE* **11**, e0164846 (2016).
25. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).
26. Bond, C. S. et al. Structure of a human lysosomal sulfatase. *Structure* **5**, 277–289 (1997).
27. Hettle, A. G. et al. The molecular basis of polysaccharide sulfatase activity and a nomenclature for catalytic subsites in this class of enzyme. *Structure* **26**, 747–758.e4 (2018).
28. Cartmell, A. et al. How members of the human gut microbiota overcome the sulfation problem posed by glycosaminoglycans. *Proc. Natl Acad. Sci. USA* **114**, 7037–7042 (2017).
29. Hehemann, J.-H. et al. Single cell fluorescence imaging of glycan uptake by intestinal bacteria. *ISME J.* **13**, 1883–1889 (2019).
30. Reintjes, G., Arnosti, C., Fuchs, B. M. & Amann, R. An alternative polysaccharide uptake mechanism of marine bacteria. *ISME J.* **11**, 1640–1650 (2017).
31. Larsbrink, J. et al. A discrete genetic locus confers xyloglucan metabolism in select human gut *Bacteroidetes*. *Nature* **506**, 498–502 (2014).
32. Lee, C. H. et al. A novel agarolytic β -galactosidase acts on agarooligosaccharides for complete hydrolysis of agarose into monomers. *Appl. Environ. Microbiol.* **80**, 5965–5973 (2014).
33. Robb, C. S., Reisky, L., Bornscheuer, U. T. & Hehemann, J.-H. Specificity and mechanism of carbohydrate demethylation by cytochrome P450 monooxygenases. *Biochem. J.* **475**, 3875–3886 (2018).
34. Reisky, L. et al. Oxidative demethylation of algal carbohydrates by cytochrome P450 monooxygenases. *Nat. Chem. Biol.* **14**, 342–344 (2018).
35. Grondin, J. M., Tamura, K., Déjean, G., Abbott, D. W. & Brumer, H. Polysaccharide utilization loci: fueling microbial communities. *J. Bacteriol.* **199**, e00860-16 (2017).
36. Luis, A. S. et al. Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic *Bacteroides*. *Nat. Microbiol.* **3**, 210–219 (2018).

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2022

Methods

Cloning, protein production and purification. BpGH86A, BuGH16C, BuGH2A, BuGH2C, BuGH117B and BuAgaSGBP-B proteins were produced and purified as described previously^{10,20}.

The genes encoding for BuGH29, BpGH29, BuS1_11, BpS1_11 and BuLGDH were amplified by PCR from *B. plebeius* DSM 17135 or *B. uniformis* NP1 genomic DNA using methods identical to those described previously^{10,20} (see Supplementary Table 2 for all oligonucleotide primer sequences). The gene fragments encoding the GH29 and S1_11 proteins from both *B. plebeius* and *B. uniformis*, BuGH2A from *B. uniformis* and BpPorSGBP-B were amplified with omission of the sequences encoding the predicted secretion signal peptides. Amplified DNA was cloned into pET 28a using incorporated restriction enzyme sites (indicated in Supplementary Table 2) to generate constructs encoding thrombin cleavable N-terminal six-histidine tags. Site directed mutations were introduced using the QuikChange method (Agilent).

Cultures of *Escherichia coli* BL21 (DE3) bearing each of the expression plasmids, except those encoding sulfatases, were grown to an optical density (OD₆₀₀) of 0.8 at 37 °C and induced with 0.5 mM final of isopropyl-1-thio-β-D-galactopyranoside (IPTG) at 16 °C. Sulfatase encoding plasmids were cotransformed with pBAD/myc-his A Rv0712 (FGE) (pBAD/myc-his A Rv0712 (FGE) was a gift from C. Bertozzi, Addgene plasmid no. 16132) into *Escherichia coli* BL21 (DE3) Star cells (Invitrogen) and grown in LB medium containing 50 μg ml⁻¹ kanamycin sulfate, 100 μg ml⁻¹ ampicillin and 50 μg ml⁻¹ chloramphenicol. Cells were grown at 37 °C until the cell density reached an OD₆₀₀ of roughly 0.5, at which time the temperature was dropped to 16 °C and FGE expression was induced with 0.02% L-arabinose. After roughly 2 h, sulfatase expression was induced with a final concentration of 0.5 mM IPTG and the culture incubated for a further 16 h.

Cultures were harvested by centrifugation and resuspended in sucrose solution (30% sucrose, 50 mM Tris pH 7.5). To this 1 mg of lysozyme (Sigma) was added and further incubated at room temperature 10 min. Lysis was completed by the addition of 2× volume of deoxycholate solution (1% deoxycholate, 0.1% Triton X-100, 50 mM Tris pH 8.0, 500 mM NaCl), followed by 5 mM final MgSO₄ and 0.5 mg of deoxyribonuclease. The lysis was clarified by centrifugation at 16,000g for 45 min, and the lysate was applied to an immobilized metal affinity chromatography column equilibrated in binding buffer (20 mM Tris pH 8.0, 500 mM NaCl). The column was washed with a linear gradient of elution buffer (Binding buffer plus 500 mM imidazole). Fractions containing protein were pooled, and the protein was concentrated before being applied to an S200 size exclusion chromatography column (GE) equilibrated binding buffer for crystallization. Protein concentrations were determined by measuring the absorbance at 280 nm using molar extinction coefficients calculated from the amino acid sequences³⁷.

Preparation and characterization of native and enriched porphyrans. Native porphyrin was extracted from nori (*Porphyra* sp.) using previously described formalin/water extraction procedures¹⁹. Native porphyrin depleted of agarose (that is, porphyrin enriched native porphyrin, PENP) was prepared by extensive digestion of 1 g of native porphyrin at 37 °C for 48 h with 10 μM each of BuGH16B, BuGH117A and BuGH2C. Native porphyrin depleted of porphyrin (that is, agarose enriched native porphyrin, AENP) was prepared by extensive digestion of 1 g of native porphyrin at 37 °C for 48 h with 10 μM each of BpGH86A, BpGH29, BpS1_11* and BuGH2A. Digested samples were heated for 30 min at 90 °C then centrifuged to pellet precipitated material. Cleared supernatants were dialyzed against water with 5-kDa cutoff membrane to remove monosaccharides and small oligosaccharide fragments followed by lyophilization. The polysaccharides were characterized by complete simultaneous enzymatic digestion with all of the *exo*- and *endo*-agarases and porphyranases followed by detection of released D- and L-galactose, and by the total sugar assay. We used the assumption of equimolar release of D- and L-galactose from porphyrin regions. Under this assumption, the difference between the actual amounts of D- and L-galactose releases would correspond to D-galactose released solely from agarose regions. Thus, the ratio of porphyrin:agarose can be estimated by the ratio of (L-galactose) to ((D-galactose) – (L-galactose)). For native porphyrin, this resulted in a porphyrin:agarose ratio of 50:50, for PENP a ratio of 75:25 and for AENP a ratio of 20:80 (Supplementary Fig. 4).

Oligosaccharide purification. One gram of PENP was then digested with the β-porphyrinases BuGH16C or BpGH86A to produce oligosaccharides with an even-numbered degree of polymerization. Undigested porphyrin was removed by filtration through a 5-kDa cutoff membrane in a stirred ultrafiltration system, and the oligosaccharide mixture contained in the flow through was lyophilized, resuspended in water and separated by chromatography on a Bio-Gel P-2 (Bio-Rad) column equilibrated with 20 mM ammonium carbonate pH 8.0. Carbohydrates were detected by thin-layer chromatography analysis using silica plates and formic acid:butanol:water (8:4:1) as eluent. The plates were charred 10 min at 110 °C after being exposed to 10% sulfuric acid in ethanol. Fractions containing pure disaccharide or tetrasaccharide were pooled, lyophilized, resuspended in water and passed over the Bio-Gel P-2 column again in D₂O. Purified oligosaccharides were then lyophilized again and stored under desiccating

conditions at room temperature. The NP2 disaccharide was generated from the BuGH16C digest and the N4 tetrasaccharide pool from the BpGH86A digest.

High-performance anion exchange chromatography coupled to pulsed amperometric detection. Ten microliters of deionized water (product no. 6442-85, Sigma-Aldrich) and the deionized water solution of the NP2 (roughly 100 μg ml⁻¹, w/v) and N4 (roughly 1.0 mg ml⁻¹) samples were injected to a Dionex ICS-3000 HPAEC-PAD system (Thermo Fisher Scientific) installed with a CarboPac PA20 guard column (3 × 30 mm) and a CarboPac PA20 column (3 × 150 mm) eluted at a flow rate of 0.4 ml min⁻¹ and under a constant high-pH background of 100 mM NaOH according to our previous report³⁸, except that the sodium acetate gradient was substituted with higher-salt (0–5 min, 0–600 mM; 5–15 min, 600–1,000 mM; 15–40 min, 1,000 mM; 40–40.1 min, 1,000–0 mM; 40.1–60 min, 0 mM) to provide enough ionic strength to elute sulfated oligosaccharides. Experiments were conducted in duplicate. For the NP2 sample, this showed a dominant peak at a retention time of around 8.3 min and a minor peak at a retention time of roughly 11.5 min (Supplementary Fig. 10), indicating the relatively high purity of the sample. The resulting N4 preparation gave multiple peaks eluting at high salt concentrations (retention time 7–25 min) indicating multiple species in the sample (Supplementary Fig. 10).

Solution-state nuclear magnetic resonance (NMR) spectroscopy. NP2 (roughly 15 mg) was dissolved in 0.5 ml of D₂O (99.9% atom D, product no. 151882, Sigma-Aldrich), and the solution was then freeze-dried against three changes of the D₂O. After the last freeze-drying step, the sample was dissolved in 0.66 ml of high purity D₂O (≥99.96% atom D, product no. 151890, Sigma-Aldrich), followed by addition of 20 μl of D₂O containing 0.75% (w/w) of sodium 3-trimethylsilyl-2,2,3,3-tetradeuteriopropionate (TSP-d4) (99.9% atom D, product no. 293040, Sigma-Aldrich). The resulting sample solution was transferred to a 5-mm NMR tube. All 1D and 2D NMR experiments were performed at 298 K on a Bruker Avance III HD 700 MHz NMR spectrometer (Bruker Corporation) operating at 700.44 and 176.14 MHz for ¹H and ¹³C NMR analysis, respectively. All measurements were carried out using a triple resonance TBO-Z probe with the outer coil tuned to ¹H and the inner coil tuned to ¹³C. 1D ¹H NMR spectrum was recorded using the standard zg30 pulse sequence with a relaxation delay of 1 s, an acquisition time of 2.3396 s, a spectrum width of 20 ppm and 1,024 scans. 1D ¹³C NMR spectra were acquired using the standard zgpg30 pulse sequence with relaxation delay, acquisition time, number of scans and spectrum width set at 2 s, 1.5729 s, 12,000 and 236 ppm, respectively. The spectral size of each 1D spectrum was 131,072. A series of 2D NMR experiments including 2D ¹H-¹H correlation spectroscopy (COSY), 2D ¹H-¹H total correlation spectroscopy (TOCSY), 2D ¹H-¹³C heteronuclear single-quantum coherence spectroscopy (HSQC) and 2D ¹H-¹³C heteronuclear multiple-bond correlation spectroscopy (HMBC) were performed using the cosygmfqc, dipsi2gpph3z, hsqcedetgpp3 and hmbcetgp3nd pulse sequences, respectively. For all the 2D experiments, 16 scans were accumulated per t1 increment and a relaxation delay of 2 s was used. The acquisition time was set to 1.2714, 1.273, 0.4499 and 2.5428 s for COSY, TOCSY, HSQC and HMBC, respectively. Mixing time for TOCSY was 80 ms. All the 2D NMR spectra had a spectrum width of 9.20 ppm and transmitter offset of 4.266 ppm in the ¹H dimension. The spectrum width in the ¹³C dimension of HSQC and HMBC spectra was 110 ppm with a transmitter offset of 53 ppm. Spectra sizes (t2 × t1) were 16,384 × 512 for COSY and TOCSY, 4,096 × 4,096 for HSQC and 16,384 × 2,048 for HMBC. All 2D NMR data were collected using nonuniform sampling set to 50% and the spectra were processed using the compressed sensing algorithm. ¹H and ¹³C chemical shifts were internally referenced to TSP-d4 as 0 ppm. Data were analyzed using MestReNova software v.14.2.1-27684 (Mestrelab Research S.L.). This allowed unambiguous structural explanation of NP2 as neoporphyrinose (L6S-G) as the dominant component of the sample (Supplementary Figs. 11–13). Full sets of ¹H chemical shifts data (Supplementary Table 7) were obtained from complete interpretation of homonuclear 2D ¹H-¹H NMR spectra including COSY (Supplementary Fig. 11) and TOCSY (spectrum not shown). The ¹H NMR spectra, peak assignment and obtained chemical shift data were in good agreement with the published ¹H NMR data of purified NP2 (ref. ¹⁹). Complete peak assignments of heteronuclear 2D ¹H-¹³C NMR spectra including HSQC (Supplementary Fig. 12) and HMBC (Supplementary Fig. 13) confirmed the ¹H chemical shift data obtained from COSY and TOCSY and enabled full sets of ¹³C chemical shift data of NP2 to be obtained (Supplementary Table 7). Notably, long-range cross-glycosidic-bond C-H correlations in the HMBC spectrum provided solid evidence for the linkage sequence of NP2 α and β anomers (Supplementary Fig. 13a,c). There were no detectable NMR signals of naturally existing O-methylated residues in the disaccharide sample.

Glycosidic linkage analysis using gas chromatography coupled to electron ionization mass spectrometry (GC-EIMS). NP2 (roughly 1 mg) was reduced by NaBD₄ as described in our previous report³⁸. The reduced product was then dissolved in deionized water, followed by passing through a Dowex 50Wx8 resin (product no. 217492, Sigma-Aldrich) prewashed with 0.1 M triethylamine hydrochloride³⁹. After evaporation to dryness by N₂, the sample was perethylated with 0.6 ml of ethyl iodide and 100 mg of NaOH powder suspended in 1 ml of

dimethyl sulfoxide with magnetic stirring for 24 h at room temperature in a sealed glass tube with head space filled with N₂ (ref. 49). The ethylation product was collected and purified according to literature⁴¹ with a modification that included solid-phase extraction (SPE) (instead of dialysis) for purifying the oligosaccharide derivatives⁴². Briefly, the reaction product was partitioned in 3 ml of dichloromethane and 3 ml of deionized water three times, and each time the lower phase was left untouched and the upper phase was transferred and pooled into a new tube and neutralized by acetic acid. Next, the lower phase was evaporated to dryness by N₂ and the pooled upper phases were passed through a Waters Sep-Pak C18 column (200 mg absorbent) preconditioned with methanol (3 ml × 2) then deionized water (3 ml × 2). The column was then washed with deionized water (3 ml × 2) and eluted with methanol (3 ml × 2). The eluents were pooled with the residue dried from the lower phase, and followed by evaporation to dryness by N₂ against three changes of methanol (3 ml). The sample was then hydrolyzed by 2 M trifluoroacetic acid (TFA) (120 °C, 2 h), reduced by NaBD₄ and peracetylated by acetic anhydride as described previously³⁸. The partially ethylated alditol acetate (PEAA) derivatives were analyzed on a Thermo Trace GC Finnigan Polaris Q GC-EIMS system (Thermo Fisher Scientific) installed with a zero polarity Zebtron ZB-1MS capillary GC column (30 m × 0.25 mm × 0.25 μm) with a constant column outlet helium flow rate of 0.8 ml min⁻¹ and with oven temperature initially set at 55 °C (hold 2 min) and then increased at 30 °C min⁻¹ to 120 °C (hold 4 min), at 8 °C min⁻¹ to 225 °C, and at 30 °C min⁻¹ to 260 °C (hold 8 min). Data were collected and analyzed by Thermo Xcalibur software. In a separate experiment, before the perethylation step, the NaBD₄-reduced sulfated oligosaccharide was selectively 6-O-desulfated according to literature⁴³ with modifications. Briefly, the reduced sulfated oligosaccharide (roughly 1 mg) dissolved in deionized water (1 ml) was passed through the Dowex 50Wx8 resin prewashed with 0.1 M pyridine hydrochloride. After evaporation to dryness by N₂, the sample was heated in a mixture of pyridine (0.5 ml) and MSTFA (0.5 ml) at 100 °C for 2 h in a sealed glass tube with head space filled with N₂, followed by evaporating the reaction mixture by N₂ and repeated evaporation to dryness in methanol (3 ml × 3). The NaBD₄-reduced, 6-O-desulfated sample was converted to PEAA derivatives then subjected to GC-EIMS analysis as described above. Each experiment was conducted in duplicate. The linkage analysis results of NaBD₄-treated NP2 showed major PEAA peaks of 3-galactitol and 6-Galp corresponding to the reduced reducing end and 6-O-sulfated nonreducing end of NP2, respectively (Supplementary Fig. 14). The NaBD₄-reduced, desulfated NP2 generated intense PEAA peaks of 3-galactitol from reduced reducing end and t-Galp from 6-O-desulfated nonreducing end of NP2. No evidence of methylation was obtained by ethylation-GC-EIMS analysis of the sample.

HPLC-MS. Samples were transferred to 1.5-ml screw capped vials and lyophilized. The dried samples were reduced with 1 M NaBH₄ in 50 mM NH₄OH for 2 h at 65 °C, after which the reaction was carefully quenched by the dropwise addition of glacial acetic acid until all fizzing stopped. The neutralized reaction mixture was loaded on preconditioned (3 ml of 80% acetonitrile (ACN) + 0.1% TFA followed by 6 ml of water) ENVI-Carb SPE cartridges (250 mg bed volume); samples were passed through the SPE cartridges dropwise, using only a minimal amount of positive pressure. The SPE cartridges were washed with 5 ml of water followed by two lots of 2.2 ml of 50% ACN containing 0.1% TFA; the latter wash that contained the reduced oligosaccharides were pooled then partially in vacuo using SpeedVac (Thermo Fisher Scientific) for roughly 3 h to remove TFA. After that, the samples were snap frozen in liquid nitrogen and lyophilized. The dry samples were reconstituted in water and analyzed by mass spectrometry.

HPLC was done using an Agilent 1290 Infinity system, supplied with a 1290 Infinity binary pump, a 1290 Infinity autosampler and a 1290 Infinity temperature-controlled sample compartment (Agilent Technologies). Mass spectrometry was performed using an Agilent 6530 quadrupole time-of-flight mass spectrometer with jet stream electrospray ionization source. A Hypercarb HPLC column (Thermo Fisher; 100 × 2.1 mm²; particle size, 3 μm) was used for the separation of glycans. Information on mass spectrometry parameters and gradient elution and solvents used for LC separation is provided in Supplementary Tables 5 and 6, respectively. Data were acquired in high resolution mode at a mass setting of 3,200 *m/z*.

MS data were analyzed using Agilent Technologies' MassHunter software (v.B.07.00); specifically, the find-by-formula (FBF) algorithm was used to obtain the total area under a chromatographic peak. The FBF algorithm takes into consideration the mass accuracy of monoisotopic masses (*m/z*), in addition to the relative heights and spacing of the different isotopic peaks associated with the given formula; the retention times are also used to give the area results along with a (confidence) score. A list of putative formulas used by FBF was generated from the known compositions of the monosaccharides in addition to substituents such as methyl and sulfate groups. One water molecule was subtracted from each formula for each glycosidic bond or sulfate ester, and a reduced reducing end (that is, alditol) was assumed for each formula. Post-MassHunter data processing was done using Microsoft Excel (Microsoft Corporation). Raw peak areas calculated for each glycan by MassHunter were normalized to the total glycan pool detected for each specific sample. Strict quality checks were maintained; for example, if any glycan had a difference of more than 10 ppm between the observed and predicted

m/z, it was not included in the results. Retention times and *m/z* ratios were used as identifiers for glycans. All the HPLC-quadrupole time-of-flight mass spectrometry runs were made on the same day after calibrating instrument with reference ions (*m/z* 112.9855, 966.0007 and 2533.8923). Throughout the runs, reference ions were also introduced to the mass spectrometer to maintain instrument resolution. Three instrumental replicates were recorded per sample; only oligosaccharides detected two or more times were included in the final processed data.

This showed that roughly 95% of the sample comprised NP2 (Hex2S1, Supplementary Table 8) and indicated negligible amounts of methylated sugar. The mixed nature of the N4 preparation was confirmed by HPLC-MS analysis of the sample (Supplementary Table 8). We were unable to detect neoagarotetraose (NA4, LA-G-LA-G) in the sample. Overall, methylated species comprised <2% of the sample, while contaminating disaccharide species made up roughly 3% of the sample. Approximately 50% of the sample comprised species of Hex4, Hex4S1 and Hex4S2, indicating the porphyrin backbone of L-G-L-G (where L is L-galactose). Approximately 80% of this (roughly 40% overall) comprised Hex2S2, which is fully sulfated neoporphyratetraose (NP4, L6S-G-L6S-G). The bulk of the remaining material (roughly 45% overall) comprised hybrid molecules of Hex3aHex1 and Hex3aHex1S1. Each species had two different retention times, indicating possible isobars of L6S-G-LA-G (NPA4) and LA-G-L6S-G (NAP4) and the low abundance L-G-LA-G and LA-G-L-G versions that lack sulfation. To assist with resolving this, we performed tandem mass spectrometry (MS/MS) on the tetrasaccharide species. MS/MS data were acquired at a fix collision energy of 40 eV in negative ion mode. Auto MS/MS mode was used with an acquisition rate of 1 spectrum per s in the *m/z* range of 100 to 3,000. The isolation width was kept at roughly 4 *m/z* (medium). Product ion spectra were annotated using GlycoWorkbench 2.1 stable build 146 (ref. 44). Product ions were assigned based on the predictions made by GlycoWorkbench. Although the hybrid species had different HPLC retention times, both had very similar fragmentation patterns and clearly terminated in a nonreducing end L6S residue (Supplementary Fig. 16b,c). The results appear to indicate that all of the Hex3aHex1S1 species is NPA4, although it remains unclear why it is associated with two different HPLC retention times.

Total sugar assay. The sugar content of polysaccharide and oligosaccharide preparations were determined essentially as described previously⁴⁵. Briefly, 5 μl 80% (v/v) of phenol was added to 200 μl of samples in dH₂O, followed by 500 μl of concentrated sulfuric acid and mixed well. Once cooled, absorbances were read at 490 nm in a SpectraMax M5 plate reader (Molecular Devices) and converted to concentration of sugar using a standard curve of D-galactose. Determinations were performed in triplicate and with different volumes of starting sample.

Enzyme assays. Qualitative analysis of BpGH16B, BpGH29, BpS1_11*, BuGH29 and BuS1_11* activity on porphyrin was performed by incubating 1 μM of enzyme and 1% native porphyrin in 20 mM Tris (pH 8) overnight at 37 °C. Samples were analyzed by fluorophore-assisted carbohydrate electrophoresis using previously described methods⁴⁶. For BuGH16C, native porphyrin (0.1% w/v final) was digested with 1 μM enzyme in 20 mM 2-(N-morpholino)ethanesulfonic acid (MES) pH 6.5, 500 mM NaCl at 37 °C. Samples were taken at intervals and stopped by heating to 90 °C for 10 min. Stopped reactions were spotted onto thin-layer chromatography plates, which were run and developed as previously described²⁰.

The activity of BpLGDH was initially qualitatively screened against a panel of monosaccharides using an excess of NAD⁺ (5 mM). The formation of reduced cofactor was followed at 340 nm using a SpectraMax5 plate reader (Molecular Devices). Significant activity was only observed for L-galactose. Subsequently, the pH optimum for BpLGDH activity was determined using 1 μM of enzyme, 600 μM of L-galactose and 10 mM of NADP⁺ in 0.1 M Bis-Tris propane buffer (with 0.5 M NaCl) over a range of pH from 6.5 to 9.5.

Reaction mixtures for the determination of NAD⁺ and NADP⁺ kinetic constants were set up in triplicate at 25 °C in 100 mM Bis-Tris propane (pH 9.0) with 0.5 M NaCl in 100-μl volumes containing 20 nM of enzyme, 40 mM L-Gal. NADP⁺ was used at a concentration range of 0–9 mM and NAD⁺ was used at 0–20 mM. Reaction mixtures for the determination of L-galactose kinetic constants were set up in triplicate at 25 °C in 100 mM Bis-Tris propane (pH 9.0) with 0.5 M NaCl in 100-μl volumes containing 20 nM of enzyme, 3 mM of NADP⁺ and 0–15 mM of L-galactose. The formation of the reaction product was followed at 340 nm using a SpectraMax5 plate reader (Molecular Devices). The rate of release was determined by linear regression using the linear part of the curve. Initial rates were converted into μM min⁻¹ using the experimentally determined extinction coefficients for NADPH and NADH in 50 mM Bis-Tris propane pH 9.0, 500 mM NaCl (4,775 and 4,036 M⁻¹ cm⁻¹, respectively).

Kinetic analysis of BuS1_11* activity on NP2 was performed as described previously in a buffer containing 0.5 mM NaCl, 1 mM 3-(N-morpholino)propane sulfonic acid (pH 7.16)²⁷. A concentration of 100 nM enzyme and concentration range of 0–2.5 mM NP2 was used. Kinetic curves were fit with the Michaelis-Menten equation and kinetic parameters determined using nonlinear regression. Values and errors represent the means and standard deviations, respectively, from experiments performed in triplicate.

Kinetic analysis of BpGH29 was performed using *p*-nitrophenyl α-L-galactopyranoside (pNP-α-L-Gal, a kind gift from H. Gilbert). The pH

optimum of the enzyme was assessed by incubating 400 μM pNP- α -L-Gal with 1 μM enzyme in 100 mM McIvaine Buffer ranging from pH 3 to 9 at 37 °C. Reactions were stopped with equal volume of 0.5 M NaOH then read at 400 nm and absorbance plotted against pH. Kinetics were determined by a stopped assay performed in 100 mM sodium acetate (pH 5.6). A concentration of 0.8 μM enzyme and concentration range of 0–2.0 mM pNP- α -L-Gal was used. Due to extremely limiting amounts of substrate, only a single determination of rate was performed at each substrate concentration. Kinetic curves were fit with the Michaelis–Menten equation and kinetic parameters determined using nonlinear regression. Reported errors are those determined from the model fitting.

The activity of BuGH2A on *p*-nitrophenyl β -D-galactopyranoside (pNP- β -gal) was qualitatively monitored by absorbance at 400 nm using reactions with 100 nM enzyme performed in 50 mM MES (pH 6.5) containing 0.5 M NaCl at 25 °C. An initial estimate of the K_m from a Michaelis–Menten plot was roughly 2.5 mM. Subsequently, Dixon plot analyses were used to estimate the inhibition of D-galactose (concentration range of 0–40 mM) and 6-O-methyl-D-galactose (concentration range of 0–40 mM) using substrate concentrations of 0.1, 0.3, 0.5 and 1.0 mM. Reactions were done with 100 nM enzyme at 25 °C in 50 mM MES (pH 6.5) containing 0.5 M NaCl. The K_i estimated for D-galactose was roughly 15 mM. The K_i for 6-O-methyl-D-galactose was very poor and beyond the accessible concentration for this assay, but we nevertheless estimated it to be roughly 100 mM, and therefore much worse than for D-galactose.

To monitor the enzymatic release of D- and L-galactose, oligo- and polysaccharide substrates were digested with enzyme cocktails containing 5 μM of each enzyme (10 μM for sulfatases) in 50 mM MES (pH 6.5) 0.5 M NaCl at 25 °C for 20 h. Any precipitated material was removed by centrifugation and the cleared supernatant transferred to a 96-well plate for galactose detection. D-Galactose was quantified by the addition of NAD⁺ to 3 mM and the galactose mutarotase/ β -galactotase dehydrogenase enzyme mix from the L-arabinose/D-galactose assay kit (Megazyme). L-Galactose was quantified by the addition of NADP⁺ to 3 mM and BpLGDH to 2 μM . Reduction of NAD⁺/NADP⁺ was monitored at 340 nm in a SpectraMax M5 plate reader until the absorbance plateaued. Final absorbance values were blanked with the no enzyme control value, and converted to nM of galactose using the experimentally determined extinction coefficients for NADH and NADPH in MES pH 6.5 (4,426 and 4,633 M⁻¹ cm⁻¹, respectively). All assays were performed in triplicate.

Affinity gel electrophoresis. Affinity electrophoresis was performed and analyzed as described previously^{47,48} using 12% (w/v) polyacrylamide gels polymerized with and without the inclusion of polysaccharide ranging in concentration from 0 to 4.5 mM (the polysaccharide concentration was based on equivalents a decasaccharide). Protein samples were mixed at a 1:1 ratio with native gel loading dye (0.014% w/v bromophenol blue, 60 mM Tris (pH 6.8), 10% (v/v) glycerol). The gels were run with native running buffer (25 mM Tris, 200 mM glycine) at 100 V for 300 min. Gels were stained Coomassie Blue R250 in 25% (v/v) methanol and 10% (v/v) acetic acid and destained with 25% (v/v) methanol and 10% (v/v) acetic acid. Binding to polysaccharide was visualized as reduced mobility of the protein on these gels relative to the nondenaturing gel lacking polysaccharide. Bovine serum albumin was used as a noninteracting reference protein. Quantitative affinity electrophoresis was performed by determining electrophoretic mobilities in 7–10 gels containing increasing concentrations of polysaccharide. We carried out the analysis as described previously⁴⁷.

Production of fluorescently labeled polysaccharides. Native porphyrin and porphyrin underwent mild acid hydrolysis in 0.1 M sulfuric acid for 2 h at 50 °C. The pH was neutralized and then polysaccharides were filtered through a 0.45- μm pore size polycarbonate filter (MilliporeSigma) to remove particles. Low molecular weight products were separated using Vivaspin 15R columns with a 5,000 Da molecular weight cutoff for 30 min at 3,000g (VS15RH11, Sartorius). Then the polysaccharides were sterile filtrated through a 0.2- μm syringe filter (MilliporeSigma). The concentration of the hydrolyzed polysaccharides was measured using the phenol-sulfuric acid method⁴⁹ using D-galactose as the standard. Then 50 μl of each polysaccharide (7 mg ml⁻¹) was used for conjugation to fluoresceinamine as described by Klassen et al.²⁹, with slight modifications. The purified FLA-PS was left in solution, and the concentration was measured using the phenol-sulfuric acid method using galactose as the standard.

Visualization and quantification of glycan uptake. *B. thetaiotaomicron* and *B. plebeius* were grown anaerobically (atmosphere: 85% N₂, 10% CO₂, 5% H₂ at 37 °C) in chopped meat broth medium overnight. Then 10 μl of the overnight cultures was used to inoculate minimal medium supplemented with beef extract (MM + BE) and 0.5% native porphyrin or porphyrin to prime the cells¹⁰. The primed cultures were harvested in the exponential phase (OD₆₀₀ 0.6–1.0) by centrifugation and washed in 2 \times minimal medium. The washing step was repeated three times. FLA-labeled native porphyrin (FLA-nPOR) and porphyrin (FLA-POR) were mixed 1:1 with resuspended culture for a final concentration of 0.2 mg ml⁻¹ FLA-PS and an OD₆₀₀ of 0.05, in triplicate. From each sample, 20 μl was collected at 0 h, 30 min, 1 h and 3 h and fixed in 1% formaldehyde for 1 h under the same atmospheric conditions. Following fixation, cells were pelleted (5,000g for 15 min)

and washed in 1 \times PBS solution. The washing step was repeated an additional three times.

For super-resolution microscopy, the washed cells were first resuspended in 2 ml of 1 \times PBS and then filtered onto a 0.2 μm pore size polycarbonate filter (Millipore) applying a gentle vacuum <200 mbar as previously described²⁹. After drying, the filters were counter-stained with 1 ng μl^{-1} 4',6-diamidino-2-phenylindole for 10 min. Subsequently, the filters were mounted onto glass slides using a Citifluor/VectaShield (4:1) mounting solution (Electron Microscopy Sciences and Vector Laboratories). Super-resolution images were taken on a Zeiss ELYRA PS.1 with an Airyscan detector (Carl Zeiss) using a $\times 63$ oil objective and a 488 and 405 nm lasers, and BP 502-538 and BP 420-480-LP 750 optical filters. The images were processed with the software Zen2011 (Carl Zeiss).

For quantification of FLA-PS uptake, the washed cells were measured using an Accuri C6 flow cytometer (BD-Accuri Cytometers, BD Bioscience) using CD Accuri C6 Plus Software as described previously²⁹. Briefly, the eight- and six-pink validation beads (Spherotech) were used as internal references, and unstained *B. thetaiotaomicron* and *B. plebeius* served as fluorescence background controls. For each treatment (FLA-nPOR and FLA-POR) and time point (0 h, 30 min, 1 h and 3 h) the fluorescence (FL1-H) of 10,000 events (cells) was measured. Subsequently, the flow cytometry data were analyzed using the FlowJo software (v.10.4.2, Tree Star). The sample statistics (counts, mean fluorescence and standard deviation) for each treatment were exported and statistically analyzed and plotted using the R software (R v.3.6.3). Statistical difference between and within the treatments were calculated by Welch's *t*-tests and paired *t*-test, respectively.

Selective growth of pure and cocultures. Pure cultures were primed as above. Growth profiles of bacterial strains on pure glycans were determined anaerobically using individual agars as sole carbon sources as previously described¹⁰. Briefly, overnight cultures were diluted in 1:100 in 2 \times MM containing no carbon source and 100 μl were inoculated into each well of a 96-well plate containing 100 μl of sterilized agars preparations (10 mg ml⁻¹) to obtain 200 μl of cultures with a final concentration of 5 mg ml⁻¹. Plates were sealed and ultraviolet absorbance was measured at 600 nm (OD₆₀₀) at 15-min intervals for 40 h using a Powerwave HT absorbance reader (Biotek Instruments). Growth of each bacterial species was measured ($n = 4$) for each glycan. Data processing was performed using Gen5 software (BioTek) and GraphPad Prism v.6 software (GraphPad Software, Inc.). Each datapoint is the mean of replicate growths and the line thickness represents the standard deviation.

For cocultures, *B. uniformis* NP1 and *B. plebeius* strains were inoculated into the same glycan mixture, in triplicate and aliquots (0.2 ml) were collected at various time points to determine the relative abundance of each bacterial species over time by quantitative PCR (qPCR) analysis of genomic DNA as described in ref.⁵⁰. To normalize for the differences in lag times between *B. uniformis* NP1 and *B. plebeius* strains a tenfold higher inoculum of *B. plebeius* was used. For qPCR enumeration of each species in the coculture samples, cells were collected by centrifugation and DNA was isolated using the BioBasic genomic DNA mini-Prep kit according to the manufacturer's instructions (BioBasic) with the addition of an RNA digestion step. DNA (2 μl of each extraction, 0.2 to 20 ng) was assayed in triplicate in a QuantStudio 6 Realtime PCR instrument (Applied Biosystems, Inc.) using PerfeCTa SYBR Green FastMix Low ROX master mix (Quantabio) and species-specific primers (Supplementary Table 3) for 40 cycles of 95 °C for 10 s, 55 °C for 30 s. Purified genomic DNA standards (range 50, 10, 2, 0.4, 0.08 and 0.016 ng) of each species were included in triplicate in each PCR run. A standard curve generated from these standards was used to calculate the relative abundances of species in each sample.

Crystallization, diffraction data collection and processing. Purified proteins were screened for crystallization by the vapor-diffusion method in 96-well sitting drop plates and crystal hits were optimized in 24-well hanging drop plates. The final crystallization conditions were as follows: BuS1_11 (60 mg ml⁻¹ in 20 mM Tris pH 8.0, 500 mM NaCl) 0.1 M HEPES (pH 7.5), 0.2 M MgCl₂, 25% (w/v) PEG 3350; BpGH29 and BpGH29_D264N (28 mg ml⁻¹ in 20 mM Tris pH 8.0, 500 mM NaCl) 0.1 M Bis-Tris (pH 5.5), 0.2 M (NH₄)₂SO₄, 25% (w/v) PEG 3350; BpLGDH (24 mg ml⁻¹ in 20 mM Tris pH 8.0), 0.2 M ammonium citrate dibasic (pH 5.5), 20% (w/v) PEG 3350; BuGH2A (10 mg ml⁻¹ in 20 mM Tris, pH 8.0), 0.2 M ammonium citrate dibasic, 20% (w/v) PEG 3350. For complexes, protein crystals were soaked in well solution supplemented with ligands at 1 mM (GIF), 5 mM (pNP- α -gal) and NP2, or 20 mM (L-gal). Crystals were cryoprotected using well solution supplemented with 30% ethylene glycol and frozen by immersion in liquid nitrogen.

Diffraction data were collected at the Canadian Light Source (CLS) on beamline 08ID-1 and the Stanford Synchrotron Radiation Lightsource (SSRL, Stanford, CA), on beamlines BL7-1, as indicated in Supplementary Table 4. Alternatively, diffraction data were collected on an 'in-house' instrument comprising either a Pilatus 200K 2D detector coupled to a MicroMax-007HF X-ray generator with a VariMaxTM-HF ArcSec Confocal Optical System and an Oxford Cryostream 800 (indicated as Pilatus in Supplementary Table 4) or a Rigaku R-Axis4++ detector coupled to a MicroMax-002 X-ray generator with Osmic 'Blue' optics and an Oxford Cryostream 700 (indicated as R-Axis in Supplementary

Table 4). All diffraction data collected at the CLS or in house using the R-Axis4++ detector were processed using MOSFLM and SCALA^{51,52}. Data were collected at the SSRL or in house using the Pilatus 200K 2D detector integrated, scaled and merged using HKL2000 (ref. 53).

The structure of BpGH29 was solved by molecular replacement using PHASER⁵⁴ and the structure of Bt2970 (PDB 2WVU, ref. 55) as search model. The same procedures were used to determine the structure of BuS1_11 and BpLGDH using the structures of Bf0881 (PDB 2qzu) and AKR11C1 (PDB 1YNP, ref. 23), respectively, as search models. In each case, BUCCANEER⁵⁶ was used to automatically build a model, which was finished by manual building with COOT⁵⁷ and refinement with REFMAC⁵⁸. The finished models were then used to solve the structures of proteins complex with substrate.

For all structures, the addition of water molecules was performed in COOT with FINDWATERS and manually checked after refinement. In all datasets, refinement procedures were monitored by flagging 5% of all observations as 'free'⁵⁹. Model validation was performed with MOLPROBITY⁶⁰. All data processing and model refinement statistics are shown in Supplementary Table 4.

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

Data availability

Structure coordinates and structure factors determined in this study have been deposited in the PDB with the accession codes of 7LHA, 7LJ2, 7LJJ, 7LK7, 7LNP, 5T98, 5T99 and 7LH6. Existing structure coordinates used in this study that are already deposited in the PDB are 2WVU, 1YNP, 5T9G, 7CWD, 4PSR, 5K9H, 4ZRX, 3GZA, 4OUE, 1ODU, 3UET and 2G2V. All other data are available by request.

References

- Gasteiger, E. et al. ExpASY: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31**, 3784–3788 (2003).
- Jones, D. R. et al. Analysis of active site architecture and reaction product linkage chemistry reveals a conserved cleavage substrate for an endo- α -mannanase within diverse yeast mannans. *J. Mol. Biol.* **432**, 1083–1097 (2020).
- Stevenson, T. T. & Furneaux, R. H. Chemical methods for the analysis of sulphated galactans from red algae. *Carbohydr. Res.* **210**, 277–298 (1991).
- Voiges, K., Adden, R., Rinken, M. & Mischnick, P. Critical re-investigation of the alditol acetate method for analysis of substituent distribution in methyl cellulose. *Cellulose* **19**, 993–1004 (2012).
- Patankar, M. S., Oehninger, S., Barnett, T., Williams, R. L. & Clark, G. F. A revised structure for fucoidan may explain some of its biological activities. *J. Biol. Chem.* **268**, 21770–21776 (1993).
- Heiss, C., Wang, Z. & Azadi, P. Sodium hydroxide permethylation of heparin disaccharides. *Rapid Commun. Mass Spectrom.* **25**, 774–778 (2011).
- Kariya, Y. et al. Preparation of completely 6-O-desulfated heparin and its ability to enhance activity of basic fibroblast growth factor. *J. Biol. Chem.* **275**, 25949–25958 (2000).
- Ceroni, A. et al. GlycoWorkbench: A tool for the computer-assisted annotation of mass spectra of glycans. *J. Proteome Res.* **7**, 1650–1659 (2008).
- Nielsen, S. S. in *Food Analysis Laboratory Manual* 137–141 (Springer, 2017).
- Robb, M., Hobbs, J. K. & Boraston, A. B. Separation and visualization of glycans by fluorophore-assisted carbohydrate electrophoresis. *Methods Mol. Biol.* **1588**, 215–221 (2017). in.
- Abbott, D. W. & Boraston, A. B. Quantitative approaches to the analysis of carbohydrate-binding module function. *Methods Enzymol.* **510**, 211–231 (2012).
- Tomme, P., Boraston, A., Kormos, J. M., Warren, R. A. J. & Kilburn, D. G. Affinity electrophoresis for the identification and characterization of soluble sugar binding by carbohydrate-binding modules. *Enzym. Microb. Technol.* **27**, 453–458 (2000).
- Masuko, T. et al. Carbohydrate analysis by a phenol–sulfuric acid method in microplate format. *Anal. Biochem.* **339**, 69–72 (2005).
- Rogowski, A. et al. Glycan complexity dictates microbial resource allocation in the large intestine. *Nat. Commun.* **6**, 7481 (2015).
- Powell, H. R. The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Cryst.* **55**, 1690–1695 (1999).
- Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Mccoy, A. J. et al. Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
- Lammerts van Bueren, A. et al. Analysis of the reaction coordinate of α -L-fucosidases: a combined structural and quantum mechanical approach. *J. Am. Chem. Soc.* **132**, 1804–1806 (2010).
- Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D. Biol. Crystallogr.* **62**, 1002–1201 (2006).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **66**, 486–501 (2010).
- Murshudov, G. N. et al. REFMAC 5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **67**, 355–367 (2011).
- Brünger, A. T. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475 (1992).
- Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **66**, 12–21 (2010).

Acknowledgements

This research was supported by Natural Sciences and Engineering Research Council of Canada Discovery Grants (nos. FRN 04355 to A.B.B. and FRN 03929 to W.F.Z.), funding from Agriculture and Agri-Food Canada (project grant no. J-002262 to D.W.A.) and Beef and Cattle Research Council (grant no. FOS.04.17 / J-001973 to D.W.A.). G.R. was supported by funding from the European Union's Horizons 2020 research and innovations program under the Marie Skłodowska-Curie grant agreement no. 840804. We thank the staff at the CLS where diffraction data were collected. The CLS is supported by the Natural Sciences and Engineering Research Council of Canada, the National Research Council of Canada, the Canadian Institutes of Health Research, the Province of Saskatchewan, Western Economic Diversification Canada and the University of Saskatchewan.

Author contributions

C.S.R. and A.B.B. initiated the project. A.B.B. and D.W.A. directed the project. C.S.R. performed structural analyses of BuS1_11 and BpGH29 and activity analyses of the S1_11 and GH29 enzymes. J.K.H. performed all galactose release assays for the enzymes in this study and kinetic assays of BpLGDH. B.P. determined the structures of BuGH2A. G.R. and G.G. performed glycan uptake studies. L.K., S.M. and C.A. performed growth assays. C.V. determined the structure of BpLGDH. A.G.H. performed kinetic analysis of BpS1_11. R.H. performed binding studies of SusE-like proteins. N., X.X., T.M. and W.F.Z. performed carbohydrate analysis by NMR, GC–MS and LC–MS. A.B.B. wrote the paper with input from C.S.R. and D.W.A. All authors read and approved the final paper.

Competing interests

The authors declare no competing interests.

Additional information

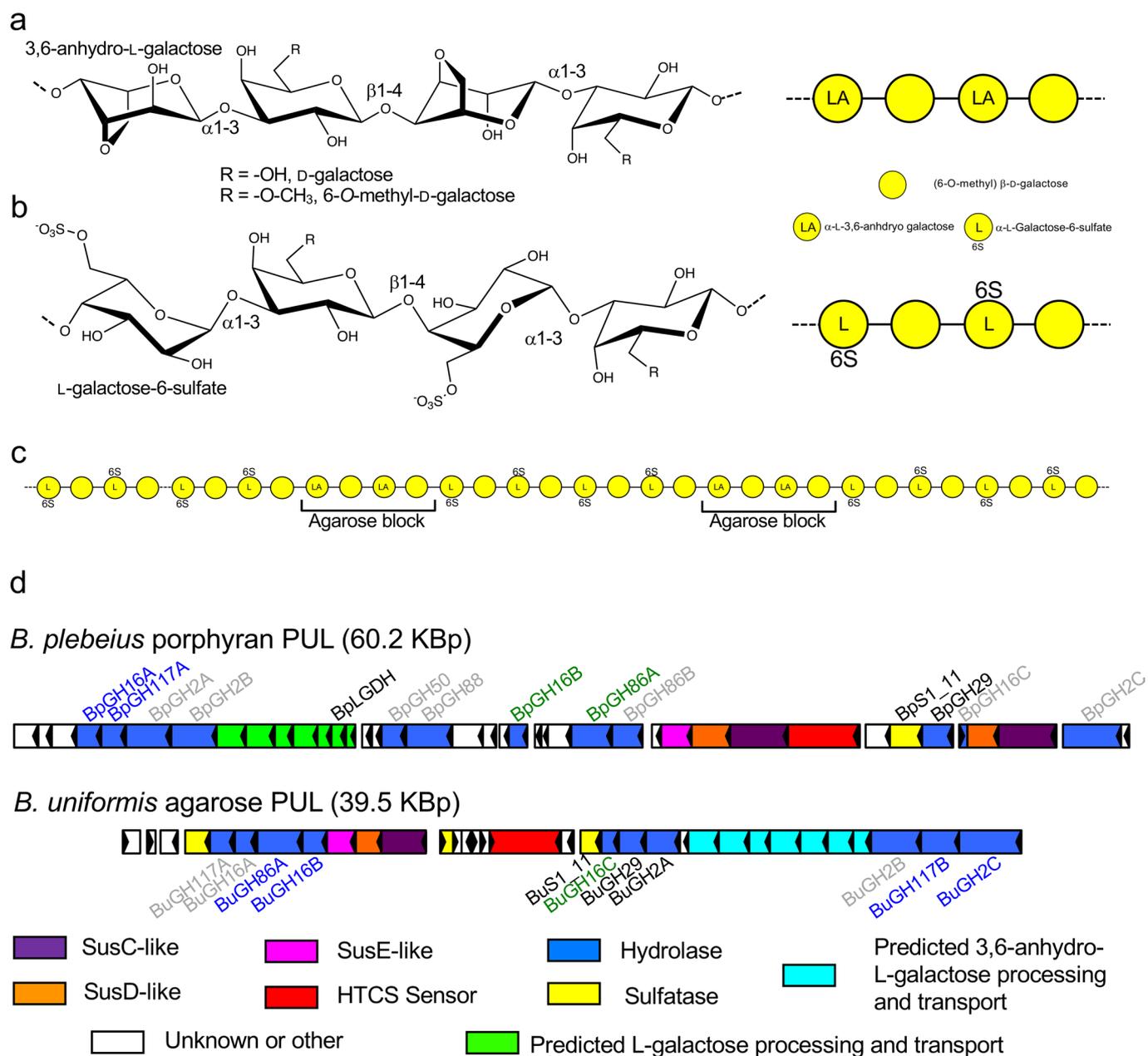
Extended data are available for this paper at <https://doi.org/10.1038/s41589-022-00983-y>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41589-022-00983-y>.

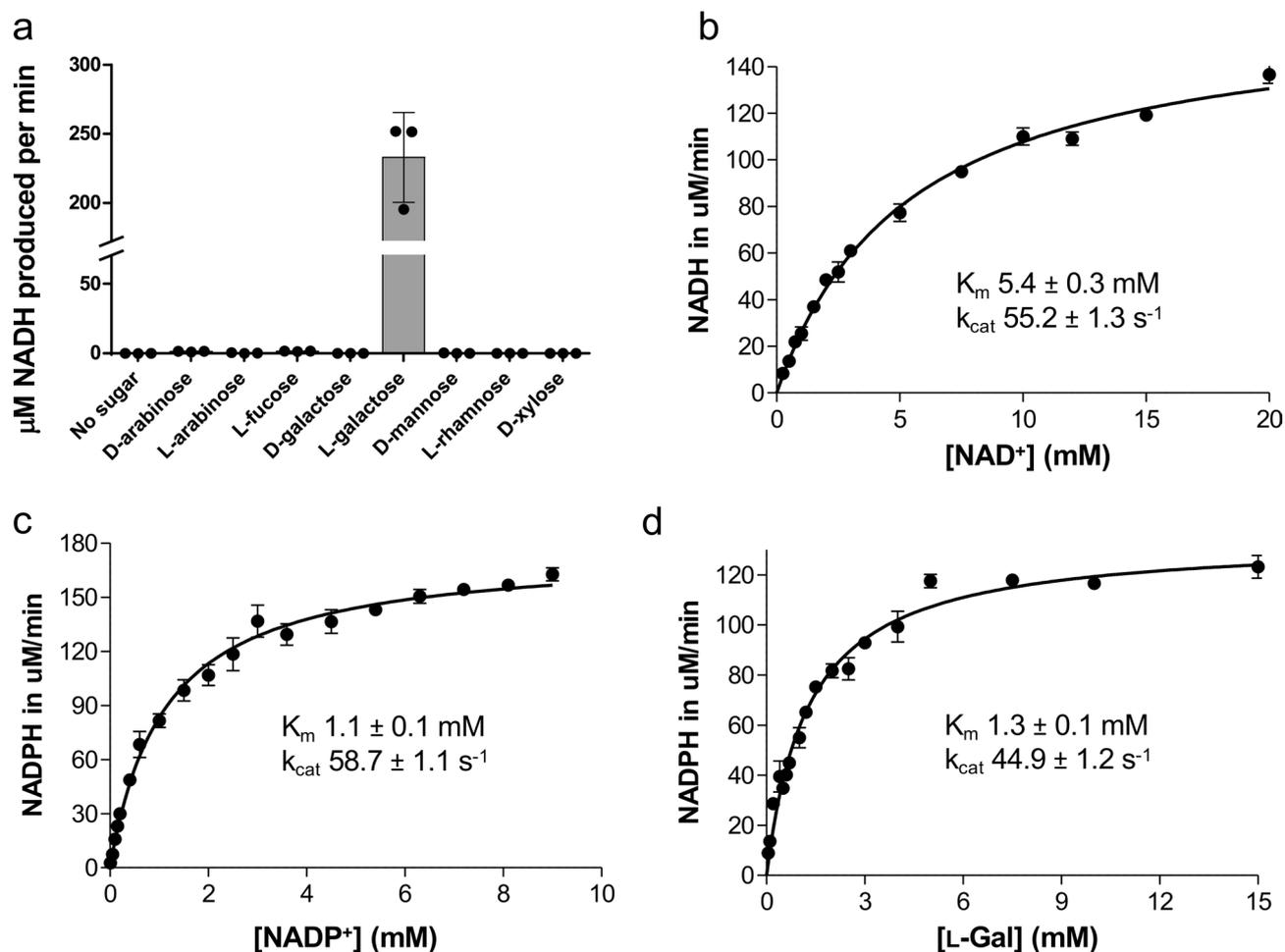
Correspondence and requests for materials should be addressed to Alisdair B. Boraston.

Peer review information *Nature Chemical Biology* thanks Uwe Bornscheuer and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

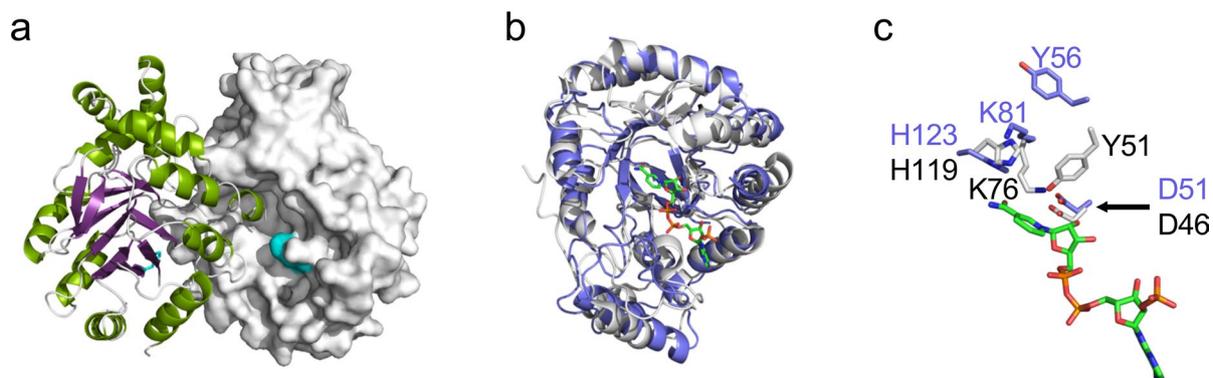
Reprints and permissions information is available at www.nature.com/reprints.



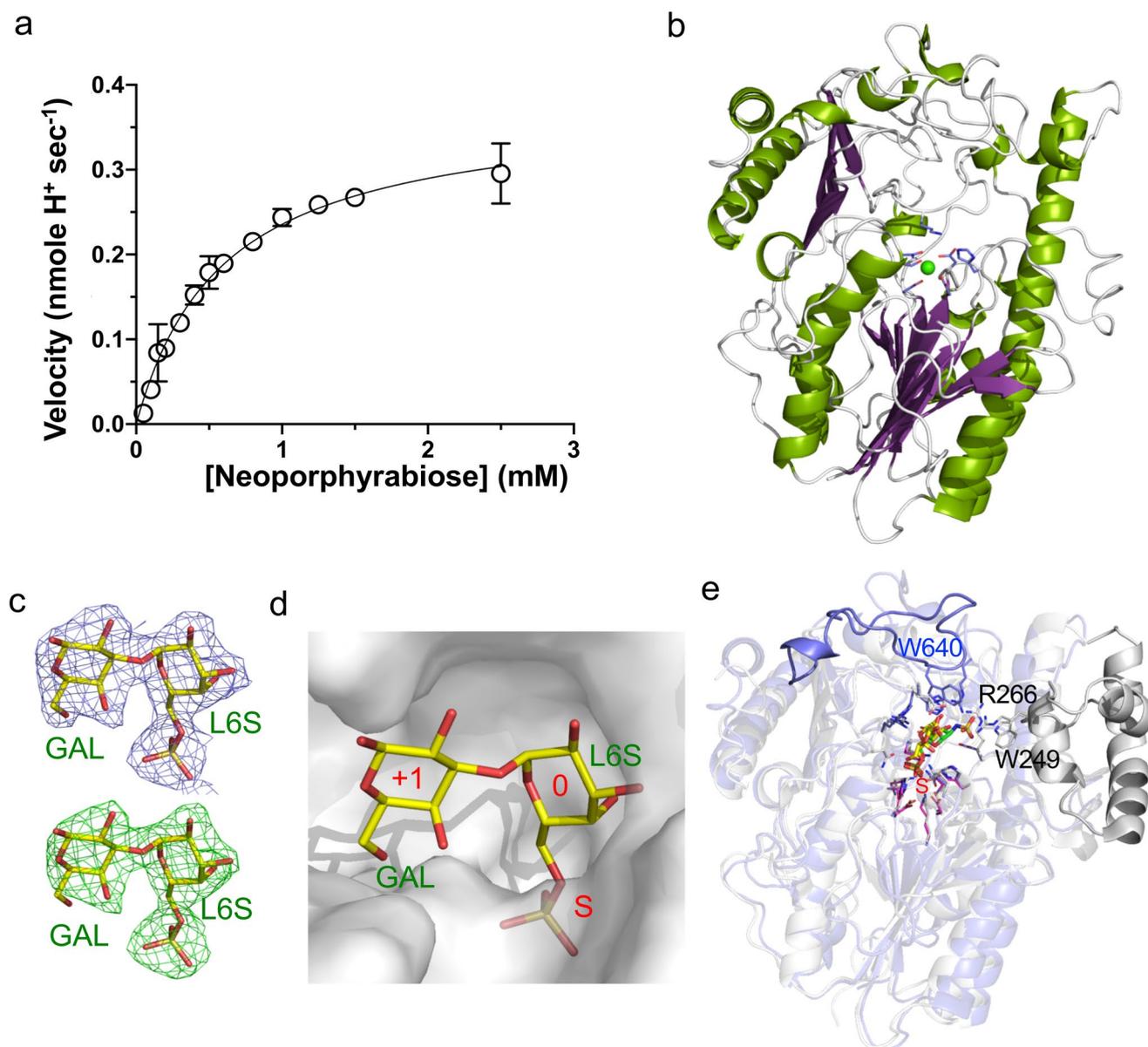
Extended Data Fig. 1 | Structures of algal galactans and the polysaccharide utilization loci (PUL) that target them. The chemical structures of **a**, agarose, **b**, porphyran, and **c**, native porphyran comprising blocks of porphyran and agarose. **d**, The PorPUL is shown on top with the AgaPUL beneath. Individual genes are labelled by general function according to the provided legend. Glycoside hydrolases with known agarose and porphyran activity have blue and green labels, respectively. Black labels indicate gene products whose activity are identified in this study. Gray labels indicate unknowns. See also Supplementary Table 1.



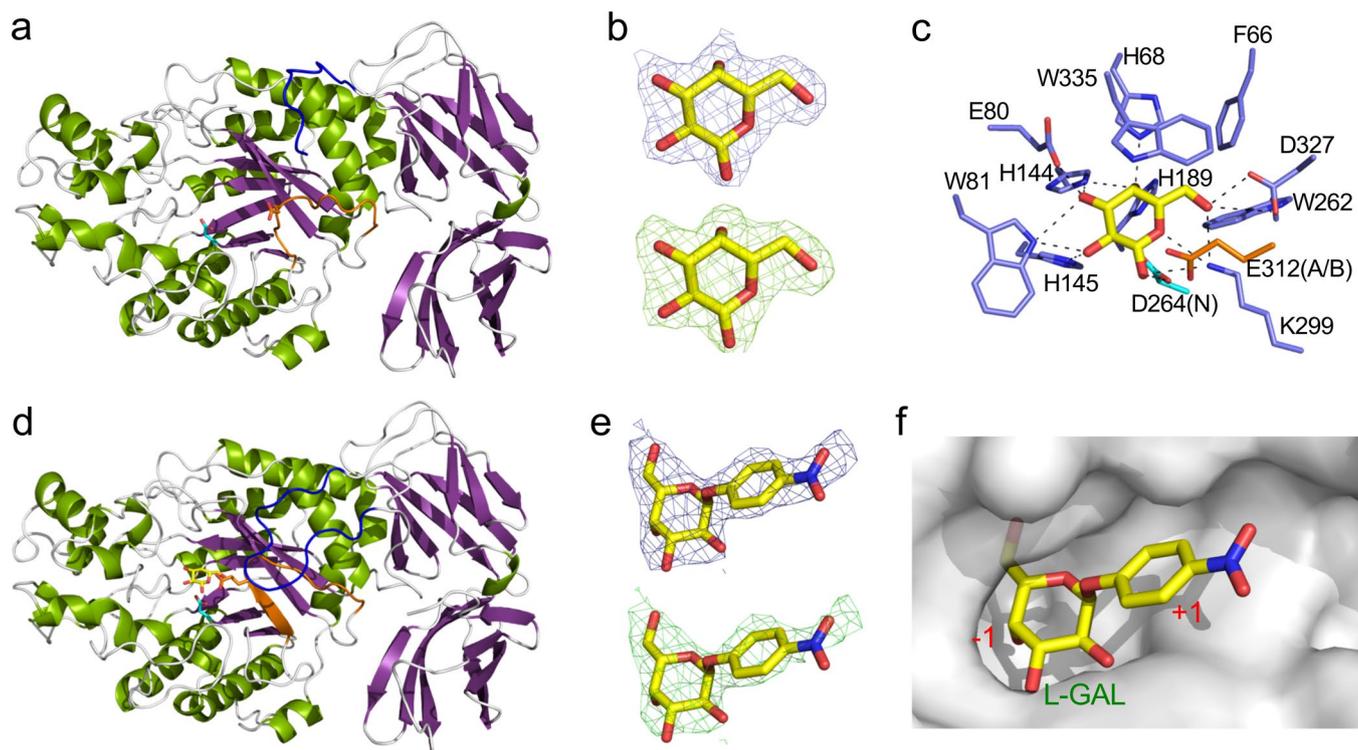
Extended Data Fig. 2 | Activity analysis of BpLGDH. **a**, screen of BpLGDH activity against various monosaccharides using NAD⁺ as a co-factor. $n=3$ independent reactions were performed with each replicate shown along with the mean \pm standard deviation. **b** and **c** show the kinetics using NAD⁺ and NADP⁺, respectively, as a co-factor when excess L-galactose is used as a substrate. **d**, the kinetics of L-galactose oxidation using excess NADP⁺ as a co-factor. The data is shown as the mean \pm standard deviation of $n=3$ independent reactions (see Supplementary Fig. 1 for independent data points). The solid line shows the best fit line to the Michaelis-Menton equation.



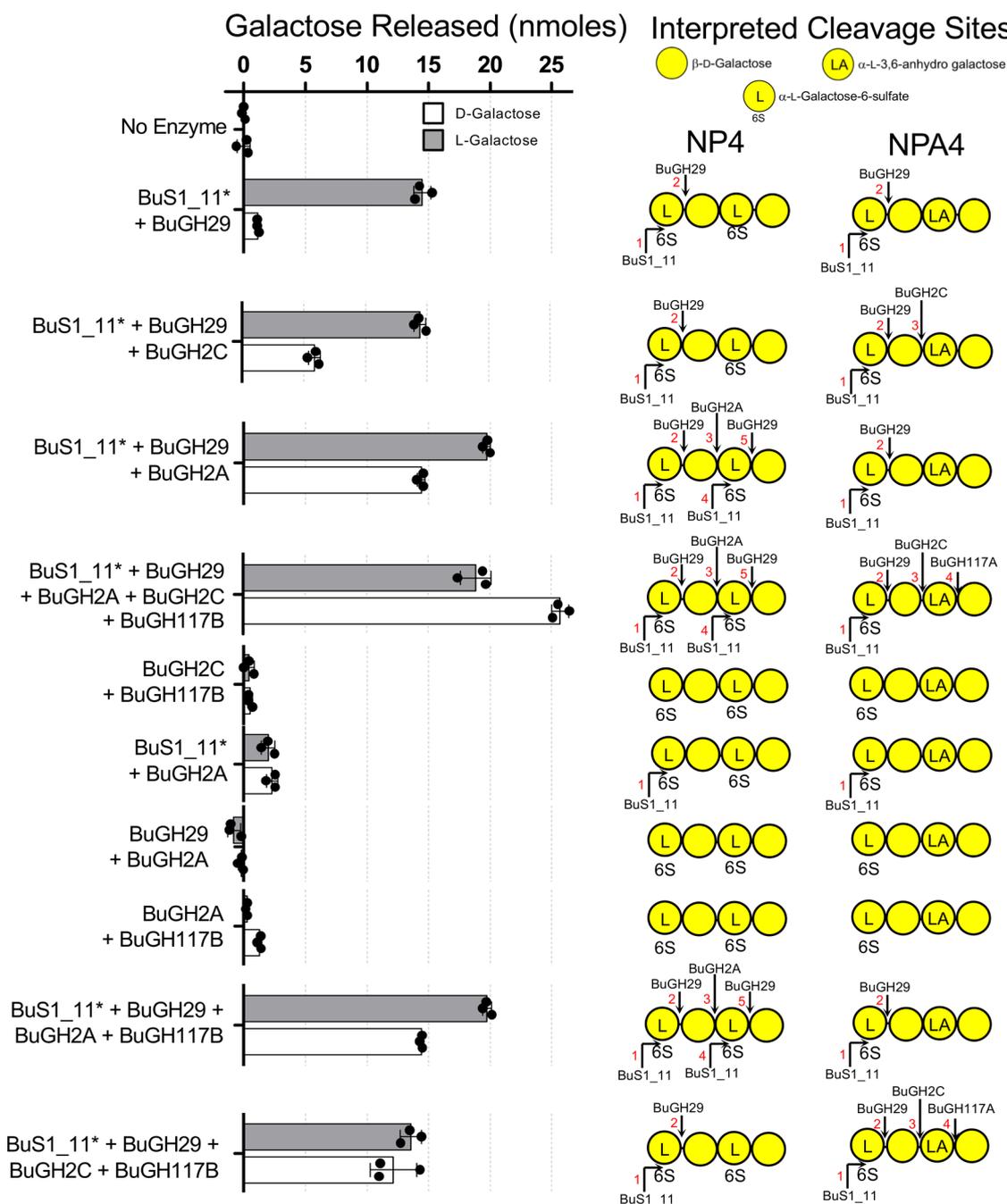
Extended Data Fig. 3 | Structural analysis of BpLGDH. **a**, structure of the BpLGDH dimer with one monomer shown in cartoon and the other as a solvent accessible surface. The putative catalytic histidine is shown in blue. **b**, overlap of BpLGDH (blue) with AKR11C1 from *Bacillus halodurans* (grey, PDB ID 1YNP). **c**, overlap of the BpLGDH (blue) active site with the active site of AKR11C1 (grey). Conserved residues, including the putative catalytic histidine (H119 in AKR11C1), are shown as sticks. In panels **c** and **d** the NADPH bound to AKR11C1 is shown in stick representation.



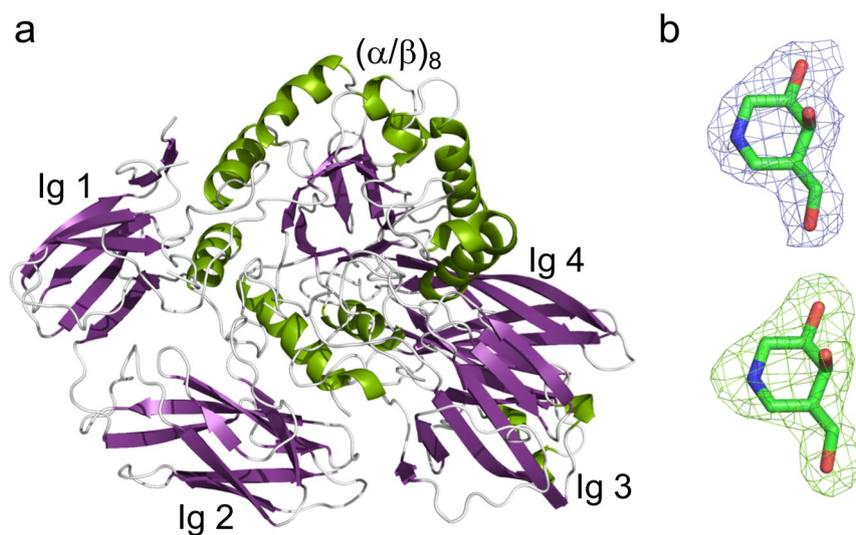
Extended Data Fig. 4 | Kinetic and structural analysis of BuS1_11. **a**, kinetic analysis of neoporphyribose hydrolysis by BpS1_11*. The data is shown as the mean \pm standard deviation of $n=3$ independent reactions (see Supplementary Fig. 2 for independent data points). The solid line shows the best fit line to the Michaelis-Menten equation. **b**, the overall fold of BpS1_11 shown as a cartoon. The bound calcium ion is shown as a green sphere and key residues in the active site are shown as blue sticks. **c**, electron density maps of NP2 (yellow sticks) bound to the BuS1_11 active site are shown with the 2Fo-Fc map at 1σ in blue (top) and the Fo-Fc omit map at 3σ in green (bottom). **d**, the solvent accessible surface of the active site is shown in transparent grey with the bound NP2 as yellow sticks. The subsites of the active site are labeled in red according to the nomenclature of Hettle *et al.*²⁷. **e**, structural overlap of BuS1_11 (blue) with BT4656 (grey, 5G2V). NP2 is shown as yellow sticks and 2-N,6-O-disulfo-D-glucosamine bound to BT4656 shown as green sticks. Relevant inserted structural motifs in each protein that contribute to carbohydrate specificity are shown as solid cartoons.



Extended Data Fig. 5 | Structural analysis of BpGH29. **a**, cartoon representation of uncomplexed BpGH29. **b**, electron density maps of L-galactose (yellow sticks) bound to the BpGH29 active site are shown with the 2Fo-Fc map at 1σ in blue (top) and the Fo-Fc omit map at 3σ in green (bottom). **c**, the specific interactions between the active site and L-galactose. The nucleophile (N) and acid/base (A/B) are colored blue and orange respectively. **d**, cartoon representation of BpGH29 in complex with L-galactose. In panels a and c the sidechain proposed to act as the nucleophile (D264) is shown as blue sticks, the sidechain proposed to act as the acid/base in orange sticks, and mobile loops that help form the active site in orange and blue. **e**, electron density maps of pNP- α -L-galactopyranoside (yellow sticks) bound to the BpGH29 D264N active site are shown with the 2Fo-Fc map at 1σ in blue (top) and the Fo-Fc omit map at 3σ in green (bottom). **f**, pNP- α -L-galactopyranoside bound to the BpGH29 D264N active site shown with the solvent accessible surface of the active site shown in grey.



Extended Data Fig. 6 | Enzymatic activity and sequencing of a purified tetrasaccharide derived from porphyran. The enzyme combinations are indicated on the left with the amounts of D-galactose and L-galactose release displayed as a bar chart. The sites of bond hydrolysis are indicated in the schematics on the right side of the figure. Thirteen nanomoles (13 nmoles) of tetrasaccharide were used in each reaction and each reaction was performed in independent triplicate ($n=3$) reactions with each replicate shown along with the mean \pm standard deviation.



Extended Data Fig. 7 | Structural analysis of BuGH2A. **a**, cartoon representation of BuGH2A showing the central $(\alpha/\beta)_8$ -barrel sitting in a nest of four Ig-like domains. **b**, electron density maps of galactoisofagmine (green sticks) bound to the BuGH2A active site are shown with the 2Fo-Fc map at 1σ in blue (top) and the Fo-Fc omit map at 3σ in green (bottom).

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection SoftMax Pro 6.2.1, HKL 3000, Blue-ice, Web-ice, MxD, CD Accuri™ C6 Plus Software (BD-Accuri Cytometers, BD Bioscience, USA), For super-resolution microscopy: Zen2011 (Carl Zeiss)

Data analysis Excel (various versions), GraphPad Prism 5.0, CCP4/CCP4i 7.0.076, MOSFLM, SCALA, PHASER 2.7.17, BUCCANEER 1.6.3, COOT 0.8.9.2, REMAC5.9.0238, AutoSol, MOLPROBITY, Phenix 1.13, Gen5 software, R version 3.6.3, Zen2011, MestReNova software version 14.2.1-27684, Thermo Xcalibur, MassHunter software (version B.07.00), GlycoWorkbench 2.1 stable build 146, FlowJo software (v10.4.2)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Structure coordinates and structure factors determined in this study have been deposited in the Protein Data Bank with the accession codes of 7HLA, 7LJ2, 7LJJ, 7LK7, 7LNP, 5T98, 5T99, and 7LH6. Existing structure coordinates used in this study that are already deposited in the Protein Data Bank are 2WVU, 1YNP, 5T9G, 7CWD, 4PSR, 5K9H, 4ZRX, 3GZA, 4OUE, 1ODU, 3UET, and 2G2V.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The field standard of triplicate or quadruplicate independent samples was used for microbial growth assays and enzyme activity experiments. No statistical analyses were performed on the results of these experiments. Flow cytometry 10,000 cells.
Data exclusions	No data was excluded.
Replication	All experiments (except structure determinations) were performed at least three times (n=3), as indicated in the manuscript. Results were reproducible.
Randomization	This is not relevant as there were no specific test groups.
Blinding	Blinding was not relevant to biochemical assays in this study as the quantitative assay readouts cannot be influenced by awareness of the investigator. Blinding is not possible for structural analyses.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

n/a	Involvement
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

n/a	Involvement
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging