A marine bacterial enzymatic cascade degrades the algal polysaccharide ulvan

Lukas Reisky, Aurelie Prechoux, Marie-Katherin Zühlke, Marcus Bäumgen, Craig Robb, Nadine Gerlach, Thomas Roret, Christian Stanetty, Robert Larocque, Gurvan Michel, et al.

A marine bacterial enzymatic cascade degrades the algal polysaccharide ulvan

Lukas Reisky,[1#] Aurélie Préchoux,[2#] Marie-Katherin Zühlke,[3,4#] Marcus Bäumgen,[1] Craig S. Robb,[5,6] Nadine Gerlach,[5,6] Thomas Roret,[7] Christian Stanetty,[8] Robert Larocque,[7] Gurvan Michel,[2] Song Tao,[5,6] Stephanie Markert,[3,4] Frank Unfried,[3,4] Marko D. Mihovilovic,[8] Anke Trautwein-Schult,[9] Dörte Becher,[9] Thomas Schweder,[3,4*] Uwe T. Bornscheuer,[1*] Jan-Hendrik Hehemann[5,6*]

[1] Department of Biotechnology & Enzyme Catalysis, Institute of Biochemistry, University Greifswald, 17487 Greifswald, Germany

[2] Sorbonne Université, CNRS, Integrative Biology of Marine Models (LBI2M), Station Biologique de Roscoff (SBR), 29680 Roscoff, Bretagne, France

[3] Pharmaceutical Biotechnology, Institute of Pharmacy, University Greifswald, 17487 Greifswald, Germany

[4] Institute of Marine Biotechnology, 17489 Greifswald, Germany

[5] Max Planck-Institute for Marine Microbiology, 28359 Bremen, Germany

[6] University of Bremen, Center for Marine Environmental Sciences (MARUM), 28359 Bremen, Germany

[7] Sorbonne Université, CNRS, FR 2424, Station Biologique de Roscoff (SBR), 29680 Roscoff, Bretagne, France

[8] Institute of Applied Synthetic Chemistry, TU Wien, 1060 Vienna, Austria

[9] Institute of Microbiology, University Greifswald, 17487 Greifswald, Germany

[#]These authors contributed equally to this work.

*To whom correspondence should be addressed. e-mail: schweder@uni-greifswald.de (TS); uwe.bornscheuer@uni-greifswald.de (UB); jhhehemann@marum.de, jheheman@mpi-bremen.de (JHH)

Abstract

Marine seaweeds increasingly grow into extensive algal blooms, which are detrimental to coastal ecosystems, tourism, and aquaculture. However, algal biomass is also emerging as sustainable raw material for bioeconomy. The potential exploitation of algae is hindered by our limited knowledge of the microbial pathways – and hence the distinct biochemical functions of the enzymes involved – that convert algal polysaccharides into oligo- and monosaccharides. Understanding these processes would be essential, however, for applications like the fermentation of algal biomass into bioethanol or other value-added compounds. Here we describe the metabolic pathway that enables the marine flavobacterium *Formosa agariphila* to degrade ulvan, the major cell wall polysaccharide of bloom-forming *Ulva* species. The pathway involves 12 biochemically characterized carbohydrate-active enzymes, including two polysaccharide lyases, three sulfatases and seven glycoside hydrolases that sequentially break down ulvan into fermentable monosaccharides. This way, the enzymes turn a previously unexploited renewable into a valuable and ecologically sustainable bioresource.

Introduction

Algal photosynthesis provides half of the global primary production[1]. Carbon dioxide is converted into carbohydrates, which are polymerized into polysaccharides to store energy, build cell walls, and perform other biological functions. Algae are furthermore considered as a promising renewable carbon source, due to their competitive growth rates and unique cell walls. Unlike plants that are rich in woody tissue, comprising the insoluble polysaccharides cellulose and the aromatic polymer lignin, which increases recalcitrance against enzymatic digestion, algal cell walls are rich in gel-forming polysaccharides that are highly hydrated[2]. Hydration and the absence of lignin make harsh chemical and physical pretreatment of cell walls unnecessary, and allow for easy access of enzymes that can digest the

polysaccharides into fermentable monosaccharides. Accordingly, recent studies showed that bioengineered microbes equipped with agarases and alginate lyases can efficiently digest and rapidly convert polysaccharides from brown and red algae into bioethanol[3].

Sessile macroalgae, such as brown algae that form kelp forests, are ecologically valuable because they provide nutrition and habitats for fish and other organisms and, consequently, harvesting them would exacerbate pressure on natural populations. However, the planktonic macroalgae *Ulva armoricana*, *Ulva rotunda* and other *Ulva* spp. that thrive in eutrophic, nutrient-rich coastal waters, grow into expansive blooms that occur with increasing frequency in recent years. They pose ecological but also economical threats when they accumulate on beaches used for recreation[4-6]. Fertilized by nitrate from agriculture that is washed into the ocean by rivers, *Ulva* blooms during summer produce up to 50–100 000 tons of biomass every year, which must be removed at high expense from the northern and western coast of France[6]. Even larger blooms occur in China[5]. Blooms of *Ulva* are thus a global phenomenon that is bound to increase with farming activities, rendering the polysaccharide ulvan, which accounts for up to 30 % of the algal dry weight[7], an emerging yet untapped resource.

Ulvan is a branched polysaccharide composed of repeating disaccharide units, in which D-glucuronic acid (GlcA) is β-1,4-linked or L-iduronic acid (IdoA) is α-1,4-linked to L-rhamnose-3-sulfate (Rha3S), which is α-1,4-linked within the main chain. Some of the uronic acids are replaced by β-1,4-linked D-xylose (Xyl), which can be sulfated at position 2 (Xyl2S). Furthermore, Rha3S can be modified by β-1,2-linked GlcA side chains and the GlcA-Rha3S or IdoA-Rha3S pattern can be interrupted by consecutive GlcA residues[7-9]. Increased interest in the enzymatic degradation of ulvan recently led to the description of several ulvan-active enzymes[10-15]. So far, and to the best of our knowledge, only two types of enzymes from different carbohydrate-active enzyme (CAZyme) families showed activity on ulvan. Ulvan polysaccharide lyases of the families PL24, PL25 and PL28 catalyze the initial cleavage between Rha3S and GlcA or IdoA, resulting in the formation of unsaturated uronic acid residues at the end of the formed oligosaccharide. Unsaturated uronic acid residues are

removed by glycoside hydrolases (GHs) from the family GH105[15,16]. Ulvan-specific degradation-related gene loci ('polysaccharide utilization loci', PULs) such as PUL H from *Formosa agariphila* encode PL28 and GH105 together with over 10 additional, putative enzymes, which were predicted to be involved in ulvan utilization. While PL28 and GH105 degrade ulvan, the other enzymes that were produced in *Escherichia coli* did not show activity[15]. This result suggested that a complex cascade of sequential enzymatic reactions is required for complete ulvan degradation[15,17].

Here, we experimentally established the complex ulvan degradation pathway of *F. agariphila* KMM 3901^T, a marine flavobacterium, which was isolated from a green alga in the Sea of Japan[18]. These degradation-related enzymes are encoded in an ulvan-specific PUL in the bacterial genome[15].

Results

Bacterial ulvan-specific PULs

To decipher the ulvan degradation pathway, we first searched microbial genomes hosted at NCBI for potential ulvan-specific PULs using the known ulvan lyase PL28 as query. We identified 12 putative ulvan PULs in 12 Bacteroidetes genomes (Fig. 1a), including the recently discovered PUL H of *F. agariphila*[15,18], a more than 75 kb long genomic region consisting of 39 genes (Fig. 1b). We verified the boundaries of PUL H with a comparative global proteome analysis of *F. agariphila* cells fed with ulvan and with control substrates (rhamnose and fructose), respectively, as sole carbon source. Ulvan promoted bacterial growth (Supplementary Fig. 1) and elicited quantitative changes of most proteins that are encoded by PUL H (Table 1, Fig. 1b, and Supplementary Fig. 2). Besides ulvan, also the monosaccharide rhamnose induced, albeit less strongly, the expression of PUL H genes. For a few proteins (P2_SusD, P3_TBDR, P8_GH2) even higher protein amounts were detected with rhamnose, compared to ulvan. The increased abundance of enzymes involved in the

degradation of ulvan-derived monosaccharides indicated a co-regulation of genes for the metabolization of ulvan and its corresponding monosaccharides (Table 1, Supplementary Figs. 2 and 3, Supplementary Data Sets 1 and 2). PUL H includes 17 potential carbohydrate-active enzymes (CAZymes) from different GH and PL families (http://www.cazy.org/[19]) and eight sulfatases from five S1 subfamilies (http://abims.sb-roscoff.fr/sulfatlas/[20]). For most of these enzymes, their role in ulvan depolymerization remains unknown. A co-occurrence analysis of putative enzymes and associated genes within the set of 12 PULs from marine Bacteroidetes identified conserved CAZymes in the putative ulvan pathways (Fig. 1c). This analysis allowed us to focus our biochemical experiments on a smaller subset of CAZymes and sulfatases, whose involvement in ulvan utilization was suggested by our proteomic results (Fig. 1b).

In addition to the two already known[15] ulvanolytic enzyme activities (ulvan lyase and unsaturated glucuronyl hydrolase, GH105) we uncovered eight so far unknown enzyme functions for the complete depolymerization of ulvan. Besides a novel PL family, we identified and characterized six GH families (GH2, GH3, GH39, GH43, GH78, GH88) and three sulfatases.

Activity-based screenings of these enzymes were used to identify their function in the ulvan degradation pathway. The selection of putative CAZymes and sulfatases for cloning, heterologous expression and characterization was guided by the co-occurrence analysis of genes in the diverse ulvan PULs (Fig. 1c).

Sulfatases active on ulvan

Ulvans feature a large structural variability, with substitution by sulfate esters at various positions. This chemical diversity is influenced by several factors such as the algal species, the environmental conditions or the seasons[7]. The studied PUL of *F. agariphila* encodes 8 formylglycine-dependent sulfatases belonging to 5 subfamilies of the SulfAtlas S1 family

(Table 1): S1_7: 3 genes; S1_8: 2 genes; S1_16: 1 gene; S1_25: 1 gene; S1_27: 1 gene (http://abims.sbroscoff.fr/sulfatlas)[20]. With such a diversity of S1 subfamilies, these sulfatases likely display significant differences in substrate recognition, even though they are all predicted to act on ulvans. We expressed 7 sulfatases in soluble form in *E. coli*. After purification, these recombinant sulfatases were incubated with ulvan polymers from three different sources (Agrival, Elicityl, and one extracted from an Altantic *Ulva* sp. collected in Roscoff, France). As shown by the HPAEC analyzes of released sulfate ions, 6 sulfatases are clearly active on ulvan polymers, although their activity varies depending on the polysaccharide sources (Supplementary Fig. 4). The sulfatase P18_S1_7 (for numbering/nomenclature see Table 1) was most active on ulvan polymers, particularly on the xylose-rich ulvan (Supplementary Figs. 4 and 5) and can desulfate oligosaccharides containing the motif Rha3S-Xyl2S-Rha3S. Thus, this sulfatase likely proceeds in an endolytic mode of action. This assumption is consistent with the “open groove” topology of the active site unraveled by the P18_S1_7 crystal structure (Fig. 2a and 2g). Interestingly, P14_S1_7 (predicted as exolytic, since this sulfatase is almost inactive on ulvan, Supplementary Fig. 4) and P18_S1_7 (predicted as endolytic) belong to the same subfamily (S1_7). Such dissimilar modes of action within the same (sub)family have been described in glycoside hydrolase and polysaccharide lyase families[21,22]. In comparison to P18_S1_7, the S1_25 sulfatase module of P36 (referred to as P36_S1_25) presents moderate activities on polymers. On oligosaccharides, P36_S1_25 was the most active enzyme. This enzyme specifically desulfates L-rhamnose at the 3-position and can act on the motif Rha3S-Xyl-Rha3S in an exolytic mode of action.

Sequence analyses revealed that P18_S1_7 (485 residues) and the S1_25 sulfatase module of P36_S1_25 (443 residues) are only distantly related (25% identity) and thus belong to two different SulfAtlas S1 subfamilies[20], S1_7 and S1_25, respectively. We determined the crystal structure of these two sulfatases, with higher resolution for P18_S1_7 (1.23 Å) and lower resolution for the sulfatase module of P36_S1_25 (2.91 Å). P18_S1_7 and P36_S1_25 adopt a similar fold with two α/β-structural domains, an *N*-terminal catalytic domain SD1

(Ser25-Asp388; P18_S1_7) separated by a structured linker (Arg389-Val397) from a *C*-terminal domain SD2 (Ala398-Pro483). Nonetheless, the sulfatase module of P36_S1_25 is a smaller protein and lacks some secondary elements, which are present in P18_S1_7 (the β-strands β6 and β8, the α-helices α5, α7, α8 and several short 3:10 helices). Notably, the helix α7 and the loops connecting it to the main part of SD1 constitute a protruding extension, which overhangs the active site (Fig. 2d). The active site of P18_S1_7 is a large, open groove with a strong basic character (Fig. 2a and 2g). This type of active site topology is consistent with the endo-character and its efficiency on polymeric ulvan (Supplementary Fig. 4). In contrast, the active site of P36_S1_25 is a pocket (Fig. 2c), which is consistent with its activity on oligosaccharides (Supplementary Fig. 4). The most similar protein in the Protein Data Bank (PDB) is the human iduronate 2-sulfatase (IDS, 31% sequence identity; PDB: 5FQL; Fig. 2b and 2e)[23]. Interestingly, IDS also displays a pocket active site topology (Fig. 2b). Therefore, different active sites (and subsequently different modes of action) can exist within the same S1 subfamily. Such differences in topology likely explain the varying efficiencies at the polymer level observed for P11_S1_7, P14_S1_7 and P18_S1_7 although they all belong to the S1_7 subfamily (Supplementary Fig. 4).

The catalytic machinery of the S1 family sulfatases[24] is well conserved in P18_S1_7 and P36_S1_25. We find the catalytic nucleophile (Cys74 and Cys58, respectively), residues involved in Ca^{2+} coordination (Asp35, Asp36, Asp312 and His313; Asp18, Asp19, Asp284 and Asn285), residues stabilizing the catalytic nucleophile (Arg78 and His128; Arg62 and Gly110), and residues of the sulfate-binding S subsite, as defined in the recent nomenclature for sulfatase-binding subsites[25] (Lys125, His213 and Lys325; Lys108, His182 and Lys297) (Fig. 2g, Supplementary Figs. 6-8). His313 in P18_S1_7 is not the most frequent residue for the coordination of the calcium ion (usually an asparagine), but a histidine at this position is found in a minority of sulfatases and is part of the updated PROSITE signature “Calcium-binding site 2”[20]. Most surprising is the replacement of His128 in P18_S1_7 by Gly100 in the sulfatase module of P36_S1_25. Indeed, a histidine at this position is supposed not only to stabilize the catalytic formylglycine, but also to abstract its Oγ2 proton at the end of the

188 catalytic cycle to induce the sulfate elimination and the aldehyde regeneration[24].
189 Nonetheless, this glycine is strictly conserved in the closest homologs of the sulfatase
190 module of P36_S1_25 (105 sequences with >50% identity; Supplementary Fig. 7),
191 suggesting that the function of the histidine at this position may not be essential in this ulvan
192 sulfatase subgroup.
193 While some sulfatases were not quantified in our metabolic labeling approach (Fig. 1b, Table
194 1), they were detected by subproteome analysis in the membrane-enriched fraction. In five
195 cases, lipoprotein signal peptides were predicted and P18_S1_7 and P36_S1_25 were
196 highly abundant in the intracellular soluble fraction (Supplementary Data Set 3). Taken
197 together, these results indicate a periplasmic localization of sulfatases, with some of them
198 putatively membrane-bound. Notably, the sulfatase P36_S1_25 activity is found in a
199 multimodular enzyme that contains also a GH78 domain. Comparative genome analyses
200 indicated multimodular enzyme structures in the ulvan PUL H of *F. agariphila*[15] and other
201 putative ulvan-degrading Bacteroidetes strains (Supplementary Data Set 4).

202

203 **Enzymatic ulvan degradation**

204 In brief, the distinct function of each enzyme was established by activity testing on ulvan and
205 on defined enzymatically produced ulvan oligomers using photometric assays, fluorophore-
206 assisted carbohydrate electrophoresis (FACE) and carbohydrate polyacrylamide gel
207 electrophoresis (C-PAGE), high performance anionic exchange chromatography with pulsed
208 amperometric detection (HPAEC-PAD) and mass-spectrometry. Detailed procedures of
209 these steps are outlined in the Online Methods section. Structures of all important
210 carbohydrate intermediates were confirmed by 1D and 2D nuclear magnetic resonance
211 (NMR) spectroscopy together with mass-spectrometry analysis.

212 We performed an initial photometric screening, which detects the unsaturated uronic acid
213 moiety (Δ) introduced by the lytic mechanism of lyases. We show that P10_PLnc and
214 P30_PL28 are both endo-acting ulvan lyases generating the same product pattern, implying

215 that they have a similar specificity (Supplementary Fig. 9). P30_PL28 accepts GlcA and IdoA
216 at the cleavage site and generates the dimer Δ-Rha3S and the tetramer Δ-Rha3S-Xyl-Rha3S
217 as main products[14]. Both ulvan lyases, P30_PL28[26] and P10_PLnc, appear to initiate ulvan
218 depolymerization outside of the bacterial cell. P30_PL28 contains an additional ulvan-binding
219 module[13] and a type IX secretion system signal that drives secretion[27], corroborating the
220 proteomic results (Supplementary Fig. 10, Supplementary Data Sets 2 and 3). P10_PLnc
221 might be associated to the outer membrane (Supplementary Data Set 3), although a
222 periplasmic localization is also possible (Supplementary Fig. 10, Supplementary Data Set 2).

223 Two variants of ulvan lyase with distinct localizations indicate synergistic functions: while
224 P30_PL28 is an extracellular enzyme catalyzing rapid dissolution of insoluble ulvan,
225 P10_PLnc most likely dissolves soluble ulvan oligomers at the cell surface, where uptake
226 proceeds through the expressed TonB-dependent receptor system into the periplasm. Here,
227 the unsaturated uronyl residue (Δ) at the non-reducing end of oligomers is removed by the
228 exo-acting unsaturated glucuronyl hydrolases (outer membrane P1_GH88 and periplasmic
229 P33_GH105) (Supplementary Figs. 11-13), thus forming 5-dehydro-4-deoxy-D-glucuronate.
230 The resulting Rha3S was purified and the structure was confirmed by NMR (Supplementary
231 Figs. 14 and 15, Supplementary Table 1). This monosaccharide is desulfated by the S1_25
232 sulfatase domain of P36_S1_25 yielding rhamnose, which can enter the cellular sugar
233 metabolism (Fig. 3, Supplementary Fig. 16). Rha3S-Xyl-Rha3S was another major
234 intermediate which was isolated (Supplementary Figs. 17 and 18, Supplementary Table 2).
235 Rha3S-Xyl-Rha3S was desulfated by the sulfatase P36_S1_25 to yield Rha-Xyl-Rha3S,
236 which was isolated to confirm the desulfation site at the non-reducing end (Supplementary
237 Figs. 19-21, Supplementary Table 3). Next, Rha-Xyl-Rha3S is converted by the periplasmic
238 P20_GH78 to Rha and Xyl-Rha3S (Fig. 3, Supplementary Fig. 22). The CBM67 domain of
239 P20_GH78 likely elevates specificity for rhamnose and contributes to substrate recognition[28].
240 Finally, the dimer Xyl-Rha3S is further cleaved by P24_GH3 or P27_GH43 to yield Xyl and
241 Rha3S, making these the first identified β-xylosidases that are active on ulvan
242 oligosaccharides (Fig. 3, Supplementary Figs. 22 and 23). Notably, only the P24_GH3 was

243 previously found to be active on 4-methylumbelliferyl-β-D-xylopyranoside (MUX) showing that
244 the two enzymes have different substrate specificity at the aglycone site[15].

245 Besides ulvan lyases, the endo-active alpha-1,4-L-rhamnosidase GH39 cleaves rhamnose
246 sections interspersed between xylose residues within the polymer. Such a function has, to
247 the best of our knowledge, not been described in this family before. Accordingly, larger
248 oligomers with consecutive Xyl-Rha3S units that are resistant to the ulvan lyases P30_PL28
249 and P10_PLnc were efficiently degraded by P31_GH39 (Supplementary Fig. 24). The
250 catalytic order of ulvan lyases and P31_GH39 was interchangeable as the larger degradation
251 products of P31_GH39 were prime substrates for both ulvan lyases (Supplementary Fig. 25).
252 The dimers Xyl-Rha3S and Xyl2S-Rha3S were isolated as the smallest products and the
253 structure was elucidated by NMR, identifying GH39 as an α-rhamnosidase active on ulvan
254 (Supplementary Figs. 26-29, Supplementary Tables 4 and 5). While Xyl-Rha3S is further
255 degraded as described above, Xyl2S-Rha3S was resistant to P24_GH3 or P27_GH43 and
256 needs to be desulfated by the P32_S1_8 sulfatase prior to enzymatic conversion by these
257 enzymes (Supplementary Fig. 30). Desulfation of Xyl2S within the trimer Rha3S-Xyl2S-
258 Rha3S, released by P30_PL28 and P33_105 digestion (Supplementary Figs. 31 and 32,
259 Supplementary Table 6), was catalyzed by the P18_S1_7 sulfatase (Supplementary Fig. 33).

260 GlcA side chains present on some *O*2 residues of Rha3S[7] are removed by P17_GH2. When
261 P17_GH2 was added to untreated ulvan, it produced a single band in FACE with the same
262 mobility as a GlcA (Supplementary Fig. 34a) while not decreasing the overall molecular
263 weight of the raw ulvan as seen by C-PAGE (Supplementary Fig. 9). To confirm this activity,
264 defined oligomers with GlcA side chains were produced from ulvan with P30_PL28 and
265 P31_GH39 with or without P33_GH105. The structure of Δ-Rha3S[2GlcA]-Xyl-Rha3S and
266 Rha3S[2GlcA]-Xyl-Rha3S, was confirmed by NMR (Supplementary Figs. 35–38,
267 Supplementary Tables 7 and 8) and these products were used as substrates for P17_GH2.
268 This enzyme was also active on these smaller oligomers (Fig. 3, Supplementary Fig. 34b).
269 This result indicates that the GlcA side chains were removed from polymeric ulvan or from

smaller intermediates (Supplementary Fig. 34c), although in *F. agariphila* we predict P17_GH2 to be localized in the periplasm and thus to be active on oligomers, which also applies to P31_GH39.

GlcA side chains partially shielded the main chain against hydrolysis by P31_GH39. When the GlcA residues were removed by P17_GH2, a higher degree of degradation was observed with P31_GH39 (Supplementary Fig. 39). The newly determined crystal structure of P17_GH2 (Supplementary Fig. 40) contains a pair of *N*-terminal β-sandwich domains, a TIM-barrel with the active site, two more β-sandwich domains and a *C*-terminal putative carbohydrate-binding module connected by an extended flexible linker at the *C*-terminus that places the CBM over the active site (Supplementary Fig. 40a). The active site pocket is at the surface of the catalytic domain; its size provides just enough space to accommodate one GlcA residue. The catalytic site of this enzyme, obscured by the aforementioned CBM, further deviates from other members of the GH2 family. In most GH2 the nucleophile and acid/base catalytic residues are approximately 200 residues apart at the *C*-terminal ends of strands 4 and 7 of the conserved $(\alpha/\beta)_8$-TIM barrel fold. In P17_GH2, the nucleophile is conserved (Glu509) but the acid/base position has a tryptophan (Trp447) (Supplementary Fig. 40d). Two alternative possibilities exist for the acid/base of P17_GH2 Glu411 found on strand 3 and Asp908 from the C-terminal domain (CTD) are both approximately 6.8 Å from Glu509 and could contribute to catalyzing hydrolysis as acid/base residues (Supplementary Fig. 40c).

Monosaccharide metabolism

Ulvan degradation releases different monosaccharides to be further utilized by *F. agariphila*. Many of the enzymes involved in monosaccharide metabolism had significantly higher relative abundances with ulvan compared to fructose or rhamnose as substrate (Table 1, Supplementary Fig. 3, Supplementary Data Set 1). Based on this result and on the MetaCyc database[29], pathways for monosaccharide utilization were deduced, which are consistent with previously proposed pathways[15]. Unlike the PUL H-encoded polysaccharide-degrading

297 proteins, these monosaccharide-utilizing proteins are randomly distributed across the *F.*
298 *agariphila* genome (Supplementary Fig. 3).

299 The spontaneous conversion of α- to β-anomer (mutarotation) of free α-L-rhamnose is a
300 relatively slow process. This rate-limiting step affects growth of L-rhamnose-utilizing
301 bacteria[30,31] because the first metabolic enzyme rhamnose isomerase (EC 5.3.1.14) is
302 specific for the β-anomer[32]. Various bacteria, such as *E. coli* and *Rhizobium leguminosarum*,
303 contain the L-rhamnose mutarotase, accelerating the rate of mutarotation of α- to β-L-
304 rhamnose[37,38]. In contrast to the proteobacterial L-rhamnose mutarotase genes, which are
305 part of small operons dedicated to the uptake and use of free L-rhamnose[30,31], the
306 P21_mutarotase gene is localized in PUL H. We solved the crystal structure of the
307 P21_mutarotase at 1.47 Å (Fig. 4, Supplementary Table 9) with one molecule in the
308 asymmetric unit. P21_mutarotase adopts a ferredoxin-like fold with an antiparallel β-sheet of
309 4 β-strands flanked by a bundle of 3 α-helices. The P21_mutarotase structure superimposed
310 with the characterized L-rhamnose mutarotases YiiL (PDB: 1x8d) and RhaU (PDB: 2qlw) with
311 rmsd on C_α of 0.76 Å and 0.73 Å, respectively[30,31]. Similar to these, the P21_mutarotase (Fig.
312 4a and 4b) formed a dimer with a large hydrophobic dimeric interface antiparallel β-sheets
313 from each monomer (Fig. 4c). All key residues of the active site are well conserved in the
314 P21_mutarotase (Fig. 4d and 4e).
315 *F. agariphila* further metabolizes the β-L-rhamnose via L-rhamnulose-1-phosphate, which is
316 then cleaved by an aldolase (putatively NP3_ or/and NP6_aldolase, Table 1) into L-
317 lactaldehyde and dihydroxyacetone phosphate (Fig. 5)[15]. The corresponding genes are
318 located directly upstream of PUL H (Supplementary Figs. 2 and 3). Glucuronic and
319 unsaturated uronic acids are stepwise converted into KDG (2-dehydro-3-deoxy-D-gluconate),
320 which enters the central metabolism via D-glyceraldehyde 3-phosphate and pyruvate (Fig. 5).
321 Corresponding genes are encoded within PUL H, PUL A or elsewhere in the genome
322 (Supplementary Figs. 3 and 41). NP8_isomerase and NP7_kinase convert D-xylose to D-
323 xylulose-5P, which is an intermediate of the pentose phosphate pathway. In addition,
324 putative monosaccharide transporters were identified (Fig. 5). A D-xylose transporter

(NP16_XylE) was quantified in the membrane fraction in the subproteome experiments (Supplementary Data Set 2). Four ATP-binding proteins of ABC transporters were more abundant with ulvan or with rhamnose in the metabolic labeling experiments (Table 1, Supplementary Data Set 1), indicating that ABC-transporters are involved in monosaccharide uptake. Specific mono- or oligosaccharides generated by the above described enzymatic steps were also verified by HPLC-ELS-ESI-MS (Supplementary Figs. 42-48).

Discussion

Using the DNA sequence of the known ulvan polysaccharide lyase PL28 as query, 12 ulvan PULs were extracted from the NCBI-GenBank, including the biochemically characterized *F. agariphila* ulvan PUL. All PULs were from Bacteroidetes, indicating that our procedure was selective for this phylum since ulvan PULs also exist in Gammaproteobacteria[33]. Interestingly, although four ulvan PULs were from the genus *Polaribacter*, they did not cluster on the heatmap (Figure 1c) indicating that ulvan PULs are diverse at the genus level. Also, within different ulvan PULs, PL28 or PLs from PLnc are over 50% identical at the pairwise amino acid sequence level. Conservation and invariable presence suggest that the first steps of the ulvan degradation cascade proceed through similar enzymes in these organisms. On the other hand, the GH88 enzyme was only present in ulvan PULs of *Flammeovirga pacifica* and *F. agariphila*. GH88 is an exo-acting, unsaturated glucuronyl hydrolase. Its absence in other ulvan PULs could be compensated for by the presence of a GH105, which has the same function. Thus, the later steps in ulvan degradation proceed in dissimilar ways in bacteria.

As shown in the protein domain distribution analysis, the most abundant proteins are sulfatases, which catalyze the removal of sulfate from ulvan. Sulfatase copy numbers ranged from 4-12. At the same time, PLs or GHs such as GH2, GH78 and GH39 in the *F. agariphila* ulvan PUL were also abundant and have several copies in the other predicted ulvan PULs. Notably, some of the proteins of the ulvan PUL, such as the sulfatase P36_GH78/S1_25, are

multimodular enzymes. Our analyses indicated similar domain structures of ulvan-degrading enzymes in other marine Bacteroidetes strains. However, the cursory inspection of gut *Bacteroides* genomes revealed no multimodular GH78 and sulfatase fusion proteins. This suggests that some gene fusions involved in polysaccharide degradation could be more abundant in the marine environment[34].

Our biochemical analyses demonstrated that six of the putative sulfatases (P11_S1_7, P12_S1_8, P18_S1_7, P19_S1_27, P32_S1_8 and P36_S1_25) are ulvan-active indeed sulfatases (Supplementary Fig. 4). However, the sulfatase P14_S1_7 was inactive on both, ulvan from Elicityl and a xylose-rich ulvan from Atlantic *Ulva* spp. and displayed only faint activity on an ulvan from Agrival. This apparent inactivity may be due to a strict exolytic character of P14_S1_7. Consequently, activity maxima are not the same for different types of ulvans. Substrate diversity may cause the variable enzyme content in Bacteroidetes (Figure 1c). This diversity may reflect an adaptation to the different types of ulvans present in *Ulva* spp. Such fine scale adaptation points towards the exploration of PUL microdiversity as a promising avenue for enzyme discovery and for the biocatalytic elucidation of ulvan structures.

Our elucidation of the enzymatic ulvan degradation cascade and characterization of 12 of its enzymes has major implications. Firstly, the conservation of CAZyme- and sulfatase-encoding genes in ulvan PULs of different bacteria underlines their importance and provides a mean to reliably predict new ulvan degradation pathways for bioengineering. Secondly, the substantially extended knowledge of the specific substrate scope of each enzyme enables the targeted use of these enzymes for the production of a variety of novel defined, tailor-made ulvan oligomers, representing useful products, e.g., for pharmaceutical or cosmetic applications. Moreover, these enzymes provide a way to deconstruct ulvan cell walls, which may facilitate the extraction of marine poly- or oligosaccharides and other valuable molecules such as proteins from *Ulva* spp. Finally, the enzymatic cascade allows for the production of

bulk monomeric sugars from the abundant, so far underexplored renewable, the green tide *Ulva*.

Acknowledgements

We thank the German Research Foundation (DFG) for funding through the Research Unit FOR2406 “Proteogenomics of Marine Polysaccharide Utilization” (POMPU) (by grants of U.T.B. (BO 1862/17-1), J.-H.H. (HE 7217/2-1), and T.S. (SCHW 595/10-1). J.-H.H. acknowledges funding by the Emmy-Noether-Program of the DFG, grant number HE 7217/1-1. G.M. is grateful to the French National Research Agency (ANR) for its support with regards to the investment expenditure program IDEALG (grant number ANR-10-BTBR-04) and the Blue Enzymes project (reference ANR-14-CE19-0020-01). M.-K.Z. and F.U. were supported by scholarships from the Institute of Marine Biotechnology e.V. We thank Cédric Leroux for mass spectrometry analyses and Mirjam Czjzek and Alisdair Boraston for helpful discussions. We are indebted to the local contacts for their support during X-ray data collection at the PROXIMA-1 and PROXIMA-2 beamlines (SOLEIL Synchrotron, Saint Aubin, France) and the P11 beamline (DESY, Hamburg, Germany). We thank Andreas Otto and Sabryna Junker for help with the metabolic labeling approach and Tjorven Hinzke for support with analyses of the proteome data. We thank Dr. Frédéric Lesourd (Agrival, Plouenan, France) for the gift of the “Agrival” ulvan sample.

Author contributions

J.-H.H., T.S., G.M. and U.T.B. initiated the study and directed the project. L.R., A.P., R.L. and M.B. cloned the genes and expressed and purified the enzymes for the degradation reactions. M.B., J.-H.H. and L.R. isolated ulvan and purified oligomers. Metabolites were analyzed by C.S. via NMR and HPLC-ELS-MS for which M.D.M. provided resources. L.R. and M.B. performed biocatalysis for the analyses in gel-based assays whereas A.P. together

with M.B. performed HPAEC-PAD analyses. M.-K.Z. with support from S.M., F.U. and A.T.-S. performed the proteome analyses for which D.B. provided the resources. N.G., C.S.R. and T.R. performed crystallographic experiments and solved the protein structures. G.M. analyzed the crystal structure of the L-rhamnose mutarotase and of the sulfatases. S.T. performed the computational analyses of PUL predictions. J.-H.H. and L.R. wrote the paper with input from U.T.B., G.M., S.M., M.-K.Z. and T.S. All authors read and approved the final manuscript and declare that there is no conflict of interest**.**

Competing financial interests

The authors declare no conflict of interest.

References

1. Field, C.B., Behrenfeld, M.J., Randerson, J.T. & Falkowski, P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**, 237-240 (1998).
2. Kloareg, B. & Quatrano, R. Structure of the cell walls of marine algae and ecophysiological functions of the matrix polysaccharides. *Oceanogr. Mar. Biol.* **26**, 259-315 (1988).
3. Wargacki, A.J. et al. An engineered microbial platform for direct biofuel production from brown macroalgae. *Science* **335**, 308-313 (2012).
4. Smetacek, V. & Zingone, A. Green and golden seaweed tides on the rise. *Nature* **504**, 84 (2013).
5. Liu, D. et al. The world's largest macroalgal bloom in the Yellow Sea, China: formation and implications. *Estuar. Coast. Shelf Sci.* **129**, 2-10 (2013).
6. Ménesguen, A. & Piriou, J.-Y. Nitrogen loadings and macroalgal (*Ulva* sp.) mass accumulation in Brittany (France). *Ophelia* **42**, 227-237 (1995).
7. Lahaye, M. & Robic, A. Structure and functional properties of ulvan, a polysaccharide from green seaweeds. *Biomacromolecules* **8**, 1765-1774 (2007).
8. Lahaye, M., Brunel, M. & Bonnin, E. Fine chemical structure analysis of oligosaccharides produced by an ulvan-lyase degradation of the water-soluble cell-wall polysaccharides from *Ulva* sp. (Ulvales, Chlorophyta). *Carbohydr. Res.* **304**, 325-333 (1997).
9. Lahaye, M. NMR spectroscopic characterisation of oligosaccharides from two *Ulva rigida* ulvan samples (Ulvales, Chlorophyta) degraded by a lyase. *Carbohydr. Res.* **314**, 1-12 (1998).
10. Ulaganathan, T. et al. New ulvan-degrading polysaccharide lyase family: structure and catalytic mechanism suggests convergent evolution of active site architecture. *ACS Chem. Biol.* **12**, 1269-1280 (2017).
11. Ulaganathan, T., Banin, E., Helbert, W. & Cygler, M. Structural and functional characterization of PL28 family ulvan lyase NLR48 from *Nonlabens ulvanivorans*. *J. Biol. Chem.* **293**, 11564-11573 (2018).

12. Ulaganathan, T., Helbert, W., Kopel, M., Banin, E. & Cygler, M. Structure–function analyses of a PL24 family ulvan lyase reveal key features and suggest its catalytic mechanism. *J. Biol. Chem.* **293**, 4026-4036 (2018).
13. Melcher, R.L., Neumann, M., Werner, J.P.F., Gröhn, F. & Moerschbacher, B.M. Revised domain structure of ulvan lyase and characterization of the first ulvan binding domain. *Sci. Rep.* **7**, 44115 (2017).
14. Reisky, L. et al. Biochemical characterization of an ulvan lyase from the marine flavobacterium *Formosa agariphila* KMM 3901^T. *Appl. Microbiol. Biotechnol.* **102**, 6987-6996 (2018).
15. Salinas, A. & French, C.E. The enzymatic ulvan depolymerisation system from the alga-associated marine flavobacterium *Formosa agariphila*. *Algal Res.* **27**, 335-344 (2017).
16. Collén, P.N. et al. A novel unsaturated β-glucuronyl hydrolase involved in ulvan degradation unveils the versatility of stereochemistry requirements in family GH105. *J. Biol. Chem.* **289**, 6199-6211 (2014).
17. Foran, E. et al. Functional characterization of a novel "ulvan utilization loci" found in *Alteromonas* sp. LOR genome. *Algal Res.* **25**, 39-46 (2017).
18. Mann, A.J. et al. The genome of the alga-associated marine flavobacterium *Formosa agariphila* KMM 3901^T reveals a broad potential for degradation of algal polysaccharides. *Appl. Environ. Microbiol.* **79**, 6813-6822 (2013).
19. 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).
20. 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).
21. Mewis, K., Lenfant, N., Lombard, V. & Henrissat, B. Dividing the large glycoside hydrolase family 43 into subfamilies: a motivation for detailed enzyme characterization. *Appl. Environ. Microbiol.* **82**, 1686-1692 (2016).
22. Thomas, F. et al. Comparative characterization of two marine alginate lyases from *Zobellia galactanivorans* reveals distinct modes of action and exquisite adaptation to their natural substrate. *J. Biol. Chem.* **288**, 23021-23037 (2013).
23. Demydchuk, M. et al. Insights into Hunter syndrome from the structure of iduronate-2-sulfatase. *Nat. Commun.* **8**, 15786 (2017).
24. Hanson, S.R., Best, M.D. & Wong, C.H. Sulfatases: structure, mechanism, biological activity, inhibition, and synthetic utility. *Angew. Chem. Int. Ed.* **43**, 5736-5763 (2004).
25. 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).
26. Kopel, M. et al. New family of ulvan lyases identified in three isolates from the Alteromonadales order. *J. Biol. Chem.* **291**, 5871-5878 (2016).
27. Sato, K. et al. A protein secretion system linked to bacteroidete gliding motility and pathogenesis. *Proc. Natl. Acad. Sci. USA* **107**, 276-281 (2010).
28. Fujimoto, Z. et al. The structure of a *Streptomyces avermitilis* α-L-rhamnosidase reveals a novel carbohydrate-binding module CBM67 within the six-domain arrangement. *J. Biol. Chem.* **288**, 12376-12385 (2013).
29. Caspi, R. et al. The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Res.* **46**, D633-D639 (2017).
30. Ryu, K.-S. et al. Structural insights into the monosaccharide specificity of *Escherichia coli* rhamnose mutarotase. *J. Mol. Biol.* **349**, 153-162 (2005).
31. Richardson, J.S. et al. RhaU of *Rhizobium leguminosarum* is a rhamnose mutarotase. *J. Bacteriol.* **190**, 2903-2910 (2008).
32. Korndörfer, I., Fessner, W.-D. & Matthews, B.W. The structure of rhamnose isomerase from *Escherichia coli* and its relation with xylose isomerase illustrates a change between inter and intra-subunit complementation during evolution. *J. Mol. Biol.* **300**, 917-933 (2000).

33. Koch, H., Freese, H.M., Hahnke, R.L., Simon, M. & Wietz, M. Adaptations of *Alteromonas* sp. 76-1 to Polysaccharide Degradation: A CAZyme Plasmid for Ulvan Degradation and Two Alginolytic Systems. *Front Microbiol* **10**, 504 (2019).
34. Hehemann, J.H. et al. Aquatic adaptation of a laterally acquired pectin degradation pathway in marine gammaproteobacteria. *Environ. Microbiol.* **19**, 2320-2333 (2017).

Figure legends

Fig. 1 | Genomic overview of putative ulvan PULs in marine Bacteroidetes and the proteomic response of the *F. agariphila* PUL to ulvan and rhamnose. **a**, Comparative genomics of ulvan PULs that contain the known PL28 ulvan lyase (connected with blue lines when over 50% identical) revealed that the enzymes are encoded by conserved genes in diverse marine Bacteroidetes genomes, including the model organism of this study, *F. agariphila* shown as **#**1; the complete list of all analyzed strains is provided in panel 1c. SusD and TBDR proteins are colored as 'other' in this panel. **b**, Ulvan and rhamnose as sole carbon source elicit quantitative changes in proteins encoded in the putative ulvan PUL in *F. agariphila*. Bars indicate relative changes between both conditions. A positive $\Delta\log_2$ value corresponds to higher protein abundance with ulvan, while a negative value corresponds to higher protein abundance with rhamnose. Stars mark proteins that were exclusively quantified in either ulvan- or rhamnose-grown cells (see Supplementary Fig. 2, Table 1 and Supplementary Data Set 1). Arrows refer to the orientation of genes that encode the respective proteins. Proteins encoded by the ulvan PUL were numbered (P1-P39) and protein function was indicated (see Table 1). In the case of glycoside hydrolases (GH) and sulfatases (S), families and subfamilies were specified[19,20], e.g. GH2 (family) or S1_7 (family and subfamily). **c**, Co-occurrence analysis of genes in the predicted 12 putative bacteroidetal ulvan PULs highlights a conserved set of ulvan-degrading enzymes. The dendograms shown above and to the left of the similarity heat map depict the pairwise similarities between rows and columns, respectively.

Figure 2 | Structural analyses of ulvan specific sulfatases. a-c, Molecular surface of P18_S1_7 (**a**) and of the human iduronate 2-sulfatase (PDB: 5FQL) (**b**) both of which belong to the S1_7 subfamily, as well as of the S1_25 sulfatase module of P36 (P36_S1_25) (**c**). These molecular surfaces are colored according to electrostatic potential ranging from deep blue, +, to red, -. **d–e**, Fold representation of P18_S1_7 (**d**), of the human iduronate 2-sulfatase (**e**) and of P36_S1_25 (**f**). The structures are shown in cartoon style. The α-helices and the β-strands are colored in cyan and magenta, respectively. **g**, Stereo view of the key conserved residues in the catalytic groove of P18_S1_7. The amino acids are presented as sticks. The calcium ion is shown as a yellow sphere. The molecular surface of P18_S1_7 is shown as semi-transparent background. **h**, Electron density around the catalytic calcium binding site of P18_S1_7. The coordination residues (Asp35, Asp36, Asp312 and His313) and the catalytic residue Cys74 are shown as sticks. Interactions with the calcium are represented by black dashed lines. The map shown is σA-weighted $2mF_o$ - DF_c maps contoured at 1.2σ (0.07 e/Å^3).

Figure 3 | Zooming into the degradation of ulvan fragments. The experimental procedure to uncover the order of enzymes for ulvan degradation is shown exemplarily for an

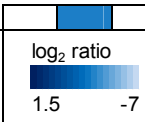
ulvan pentamer. All other investigated enzyme activities are shown in the Supplementary Information. All intermediate products were purified and their structures were confirmed by NMR and MS. MS spectra for individual oligomers are shown on the left next to the respective oligomer. Full spectra for all purified oligomers are shown in the Supplementary Information together with the corresponding NMR spectra. Red arrows indicate cleavage points of the following step. FACE gels for the analysis of the enzymatic interconversion steps are displayed on the right next to the respective enzyme. Full gel images including standards are shown in the Supplementary Information. The desulfation of Rha3S was detected by HPAEC-PAD and the full chromatograms are shown in the Supplementary Information. Numbers with "S" attached to the sugar symbols indicate the position of sulfate groups.

Figure 4 | Structure of the L-rhamnose mutarotase P21_mutarotase. a, Stereo view of the P21_mutarotase dimer shown in cartoon style. **b**, Stereo view of the molecular surface of the P21_mutarotase dimer color coded according to electrostatic potential ranging from deep blue, +, to red, -. **c**, Electron density around the inter-subunit β-sheet in the mutarotase P21_mutarotase dimer. The β4-strand found at the *C*-terminal extremity of the subunit B is involved in β-sheet formation with the subunit A through hydrogen bonding with the β2-strand. Subunits A and B are green and yellow, respectively. Hydrogen bonds between β2 and β4 are shown as black dashed line. The map shown is σA-weighted $2mF_o$ - DF_c maps contoured at 1.2σ (0.12 e/Å^3). **d-e,** Stereo view of the active site of P21_mutarotase (**d**) and of YiiL bound to an L-rhamnose (**e**). The amino acids are presented as sticks. The carbon atoms are colored in yellow and in cyan in P21_mutarotase and YiiL, respectively. The small red spheres are water molecules in the P21_mutarotase structure.

Figure 5 | Model of the ulvan degradation pathway in *F. agariphila* as suggested by the proteogenomic, biochemical and structural biological analyses in this study. Redundant pathways are omitted to maintain clarity. The ulvan molecule on top represents a part within the larger ulvan chain where rhamnose and iduronate are α- while xylose and glucuronate are β-configured. The formed products – at both ends of the initial ulvan molecule after cleavage with P30_PL28 – are not shown in the downstream degradation pathway. Activity of ulvan lyases P30_PL28 and P10_PLnc will form an unsaturated uronic acid residue from glucuronic acid or iduronic acid at the non-reducing end of the products. Numbers with "S" attached to the sugar symbols indicate the position of sulfate groups. Black arrows indicate pathways elucidated by proteogenomic, biochemical and structural biological analyses, while grey arrows only refer to proteome analyses or additional structural analyses in the case of P21_mutarotase. For numbering/nomenclature see Table 1. For reasons of simplicity, the linkage of the TBDRs to the TonB-ExbBD complex or a putative membrane association of certain enzymes were not included. KDG: 2-dehydro-3-deoxy-D-gluconate; DKI: 5-dehydro-4-deoxy-D-glucuronate; DKII: 3-deoxy-D-glycero-2,5-hexodiulosonate.

Table 1 | List of PUL H-encoded and relevant non-PUL H-encoded proteins with abbreviations used in the text, corresponding locus tags and functional annotation as well as their relative abundance (mean log_2 ratio) with the respective carbon source. Empty/white squares refer to non-quantified proteins while grey squares indicate OFF-proteins that could not be quantified due to a lack of ^{14}N signals (see Online Methods)

PUL H-encoded proteins (for ulvan and ulvan-derived monosaccharide utilization)			**log_2 ratio**		
Abbreviation	**Locus tag**	**Functional annotation**	**fru**	**rha**	**ulv**
P1_GH88	*21900	unsaturated glucuronyl hydrolase (GH88)			
P2_SusD	*21910	SusD-like protein			
P3_TBDR	*21920	TonB-dependent receptor			
P4_HK	*21930	histidine kinase			
P5_isomerase	*21940	4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase			
P6_dehydrogenase	*21950	2-deoxy-D-gluconate 3-dehydrogenase			
P7	*21960	conserved hypothetical protein			
P8_GH2	*21970	beta-galactosidase (GH2)			
P9_lactonase	*21980	6-phosphogluconolactonase			
P10_PLnc	*21990	ulvan lyase (PLnc)			
P11_S1_7	*22000	iduronate-2-sulfatase (S1_7)			
P12_S1_8	*22010	arylsulfatase (S1_8)			
P13_S1_16	*22020	arylsulfatase (S1_16)			
P14_S1_7	*22030	arylsulfatase (S1_7)			
P15_GH2	*22040	glycoside hydrolase (GH2)			
P16_GH2	*22050	beta-galactosidase (GH2)			
P17_GH2	*22060	beta-galactosidase (GH2)			
P18_S1_7	*22070	arylsulfatase (S1_7)			
P19_S1_27	*22080	sulfatase (S1_27)			
P20_GH78	*22090	alpha-L-rhamnosidase (GH78)			
P21_mutarotase	*22100	L-rhamnose mutarotase			
P22	*22110	conserved hypothetical protein			
P23	*22120	conserved hypothetical protein			
P24_GH3	*22130	beta-glucosidase (GH3)			
P25_SusD	*22140	SusD-like protein			
P26_TBDR	*22150	TonB-dependent receptor			
P27_GH43	*22160	beta-xylosidase (GH43)			
P28_GH78	*22170	alpha-L-rhamnosidase (GH78)			
P29	*22180	conserved hypothetical protein			
P30_PL28	*22190	ulvan lyase (PL28)			
P31_GH39	*22200	glycoside hydrolase (GH39)			
P32_S1_8	*22210	arylsulfatase (S1_8)			
P33_GH105	*22220	glycoside hydrolase (GH105)			
P34_GH3	*22230	beta-glucosidase (GH3)			
P35_oxidoreductase	*22240	oxidoreductase			
P36_GH78/S1_25	*22250	alpha-L-rhamnosidase/sulfatase (GH78/S1_25)			
P37	*22260	hypothetical protein			
P38_SusD	*22270	SusD-like protein			
P39_TBDR	*22280	TonB-dependent receptor			
Non-PUL H-encoded proteins (for ulvan-derived monosaccharide utilization)			**log_2 ratio**		
Abbreviation	**Locus tag**	**Functional annotation**	**fru**	**rha**	**ulv**
NP1_dehydrogenase	*21840	aldehyde dehydrogenase A			
NP2_dehydrogenase	*21850	L-lactate dehydrogenase			
NP3_aldolase	*21860	class II aldolase/adducin family protein			
NP4_kinase	*21870	pentulose/hexulose kinase			
NP5_isomerase	*21880	rhamnose isomerase[a]			
NP6_aldolase	*21890	rhamnulose-1-phosphate aldolase			
NP7_kinase	*160	xylulose kinase			
NP8_isomerase	*170	xylose isomerase			
NP9_oxidoreductase	*9410	D-mannonate oxidoreductase			
NP10_dehydratase	*9420	mannonate dehydratase			
NP11_isomerase	*9430	uronate isomerase			
NP12_kinase	*9800	2-dehydro-3-deoxygluconate kinase			
NP13_aldolase	*9820	aldolase[b]			
NP14_kinase	*11640	2-dehydro-3-deoxygluconate kinase			
NP15_kinase	*16400	2-dehydro-3-deoxygluconate kinase			
NP16_XylE	*180	D-xylose transporter XylE[c]			
NP17_ABC	*11090	ABC transporter, ATP-binding protein			
NP18_ABC	*25150	ABC transporter, ATP-binding protein			
NP19_ABC	*7480	ABC transporter, ATP-binding protein			

NP20_ABC	*12820	ABC transporter, ATP-binding protein	
Proteins were numbered (P1 - P39: PUL H-encoded proteins, NP1 - NP20: non-PUL H-encoded proteins) and protein function was indicated. In the case of glycoside hydrolases (GH) and sulfatases (S), families and subfamilies were specified[19,20], e.g. GH2 (family) or S1_7 (family and subfamily). *BN863_, e.g. "*21900" refers to locus tag BN863_21800; [a]identified by BLAST against the Uniprot database, previously annotated as xylose isomerase-like TIM barrel domain protein, [b]4-hydroxy-2-oxoglutarate aldolase / 2-dehydro-3-deoxyphosphogluconate aldolase, [c]only captured by subproteome analysis of ulvan-grown cells (Supplementary Data Set 2), fru: fructose, rha: rhamnose, ulv: ulvan		

Online Methods

Prediction of ulvan PULs

118,981 bacterial genomes were downloaded from the NCBI-GenBank using an in-house script (updated in 2018.09.10). Hmmer 3.0 was used to identify proteins with a PL28 or sulfatase domain, using a cut-off value of 1e-10[35]. Hidden Markov models of PL28 and sulfatase were obtained from dbCAN2 and the pfam database, respectively[36,37]. Models for the new PLnc family have not been released, thus blastp was used to identify its homologs, using 1e-50 and 30% sequence identity as cut-off values[38]. In each bacterial genome, if the adjacent 50 proteins to the afore-mentioned marker genes contained three marker genes (PL28, PLnc and sulfatase), this locus was considered as a potential ulvan PUL hit. To further determine PUL boundaries, 100 proteins surrounding the predicted ulvan PUL were collected and then locally annotated using pfam and dbCAN Hidden Marikov models (cut value 1e-10). Firstly, PL28 or PLnc families were set as boundaries, which were extended if adjacent genes are annotated as sugar utilization proteins, such as GH, PL, sugar transporter and transcription factors. In cases where five continuous genes were not related to sugar utilization or ulvan degradation, the last functionally relevant protein was taken as the putative ulvan PUL boundary. Protein sequences within putative ulvan PULs were collected for further analysis. Circos was used to visualize the different ulvan PULs[39]. Blastp was used to calculate the identity between PL28 sequences from different ulvan PULs (cut-off value: 1e-10, over 50% identity). To simplify and reduce non-conserved proteins, domains with less than 80% presence among the predicted ulvan PULs were excluded. Domain numbers in each PUL were counted, summarized and displayed in R studio.

Proteome analyses

Whole cell proteome – metabolic labeling

A $^{14}N/^{15}N$ relative quantification approach, based on metabolic labeling, was used for protein quantification as described previously[40]. For this purpose, *F. agariphila* KMM 3901^T was cultivated in MPM salts[41] containing either ^{14}N- or ^{15}N-ammonium chloride, supplemented with 0.2% of the individual carbon source: ulvan, rhamnose or fructose. Cultivation (21°C, 170 rpm) comprised three steps: (i) 24 h of marine broth 2216-cultivation and subsequent (ii) pre-cultures as well as (iii) main cultures in the above-described minimal medium. At an OD_{600nm} of 0.5, cells were harvested from main cultures by centrifugation (30 min, 9,384 x *g*, 4°C). Cell pellets were suspended in TE-buffer (10 mM Tris, 10 mM EDTA) and cells were disrupted by sonication (4 cycles of 25 s at 5 m/s). Cell debris and protein extract were separated by centrifugation (10 min, 21,460 x *g*, 4 °C). In case of the ^{15}N-labeled samples, protein extracts of all samples from all 3 carbon sources were combined to form the ^{15}N-labeled reference pool, which served as an internal standard. Protein concentration was determined using the Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific). 12.5 µg of protein of each ^{14}N-sample was combined with 12.5 µg protein of the ^{15}N-labeled pool. These mixtures were separated by 1D-SDS-PAGE. Protein lanes were cut into ten equal-sized pieces, destained and proteins were in-gel-digested with a 1 µg mL^{-1} trypsin solution[42]. Peptides were separated by RP chromatography and analyzed in an LTQ-Orbitrap Classic mass spectrometer equipped with a nanoelectrospray ion source[43]. Data represent three independent experiments (n=3).

MS data were searched with Sorcerer SEQUEST v.27, rev.11 (Thermo-Finnigan, Thermo Fisher Scientific, Germany) against a target decoy database including all *F. agariphila* KMM 3901^T protein sequences, corresponding reversed sequences (decoys) as well as common laboratory contaminants (total 7224 entries)as described previously[43], but using a false positive rate of 0.05. In brief, peak intensities of the ^{14}N-peptide ions of a protein versus its ^{15}N-peptide ions were compared to calculate a regression ratio. Only unique peptides and

peptides with an R^2 above 0.7 were taken into account. Non-quantified peptides were manually checked. Average regression ratios were then exported. Proteins with at least two quantified peptides were considered for the following calculations: ratios were median-centered and log-transformed, termed as $\log_2$ ratios, per sample. If proteins were quantified in at least two of the three replicates, means and standard deviations (SD) were calculated from these values. In order to identify relative changes between the different carbon sources, $\log_2$ ratios of fructose- or rhamnose-cultivated cells were subtracted from $\log_2$ ratios of ulvan-cultivated cells, termed $\Delta\log_2$. Fold-changes correspond to the exponentials of these $\Delta\log_2$ values. Statistical analyses were performed with Welch's two-sided t-test (permutation-based false discovery rate 0.01) using Perseus v1.6.0.7[44], based on the $\log_2$ ratios. Putative ON/OFF proteins were marked with 15N (OFF) or 14N (ON) in Supplementary Data Set 1, but were not included in any of the calculations. Only if a protein was identified as an ON/OFF protein in all three replicates, it was assigned to a fixed value (10/-10), to highlight these proteins.

Subproteome fractionation

F. agariphila KMM 3901^T was cultivated as described above, except that no ^{15}N-labeling was performed and only ulvan was applied as a carbon source.

For the surface proteome (trypsin-shaving approach), 1.5 mL of cell suspension was removed from the culture and centrifuged (5 min, 5,867 x *g*, 4 °C). Cells were washed with 50 mM triethylammoniumbicarbonate-buffer (TEAB) and finally resuspended in 45 µL TEAB-buffer. In order to cleave proteins from the cell surface, 5 µL of a 1 µg mL^{-1} trypsin solution was added. The solution was transferred onto a 0.22 µm cellulose-acetate spin-column and incubated for 15 min at 900 rpm and 37 °C. The flow-through was collected by centrifugation (10 min, 4,000 x *g*, 4 °C), another 1 µL of trypsin was added and the sample was incubated at 900 rpm and 37 °C overnight. The peptide mixture was desalted using C18 StageTips. The following solutions were used: 0.1% (v/v) acetic acid in ultra-pure water (buffer A) and 0.1% (v/v) acetic acid in acetonitrile (buffer B). Before the sample was added, C18 material

672 was rinsed and equilibrated with buffer A and washed with buffer B in between these steps.
673 After the sample was added, buffer A was used for washing and buffer B for elution.

674 In the case of cytosolic, membrane-associated and extracellular protein fractions, 100 mL of
675 cell suspension was harvested by centrifugation (30 min, 9,384 x *g*, 4 °C). Cell pellets and
676 supernatants were processed separately as previously described[45,46]. 1D-SDS-PAGE, in-gel-
677 digestion and LC-MS/MS analysis were performed as described above. Experiments were
678 carried out in triplicates (n=3).

679 Database searches were done with Sorcerer SEQUEST v.27, rev.11 (see above). Results
680 were summarized and filtered using Scaffold 4.4.1.1 (Proteome Software, Portland, OR,
681 USA): protein and peptide false discovery rate was set to 0.01 and protein identification
682 required two peptides minimum. For protein quantification, the normalized spectral
683 abundance factor was calculated for each protein giving the percentage (%NSAF) of all
684 proteins in the same sample[47]. If proteins were identified in at least two of the three
685 replicates, they were considered for further calculations.

686 **Gene cloning and expression**

687 Expression constructs were prepared using the FastCloning strategy[48] with genomic DNA
688 from *F. agariphila* KMM 3901^T (collection number DSM15362 at DSMZ, Braunschweig,
689 Germany) as template for the amplification of the inserts. Generally, the pET28 constructs
690 were prepared as described previously[14] with the gene primers shown in Supplementary
691 Table 10. To clone the gene for the formylglycine-generating enzyme (FGE) from *F.*
692 *agariphila*, the vector backbone was amplified with the primers 5'-AATA GCGC CGTC GACC
693 ATCA TCAT CATC ATCAT-3' and 5'-CATG GTTA ATTC CTCC TGTT AGCC CAAA AA-3'
694 from pBAD/myc-his A. For the pFA, constructs were cloned and overexpressed as previously
695 described[49]. Briefly, genes were PCR-amplified using the NEB Q5 High-Fidelity DNA
696 Polymerase system. PCR reactions were done with 30 cycles (denaturation: 95 °C;
697 annealing: 60 °C; elongation: 72 °C) using 0.5 units of enzyme in a total reaction of 50 µL

using the primers shown in Supplementary Table 10. Amplicons were cleaned up using the QIAquick PCR Purification Kit (Qiagen) and digested with the appropriate restriction endonucleases. All ligations were done in the linearized T7 system vector pFO4.

Genes encoding the sulfatases P18_S1_7, P19_S1_27, P32_S1_8 and P36_S1_25 were ordered codon-optimized for *E. coli* and sub-cloned into pET28 with NheI and XhoI from Genscript. The optimized nucleotide sequences are shown in the Supplementary Information.

Escherichia coli BL21(DE3) was transformed with pET28-based plasmids. For the overexpression, 50 mL ZYP-5052[50] with 100 µg mL^{-1} kanamycin were inoculated from an overnight culture in LB containing 50 µg mL^{-1} kanamycin. The culture was grown at 30 °C and 180 rpm until the OD_{600nm} reached 1.0 and was then cooled to 20 °C for 48 h. In the case of sulfatases, the formylglycine-generating enzyme (FGE) from *F. agariphila* was co-expressed. LB medium with 100 µg mL^{-1} ampicillin and 50 µg mL^{-1} kanamycin was inoculated from an overnight culture in the same medium and incubated at 37 °C and 180 rpm until the OD_{600nm} reached 0.3 to 0.5. After the addition of 1.5 mM L-arabinose and incubation for 90 min at 37 °C, the culture was cooled to 18 °C for 2 h before 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added and the culture was incubated overnight at 18 °C. Alternatively, sulfatases were expressed from the pFA constructs in *E. coli* BL21(DE3) cells grown in LB medium supplemented with 15 µg mL^{-1} ampicillin, at 37 °C, until reaching an $OD_{600\ nm}$ of 0.8. Expression was induced with 0.1 mM IPTG overnight at 18 °C. For crystallization screening, *E. coli* BL21(DE3) cells were transformed with the plasmids containing the gene fragment of interest, then grown in the autoinduction Zyp-5052 medium (200 µg mL^{-1} ampicillin, 20 °C, 72 h). Cells were harvested by centrifugation (10,000 x *g*, 4 °C, 20 min) and the cell pellets were stored at -20 °C until further use.

Samples from the cultivations equivalent to 1 mL of culture with an OD_{600nm} of 7 were taken before harvest and the cells were collected by centrifugation (13,000 x *g*, 4 °C, 2 min). Pellets were resuspended in 500 µL 50 mM HEPES with 100 mM NaCl (pH 7.4). After

chemical lysis with BugBuster (Merck, Darmstadt, Germany), whole cell protein (W) samples were obtained prior to removal of the cell debris by centrifugation (13,000 x *g*, 4 °C, 10 min). Samples of the soluble protein fraction (S) were taken from the respective supernatant.

Enzyme purification

Cell pellets were thawed and resuspended in 50 mM NaPi with 300 mM NaCl (pH 8.0) and lysed by three cycles of sonication (2.0 min, 30% pulse, 50% power). After centrifugation (10,000 x *g*, 4 °C, 20 min), the supernatant was filtered (0.45 µm) and loaded onto a 5 mL HisTrap FF crude column (GE Healthcare, Freiburg, Germany) equilibrated with lysis buffer. Alternatively, Rotigarose-His/Ni beads (Karl Roth, Karlsruhe, Germany) were used in gravity flow columns. After washing, the protein was eluted with 50 mM NaPi and 300 mM NaCl containing 300 mM imidazole (pH 8.0). Fractions containing the protein of interest were pooled and desalted using PD-10 columns (GE Healthcare, Freiburg, Germany) equilibrated with 50 mM NaPi pH 7.4. Proteins were analyzed by SDS-PAGE on 12.5% acrylamide gels. 1% (v/v) 2,2,2-trichloroethanol was used for the visualization of proteins under UV light[51]. Alternatively, proteins were stained with Coomassie Blue (PhastGel® Blue R). All enzymes were used undiluted, or in dilutions of 1:5, 1:10 or 1:20 with enzyme storage buffer (Supplementary Table 11).

Alternatively, cells were subjected to mechanical lysis and cytoplasmic extracts were loaded onto an Histrap column (5ml, GE Healthcare) equilibrated with 50 mM Tris, 0.2 M NaCl, 20 mM imidazol, 1 mM $CaCl_2$ at pH 8.0. Recombinant proteins were eluted with around 250 mM imidazole and then loaded onto a Hiprep Desalting column (26/10, 53ml, GE Healthcare) in order to eliminate the imidazole, which notably interfered with sulfatase activity. Purified enzymes were concentrated (Amicon® Ultra Centrifugal Filter, 30 kDa) to a concentration of 1 mg mL^{-1} (Nanodrop).

Purification of ulvan

Green tide *Ulva* sp. was collected near Roscoff (France) and dried. Alternatively, dried *Ulva* biomass from the Atlantic coast in Spain was purchased as organic sea lettuce (Kulau, Berlin, Germany). Ulvan was extracted according to the literature[52]. The dialysis step was exchanged by precipitation with acetone (80% (v/v) final concentration). After washing, acetone-precipitated ulvan was dissolved in deionized water and freeze-dried. Alternatively, ulvan was obtained from Agrival (Plouenan, France) or Elicityl (Grenoble, France).

Enzyme assays

Generally, reactions were performed in 50 mM HEPES pH 7.4 with 100 mM NaCl or 35 mM Tris pH 8.0 with 50 mM NaCl. Initial degradation of ulvan into larger oligomers was monitored by C-PAGE, while smaller degradation products and the conversion of purified oligomers was analyzed by FACE. For lyases, the increase in absorbance at 235 nm was recorded over time. For unsaturated uronyl hydrolases (GH88 and GH105), the decrease in absorbance at 235 nm of ulvan lyase products was monitored. For screening reactions, 10% (v/v) clarified lysate as used for the SDS-PAGE was added. Untreated ulvan was generally used at a concentration of 1 g L^{-1} while purified oligomers were used at 0.25 mg mL^{-1}. Incubation was performed overnight at room temperature.

Sulfatase activity assay on ulvan polymers

Activity assays were conducted on three different ulvan polymers from *Ulva* species: a commercial ulvan from Elicityl (Grenoble, France), an ulvan which was a gift from the company Agrival (Plouenan, France), and an ulvan extracted from *Ulva* sp. harvested on Brittany north coast (Roscoff, France). 10 µL of each ulvan solution (1% w/v in H_2O) was incubated with 10 µL of purified sulfatase (1 mg mL^{-1}) in a final volume of 80 µL of 25 mM Tris-HCl, 0.1 M NaCl, 0.5 mM $CaCl_2$, pH 8.0 buffer mix, for 18 h at 37 °C. For each reaction, a control sample was prepared using similar conditions but with an inactivated enzyme

775 (100 °C, 10 min). Reaction mixtures and blanks were then filtered (10 kDa, Amicon® Ultra,
776 *Millipore*) to measure the amount of free sulfate in the filtrates.

777 Ulvan-specific sulfatase activity was measured by high-performance anion-exchange
778 chromatography (HPAEC). Using an ICS5000 system (*Thermo Scientific Dionex*), anions
779 from reaction mixture filtrates were injected (AS-AP Autosampler) and separated using an
780 AG11-HC guard column (4x50 mm) mounted in series with an AS11-HC anion-exchange
781 column (4x250 mm). Elutions were performed with isocratic 12 mM NaOH at a flow rate of 1
782 mL min^{-1} (Single Pump-5), and the detection of anions was leaded by an Analytical CD
783 Conductivity Detector associated to a suppressor (ASRS 500, 4 mm) running at 50 mA.
784 Using a standard curve of sulfate, concentration of sulfate released by the enzymatic
785 reaction was calculated from the difference of the amount of sulfate between samples and
786 the associated blanks.

787 **Sulfatase activity assay on characterized ulvan oligosaccharides**

788 10 µL of ulvan oligosaccharides (0.5-1% w/v in H_2O) were incubated with 15 µL of purified
789 sulfatase (0.5 mg mL^{-1}) in a final volume of 75 µL of 5 mM Tris-HCl, 10 mM NaCl, 0.5 mM
790 $CaCl_2$, pH 8.0 buffer, for 18 h at 37 °C. The recombinant enzymes P33_GH105 or
791 P36_GH78 were added (2 µL – 3 mg mL^{-1}). Each reaction mixture was centrifuged (14,000 x
792 *g* for 10 min) before injection. Oligosaccharide detection was realized by HPAEC analyzes
793 on the same ICS 5000 system described for the sulfate quantification. Elutions were
794 performed at a flow rate of 0.5 mL min^{-1} using a NaOH multistep gradient from 8 to 280 mM
795 (45 min). Oligosaccharides were detected by conductivity mode under a current suppression
796 of 50-300 mA.

797 **Carbohydrate polyacrylamide gel electrophoresis**

798 Fluorophore-assisted carbohydrate electrophoresis (FACE) was performed with 2-
799 aminoacridone (AMAC) as fluorophore[53].

For carbohydrate polyacrylamide gel electrophoresis (C-PAGE), samples were mixed with an equal volume of FACE loading buffer[53]. Gels and running conditions were identical to FACE. Carbohydrates were visualized by staining with Stains-All solution (0.25 g L^{-1} in 1.7 mM Tris-HCl pH 7.5 with 25% (v/v) isopropanol). The gels were destained with 25% (v/v) isopropanol in deionized water.

Purification of oligomers and structure determination

Ulvan was digested with purified enzymes in Tris-HCl pH 8.5 at room temperature. Oligomers were separated on two XK 26/100 (GE Healthcare, Freiburg, Germany) in series filled with Bio-Gel P-2 (Rio-Rad, Munich, Germany) using 100 mM $(NH_4)_2CO_3$ as mobile phase at a flow rate of 1 mL min^{-1}. After lyophilization of the fractions containing the products, oligomers were dissolved in D_2O and lyophilized three times before NMR spectra were recorded on a Bruker Avance III HD 600 (600 MHz) spectrometer (Bruker, Billerica, USA) in D_2O solutions. The structures were independently elucidated based on 1D and 2D (COSY, HSQC, HMBC, TOCSY) methods and the assigned ^{1}H and ^{13}C-NMR signals were then compared with literature data, showing excellent consistency[8,9]. For samples containing uronic acid structures, it was required to neutralize the otherwise acidic NMR samples with Na_2HPO_4 to pH 7-8 (pH-electrode calibrated to H^+) in order to achieve fully resolved signals for the carboxylic acid and neighboring positions (^{13}C). HPLC−ELS-MS analysis was performed by injection of ~0.1% solutions (1−5 µL) on a Nexera UHPLC system from Shimadzu (equipped with two binary LC-30AD pumps plus degassers, a CTO-20 column oven) and a LCMS-2200 EV MS-detector and an additional ELS-detector (JASCO ELS-2041). Analysis was performed with mobile phase A = H_2O (0.1% HCOOH) and mobile phase B = CH_3CN on a C18 column (XSelect CSH XP C18 2.5 µm 3 x 50 mm) at 40 °C. Flow rate was 1.3 mL min^{-1} (0−3 min) with 5% B from 0−0.15 min, 5−98% B from 0.15-2.2 min and 98%-5% B from 2.2−2.5 min.

Crystallization of proteins and structure determination

826 Crystallization trials of P18_S1_7 (pFA13 construct) and of the family S1_25 sulfatase
827 module of the bimodular GH78 L-rhamnosidase P36 (pET28 construct, referred to as
828 P36_S1_25) were undertaken at room temperature using the vapor-diffusion method in
829 sitting drops containing a 2:1 ratio of pure protein (12.9 and 13.0 mg mL^{-1}, respectively) and
830 of precipitant solution. P18_S1_7 and P36_S1_25 were mixed with reservoir solution
831 containing 100 mM MIB pH 5.0 and 25 % PEG 1,500 and 100 mM MES pH 6.5 and 25 %
832 PEG 2,000 MME, respectively. Crystals of the L-rhamnose mutarotase P21_mutarotase
833 (pFA16 construct, concentration: 14.9 mg mL^{-1}) were obtained by the hanging-drop vapor-
834 diffusion method at room temperature and also at a 2:1 protein/precipitant ratio with a
835 reservoir solution containing 100 mM sodium acetate pH 4.6 and 4.3 M sodium formate.
836 Crystals of P18_S1_7, P21_mutarotase and P36_S1_25 were cryoprotected with 10%, 14%
837 and 14% glycerol, respectively, and flash-frozen in liquid nitrogen. X-ray diffraction
838 experiments were carried out at 100 K at beamlines PROXIMA-1 (PX1) for P18_S1_7 and
839 P21_mutarotase and PROXIMA-2 for P36_S1_25 (SOLEIL Synchrotron, GIF-sur-YVETTE,
840 France). Diffraction data of P18_S1_7, P21_mutarotase and P36_S1_25 were obtained at
841 1.23, 1.47 and 2.91 Å, respectively, and were processed using XDS[54]. Scaling and merging
842 were performed using the program Aimless from the CCP4 package[55]. The structure of
843 P21_mutarotase (a dimer of 2 x 115 residues), P18_S1_7 (475 residues), and P36_S1_25
844 (467 residues) were solved by molecular replacement with the CCP4 suite program MolRep[56]
845 using the structures of the rhamnose mutarotase RhaU from *Rhizobium leguminosarum*
846 (PDB entry: 2QLX)[31], of the human iduronate-2-sulfatase (5FQL)[23] and of the putative
847 sulfatase YidJ from *Bacteroides fragilis* (2QZU) as starting models, respectively. Refinement
848 and model building of P18_S1_7 and P21_mutarotase were undertaken using the PHENIX
849 program suite[57] and the Coot software[58]. Initial refinement of the P36_S1_25 structure was
850 performed with BUSTER[59] and PHENIX[57], and then manual examination and rebuilding of
851 the refined coordinates were carried out in Coot[58]. Structural validation was undertaken using
852 MOLPROBITY[60].

SEC-purified P17_GH2 crystallized in 1:1 ratio of 7 mg mL^{-1} protein in 20 mM Tris pH 8.0 and mother liquor in the JBScreen PACT ++ HTS and JBScreen Classic HTS I (Jena Bioscience). A single crystal from the screen grown in 20 % PEG 3350, 0.1 M Bis-Tris pH 7.5, 0.2 M sodium bromide was cryo-protected in 30 % glycerol prior X-ray crystallography. The diffraction data were collected at DESY P11 automatically integrated in XDS and scaled and merged in Aimless[54,61].

The structure of P17_GH2 was solved by molecular replacement using 5dmy as search model in phaser[62]. The structure was built automatically using buccaneer and manually in Coot building directly into the 2Fo-Fc maps[58,63]. Structural validation was carried out using MOLPROBITY[60].

Data availability

All data that support the findings of this study are available from the corresponding authors upon reasonable request. The protein structures are deposited in the PDB under 6HHM, 6HHN, 6HPD and 6HR5. Mass spectrometry data were deposited to the ProteomeXchange Consortium via the PRIDE partner repository[64] with the dataset identifier PXD009299. The sequences of the newly characterized ulvan-degrading enzymes can be found in the GenBank sequence database under the respective locus tags mentioned in Table 1.

References

35. Finn, R.D., Clements, J. & Eddy, S.R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29-W37 (2011).
36. Zhang, H. et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **46**, W95-W101 (2018).
37. Finn, R.D. et al. Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222-D230 (2013).
38. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
39. Krzywinski, M.I. et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* **9**, 1639-1645 (2009).
40. MacCoss, M.J., Wu, C.C., Liu, H., Sadygov, R. & Yates, J.R., 3rd. A correlation algorithm for the automated quantitative analysis of shotgun proteomics data. *Anal. Chem.* **75**, 6912-6921 (2003).
41. Schut, F. et al. Isolation of typical marine bacteria by dilution culture: growth, maintenance, and characteristics of isolates under laboratory conditions. *Appl. Environ. Microbiol.* **59**, 2150-2160 (1993).

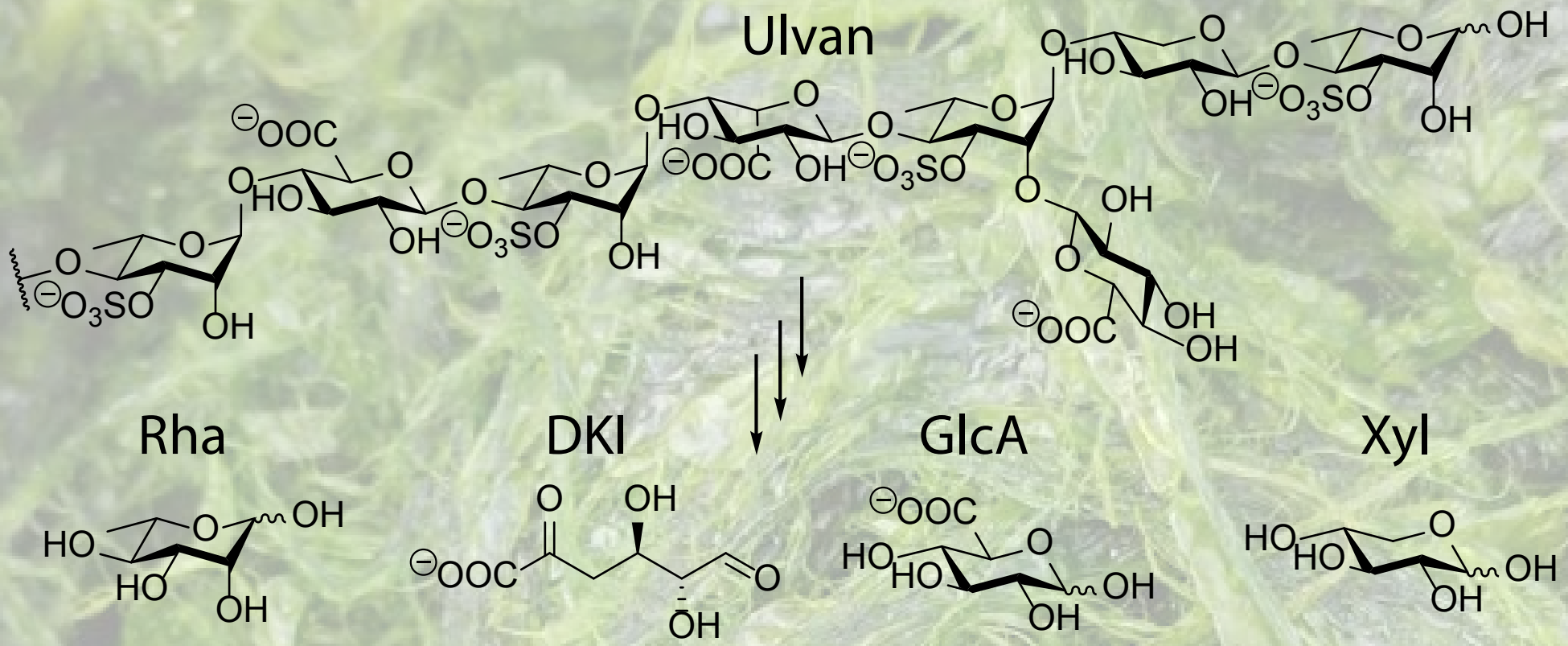
42. Grube, M. et al. Exploring functional contexts of symbiotic sustain within lichen-associated bacteria by comparative omics. *ISME J.* **9**, 412-424 (2015).
43. Otto, A. et al. Systems-wide temporal proteomic profiling in glucose-starved *Bacillus subtilis*. *Nat. Commun.* **1**, 137 (2010).
44. Tyanova, S. & Cox, J. Perseus: a bioinformatics platform for integrative analysis of proteomics data in cancer research. *Methods Mol. Biol.* **1711**, 133-148 (2018).
45. Antelmann, H. et al. A proteomic view on genome-based signal peptide predictions. *Genome Res.* **11**, 1484-1502 (2001).
46. Eymann, C. et al. A comprehensive proteome map of growing *Bacillus subtilis* cells. *Proteomics* **4**, 2849-2876 (2004).
47. Zybailov, B. et al. Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J. Proteome Res.* **5**, 2339-2347 (2006).
48. Li, C. et al. FastCloning: a highly simplified, purification-free, sequence-and ligation-independent PCR cloning method. *BMC Biotechnol.* **11**, 92 (2011).
49. Groisillier, A. et al. MARINE-EXPRESS: taking advantage of high throughput cloning and expression strategies for the post-genomic analysis of marine organisms. *Microb. Cell Fact.* **9**, 45 (2010).
50. Studier, F.W. Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **41**, 207-234 (2005).
51. Ladner, C.L., Yang, J., Turner, R.J. & Edwards, R.A. Visible fluorescent detection of proteins in polyacrylamide gels without staining. *Anal. Biochem.* **326**, 13-20 (2004).
52. Robic, A., Gaillard, C., Sassi, J.F., Lerat, Y. & Lahaye, M. Ultrastructure of ulvan: a polysaccharide from green seaweeds. *Biopolymers* **91**, 652-664 (2009).
53. Hehemann, J.H. et al. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464**, 908-912 (2010).
54. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 133-144 (2010).
55. Winn, M.D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235-242 (2011).
56. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 22-25 (2010).
57. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213-221 (2010).
58. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486-501 (2010).
59. Smart, O.S. et al. Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 368-380 (2012).
60. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12-21 (2010).
61. Evans, P.R. & Murshudov, G.N. How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204-1214 (2013).
62. McCoy, A.J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658-674 (2007).
63. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002-1011 (2006).
64. Vizcaíno, J.A. et al. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44**, D447-D456 (2016).

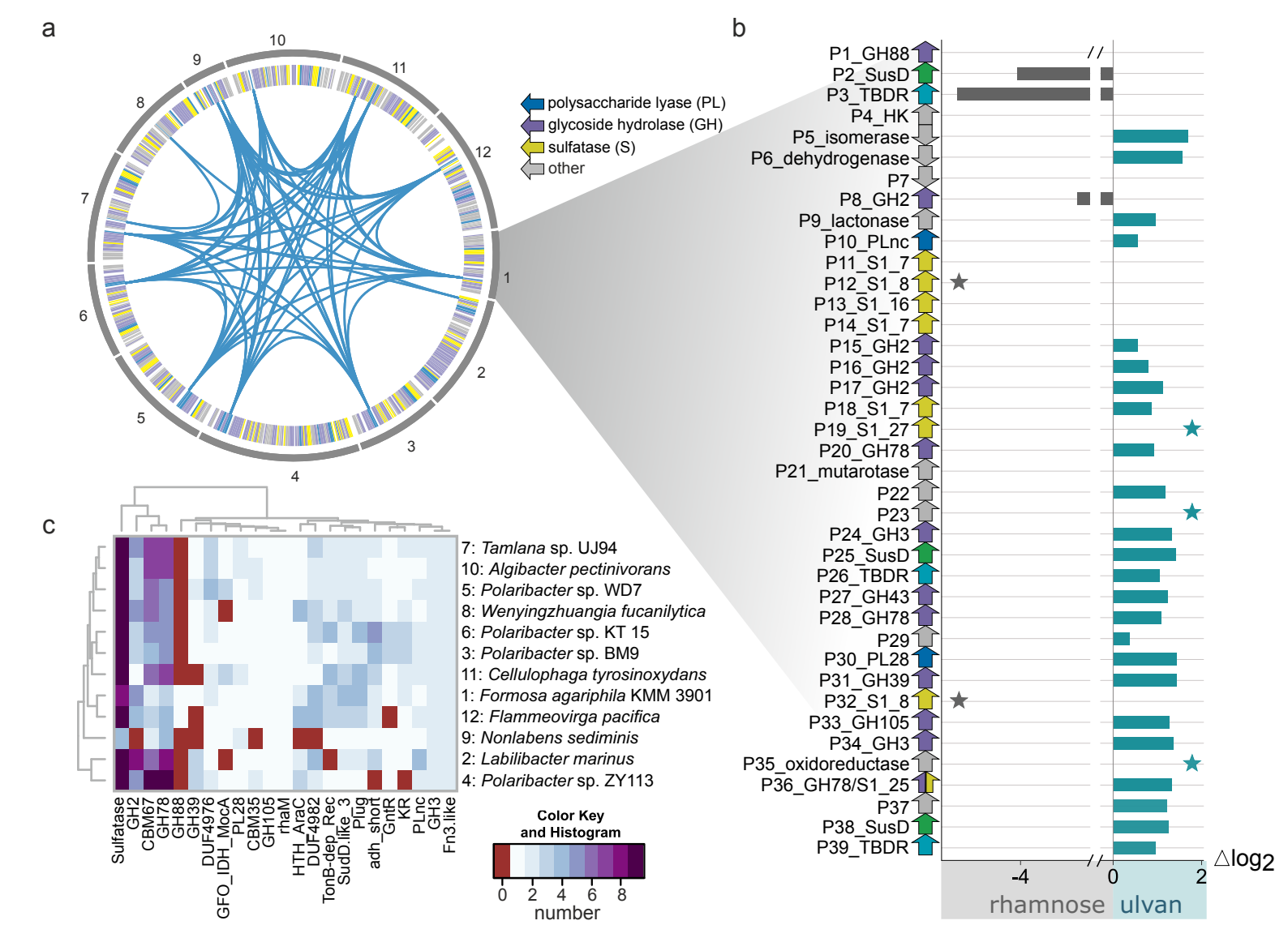
Table 1 | List of PUL H-encoded and relevant non-PUL H-encoded proteins with abbreviations used in the text, corresponding locus tags and functional annotation as well as their relative abundance (mean log_2 ratio) with the respective carbon source. Empty/white squares refer to non-quantified proteins while grey squares indicate OFF-proteins that could not be quantified due to a lack of ^{14}N signals (see Online Methods)

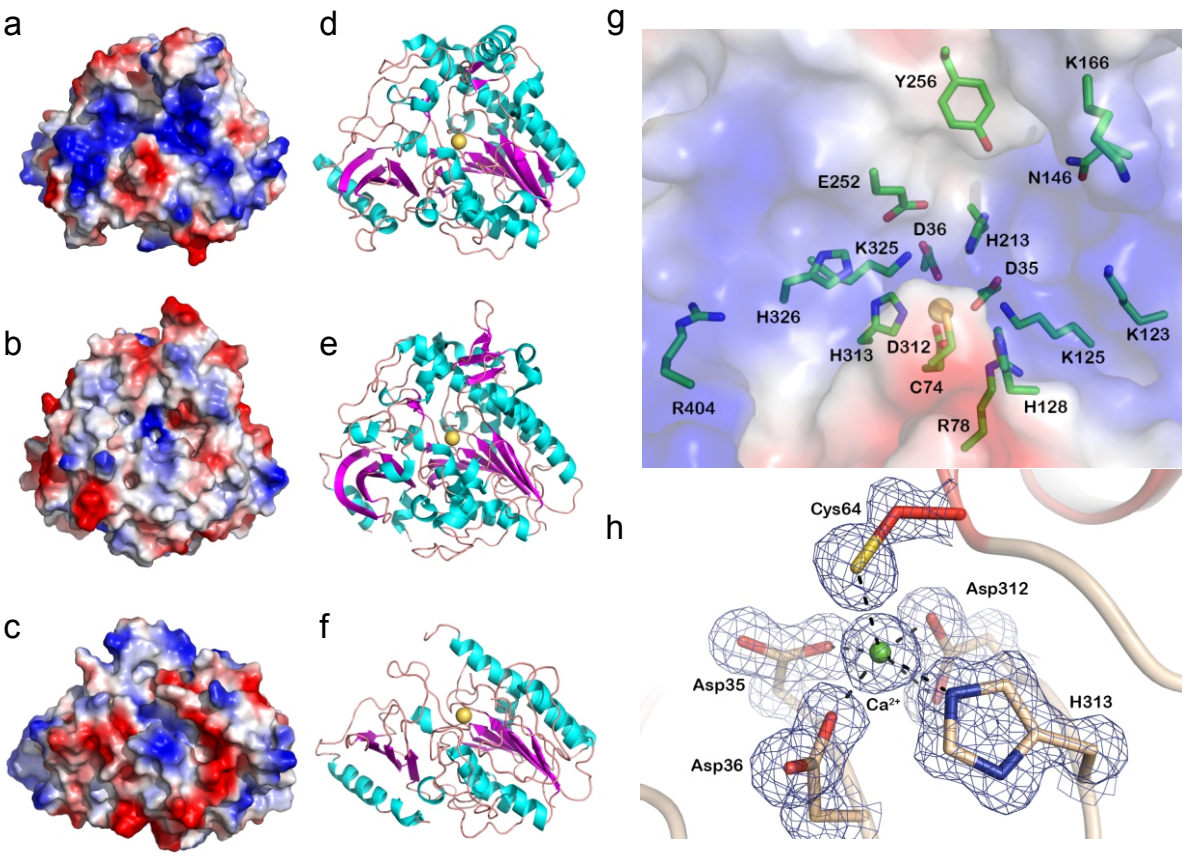
PUL H-encoded proteins (for ulvan and ulvan-derived monosaccharide utilization)			log$_2$ ratio		
Abbreviation	**Locus tag**	**Functional annotation**	**fru**	**rha**	**ulv**
P1_GH88	*21900	unsaturated glucuronyl hydrolase (GH88)			
P2_SusD	*21910	SusD-like protein			
P3_TBDR	*21920	TonB-dependent receptor			
P4_HK	*21930	histidine kinase			
P5_isomerase	*21940	4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase			
P6_dehydrogenase	*21950	2-deoxy-D-gluconate 3-dehydrogenase			
P7	*21960	conserved hypothetical protein			
P8_GH2	*21970	beta-galactosidase (GH2)			
P9_lactonase	*21980	6-phosphogluconolactonase			
P10_PLnc	*21990	ulvan lyase (PLnc)			
P11_S1_7	*22000	iduronate-2-sulfatase (S1_7)			
P12_S1_8	*22010	arylsulfatase (S1_8)			
P13_S1_16	*22020	arylsulfatase (S1_16)			
P14_S1_7	*22030	arylsulfatase (S1_7)			
P15_GH2	*22040	glycoside hydrolase (GH2)			
P16_GH2	*22050	beta-galactosidase (GH2)			
P17_GH2	*22060	beta-galactosidase (GH2)			
P18_S1_7	*22070	arylsulfatase (S1_7)			
P19_S1_27	*22080	sulfatase (S1_27)			
P20_GH78	*22090	alpha-L-rhamnosidase (GH78)			
P21_mutarotase	*22100	L-rhamnose mutarotase			
P22	*22110	conserved hypothetical protein			
P23	*22120	conserved hypothetical protein			
P24_GH3	*22130	beta-glucosidase (GH3)			
P25_SusD	*22140	SusD-like protein			
P26_TBDR	*22150	TonB-dependent receptor			
P27_GH43	*22160	beta-xylosidase (GH43)			
P28_GH78	*22170	alpha-L-rhamnosidase (GH78)			
P29	*22180	conserved hypothetical protein			
P30_PL28	*22190	ulvan lyase (PL28)			
P31_GH39	*22200	glycoside hydrolase (GH39)			
P32_S1_8	*22210	arylsulfatase (S1_8)			
P33_GH105	*22220	glycoside hydrolase (GH105)			
P34_GH3	*22230	beta-glucosidase (GH3)			
P35_oxidoreductase	*22240	oxidoreductase			
P36_GH78/S1_25	*22250	alpha-L-rhamnosidase/sulfatase (GH78/S1_25)			
P37	*22260	hypothetical protein			
P38_SusD	*22270	SusD-like protein			
P39_TBDR	*22280	TonB-dependent receptor			
Non-PUL H-encoded proteins (for ulvan-derived monosaccharide utilization)			**log$_2$ ratio**		
Abbreviation	**Locus tag**	**Functional annotation**	**fru**	**rha**	**ulv**
NP1_dehydrogenase	*21840	aldehyde dehydrogenase A			
NP2_dehydrogenase	*21850	L-lactate dehydrogenase			
NP3_aldolase	*21860	class II aldolase/adducin family protein			
NP4_kinase	*21870	pentulose/hexulose kinase			
NP5_isomerase	*21880	rhamnose isomerase[a]			
NP6_aldolase	*21890	rhamnulose-1-phosphate aldolase			
NP7_kinase	*160	xylulose kinase			
NP8_isomerase	*170	xylose isomerase			
NP9_oxidoreductase	*9410	D-mannonate oxidoreductase			
NP10_dehydratase	*9420	mannonate dehydratase			
NP11_isomerase	*9430	uronate isomerase			
NP12_kinase	*9800	2-dehydro-3-deoxygluconate kinase			
NP13_aldolase	*9820	aldolase[b]			
NP14_kinase	*11640	2-dehydro-3-deoxygluconate kinase			
NP15_kinase	*16400	2-dehydro-3-deoxygluconate kinase			
NP16_XylE	*180	D-xylose transporter XylE[c]			
NP17_ABC	*11090	ABC transporter, ATP-binding protein			
NP18_ABC	*25150	ABC transporter, ATP-binding protein			
NP19_ABC	*7480	ABC transporter, ATP-binding protein			
NP20_ABC	*12820	ABC transporter, ATP-binding protein			

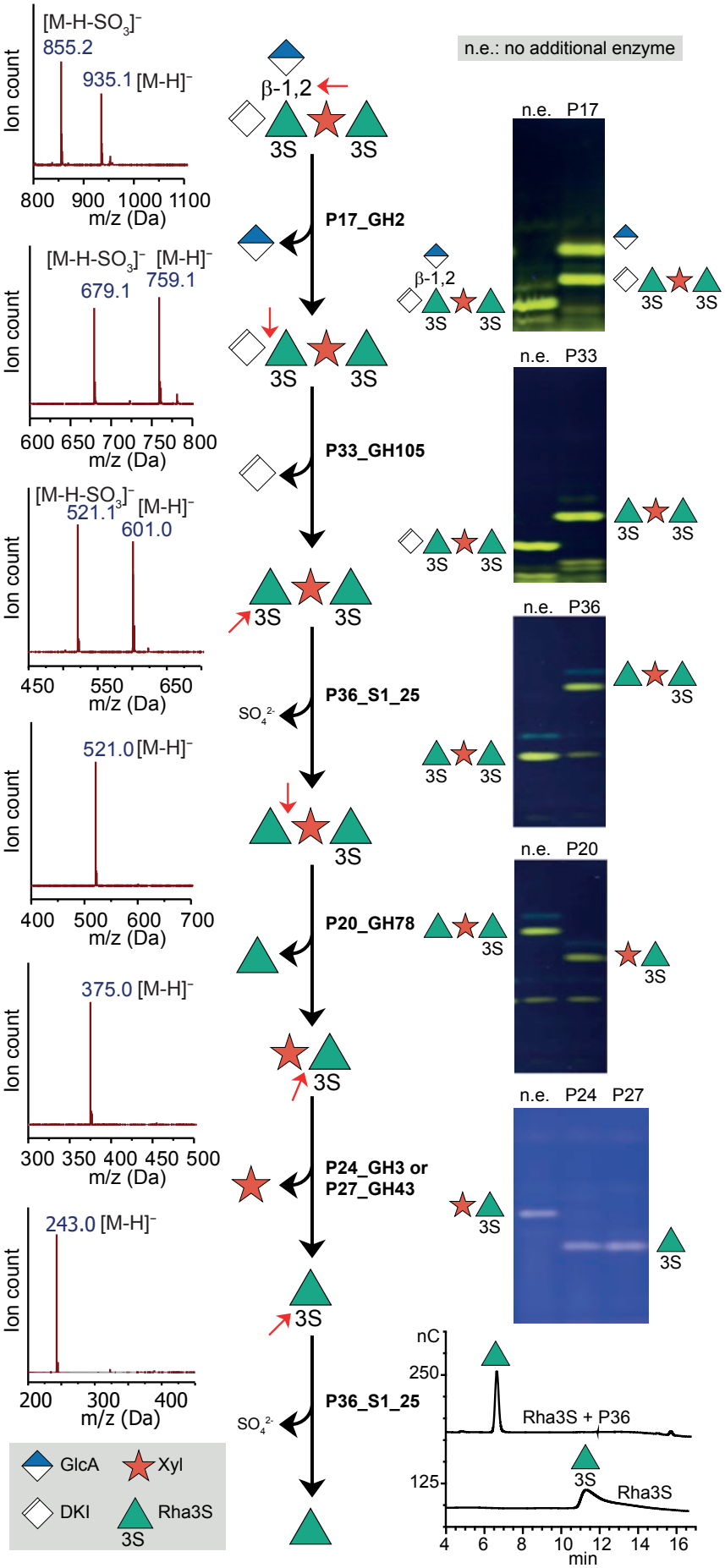
Proteins were numbered (P1 - P39: PUL H-encoded proteins, NP1 - NP20: non-PUL H-encoded proteins) and protein function was indicated. In the case of glycoside hydrolases (GH) and sulfatases (S), families and subfamilies were specified[21,22], e.g. GH2 (family) or S1_7 (family and subfamily). *BN863_, e.g. "*21900" refers to locus tag BN863_21800; [a]identified by BLAST against the Uniprot database, previously annotated as xylose isomerase-like TIM barrel domain protein, [b]4-hydroxy-2-oxoglutarate aldolase / 2-dehydro-3-deoxyphosphogluconate aldolase, [c]only captured by subproteome analysis of ulvan-grown cells (Supplementary Data Set 2), fru: fructose, rha: rhamnose, ulv: ulvan

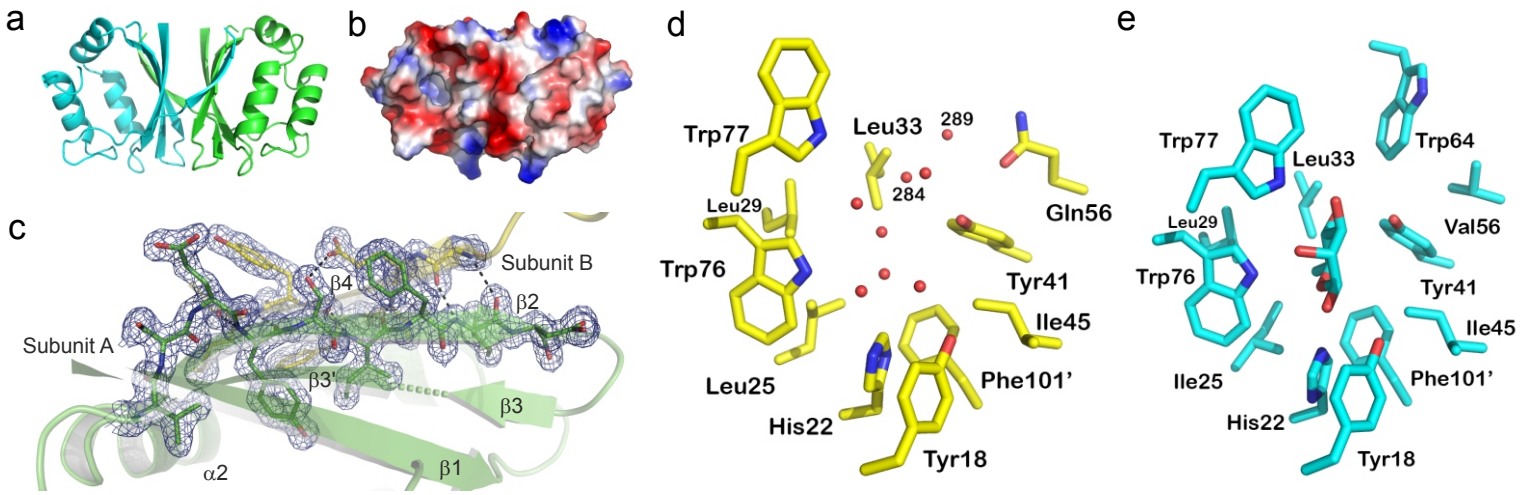
log$_2$ ratio
1.5 -7


Ulvan
Rha
DKI
GlcA
Xyl


a
b
c
polysaccharide lyase (PL)
glycoside hydrolase (GH)
sulfatase (S)
other
P1_GH88
P2_SusD
P3_TBDR
P4_HK
P5_isomerase
P6_dehydrogenase
P7
P8_GH2
P9_lactonase
P10_PLnc
P11_S1_7
P12_S1_8
P13_S1_16
P14_S1_7
P15_GH2
P16_GH2
P17_GH2
P18_S1_7
P19_S1_27
P20_GH78
P21_mutarotase
P22
P23
P24_GH3
P25_SusD
P26_TBDR
P27_GH43
P28_GH78
P29
P30_PL28
P31_GH39
P32_S1_8
P33_GH105
P34_GH3
P35_oxidoreductase
P36_GH78/S1_25
P37
P38_SusD
P39_TBDR
-4
0
2
△log2
rhamnose
ulvan
7: Tamlana sp. UJ94
10: Algibacter pectinivorans
5: Polaribacter sp. WD7
8: Wenyingzhuangia fucanilytica
6: Polaribacter sp. KT 15
3: Polaribacter sp. BM9
11: Cellulophaga tyrosinoxydans
1: Formosa agariphila KMM 3901
12: Flammeovirga pacifica
9: Nonlabens sediminis
2: Labilibacter marinus
4: Polaribacter sp. ZY113
Sulfatase
GH2
CBM67
GH78
GH88
GH39
DUF4976
GFO_IDH_MocA
PL28
CBM35
GH105
rhaM
HTH_AraC
DUF4982
TonB-dep_Rec
SusD_like_3
Plug
adh_short
GntR
KR
PLnc
GH3
Fn3_like
Color Key and Histogram
0 2 4 6 8
number


a
d
g
b
e
c
f
h
Y256
K166
E252
N146
D36
H213
K325
D35
H326
K123
H313
D312
K125
C74
R404
H128
R78
Cys64
Asp312
Asp35
Ca2+
H313
Asp36


[M-H-SO3]⁻
855.2
935.1 [M-H]⁻
Ion count
800 900 1000 1100
m/z (Da)
[M-H-SO3]⁻ [M-H]⁻
679.1
759.1
600 650 700 750 800
[M-H-SO3]⁻ [M-H]⁻
521.1
601.0
450 550 650
521.0 [M-H]⁻
400 500 600 700
375.0 [M-H]⁻
300 350 400 450 500
243.0 [M-H]⁻
200 300 400
β-1,2
3S
P17_GH2
P33_GH105
P36_S1_25
SO4 2-
P20_GH78
P24_GH3 or
P27_GH43
n.e.: no additional enzyme
n.e. P17
n.e. P33
n.e. P36
n.e. P20
n.e. P24 P27
nC
250
125
Rha3S + P36
Rha3S
4 6 8 10 12 14 16
min
GlcA
Xyl
DKI
Rha3S


a
b
c
d
e
Subunit A
Subunit B
β4
β2
β3'
β3
α2
β1
Trp77
Leu33
289
284
Gln56
Leu29
Trp76
Tyr41
Ile45
Leu25
Phe101'
His22
Tyr18
Trp64
Val56
Ile25

e

ulvan into oligosaccharides

uptake into periplasm

oligosaccharides into monosaccharides

uptake into cytosol

monosaccharides into central intermediates

β-1,2

3S 3S 3S 3S 3S 2S 3S 3S 3S 2S 3S 3S 3S 3S 3S

P30_PL28

3S 3S

3S 3S 2S 3S 3S 3S 2S 3S

β-1,2

3S 3S 3S 3S

P10_PLnc

3S

3S 3S 2S 3S 3S

3S 2S 3S

3S 3S

β-1,2

3S 3S

P25_ SusD

P26_ TBDR

P38_ SusD

P39_ TBDR

outer membrane

3S

3S 3S 2S 3S 3S

3S 2S 3S

β-1,2

3S 3S

P18_S1_7

P17_GH2

P31_GH39

SO_4^{2-}

P33_GH105

2S 3S

P32_S1_8

SO_4^{2-}

3S 3S

P33_GH105

P20_GH78

P36_S1_25

3S

3S

3S

3S 3S

P24_GH3 P27_GH43

P36_S1_25

SO_4^{2-}

SO_4^{2-}

DKI

NP16_ XylE

NP17_ NP18_ ABC

NP19_ NP20_ ABC

inner membrane

P21_mutarotase

α → β

NP8_isomerase

NP11_ isomerase

DKI

P5_isomerase

NP5_isomerase

D-xylulose

D-fructuronate

DKII

L-rhamnulose

NP7_kinase

NP9_oxidoreductase

P6_dehydrogenase

NP4_kinase

D-xylulose-5P

D-mannonate

L-rhamnulose-1P

pentose phosphate pathway

NP10_dehydratase

NP3_/NP6_ aldolase

D-glyceraldehyde-3P

KDG

L-lactaldehyde

dihydroxyacetone-P

NP12_/NP14_/NP15_ kinase

NP1_dehydrogenase

KDG-6P

L-lactate

NP13_aldolase

NP2_dehydrogenase

D-glyceraldehyde-3P + pyruvate

pyruvate

L-rhamnose D-xylose D-glucuronic acid L-iduronic acid unsaturated uronic acid