



HAL
open science

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.

► To cite this version:

Lukas Reisky, Aurelie Prechoux, Marie-Katherin Zühlke, Marcus Bäumgen, Craig Robb, et al.. A marine bacterial enzymatic cascade degrades the algal polysaccharide ulvan. *Nature Chemical Biology*, 2019, 15 (8), pp.803-812. 10.1038/s41589-019-0311-9. hal-02347779

HAL Id: hal-02347779

<https://hal.science/hal-02347779v1>

Submitted on 5 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **A marine bacterial enzymatic cascade degrades the algal polysaccharide ulvan**

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

7

8 ¹ Department of Biotechnology & Enzyme Catalysis, Institute of Biochemistry, University
9 Greifswald, 17487 Greifswald, Germany

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

12 ³ Pharmaceutical Biotechnology, Institute of Pharmacy, University Greifswald, 17487
13 Greifswald, Germany

14 ⁴ Institute of Marine Biotechnology, 17489 Greifswald, Germany

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

16 ⁶ University of Bremen, Center for Marine Environmental Sciences (MARUM), 28359
17 Bremen, Germany

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

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

21 ⁹ Institute of Microbiology, University Greifswald, 17487 Greifswald, Germany

22

23 #These authors contributed equally to this work.

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

27 **Abstract**

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

42

43 **Introduction**

44 Algal photosynthesis provides half of the global primary production¹. Carbon dioxide is
45 converted into carbohydrates, which are polymerized into polysaccharides to store energy,
46 build cell walls, and perform other biological functions. Algae are furthermore considered as
47 a promising renewable carbon source, due to their competitive growth rates and unique cell
48 walls. Unlike plants that are rich in woody tissue, comprising the insoluble polysaccharides
49 cellulose and the aromatic polymer lignin, which increases recalcitrance against enzymatic
50 digestion, algal cell walls are rich in gel-forming polysaccharides that are highly hydrated².
51 Hydration and the absence of lignin make harsh chemical and physical pretreatment of cell
52 walls unnecessary, and allow for easy access of enzymes that can digest the

53 polysaccharides into fermentable monosaccharides. Accordingly, recent studies showed that
54 bioengineered microbes equipped with agarases and alginate lyases can efficiently digest
55 and rapidly convert polysaccharides from brown and red algae into bioethanol³.

56 Sessile macroalgae, such as brown algae that form kelp forests, are ecologically valuable
57 because they provide nutrition and habitats for fish and other organisms and, consequently,
58 harvesting them would exacerbate pressure on natural populations. However, the planktonic
59 macroalgae *Ulva armoricana*, *Ulva rotunda* and other *Ulva* spp. that thrive in eutrophic,
60 nutrient-rich coastal waters, grow into expansive blooms that occur with increasing frequency
61 in recent years. They pose ecological but also economical threats when they accumulate on
62 beaches used for recreation⁴⁻⁶. Fertilized by nitrate from agriculture that is washed into the
63 ocean by rivers, *Ulva* blooms during summer produce up to 50–100 000 tons of biomass
64 every year, which must be removed at high expense from the northern and western coast of
65 France⁶. Even larger blooms occur in China⁵. Blooms of *Ulva* are thus a global phenomenon
66 that is bound to increase with farming activities, rendering the polysaccharide ulvan, which
67 accounts for up to 30 % of the algal dry weight⁷, an emerging yet untapped resource.

68 Ulvan is a branched polysaccharide composed of repeating disaccharide units, in which D-
69 glucuronic acid (GlcA) is β -1,4-linked or L-iduronic acid (IdoA) is α -1,4-linked to L-rhamnose-
70 3-sulfate (Rha3S), which is α -1,4-linked within the main chain. Some of the uronic acids are
71 replaced by β -1,4-linked D-xylose (Xyl), which can be sulfated at position 2 (Xyl2S).
72 Furthermore, Rha3S can be modified by β -1,2-linked GlcA side chains and the GlcA-Rha3S
73 or IdoA-Rha3S pattern can be interrupted by consecutive GlcA residues⁷⁻⁹. Increased interest
74 in the enzymatic degradation of ulvan recently led to the description of several ulvan-active
75 enzymes¹⁰⁻¹⁵. So far, and to the best of our knowledge, only two types of enzymes from
76 different carbohydrate-active enzyme (CAZyme) families showed activity on ulvan. Ulvan
77 polysaccharide lyases of the families PL24, PL25 and PL28 catalyze the initial cleavage
78 between Rha3S and GlcA or IdoA, resulting in the formation of unsaturated uronic acid
79 residues at the end of the formed oligosaccharide. Unsaturated uronic acid residues are

80 removed by glycoside hydrolases (GHs) from the family GH105^{15,16}. Ulvan-specific
81 degradation-related gene loci ('polysaccharide utilization loci', PULs) such as PUL H from
82 *Formosa agariphila* encode PL28 and GH105 together with over 10 additional, putative
83 enzymes, which were predicted to be involved in ulvan utilization. While PL28 and GH105
84 degrade ulvan, the other enzymes that were produced in *Escherichia coli* did not show
85 activity¹⁵. This result suggested that a complex cascade of sequential enzymatic reactions is
86 required for complete ulvan degradation^{15,17}.

87 Here, we experimentally established the complex ulvan degradation pathway of *F. agariphila*
88 KMM 3901^T, a marine flavobacterium, which was isolated from a green alga in the Sea of
89 Japan¹⁸. These degradation-related enzymes are encoded in an ulvan-specific PUL in the
90 bacterial genome¹⁵.

91

92 **Results**

93 **Bacterial ulvan-specific PULs**

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

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

117 In addition to the two already known¹⁵ ulvanolytic enzyme activities (ulvan lyase and
118 unsaturated glucuronyl hydrolase, GH105) we uncovered eight so far unknown enzyme
119 functions for the complete depolymerization of ulvan. Besides a novel PL family, we identified
120 and characterized six GH families (GH2, GH3, GH39, GH43, GH78, GH88) and three
121 sulfatases.

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

126

127 **Sulfatases active on ulvan**

128 Ulvans feature a large structural variability, with substitution by sulfate esters at various
129 positions. This chemical diversity is influenced by several factors such as the algal species,
130 the environmental conditions or the seasons⁷. The studied PUL of *F. agariphila* encodes 8
131 formylglycine-dependent sulfatases belonging to 5 subfamilies of the SulfAtlas S1 family

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

154 Sequence analyses revealed that P18_S1_7 (485 residues) and the S1_25 sulfatase module
155 of P36_S1_25 (443 residues) are only distantly related (25% identity) and thus belong to two
156 different SulfAtlas S1 subfamilies²⁰, S1_7 and S1_25, respectively. We determined the
157 crystal structure of these two sulfatases, with higher resolution for P18_S1_7 (1.23 Å) and
158 lower resolution for the sulfatase module of P36_S1_25 (2.91 Å). P18_S1_7 and P36_S1_25
159 adopt a similar fold with two α/β -structural domains, an *N*-terminal catalytic domain SD1

160 (Ser25-Asp388; P18_S1_7) separated by a structured linker (Arg389-Val397) from a C-
161 terminal domain SD2 (Ala398-Pro483). Nonetheless, the sulfatase module of P36_S1_25 is
162 a smaller protein and lacks some secondary elements, which are present in P18_S1_7 (the
163 β -strands β 6 and β 8, the α -helices α 5, α 7, α 8 and several short 3:10 helices). Notably, the
164 helix α 7 and the loops connecting it to the main part of SD1 constitute a protruding extension,
165 which overhangs the active site (Fig. 2d). The active site of P18_S1_7 is a large, open
166 groove with a strong basic character (Fig. 2a and 2g). This type of active site topology is
167 consistent with the endo-character and its efficiency on polymeric ulvan (Supplementary Fig.
168 4). In contrast, the active site of P36_S1_25 is a pocket (Fig. 2c), which is consistent with its
169 activity on oligosaccharides (Supplementary Fig. 4). The most similar protein in the Protein
170 Data Bank (PDB) is the human iduronate 2-sulfatase (IDS, 31% sequence identity; PDB:
171 5FQL; Fig. 2b and 2e)²³. Interestingly, IDS also displays a pocket active site topology (Fig.
172 2b). Therefore, different active sites (and subsequently different modes of action) can exist
173 within the same S1 subfamily. Such differences in topology likely explain the varying
174 efficiencies at the polymer level observed for P11_S1_7, P14_S1_7 and P18_S1_7 although
175 they all belong to the S1_7 subfamily (Supplementary Fig. 4).

176 The catalytic machinery of the S1 family sulfatases²⁴ is well conserved in P18_S1_7 and
177 P36_S1_25. We find the catalytic nucleophile (Cys74 and Cys58, respectively), residues
178 involved in Ca^{2+} coordination (Asp35, Asp36, Asp312 and His313; Asp18, Asp19, Asp284
179 and Asn285), residues stabilizing the catalytic nucleophile (Arg78 and His128; Arg62 and
180 Gly110), and residues of the sulfate-binding S subsite, as defined in the recent nomenclature
181 for sulfatase-binding subsites²⁵ (Lys125, His213 and Lys325; Lys108, His182 and Lys297)
182 (Fig. 2g, Supplementary Figs. 6-8). His313 in P18_S1_7 is not the most frequent residue for
183 the coordination of the calcium ion (usually an asparagine), but a histidine at this position is
184 found in a minority of sulfatases and is part of the updated PROSITE signature "Calcium-
185 binding site 2"²⁰. Most surprising is the replacement of His128 in P18_S1_7 by Gly100 in the
186 sulfatase module of P36_S1_25. Indeed, a histidine at this position is supposed not only to
187 stabilize the catalytic formylglycine, but also to abstract its O γ 2 proton at the end of the

188 catalytic cycle to induce the sulfate elimination and the aldehyde regeneration²⁴.
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. agariophila*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 products¹⁴. Both ulvan lyases, P30_PL28²⁶ and P10_PLnc, appear to initiate ulvan
218 depolymerization outside of the bacterial cell. P30_PL28 contains an additional ulvan-binding
219 module¹³ and a type IX secretion system signal that drives secretion²⁷, 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 recognition²⁸.
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 site¹⁵.

245 Besides ulvan lyases, the endo-active α -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 O2 residues of Rha3S⁷ 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

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

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

290 **Monosaccharide metabolism**

291 Ulvan degradation releases different monosaccharides to be further utilized by *F. agariphila*.
292 Many of the enzymes involved in monosaccharide metabolism had significantly higher
293 relative abundances with ulvan compared to fructose or rhamnose as substrate (Table 1,
294 Supplementary Fig. 3, Supplementary Data Set 1). Based on this result and on the MetaCyc
295 database²⁹, pathways for monosaccharide utilization were deduced, which are consistent
296 with previously proposed pathways¹⁵. 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 -anomer³². 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

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

331

332 **Discussion**

333 Using the DNA sequence of the known ulvan polysaccharide lyase PL28 as query, 12 ulvan
334 PULs were extracted from the NCBI-GenBank, including the biochemically characterized *F.*
335 *agariphila* ulvan PUL. All PULs were from Bacteroidetes, indicating that our procedure was
336 selective for this phylum since ulvan PULs also exist in Gammaproteobacteria³³.
337 Interestingly, although four ulvan PULs were from the genus *Polaribacter*, they did not cluster
338 on the heatmap (Figure 1c) indicating that ulvan PULs are diverse at the genus level. Also,
339 within different ulvan PULs, PL28 or PLs from PLnc are over 50% identical at the pairwise
340 amino acid sequence level. Conservation and invariable presence suggest that the first steps
341 of the ulvan degradation cascade proceed through similar enzymes in these organisms. On
342 the other hand, the GH88 enzyme was only present in ulvan PULs of *Flammeovirga pacifica*
343 and *F. agariphila*. GH88 is an exo-acting, unsaturated glucuronyl hydrolase. Its absence in
344 other ulvan PULs could be compensated for by the presence of a GH105, which has the
345 same function. Thus, the later steps in ulvan degradation proceed in dissimilar ways in
346 bacteria.

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

352 multimodular enzymes. Our analyses indicated similar domain structures of ulvan-degrading
353 enzymes in other marine Bacteroidetes strains. However, the cursory inspection of gut
354 *Bacteroides* genomes revealed no multimodular GH78 and sulfatase fusion proteins. This
355 suggests that some gene fusions involved in polysaccharide degradation could be more
356 abundant in the marine environment³⁴.

357 Our biochemical analyses demonstrated that six of the putative sulfatases (P11_S1_7,
358 P12_S1_8, P18_S1_7, P19_S1_27, P32_S1_8 and P36_S1_25) are ulvan-active indeed
359 sulfatases (Supplementary Fig. 4). However, the sulfatase P14_S1_7 was inactive on both,
360 ulvan from Elicityl and a xylose-rich ulvan from Atlantic *Ulva* spp. and displayed only faint
361 activity on an ulvan from Agrival. This apparent inactivity may be due to a strict exolytic
362 character of P14_S1_7. Consequently, activity maxima are not the same for different types of
363 ulvans. Substrate diversity may cause the variable enzyme content in Bacteroidetes (Figure
364 1c). This diversity may reflect an adaptation to the different types of ulvans present in *Ulva*
365 spp. Such fine scale adaptation points towards the exploration of PUL microdiversity as a
366 promising avenue for enzyme discovery and for the biocatalytic elucidation of ulvan
367 structures.

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

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

380

381 **Acknowledgements**

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

397

398 **Author contributions**

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

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

411

412 **Competing financial interests**

413 The authors declare no conflict of interest.

414

415 **References**

416

- 417 1. Field, C.B., Behrenfeld, M.J., Randerson, J.T. & Falkowski, P. Primary production of
418 the biosphere: integrating terrestrial and oceanic components. *Science* **281**, 237-240
419 (1998).
- 420 2. Kloareg, B. & Quatrano, R. Structure of the cell walls of marine algae and
421 ecophysiological functions of the matrix polysaccharides. *Oceanogr. Mar. Biol.* **26**,
422 259-315 (1988).
- 423 3. Wargacki, A.J. et al. An engineered microbial platform for direct biofuel production
424 from brown macroalgae. *Science* **335**, 308-313 (2012).
- 425 4. Smetacek, V. & Zingone, A. Green and golden seaweed tides on the rise. *Nature*
426 **504**, 84 (2013).
- 427 5. Liu, D. et al. The world's largest macroalgal bloom in the Yellow Sea, China:
428 formation and implications. *Estuar. Coast. Shelf Sci.* **129**, 2-10 (2013).
- 429 6. Ménesguen, A. & Piriou, J.-Y. Nitrogen loadings and macroalgal (*Ulva* sp.) mass
430 accumulation in Brittany (France). *Ophelia* **42**, 227-237 (1995).
- 431 7. Lahaye, M. & Robic, A. Structure and functional properties of ulvan, a polysaccharide
432 from green seaweeds. *Biomacromolecules* **8**, 1765-1774 (2007).
- 433 8. Lahaye, M., Brunel, M. & Bonnin, E. Fine chemical structure analysis of
434 oligosaccharides produced by an ulvan-lyase degradation of the water-soluble cell-
435 wall polysaccharides from *Ulva* sp. (Ulvales, Chlorophyta). *Carbohydr. Res.* **304**, 325-
436 333 (1997).
- 437 9. Lahaye, M. NMR spectroscopic characterisation of oligosaccharides from two *Ulva*
438 *rigida* ulvan samples (Ulvales, Chlorophyta) degraded by a lyase. *Carbohydr. Res.*
439 **314**, 1-12 (1998).
- 440 10. Ulaganathan, T. et al. New ulvan-degrading polysaccharide lyase family: structure
441 and catalytic mechanism suggests convergent evolution of active site architecture.
442 *ACS Chem. Biol.* **12**, 1269-1280 (2017).
- 443 11. Ulaganathan, T., Banin, E., Helbert, W. & Cygler, M. Structural and functional
444 characterization of PL28 family ulvan lyase NLR48 from *Nonlabens ulvanivorans*. *J.*
445 *Biol. Chem.* **293**, 11564-11573 (2018).

- 446 12. Ulaganathan, T., Helbert, W., Kopel, M., Banin, E. & Cygler, M. Structure–function
447 analyses of a PL24 family ulvan lyase reveal key features and suggest its catalytic
448 mechanism. *J. Biol. Chem.* **293**, 4026-4036 (2018).
- 449 13. Melcher, R.L., Neumann, M., Werner, J.P.F., Gröhn, F. & Moerschbacher, B.M.
450 Revised domain structure of ulvan lyase and characterization of the first ulvan binding
451 domain. *Sci. Rep.* **7**, 44115 (2017).
- 452 14. Reisky, L. et al. Biochemical characterization of an ulvan lyase from the marine
453 flavobacterium *Formosa agariphila* KMM 3901^T. *Appl. Microbiol. Biotechnol.* **102**,
454 6987-6996 (2018).
- 455 15. Salinas, A. & French, C.E. The enzymatic ulvan depolymerisation system from the
456 alga-associated marine flavobacterium *Formosa agariphila*. *Algal Res.* **27**, 335-344
457 (2017).
- 458 16. Collén, P.N. et al. A novel unsaturated β -glucuronyl hydrolase involved in ulvan
459 degradation unveils the versatility of stereochemistry requirements in family GH105.
460 *J. Biol. Chem.* **289**, 6199-6211 (2014).
- 461 17. Foran, E. et al. Functional characterization of a novel "ulvan utilization loci" found in
462 *Alteromonas* sp. LOR genome. *Algal Res.* **25**, 39-46 (2017).
- 463 18. Mann, A.J. et al. The genome of the alga-associated marine flavobacterium *Formosa*
464 *agariphila* KMM 3901^T reveals a broad potential for degradation of algal
465 polysaccharides. *Appl. Environ. Microbiol.* **79**, 6813-6822 (2013).
- 466 19. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M. & Henrissat, B. The
467 carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**,
468 D490-D495 (2014).
- 469 20. Barbeyron, T. et al. Matching the diversity of sulfated biomolecules: creation of a
470 classification database for sulfatases reflecting their substrate specificity. *PLoS One*
471 **11**, e0164846 (2016).
- 472 21. Mewis, K., Lenfant, N., Lombard, V. & Henrissat, B. Dividing the large glycoside
473 hydrolase family 43 into subfamilies: a motivation for detailed enzyme
474 characterization. *Appl. Environ. Microbiol.* **82**, 1686-1692 (2016).
- 475 22. Thomas, F. et al. Comparative characterization of two marine alginate lyases from
476 *Zobellia galactanivorans* reveals distinct modes of action and exquisite adaptation to
477 their natural substrate. *J. Biol. Chem.* **288**, 23021-23037 (2013).
- 478 23. Demydchuk, M. et al. Insights into Hunter syndrome from the structure of iduronate-2-
479 sulfatase. *Nat. Commun.* **8**, 15786 (2017).
- 480 24. Hanson, S.R., Best, M.D. & Wong, C.H. Sulfatases: structure, mechanism, biological
481 activity, inhibition, and synthetic utility. *Angew. Chem. Int. Ed.* **43**, 5736-5763 (2004).
- 482 25. Hettle, A.G. et al. The molecular basis of polysaccharide sulfatase activity and a
483 nomenclature for catalytic subsites in this class of enzyme. *Structure* **26**, 747-758. e4
484 (2018).
- 485 26. Kopel, M. et al. New family of ulvan lyases identified in three isolates from the
486 Alteromonadales order. *J. Biol. Chem.* **291**, 5871-5878 (2016).
- 487 27. Sato, K. et al. A protein secretion system linked to bacteroidete gliding motility and
488 pathogenesis. *Proc. Natl. Acad. Sci. USA* **107**, 276-281 (2010).
- 489 28. Fujimoto, Z. et al. The structure of a *Streptomyces avermitilis* α -L-rhamnosidase
490 reveals a novel carbohydrate-binding module CBM67 within the six-domain
491 arrangement. *J. Biol. Chem.* **288**, 12376-12385 (2013).
- 492 29. Caspi, R. et al. The MetaCyc database of metabolic pathways and enzymes. *Nucleic*
493 *Acids Res.* **46**, D633-D639 (2017).
- 494 30. Ryu, K.-S. et al. Structural insights into the monosaccharide specificity of *Escherichia*
495 *coli* rhamnose mutarotase. *J. Mol. Biol.* **349**, 153-162 (2005).
- 496 31. Richardson, J.S. et al. RhaU of *Rhizobium leguminosarum* is a rhamnose mutarotase.
497 *J. Bacteriol.* **190**, 2903-2910 (2008).
- 498 32. Korndörfer, I., Fessner, W.-D. & Matthews, B.W. The structure of rhamnose
499 isomerase from *Escherichia coli* and its relation with xylose isomerase illustrates a
500 change between inter and intra-subunit complementation during evolution. *J. Mol.*
501 *Biol.* **300**, 917-933 (2000).

- 502 33. Koch, H., Freese, H.M., Hahnke, R.L., Simon, M. & Wietz, M. Adaptations of
503 *Alteromonas* sp. 76-1 to Polysaccharide Degradation: A CAZyme Plasmid for Ulvan
504 Degradation and Two Alginolytic Systems. *Front Microbiol* **10**, 504 (2019).
505 34. Hehemann, J.H. et al. Aquatic adaptation of a laterally acquired pectin degradation
506 pathway in marine gammaproteobacteria. *Environ. Microbiol.* **19**, 2320-2333 (2017).
507

508 Figure legends

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

529

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

544

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

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

558

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

571

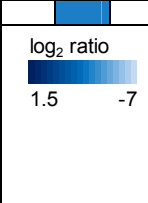
572 **Figure 5 | Model of the ulvan degradation pathway in *F. agariphila* as suggested by the**
573 **proteogenomic, biochemical and structural biological analyses in this study.**
574 Redundant pathways are omitted to maintain clarity. The ulvan molecule on top represents a
575 part within the larger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

587

588

589 **Table 1 | List of PUL H-encoded and relevant non-PUL H-encoded proteins** with
590 abbreviations used in the text, corresponding locus tags and functional annotation as well as
591 their relative abundance (mean log₂ ratio) with the respective carbon source. Empty/white
592 squares refer to non-quantified proteins while grey squares indicate OFF-proteins that could
593 not be quantified due to a lack of ¹⁴N 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			

594

595

596 Online Methods

597 Prediction of ulvan PULs

598 118,981 bacterial genomes were downloaded from the NCBI-GenBank using an in-house
 599 script (updated in 2018.09.10). Hmmer 3.0 was used to identify proteins with a PL28 or
 600 sulfatase domain, using a cut-off value of $1e-10^{35}$. Hidden Markov models of PL28 and
 601 sulfatase were obtained from dbCAN2 and the pfam database, respectively^{36,37}. Models for
 602 the new PLnc family have not been released, thus blastp was used to identify its homologs,
 603 using $1e-50$ and 30% sequence identity as cut-off values³⁸. In each bacterial genome, if the
 604 adjacent 50 proteins to the afore-mentioned marker genes contained three marker genes
 605 (PL28, PLnc and sulfatase), this locus was considered as a potential ulvan PUL hit. To
 606 further determine PUL boundaries, 100 proteins surrounding the predicted ulvan PUL were
 607 collected and then locally annotated using pfam and dbCAN Hidden Markov models (cut
 608 value $1e-10$). Firstly, PL28 or PLnc families were set as boundaries, which were extended if
 609 adjacent genes are annotated as sugar utilization proteins, such as GH, PL, sugar
 610 transporter and transcription factors. In cases where five continuous genes were not related
 611 to sugar utilization or ulvan degradation, the last functionally relevant protein was taken as
 612 the putative ulvan PUL boundary. Protein sequences within putative ulvan PULs were
 613 collected for further analysis. Circos was used to visualize the different ulvan PULs³⁹. Blastp
 614 was used to calculate the identity between PL28 sequences from different ulvan PULs (cut-
 615 off value: $1e-10$, over 50% identity). To simplify and reduce non-conserved proteins, domains
 616 with less than 80% presence among the predicted ulvan PULs were excluded. Domain
 617 numbers in each PUL were counted, summarized and displayed in R studio.

618 **Proteome analyses**

619 *Whole cell proteome – metabolic labeling*

620 A $^{14}\text{N}/^{15}\text{N}$ relative quantification approach, based on metabolic labeling, was used for protein
621 quantification as described previously⁴⁰. For this purpose, *F. agariphila* KMM 3901^T was
622 cultivated in MPM salts⁴¹ containing either ^{14}N - or ^{15}N -ammonium chloride, supplemented
623 with 0.2% of the individual carbon source: ulvan, rhamnose or fructose. Cultivation (21°C,
624 170 rpm) comprised three steps: (i) 24 h of marine broth 2216-cultivation and subsequent (ii)
625 pre-cultures as well as (iii) main cultures in the above-described minimal medium. At an
626 $\text{OD}_{600\text{nm}}$ of 0.5, cells were harvested from main cultures by centrifugation (30 min, 9,384 x g,
627 4°C). Cell pellets were suspended in TE-buffer (10 mM Tris, 10 mM EDTA) and cells were
628 disrupted by sonication (4 cycles of 25 s at 5 m/s). Cell debris and protein extract were
629 separated by centrifugation (10 min, 21,460 x g, 4 °C). In case of the ^{15}N -labeled samples,
630 protein extracts of all samples from all 3 carbon sources were combined to form the ^{15}N -
631 labeled reference pool, which served as an internal standard. Protein concentration was
632 determined using the PierceTM BCA Protein Assay Kit (Thermo Fisher Scientific). 12.5 µg of
633 protein of each ^{14}N -sample was combined with 12.5 µg protein of the ^{15}N -labeled pool. These
634 mixtures were separated by 1D-SDS-PAGE. Protein lanes were cut into ten equal-sized
635 pieces, destained and proteins were in-gel-digested with a 1 µg mL⁻¹ trypsin solution⁴².
636 Peptides were separated by RP chromatography and analyzed in an LTQ-Orbitrap Classic
637 mass spectrometer equipped with a nanoelectrospray ion source⁴³. Data represent three
638 independent experiments (n=3).

639 MS data were searched with Sorcerer SEQUEST v.27, rev.11 (Thermo-Finnigan, Thermo
640 Fisher Scientific, Germany) against a target decoy database including all *F. agariphila* KMM
641 3901^T protein sequences, corresponding reversed sequences (decoys) as well as common
642 laboratory contaminants (total 7224 entries) as described previously⁴³, but using a false
643 positive rate of 0.05. In brief, peak intensities of the ^{14}N -peptide ions of a protein versus its
644 ^{15}N -peptide ions were compared to calculate a regression ratio. Only unique peptides and

645 peptides with an R^2 above 0.7 were taken into account. Non-quantified peptides were
646 manually checked. Average regression ratios were then exported. Proteins with at least two
647 quantified peptides were considered for the following calculations: ratios were median-
648 centered and log-transformed, termed as \log_2 ratios, per sample. If proteins were quantified
649 in at least two of the three replicates, means and standard deviations (SD) were calculated
650 from these values. In order to identify relative changes between the different carbon sources,
651 \log_2 ratios of fructose- or rhamnose-cultivated cells were subtracted from \log_2 ratios of ulvan-
652 cultivated cells, termed $\Delta\log_2$. Fold-changes correspond to the exponentials of these $\Delta\log_2$
653 values. Statistical analyses were performed with Welch's two-sided t-test (permutation-based
654 false discovery rate 0.01) using Perseus v1.6.0.7⁴⁴, based on the \log_2 ratios. Putative
655 ON/OFF proteins were marked with 15N (OFF) or 14N (ON) in Supplementary Data Set 1,
656 but were not included in any of the calculations. Only if a protein was identified as an
657 ON/OFF protein in all three replicates, it was assigned to a fixed value (10/-10), to highlight
658 these proteins.

659 *Subproteome fractionation*

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

662 For the surface proteome (trypsin-shaving approach), 1.5 mL of cell suspension was
663 removed from the culture and centrifuged (5 min, 5,867 x g, 4 °C). Cells were washed with
664 50 mM triethylammoniumbicarbonate-buffer (TEAB) and finally resuspended in 45 μ L TEAB-
665 buffer. In order to cleave proteins from the cell surface, 5 μ L of a 1 μ g mL⁻¹ trypsin solution
666 was added. The solution was transferred onto a 0.22 μ m cellulose-acetate spin-column and
667 incubated for 15 min at 900 rpm and 37 °C. The flow-through was collected by centrifugation
668 (10 min, 4,000 x g, 4 °C), another 1 μ L of trypsin was added and the sample was incubated
669 at 900 rpm and 37 °C overnight. The peptide mixture was desalted using C18 StageTips.
670 The following solutions were used: 0.1% (v/v) acetic acid in ultra-pure water (buffer A) and
671 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 sample⁴⁷. 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 strategy⁴⁸ 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 previously¹⁴ 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 described⁴⁹. 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

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

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

705 *Escherichia coli* BL21(DE3) was transformed with pET28-based plasmids. For the
706 overexpression, 50 mL ZYP-5052⁵⁰ with 100 µg mL⁻¹ kanamycin were inoculated from an
707 overnight culture in LB containing 50 µg mL⁻¹ kanamycin. The culture was grown at 30 °C
708 and 180 rpm until the OD_{600nm} reached 1.0 and was then cooled to 20 °C for 48 h. In the case
709 of sulfatases, the formylglycine-generating enzyme (FGE) from *F. agariphila* was co-
710 expressed. LB medium with 100 µg mL⁻¹ ampicillin and 50 µg mL⁻¹ kanamycin was inoculated
711 from an overnight culture in the same medium and incubated at 37 °C and 180 rpm until the
712 OD_{600nm} reached 0.3 to 0.5. After the addition of 1.5 mM L-arabinose and incubation for 90
713 min at 37 °C, the culture was cooled to 18 °C for 2 h before 0.5 mM isopropyl β-D-1-
714 thiogalactopyranoside (IPTG) was added and the culture was incubated overnight at 18 °C.
715 Alternatively, sulfatases were expressed from the pFA constructs in *E. coli* BL21(DE3) cells
716 grown in LB medium supplemented with 15 µg mL⁻¹ ampicillin, at 37 °C, until reaching an
717 OD_{600 nm} of 0.8. Expression was induced with 0.1 mM IPTG overnight at 18 °C. For
718 crystallization screening, *E. coli* BL21(DE3) cells were transformed with the plasmids
719 containing the gene fragment of interest, then grown in the autoinduction Zyp-5052 medium
720 (200 µg mL⁻¹ ampicillin, 20 °C, 72 h). Cells were harvested by centrifugation (10,000 x g, 4
721 °C, 20 min) and the cell pellets were stored at -20 °C until further use.

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

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

728 **Enzyme purification**

729 Cell pellets were thawed and resuspended in 50 mM NaPi with 300 mM NaCl (pH 8.0) and
730 lysed by three cycles of sonication (2.0 min, 30% pulse, 50% power). After centrifugation
731 (10,000 x g, 4 °C, 20 min), the supernatant was filtered (0.45 µm) and loaded onto a 5 mL
732 HisTrap FF crude column (GE Healthcare, Freiburg, Germany) equilibrated with lysis buffer.
733 Alternatively, Rotigarose-His/Ni beads (Karl Roth, Karlsruhe, Germany) were used in gravity
734 flow columns. After washing, the protein was eluted with 50 mM NaPi and 300 mM NaCl
735 containing 300 mM imidazole (pH 8.0). Fractions containing the protein of interest were
736 pooled and desalted using PD-10 columns (GE Healthcare, Freiburg, Germany) equilibrated
737 with 50 mM NaPi pH 7.4. Proteins were analyzed by SDS-PAGE on 12.5% acrylamide gels.
738 1% (v/v) 2,2,2-trichloroethanol was used for the visualization of proteins under UV light⁵¹.
739 Alternatively, proteins were stained with Coomassie Blue (PhastGel[®] Blue R). All enzymes
740 were used undiluted, or in dilutions of 1:5, 1:10 or 1:20 with enzyme storage buffer
741 (Supplementary Table 11).

742 Alternatively, cells were subjected to mechanical lysis and cytoplasmic extracts were loaded
743 onto an Histrap column (5ml, GE Healthcare) equilibrated with 50 mM Tris, 0.2 M NaCl, 20
744 mM imidazol, 1 mM CaCl₂ at pH 8.0. Recombinant proteins were eluted with around 250 mM
745 imidazole and then loaded onto a Hiprep Desalting column (26/10, 53ml, GE Healthcare) in
746 order to eliminate the imidazole, which notably interfered with sulfatase activity. Purified
747 enzymes were concentrated (Amicon[®] Ultra Centrifugal Filter, 30 kDa) to a concentration of
748 1 mg mL⁻¹ (Nanodrop).

749

750 **Purification of ulvan**

751 Green tide *Ulva* sp. was collected near Roscoff (France) and dried. Alternatively, dried *Ulva*
752 biomass from the Atlantic coast in Spain was purchased as organic sea lettuce (Kulau,
753 Berlin, Germany). Ulvan was extracted according to the literature⁵². The dialysis step was
754 exchanged by precipitation with acetone (80% (v/v) final concentration). After washing,
755 acetone-precipitated ulvan was dissolved in deionized water and freeze-dried. Alternatively,
756 ulvan was obtained from Agrival (Plouenan, France) or Elicityl (Grenoble, France).

757 **Enzyme assays**

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

767 **Sulfatase activity assay on ulvan polymers**

768 Activity assays were conducted on three different ulvan polymers from *Ulva* species: a
769 commercial ulvan from Elicityl (Grenoble, France), an ulvan which was a gift from the
770 company Agrival (Plouenan, France), and an ulvan extracted from *Ulva* sp. harvested on
771 Brittany north coast (Roscoff, France). 10 µL of each ulvan solution (1% w/v in H₂O) was
772 incubated with 10 µL of purified sulfatase (1 mg mL⁻¹) in a final volume of 80 µL of 25 mM
773 Tris-HCl, 0.1 M NaCl, 0.5 mM CaCl₂, pH 8.0 buffer mix, for 18 h at 37 °C. For each reaction,
774 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 H₂O) 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 CaCl₂, 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 fluorophore⁵³.

800 For carbohydrate polyacrylamide gel electrophoresis (C-PAGE), samples were mixed with an
801 equal volume of FACE loading buffer⁵³. Gels and running conditions were identical to FACE.
802 Carbohydrates were visualized by staining with Stains-All solution (0.25 g L⁻¹ in 1.7 mM Tris-
803 HCl pH 7.5 with 25% (v/v) isopropanol). The gels were destained with 25% (v/v) isopropanol
804 in deionized water.

805 **Purification of oligomers and structure determination**

806 Ulvan was digested with purified enzymes in Tris-HCl pH 8.5 at room temperature.
807 Oligomers were separated on two XK 26/100 (GE Healthcare, Freiburg, Germany) in series
808 filled with Bio-Gel P-2 (Rio-Rad, Munich, Germany) using 100 mM (NH₄)₂CO₃ as mobile
809 phase at a flow rate of 1 mL min⁻¹. After lyophilization of the fractions containing the products,
810 oligomers were dissolved in D₂O and lyophilized three times before NMR spectra were
811 recorded on a Bruker Avance III HD 600 (600 MHz) spectrometer (Bruker, Billerica, USA) in
812 D₂O solutions. The structures were independently elucidated based on 1D and 2D (COSY,
813 HSQC, HMBC, TOCSY) methods and the assigned ¹H and ¹³C-NMR signals were then
814 compared with literature data, showing excellent consistency^{8,9}. For samples containing
815 uronic acid structures, it was required to neutralize the otherwise acidic NMR samples with
816 Na₂HPO₄ to pH 7-8 (pH-electrode calibrated to H⁺) in order to achieve fully resolved signals
817 for the carboxylic acid and neighboring positions (¹³C). HPLC-ELS-MS analysis was
818 performed by injection of ~0.1% solutions (1-5 μL) on a Nexera UHPLC system from
819 Shimadzu (equipped with two binary LC-30AD pumps plus degassers, a CTO-20 column
820 oven) and a LCMS-2200 EV MS-detector and an additional ELS-detector (JASCO ELS-
821 2041). Analysis was performed with mobile phase A = H₂O (0.1% HCOOH) and mobile
822 phase B = CH₃CN on a C18 column (XSelect CSH XP C18 2.5 μm 3 x 50 mm) at 40 °C.
823 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
824 min and 98%-5% B from 2.2-2.5 min.

825 **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 XDS⁵⁴. Scaling and merging
842 were performed using the program Aimless from the CCP4 package⁵⁵. 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 MolRep⁵⁶
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 suite⁵⁷ and the Coot software⁵⁸. Initial refinement of the P36_S1_25 structure was
850 performed with BUSTER⁵⁹ and PHENIX⁵⁷, and then manual examination and rebuilding of
851 the refined coordinates were carried out in Coot⁵⁸. Structural validation was undertaken using
852 MOLPROBITY⁶⁰.

853 SEC-purified P17_GH2 crystallized in 1:1 ratio of 7 mg mL⁻¹ protein in 20 mM Tris pH 8.0 and
854 mother liquor in the JBScreen PACT ++ HTS and JBScreen Classic HTS I (Jena
855 Bioscience). A single crystal from the screen grown in 20 % PEG 3350, 0.1 M Bis-Tris pH
856 7.5, 0.2 M sodium bromide was cryo-protected in 30 % glycerol prior X-ray crystallography.
857 The diffraction data were collected at DESY P11 automatically integrated in XDS and scaled
858 and merged in Aimless^{54,61}.

859 The structure of P17_GH2 was solved by molecular replacement using 5dmy as search
860 model in phaser⁶². The structure was built automatically using buccaneer and manually in
861 Coot building directly into the 2Fo-Fc maps^{58,63}. Structural validation was carried out using
862 MOLPROBITY⁶⁰.

863 **Data availability**

864 All data that support the findings of this study are available from the corresponding authors
865 upon reasonable request. The protein structures are deposited in the PDB under 6HHM,
866 6HHN, 6HPD and 6HR5. Mass spectrometry data were deposited to the ProteomeXchange
867 Consortium via the PRIDE partner repository⁶⁴ with the dataset identifier PXD009299. The
868 sequences of the newly characterized ulvan-degrading enzymes can be found in the
869 GenBank sequence database under the respective locus tags mentioned in Table 1.

870 **References**

- 871 35. Finn, R.D., Clements, J. & Eddy, S.R. HMMER web server: interactive sequence
872 similarity searching. *Nucleic Acids Res.* **39**, W29-W37 (2011).
873 36. Zhang, H. et al. dbCAN2: a meta server for automated carbohydrate-active enzyme
874 annotation. *Nucleic Acids Res.* **46**, W95-W101 (2018).
875 37. Finn, R.D. et al. Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222-
876 D230 (2013).
877 38. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinformatics* **10**,
878 421 (2009).
879 39. Krzywinski, M.I. et al. Circos: an information aesthetic for comparative genomics.
880 *Genome Res.* **9**, 1639-1645 (2009).
881 40. MacCoss, M.J., Wu, C.C., Liu, H., Sadygov, R. & Yates, J.R., 3rd. A correlation
882 algorithm for the automated quantitative analysis of shotgun proteomics data. *Anal.*
883 *Chem.* **75**, 6912-6921 (2003).
884 41. Schut, F. et al. Isolation of typical marine bacteria by dilution culture: growth,
885 maintenance, and characteristics of isolates under laboratory conditions. *Appl.*
886 *Environ. Microbiol.* **59**, 2150-2160 (1993).

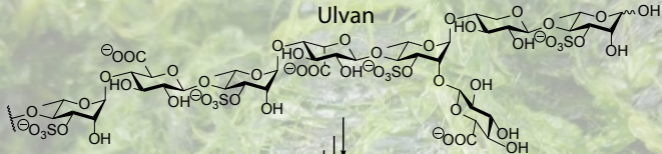
- 887 42. Grube, M. et al. Exploring functional contexts of symbiotic sustain within lichen-
888 associated bacteria by comparative omics. *ISME J.* **9**, 412-424 (2015).
- 889 43. Otto, A. et al. Systems-wide temporal proteomic profiling in glucose-starved *Bacillus*
890 *subtilis*. *Nat. Commun.* **1**, 137 (2010).
- 891 44. Tyanova, S. & Cox, J. Perseus: a bioinformatics platform for integrative analysis of
892 proteomics data in cancer research. *Methods Mol. Biol.* **1711**, 133-148 (2018).
- 893 45. Antelmann, H. et al. A proteomic view on genome-based signal peptide predictions.
894 *Genome Res.* **11**, 1484-1502 (2001).
- 895 46. Eymann, C. et al. A comprehensive proteome map of growing *Bacillus subtilis* cells.
896 *Proteomics* **4**, 2849-2876 (2004).
- 897 47. Zybaylov, B. et al. Statistical analysis of membrane proteome expression changes in
898 *Saccharomyces cerevisiae*. *J. Proteome Res.* **5**, 2339-2347 (2006).
- 899 48. Li, C. et al. FastCloning: a highly simplified, purification-free, sequence-and ligation-
900 independent PCR cloning method. *BMC Biotechnol.* **11**, 92 (2011).
- 901 49. Groisillier, A. et al. MARINE-EXPRESS: taking advantage of high throughput cloning
902 and expression strategies for the post-genomic analysis of marine organisms. *Microb.*
903 *Cell Fact.* **9**, 45 (2010).
- 904 50. Studier, F.W. Protein production by auto-induction in high-density shaking cultures.
905 *Protein Expr. Purif.* **41**, 207-234 (2005).
- 906 51. Ladner, C.L., Yang, J., Turner, R.J. & Edwards, R.A. Visible fluorescent detection of
907 proteins in polyacrylamide gels without staining. *Anal. Biochem.* **326**, 13-20 (2004).
- 908 52. Robic, A., Gaillard, C., Sassi, J.F., Lerat, Y. & Lahaye, M. Ultrastructure of ulvan: a
909 polysaccharide from green seaweeds. *Biopolymers* **91**, 652-664 (2009).
- 910 53. Hehemann, J.H. et al. Transfer of carbohydrate-active enzymes from marine bacteria
911 to Japanese gut microbiota. *Nature* **464**, 908-912 (2010).
- 912 54. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta*
913 *Crystallogr. D Biol. Crystallogr.* **66**, 133-144 (2010).
- 914 55. Winn, M.D. et al. Overview of the CCP4 suite and current developments. *Acta*
915 *Crystallogr. D Biol. Crystallogr.* **67**, 235-242 (2011).
- 916 56. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D*
917 *Biol. Crystallogr.* **66**, 22-25 (2010).
- 918 57. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for
919 macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213-221
920 (2010).
- 921 58. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of
922 Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486-501 (2010).
- 923 59. Smart, O.S. et al. Exploiting structure similarity in refinement: automated NCS and
924 target-structure restraints in BUSTER. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 368-
925 380 (2012).
- 926 60. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular
927 crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12-21 (2010).
- 928 61. Evans, P.R. & Murshudov, G.N. How good are my data and what is the resolution?
929 *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204-1214 (2013).
- 930 62. McCoy, A.J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658-674
931 (2007).
- 932 63. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein
933 chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002-1011 (2006).
- 934 64. Vizcaíno, J.A. et al. 2016 update of the PRIDE database and its related tools. *Nucleic*
935 *Acids Res.* **44**, D447-D456 (2016).

936

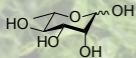
Table 1 | List of PUL H-encoded and relevant non-PUL H-encoded proteins with abbreviations used in the text, corresponding locus tags and functional annotation as well as their relative abundance (mean \log_2 ratio) with the respective carbon source. Empty/white squares refer to non-quantified proteins 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		

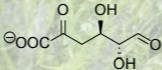
Ulvan



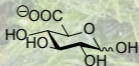
Rha



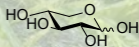
DKI

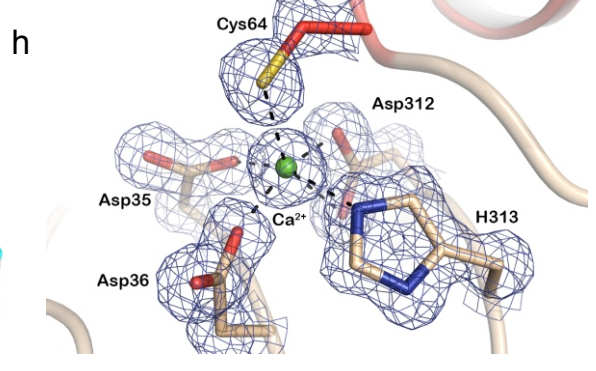
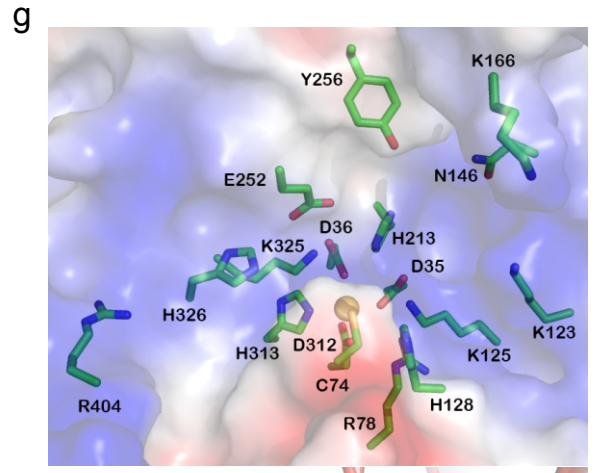
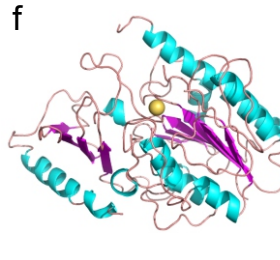
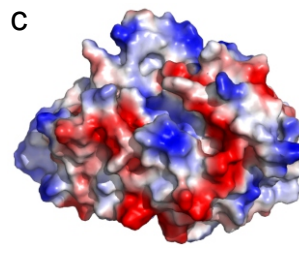
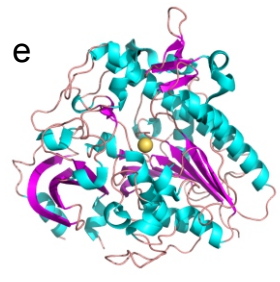
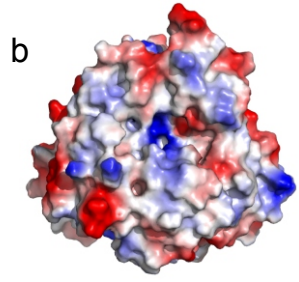
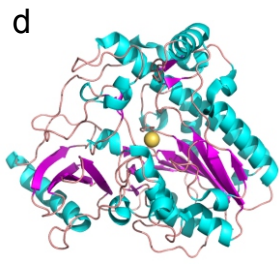
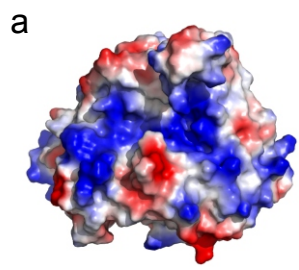


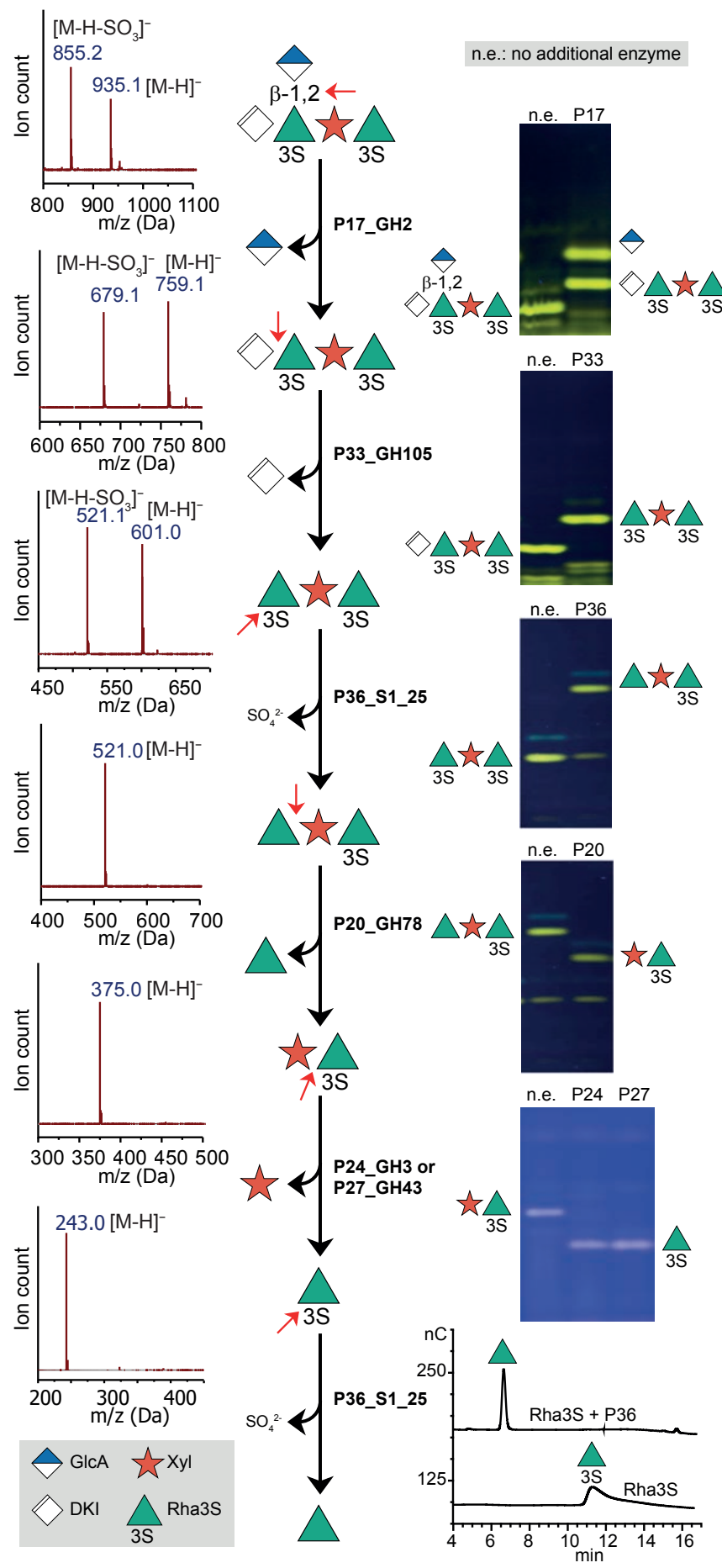
GlcA

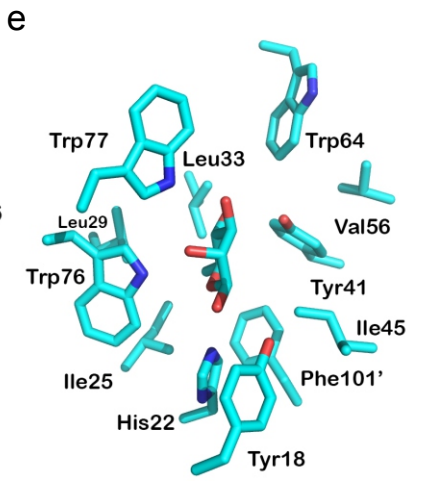
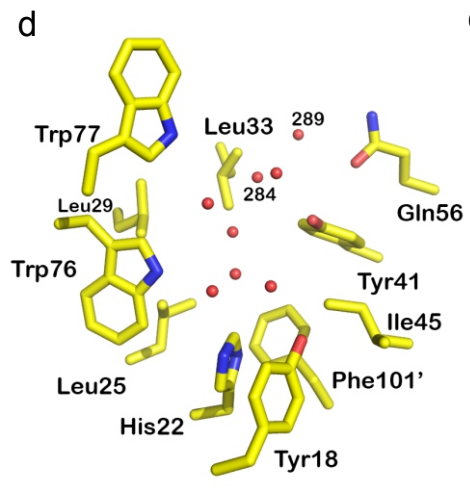
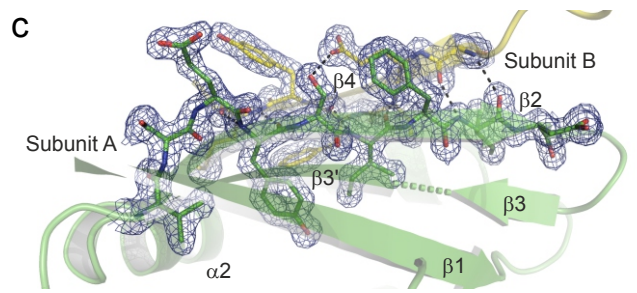
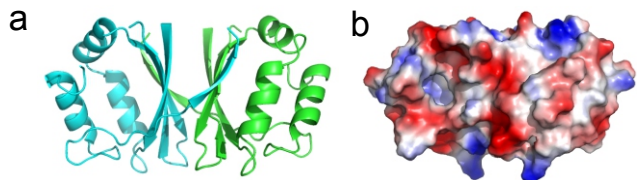


Xyl

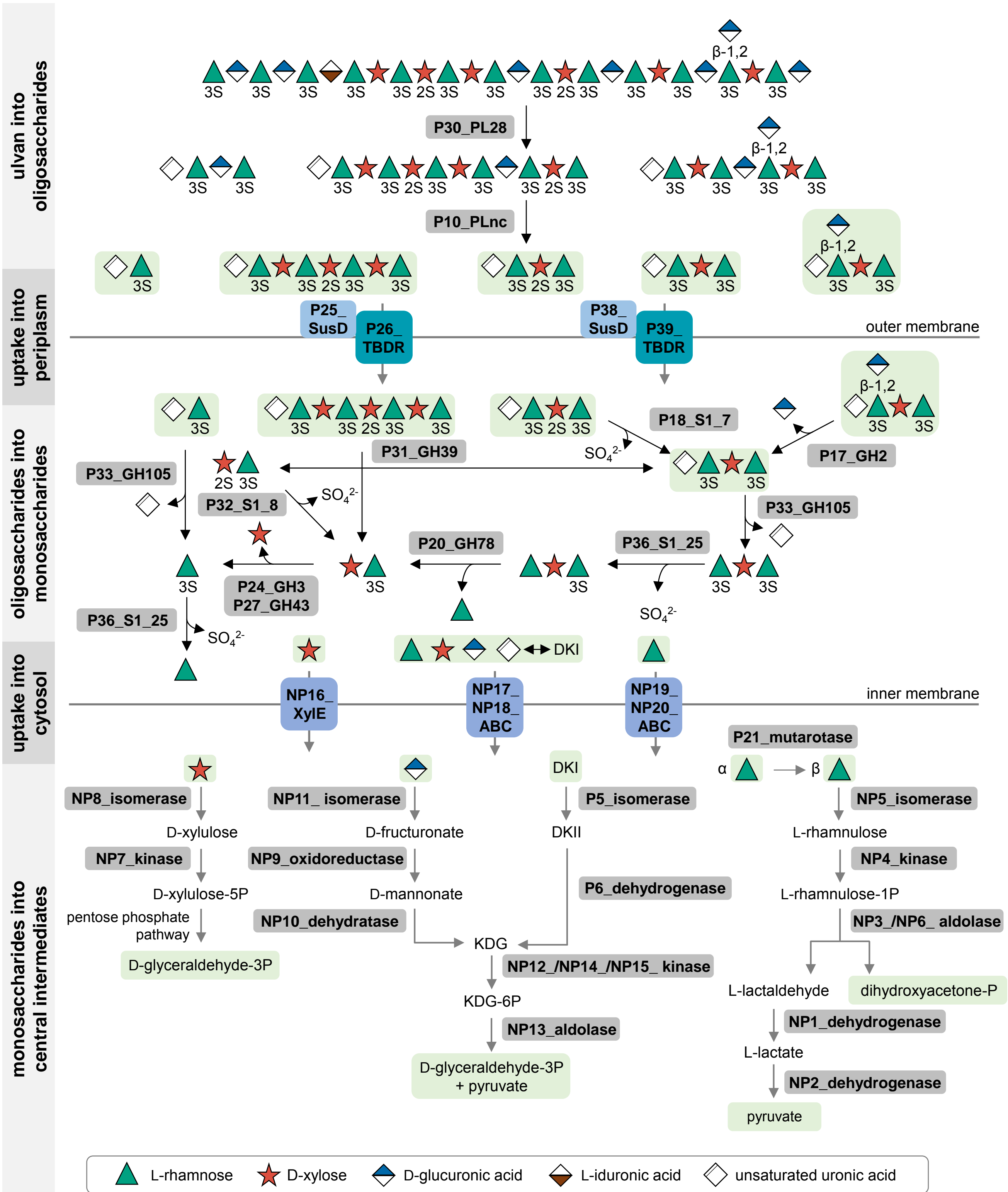








e



▲ L-rhamnose
 ★ D-xylose
 ◆ D-glucuronic acid
 ◆ L-iduronic acid
 ◇ unsaturated uronic acid