



**HAL**  
open science

## Novel laminarin-binding CBMs in multimodular proteins of marine Bacteroidota feature prominently in phytoplankton blooms

Marie-Katherin Zühlke, Elizabeth Ficko-Blean, Daniel Bartosik, Nicolas Terrapon, Alexandra Jeudy, Murielle Jam, Fengqing Wang, Norma Welsch, Robert Larocque, Diane Jouanneau, et al.

### ► To cite this version:

Marie-Katherin Zühlke, Elizabeth Ficko-Blean, Daniel Bartosik, Nicolas Terrapon, Alexandra Jeudy, et al.. Novel laminarin-binding CBMs in multimodular proteins of marine Bacteroidota feature prominently in phytoplankton blooms. 2023. hal-04260572

**HAL Id: hal-04260572**

**<https://hal.science/hal-04260572v1>**

Preprint submitted on 26 Oct 2023

**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 **Novel laminarin-binding CBMs in multimodular proteins of marine *Bacteroidota***  
2 **feature prominently in phytoplankton blooms**

3 Marie-Katherin Zühlke<sup>1,2\*</sup>, Elizabeth Ficko-Blean<sup>3</sup>, Daniel Bartosik<sup>1,2</sup>, Nicolas Terrapon<sup>4</sup>,  
4 Alexandra Jeudy<sup>3</sup>, Murielle Jam<sup>3</sup>, Fengqing Wang<sup>5</sup>, Norma Welsch<sup>1,2</sup>, Robert Larocque<sup>3</sup>,  
5 Diane Jouanneau<sup>3</sup>, Tom Eisenack<sup>1</sup>, François Thomas<sup>3</sup>, Anke Trautwein-Schult<sup>6</sup>, Hanno  
6 Teeling<sup>5</sup>, Dörte Becher<sup>6</sup>, Thomas Schweder<sup>1,2</sup>, Mirjam Czjzek<sup>3\*</sup>

7

8 <sup>1</sup> Pharmaceutical Biotechnology, Institute of Pharmacy, University Greifswald, Greifswald,  
9 17487, Germany

10 <sup>2</sup> Institute of Marine Biotechnology, 17487 Greifswald, Germany

11 <sup>3</sup> Integrative Biology of Marine Models (LBI2M), Station Biologique de Roscoff (SBR),  
12 Sorbonne Université, CNRS, 29680 Roscoff, Bretagne, France

13 <sup>4</sup> Architecture et Fonction des Macromolécules Biologiques (AFMB), Aix-Marseille Université  
14 (AMU, UMR7257), CNRS, 13288 Marseille cedex 9, Marseille, France

15 <sup>5</sup> Max Planck Institute for Marine Microbiology, Celsiusstraße 1, 28359 Bremen, Germany

16 <sup>6</sup> Microbial Proteomics, Institute of Microbiology, University Greifswald, 17487 Greifswald,  
17 Germany

18

19 \* corresponding authors

20 Mirjam Czjzek, e-mail: [czjzek@sb-roscoff.fr](mailto:czjzek@sb-roscoff.fr)

21 Marie-Katherin Zühlke, e-mail: [marie-katherin.zuehlke@uni-greifswald.de](mailto:marie-katherin.zuehlke@uni-greifswald.de)

22

23

24

25

26 **Key words**

27 carbohydrate-binding modules / glycan scavenging / laminarin / multimodular proteins /  
28 surface glycan-binding proteins

## 29 **Abstract**

30 The  $\beta$ -(1,3)-glucan laminarin functions as storage polysaccharide in marine stramenophiles  
31 such as diatoms. Laminarin is abundant, water-soluble and structured simply, making it an  
32 attractive substrate for marine bacteria. As a consequence, many marine bacteria have  
33 developed competitive strategies to scavenge and decompose laminarin, which involves  
34 carbohydrate-binding modules (CBMs) as key players. We therefore functionally and  
35 structurally characterized two yet unassigned domains as laminarin-binding CBMs in  
36 multimodular proteins from our model bacterium *Christiangramia forsetii* KT0803<sup>T</sup>, hereby  
37 unveiling the novel laminarin-binding CBM families CBMxx and CBMyy (official CAZy  
38 numbering will be provided upon acceptance of the manuscript in a peer-reviewed journal).  
39 We discovered four CBMxx repeats in a surface glycan-binding protein (SGBP) and a single  
40 CBMyy combined with a glycoside hydrolase module from family 16 (GH16\_3). Our analyses  
41 revealed that both modular proteins have an elongated shape, and that the GH16\_3 displayed  
42 a higher flexibility than the SGBP. While motility of both polypeptide chains may facilitate  
43 recognition and/or degradation of laminarin, constraints in the SGBP may support docking of  
44 laminarin onto the bacterial surface. The exploration of bacterial metagenome-assembled  
45 genomes (MAGs) from phytoplankton blooms in the North Sea revealed that both laminarin-  
46 binding CBM families are widely distributed among marine *Bacteroidota*, illustrating the high  
47 adaptability of modularity in sugar-binding and -degrading proteins. High expression of CBMxx-  
48 and CBMyy-containing proteins during phytoplankton blooms further underpins their  
49 importance in marine laminarin usage.

## 50 **Introduction**

51 Laminarin is an important energy and carbon storage compound in marine stramenophiles,  
52 including abundant diatoms. Therefore, huge amounts of laminarin are produced during  
53 common diatom-dominated phytoplankton blooms (1). Laminarin is composed of a  $\beta$ -(1,3)-  
54 linked glucose backbone with occasional  $\beta$ -(1,6)-linked glucose branches that differ in length  
55 and frequency (2, 3). Due to its abundance, water solubility and simple structure, marine

56 bacteria often prefer laminarin as a substrate over less soluble and more complex  
57 polysaccharides (4, 5). Polysaccharide utilization loci (PULs) dedicated to laminarin  
58 breakdown are thus both widely distributed (6, 7) and are highly expressed (4, 8-10) in marine  
59 bacteria that thrive during phytoplankton blooms. PULs are distinct genetic loci that encode  
60 proteins to sense, digest and transport sugars. *Bacteroidota* are key players in marine  
61 laminarin turnover and many glycoside hydrolases (GHs) active on laminarin from members of  
62 this phylum have been characterized, such as GH16 and GH17 family representatives that act  
63 on the  $\beta$ -(1,3)-main chain, or GH30 family representatives that remove  $\beta$ -(1,6)-side chains (11-  
64 14). Some of these GHs are bound to the cell surface, where they degrade laminarin into  
65 smaller-sized oligosaccharides suitable for uptake and transport (11). In *Bacteroidota*, the  
66 oligosaccharide shuttle to the periplasm is mediated by an accessory sugar-binding SusD-like  
67 protein, tethered to the outer membrane, and a SusC-like TonB-dependent transporter (TBDT)  
68 (reviewed in 15, 16). In complex with their target substrate, SusD-like proteins act as a hinge  
69 to relay their cargo into the SusC-like TBDT (17). In addition to SusD-like proteins, bacteria  
70 may have additional surface glycan-binding proteins (SGBPs), which are often encoded  
71 directly downstream of the *susCD*-like gene tandem and are sometimes termed ‘SusE-  
72 positioned’ proteins. The designations ‘-like’ *versus* ‘-positioned’ intend to distinguish between  
73 homologous *versus* analogous proteins, and refer to the starch utilization system (Sus) (18,  
74 19), which marked the beginning of the era of PUL exploration. While SusD-like proteins have  
75 a single binding site and are almost entirely  $\alpha$ -helical, SusE-positioned SGBPs may have  
76 several carbohydrate-binding modules (CBMs), which often display a  $\beta$ -sandwich fold structure  
77 and function as carbohydrate docking sites. Since SusD-like proteins also classify as SGBPs,  
78 they have sometimes been named ‘SGBP-A’ in the literature, whereas CBM-containing SGBPs  
79 are named ‘SGBP-B’ (20). Some examples of mixed-linkage  $\beta$ -glucan and laminarin-related  
80 SusD-like proteins and CBM-containing SGBPs have recently been thoroughly analyzed in gut  
81 strains from the *Bacteroides* genus (21). By comparison, there is only one structure for a  
82 marine laminarin-binding SusD-like protein, which recognizes  $\beta$ -(1,6)-branches in laminarin  
83 (22), and marine examples of CBM-containing SGBPs that bind laminarin are lacking. As

84    aforementioned, unlike SusD-like proteins, CBM-containing SGBPs rarely show detectable  
85    similarity/homology (20, 21) and are thus difficult to predict. Therefore, CBM-containing SGBPs  
86    are largely overlooked and probably underestimated. In addition to SusD-like proteins and  
87    CBM-containing SGBPs, laminarin-binding CBMs are often linked to enzymes within  
88    multidomain structures (23, 24), which further underlines the importance of CBMs in marine  
89    laminarin degradation.

90    The laminarin-consuming *Christiangramia forsetii* KT0803<sup>T</sup> (formerly *Gramella forsetii* (25)), a  
91    member of the *Bacteroidota*, has been isolated during a phytoplankton bloom in the North Sea  
92    (26). A GH16\_3 and a 'conserved hypothetical protein', encoded in the *C. forsetii* laminarin  
93    PUL, were among the most abundantly produced proteins when growing on laminarin as a  
94    sole carbon source (27). Based on subproteome analyses it was suggested that the two  
95    proteins are tethered to the outer membrane and it was speculated that the 'conserved  
96    hypothetical protein', which is SusE-positioned, might function as an SGBP (27).

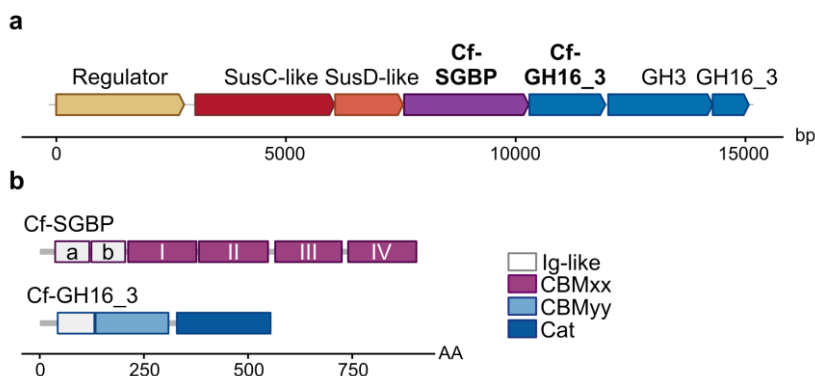
97    This study aimed to illuminate the relevance of laminarin-binding CBMs in the scavenging and  
98    processing of laminarin, which is abundantly produced during phytoplankton blooms. We  
99    therefore structurally and functionally characterized an SGBP and a GH16\_3 from the marine  
100    model bacterium *C. forsetii*, both of which display novel laminarin-binding CBMs in their  
101    multimodular structure. This gave insights into specific adaptations of a non-catalytic and a  
102    catalytic protein, respectively, with respect to their roles in first steps of laminarin utilization. It  
103    further allowed us to illustrate the distribution of these CBMs in marine bacteria and their  
104    expression as a response to algae proliferation.

105

## 106 Results

### 107 *Multidomain proteins Cf-SGBP and Cf-GH16\_3*

108 We investigated two proteins that are encoded in a laminarin PUL and that were highly  
109 produced in laminarin-grown cells of *C. forsetii* (27): a putative CBM-containing SGBP (Cf-  
110 SGBP, locus tag GFO\_RS17395) and a protein containing a GH16\_3 catalytic module (Cf-  
111 GH16\_3, locus tag GFO\_RS16360) (Fig. 1a, further information on domain nomenclature is  
112 provided in the Materials and Methods or Fig. 1b). Subproteome analyses (27) and the  
113 presence of a lipoprotein signal peptide support that both proteins are tethered to the outer  
114 membrane. Structural prediction using Alphafold2 (AF) (28) and Phyre2 (29) uncovered that  
115 the Cf-SGBP displays two Ig-like folded ( $\beta$ -sandwich) modules at its N-terminus followed by  
116 four putative CBMs (I-IV) (Fig. 1b) with high sequence-similarity (47-84%), suggesting internal  
117 duplication events. The Cf-GH16\_3 on the other hand is predicted to feature a single N-  
118 terminal Ig-like folded module followed by a putative CBM, with no significant similarity to CBMs  
119 of the Cf-SGBP, and the C-terminal GH16\_3 catalytic domain at the C-terminus.

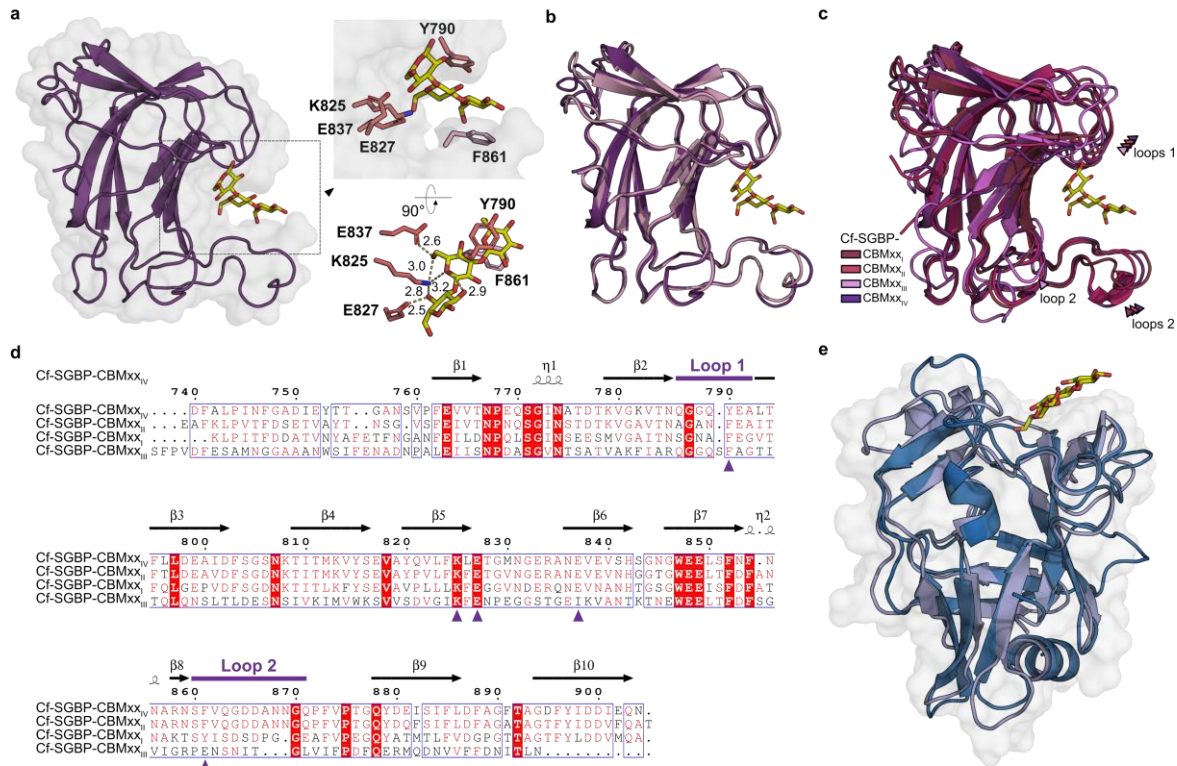


120

121 **Fig. 1 The laminarin PUL of *Christiangramia forsetii* encodes the multimodular Cf-SGBP and Cf-GH16\_3**  
122 **containing the novel CBMxx and CBMyy, respectively.** **a** The *C. forsetii* laminarin PUL encodes multidomain  
123 proteins, including a novel CBM-containing SGBP (Cf-SGBP) and a GH16\_3 (Cf-GH16\_3), highlighted bold. **b**  
124 Domain organization of the Cf-SGBP and Cf-GH16\_3. Here, gray lines represent N-terminal lipoprotein signal  
125 peptides or linkers between modules. The Cf-SGBP displays two Ig-like fold modules (a,b) followed by four CBMxxs  
126 (I-IV). The Cf-GH16\_3 comprises an Ig-like fold module and CBMyy in addition to its catalytic module (Cat).

127 All putative CBM sequences were submitted to Dali searches (30), which did not reveal any  
128 close relatives with solved 3D structures. The four CBMs from the Cf-SGBP were only remotely  
129 similar to a module attached to a GH50 from *Saccharophagus degradans* (PDB ID 4BQ2) (31),  
130 with a z-score >12 and 10-15% identity. To date, this module is not classified in any CBM

131 family in the CAZy classification, in the absence of a characterized close homolog. Similarly,  
132 the CBM of the Cf-GH16\_3 returned uncharacterized proteins, including hypothetical CBMs,  
133 as best hits with z-scores ~8 and 9-12% sequence identity (PDB ID 4L3R and 4QRL).  
134 Sequence analysis confirmed that the putative *C. forsetii* CBMs did not exhibit any domains  
135 belonging to a known CBM family. Hence, the functional characterization of the CBMs of the  
136 Cf-SGBP and Cf-GH16\_3 (cf two following sections below) provided the basis for founding the  
137 novel CAZy (32) CBM families CBMxx and CBMyy, respectively (official CAZy numbering will  
138 be provided upon acceptance of the manuscript in a peer-reviewed journal). We determined  
139 the 3D crystal structure at 2.0 Å resolution for the isolated C-terminal CBMxx, Cf-SGBP-  
140 CBMxx<sub>IV</sub>, in complex with laminaritriose by protein crystallography (Fig. 2a, Table S1). Least  
141 square superimposition of the corresponding AF predicted structure to the 3D crystal structure  
142 using COOT (33) and matching all atoms leads to an RMSD of 1.765. This high agreement  
143 (Fig. 2b) supported reliability for the AF predictions for the other CBMxxs, which further  
144 indicated two things: the four CBMxxs are highly similar in fold and CBMyy differs from CBMxx.  
145 The crystal structure of the Cf-SGBP-CBMxx<sub>IV</sub> reveals a typical CBM β-sandwich fold,  
146 consisting of two antiparallel β-sheets composed of four or five β-strands that form a concave  
147 and convex face, respectively (Fig. 2a). The concave face forms a deep and narrow binding  
148 cleft, also established by two expansive loops. While the concave face contains residues  
149 mediating polar interactions with the substrate, distal loops contain aromatic residues (Fig. 2a)  
150 to establish CH-π interactions (34). AF predicted a similar fold for all four CBMxxs (Fig. 2c)  
151 and residues that mediate interaction with laminaritriose in the Cf-SGBP-CBMxx<sub>IV</sub> align (Fig.  
152 2d). Remarkably, differences are seen in Cf-SGBP-CBMxx<sub>III</sub> loop 2, which results in a wider  
153 and very shallow binding cleft (Fig. 2c and d, Fig. S1).  
154 Compared to the β-sandwich-fold of CBMxx, CBMyy rather displays a β-barrel structure with  
155 two additional α-helices and an additional β-strand pair. Together with a lacking binding  
156 groove, the *C. forsetii* CBMyy is structurally related to a CBM present in a *Bacteroidetes fluxus*  
157 SGBP (Fig. 2e, Fig. S2) (21) devoid of a catalytic GH16\_3. Our analyses reveal that the *B.*  
158 *fluxus* CBM classifies as a CBMyy.



159

160 **Fig. 2 CBMxx repeats of the Cf-SGBP and CBMy of the Cf-GH16\_3: binding clefts versus platform.** **a** 3D  
 161 crystal structure of the Cf-SGBP-CBMxx<sub>IV</sub> with laminaritriose, highlighting interacting residues. **b** Overlay of the  
 162 crystal structure of the Cf-SGBP-CBMxx<sub>IV</sub> (violet) and the corresponding AF prediction (light pink), supporting the  
 163 accuracy of AF. **c** Overlay of AF predicted CBMxx repeats (I-III) with the crystal structure of the Cf-SGBP-CBMxx<sub>IV</sub>.  
 164 In particular Cf-SGBP-CBMxx<sub>III</sub> differs in one of the loops delimiting the binding groove, which results in a flat binding  
 165 region flanked only by one loop (see also Fig. S1). **d** Alignment of the Cf-SGBP-CBMxxs visualized using ESPript  
 166 3.0 (<https://espriti.ibcp.fr>) (35). Identical residues are highlighted in red boxes, similar residues using red letters.  
 167 Residues involved in substrate binding in the Cf-SGBP-CBMxx<sub>IV</sub> are indicated by violet arrow heads; the two loops  
 168 flanking the substrate binding groove are indicated by violet lines above the alignment. **e** CBMy of the Cf-GH16\_3  
 169 displays a  $\beta$ -barrel structure, missing a distinct binding cleft compared to CBMxx. The AF predicted model (blue)  
 170 superimposes very closely to a CBM of an SGBP from *Bacteroidetes fluxus* (21) binding laminarin and mixed-  
 171 linkage  $\beta$ (1,3)/ $\beta$ (1,4)-glucan on a platform (lightblue, here laminaritriose, PDB ID 7KV7, see also Fig. S2).

## 172 *CBMxx and CBMy bind laminarin*

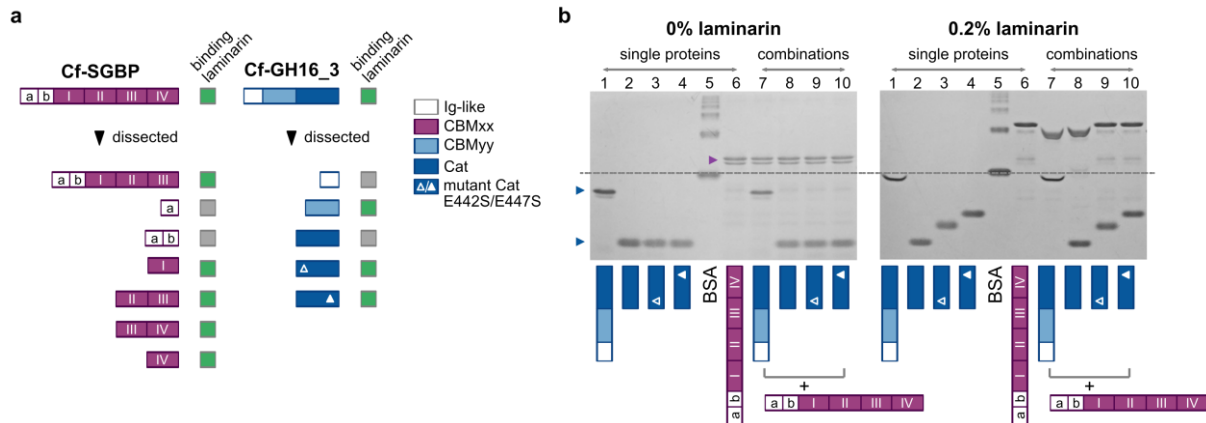
173 Affinity gel electrophoresis (AGE) demonstrated that the Cf-SGBP and Cf-GH16\_3 bind  
 174 laminarin. We narrowed down binding to the CBMs of both proteins, and excluded binding of  
 175 the N-terminal Ig-like folded modules (Fig. 3a, Fig. S3 and S4). Binding to laminarin induced a  
 176 distinct shift of the band for the full-length Cf-SGBP (Fig. 3b, lane 6), whereas it induced a  
 177 small but distinct shift for the full-length Cf-GH16\_3 with an altered, 'mustache'-like band shape  
 178 (Fig. 3b, lane 1). In comparison, the binding of the isolated catalytic domain Cf-GH16\_3-Cat  
 179 could not be inferred from AGE, nor was an altered band shape detected (Fig. 3b, lane 2). We  
 180 hypothesize that the catalytic module releases smaller laminarin oligosaccharides that diffuse



181 from the gel, which is why no shift is obtained for the dissected active catalytic module.  
182 Accordingly, the shift was restored with mutants of the dissected catalytic domain (Cf-GH16\_3-  
183 Cat<sub>E442S</sub> and Cf-GH16\_3-Cat<sub>E447S</sub>), but without the detection of the altered band shape (Fig. 3b,  
184 lanes 3 and 4). Unexpectedly, we found that the GH16\_3-catalytic activity reduced the gel shift  
185 of the Cf-SGBP and resulted in a mustache-like band shape of the Cf-SGBP wherever the  
186 active enzyme was present (Fig. 3b, lanes 7 and 8). This alteration was not obtained for the  
187 Cf-SGBP loaded together with the catalytic mutants (Fig. 3b, lanes 9 and 10). We suggest that  
188 in the presence of active enzyme, the CBM-containing polypeptides (Fig. 3b, lanes 1, 7 and 8)  
189 bind to residual laminarin in-between the gel lanes, thereby modifying the shape of the bands  
190 and resulting in the 'mustache' effect.

191 In addition to laminarin binding, AGE also demonstrated binding of the mixed-linkage  
192  $\beta(1,3)/\beta(1,4)$ -glucan (MLG) lichenan by Cf-GH16\_3-CBMyy (Fig. S5). Activity of the Cf-  
193 GH16\_3 on lichenan was not detectable by reducing sugar assay (RSA); however,  
194 fluorophore-assisted carbohydrate electrophoresis (FACE) analyses demonstrated weak  
195 activity (Fig. S6).

196



197

198 **Fig. 3 Identification of laminarin-binding domains in the Cf-SGBP and Cf-GH16\_3.** **a** Summary of AGE runs,  
 199 where green boxes indicate binding, which was concluded from corresponding gels in Fig. 3b, Fig. S3 and S4. **b**  
 200 AGE of the Cf-GH16\_3 revealed catalytic activity during runs. The Cf-GH16\_3, the dissected catalytic module Cf-  
 201 GH16\_3-Cat and two mutants thereof (Cf-GH16\_3-Cat<sub>E442S</sub> and Cf-GH16\_3-Cat<sub>E447S</sub>) were loaded individually  
 202 (lanes 1-4) or in combination with the non-catalytic Cf-SGBP (lanes 7-10), compared to the Cf-SGBP alone (lane  
 203 6). While a shift in migration in laminarin gels indicated binding (using BSA as a control, lane 5), an altered  
 204 'mustache'-like band shape of CBM-containing proteins in laminarin gels (lane 1, 7 and 8) revealed the presence  
 205 of GH16\_3 activity within the lane. Detailed information is provided in the main text.

206 *Multiple CBMxx domains in the Cf-SGBP: is 'more' really 'better'?*

207 While the *C. forsetii* Cf-GH16\_3 contains a single CBMyy to bind laminarin, the Cf-SGBP  
 208 contains four consecutive CBMxxs. AGE did not provide indications on whether the higher  
 209 number of CBMs in the Cf-SGBP translates into a higher binding efficiency (Fig. S3). In order  
 210 to explore putative benefits of multiple CBMxx domains in more detail, we investigated  
 211 laminarin binding by isothermal titration calorimetry (ITC). We selected the full-length Cf-  
 212 SGBP, the tandems Cf-SGBP-CBMxx<sub>II/III</sub> and Cf-SGBP-CBMxx<sub>III/IV</sub>, as well as the single  
 213 module Cf-SGBP-CBMxx<sub>IV</sub>, for which we had a 3D crystal structure, to cover a range of CBMxx  
 214 repetitions for comparisons. Other constructs were either insufficiently stable or insufficiently  
 215 produced and thus unfortunately not available for ITC measurements.

216 ITC confirmed laminarin-binding for the selected proteins (Table 1, Fig. S7a). Unexpectedly,  
 217 the highest affinity for laminarin was determined for the individual Cf-SGBP-CBMxx<sub>IV</sub> ( $1.04 \times$   
 218  $10^5 \text{ M}^{-1}$ ), which exceeded affinity of the other constructs by an order of magnitude (Table 1)  
 219 and is in the same range as described for the N-terminal CBM6 of a GH16\_3 (LamC) in *Zobellia*  
 220 *galactanivorans* Dsij<sup>T</sup> (23). A binding stoichiometry (n) of close to 0.5 indicates two CBMxx<sub>IV</sub>s  
 221 are able to interact with one ligand molecule. The binding stoichiometry is increased to ~0.75

222 for the double module constructs and the full-length Cf-SGBP binds 1.6 laminarin molecules  
 223 per polypeptide chain, though no significantly increased affinity was seen. Instead, affinities of  
 224 the full-length Cf-SGBP and the two CBMxx tandems were relatively similar (Cf-SGBP  $3.09 \times 10^4 \text{ M}^{-1}$   
 225  $10^4 \text{ M}^{-1}$ , Cf-SGBP-CBMxx<sub>II/III</sub>  $2.76 \times 10^4 \text{ M}^{-1}$  and Cf-SGBP-CBMxx<sub>III/IV</sub>  $3.63 \times 10^4 \text{ M}^{-1}$ ). We  
 226 speculate that the increased affinity of the individual CBMxx<sub>IV</sub> for laminarin relative to the  
 227 tandems is provided by the absence of steric/spatial inhibition of binding from the other CBMxx.  
 228 Thus, in the absence of the other modules, two individual CBMxx<sub>IV</sub> are more optimally  
 229 accommodated on the same laminarin chain. Conversely, the steric hindrance of successive  
 230 CBMs in the Cf-SGBP may contribute to partial binding site inaccessibility. Furthermore, the  
 231 longer constructs pay more entropic penalties than the single Cf-SGBP-CBMxx<sub>IV</sub>, probably  
 232 related to the loss of conformational freedom upon ligand binding. Notably, the Cf-SGBP-  
 233 CBMxx<sub>IV</sub> also bound laminarin-derived oligosaccharides DP7 with a higher affinity than DP5,  
 234 although with a lower affinity than laminarin (Table 1, Fig. S7b). This is coherent with increased  
 235 affinity of the individual Cf-SGBP-CBMxx<sub>IV</sub> on laminarin, as it suggests that this CBM can  
 236 accommodate more than DP7 in the ligand binding site. For all interactions, binding was  
 237 enthalpy-driven, with an entropic penalty (Table 1). Confirming our results from AGE, binding  
 238 of laminarin by the Cf-SGBP-Ig<sub>a</sub> was not detected.

239 **Table 1. Affinity of laminarin (Lam) and its oligosaccharides (DP7 and DP5) to the Cf-SGBP and/or dissected**  
 240 **modules thereof.** Mean values and standard deviation ( $\pm$ ) of ITC experiments ( $n \geq 3$ ). Measurements were  
 241 performed at 293.15 K. For each run,  $\Delta G$  was calculated (Gibbs-Helmholtz equation,  $\Delta G = \Delta H - T\Delta S$ ).

<b>Cf-SGBP and selected modules - laminarin</b>					
	$K_a$ [ $\text{M}^{-1}$ ]	$\Delta H$ [kcal mol <sup>-1</sup> ]	$\Delta S$ [cal mol <sup>-1</sup> K <sup>-1</sup> ]	N [sites]	$\Delta G$ [kcal mol <sup>-1</sup> ]
Cf-SGBP	$3.09 (\pm 0.14) \times 10^4$	$-17.3 \pm 0.11$	$-38.53 \pm 0.46$	$1.60 \pm 0.05$	$-6.02 \pm 0.03$
Cf-SGBP-CBMxx <sub>II/III</sub>	$2.76 (\pm 0.22) \times 10^4$	$-19.2 \pm 1.11$	$-45.17 \pm 3.95$	$0.73 \pm 0.03$	$-5.95 \pm 0.05$
Cf-SGBP-CBMxx <sub>III/IV</sub>	$3.63 (\pm 0.44) \times 10^4$	$-18.6 \pm 1.12$	$-42.45 \pm 4.04$	$0.76 \pm 0.04$	$-6.11 \pm 0.06$
Cf-SGBP-CBMxx <sub>IV</sub>	$1.04 (\pm 0.03) \times 10^5$	$-14.0 \pm 0.12$	$-24.78 \pm 0.46$	$0.49 \pm 0.01$	$-6.73 \pm 0.03$
<b>Cf-SGBP-CBMxx<sub>IV</sub> - laminarin-derived oligosaccharides (DP7 and DP5)</b>					
	$K_a$ [ $\text{M}^{-1}$ ]	$\Delta H$ [kcal mol <sup>-1</sup> ]	$\Delta S$ [cal mol <sup>-1</sup> K <sup>-1</sup> ]	N [sites]	$\Delta G$ [kcal mol <sup>-1</sup> ]
Cf-SGBP-CBMxx <sub>IV</sub> DP7	$1.87 (\pm 0.17) \times 10^4$	$-7.99 \pm 0.02$	$-7.72 \pm 0.19$	$0.66 \pm 0.02$	$-5.73 \pm 0.05$
Cf-SGBP-CBMxx <sub>IV</sub> DP5	$1.39 (\pm 0.07) \times 10^4$	$-9.14 \pm 0.18$	$-12.20 \pm 0.70$	$0.61 \pm 0.04$	$-5.56 \pm 0.04$

242

243

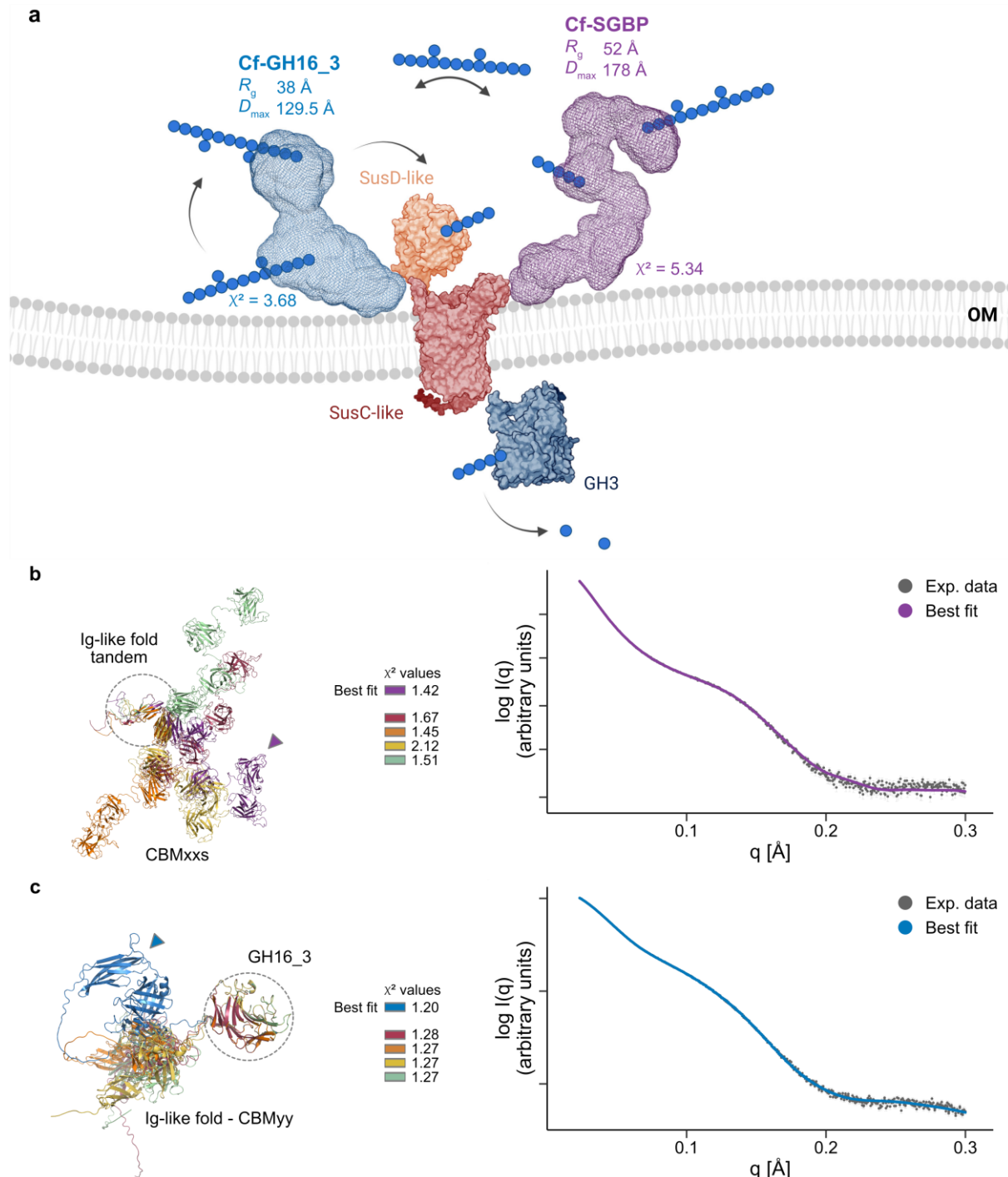
244 *Two elongated multidomain proteins allow for laminarin recognition and scavenging*

245 The multidomain structure of the *C. forsetii* proteins and corresponding ITC experiments for  
246 the Cf-SGBP constructs raised the question about the spatial organization of their polypeptide  
247 chains. AF predictions were highly uncertain for the linkers connecting the individual modules.  
248 Therefore, we analyzed the Cf-SGBP and the Cf-GH16\_3 by small-angle X-ray scattering  
249 (SAXS), which revealed that both multidomain proteins feature an elongated structural  
250 organization (Fig. 4a, Fig. S8, Table S2).

251 The values for the radius of gyration ( $R_g$ ) and the maximum particle distance ( $D_{max}$ ) derived  
252 from the scattering curve are  $R_g$  52 Å and  $D_{max}$  178 Å for the Cf-SGBP and  $R_g$  38 Å and  $D_{max}$   
253 129.5 Å for the Cf-GH16\_3 (Table S2, Fig. S8). The elongated structures were confirmed by  
254 the GASBORi (36) *ab initio* calculated envelopes (Fig. 4a, Fig. S9). All calculations for the Cf-  
255 GH16\_3 returned a bent shape, while much more variable shapes were calculated for the Cf-  
256 SGBP (Fig. S9). However, the relatively high  $\chi^2$  values (all >3) of calculated individual shapes  
257 for both proteins potentially indicated that the experimental curves were not well fitted by a  
258 single envelope, which implicated possible conformational variability. Nevertheless, and using  
259 DADIMODO, spatial arrangements of the successive domains based on those predicted by  
260 AF could be determined for both the Cf-SGBP and Cf-GH16\_3 that individually fitted the  
261 experimental curve (Fig. 4b and c, right panels). Superimposition of the individual models,  
262 however, covered variable arrangements (Fig. 4b and c, left panels). To further analyze the  
263 potential conformational flexibility of these proteins in solution, we used the ensemble  
264 optimization method (EOM) (38, 39). Here, an ensemble is selected from a large pool of  
265 theoretical conformers, for each of which a corresponding theoretical scattering curve is  
266 generated. The selected ensembles for both proteins were able to fit the experimental data,  
267 but in contrast to the Cf-GH16\_3, the  $\chi^2$  value of the EOM fit of the Cf-SGBP (10.85 *versus*  
268 4.57 for the Cf-GH16\_3, Fig. S10) was much higher than  $\chi^2$  values of the GASBORi (Fig. 4a,  
269 Fig. S9) or DADIMODO single fits (Fig. 4b and c), indicative of lower conformational flexibility.  
270 In addition, the calculated  $R_{flex}$  for the selected ensemble of the Cf-SGBP (77%) was lower  
271 compared to that of the Cf-GH16\_3 (81%), although the  $R_{flex}$  of the pool and the final  $R_{Sigma}$

272 were almost the same (Table 2).  $R_{flex}$  and  $R_{Sigma}$  are values to describe the flexibility in the  
273 EOM approach, where 100% or 1 correspond to a fully flexible system, respectively. Finally, a  
274 higher flexibility of the Cf-GH16\_3 was supported by shifts to higher  $R_g$  and  $D_{max}$  values for the  
275 selected ensemble compared to the pool, whereas they were biased to lower values for the  
276 Cf-SGBP (Fig. S11).

277



278

279 **Fig. 4 Two elongated multidomain proteins to sense and capture laminarin.** **a**  $R_g$  and  $D_{max}$  values calculated  
 280 from the experimental SAXS data revealed an elongated structure for the Cf-SGBP and the Cf-GH16\_3, which is  
 281 also reflected in two calculated models for protein envelopes using GASBORi. The figure was created with  
 282 BioRender using an adapted tentative model of laminarin utilization from (27). The spatial organization of domains  
 283 in the **b** Cf-SGBP and the **c** Cf-GH16\_3 was investigated using DADIMODO. Shown are five models (left graphs)  
 284 in which domains used for superimposition are circled. The best fit (highest  $\chi^2$ ) is highlighted with an arrow head  
 285 and was plotted against the experimental data (right panels). Further information on domains is provided in Fig. 1b.

286

287 **Table 2 Summary of the EOM-analyses to determine the flexibility of the Cf-SGBP and the Cf-GH16\_3 in**  
 288 **solution.** Given are the final  $R_g$  and  $D_{max}$  derived from the unique models for the selected ensembles, averaged  $R_g$   
 289 and  $D_{max}$  of the selected ensembles and the pool as well as  $R_{flex}$  and  $R_{sigma}$  to describe the flexibility.  $R_{flex}$  of 100%  
 290 or  $R_{sigma}$  of 1 corresponds to a fully flexible system.

	$R_g$ [Å] final	$D_{max}$ [Å] final	$R_g$ [Å] ensemble	$D_{max}$ [Å] ensemble	$R_g$ [Å] pool	$D_{max}$ [Å] pool	$R_{flex}$ ensemble/pool	$R_{sigma}$
Cf-SGBP	52.6	175.2	53.8	175.7	58.0	189.3	77%/ 86%	0.93
Cf-GH16_3	37.7	127.6	37.6	126.6	35.6	118.7	81%/ 85%	0.94

291  
 292 *CBMxx and CBMyy are widely distributed in Bacteroidota that thrive during phytoplankton*  
 293 *blooms*

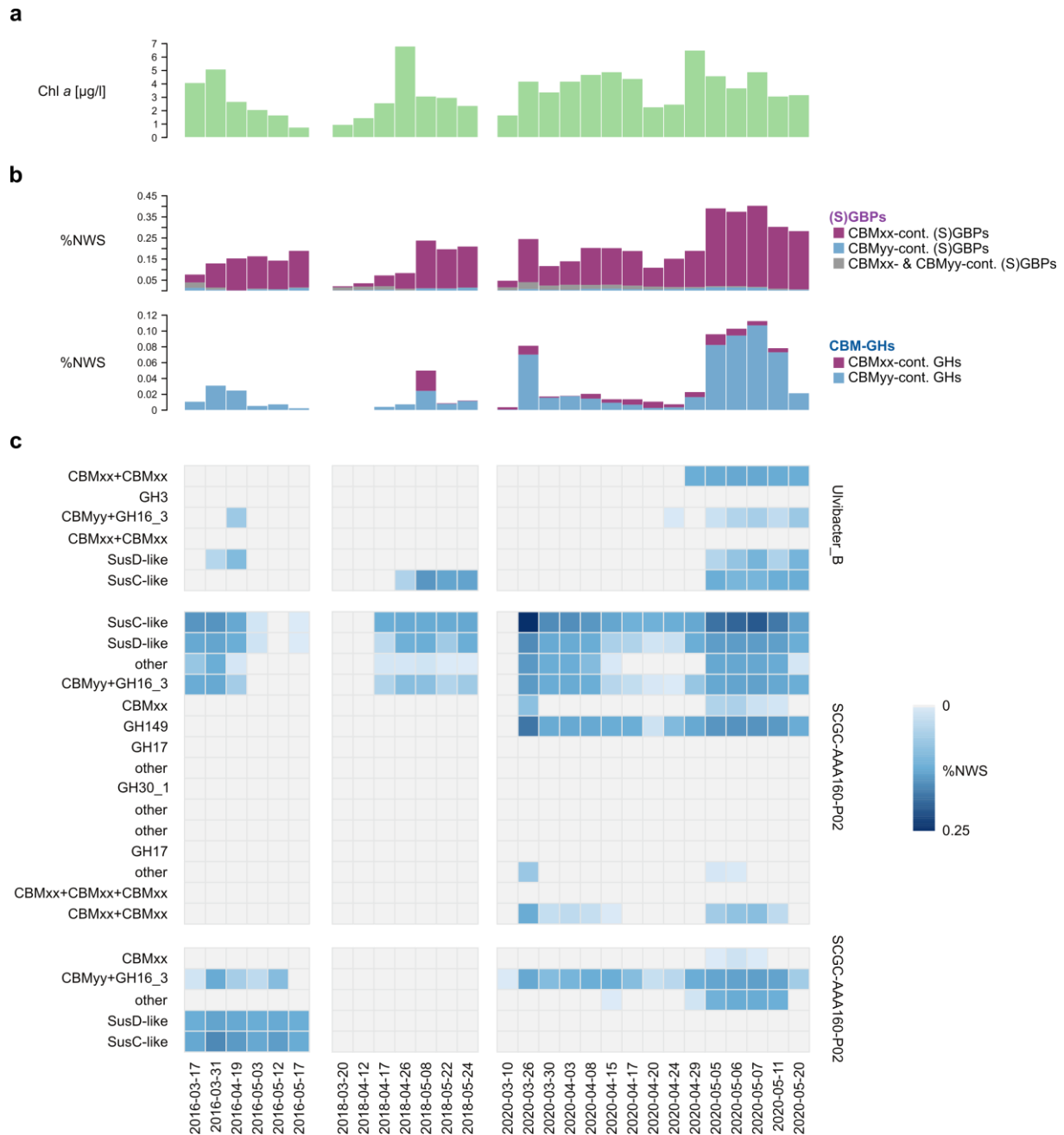
294 Screening of bacterial metagenome-assembled genomes (MAGs) obtained from North Sea  
 295 phytoplankton bloom-associated microbial biomass samples at Helgoland Roads in 2016,  
 296 2018 and 2020, yielded 427 sequences coding for CBMxx- and CBMyy-containing protein  
 297 sequences (555 representative MAGs of which 201 belonged to *Bacteroidota*). The proportion  
 298 of CBMxx-containing proteins was higher (362) compared to CBMyy (82), whereas only some  
 299 sequences contained CBMxx and CBMyy (17) (Fig. 5). We found a total of 22 domain  
 300 combinations, in which CBMxx and/or CBMyy were detected in protein-coding sequences with  
 301 or without known catalytic modules, e.g., linked to CBM4, CBM6, GH16\_3 or rare examples of  
 302 GH2, GH5\_34, GH5\_46, GH81 and GH162. The most frequent organizations included CBMxx-  
 303 or CBMyy-only-proteins (72% for CBMxx with up to eight repeats; 52% for CBMyy, always  
 304 alone) followed by CBMxx or CBMyy associated with a GH16\_3 (20% for CBMxx, 26% for  
 305 CBMyy). In addition, corresponding sequences were frequently co-localized with genes coding  
 306 for known laminarin-active enzymes, e.g., GH16\_3, GH17 or GH30 family members, but also  
 307 with *susCD*-like genes. Similar pictures of domain organization and genomic context for both  
 308 CBMs emerged from screening draft genomes of 53 published North Sea *Bacteroidota* (6),  
 309 including *C. forsetii*, as well as of the CAZy PULDB for both CBMs (as of May 2023). A total of  
 310 92 CBMxx- and/or CBMyy-containing sequences were detected in 38 of the strains from the  
 311 North Sea (Fig. S12) and >400 over ~2,000 *Bacteroidota* genomes with additional relevant  
 312 partners such as GH3, GH149 and GH5\_46 using CAZy, strengthening this linkage. In the  
 313 bloom data set, CBMxx was predominant in *Bacteroidota*, but also occurred in *Proteobacteria*  
 314 or as rare exceptions in *Actinobacteriota* and *Myxococcota*, whereas CBMyy was exclusive to





330 *Protein expression correlated with the course of the bloom*

331 Metaproteome analyses showed that the overall expression of CBMxx and CBMyy-containing  
332 proteins correlated with bloom intensities (Fig. 6a and b), as assessed by chlorophyll *a*  
333 measurements (40). CBM-containing (S)GBPs were more prevalent than CBM-containing  
334 enzymes (mainly GH16\_3s), possibly due to the higher number of putative (S)GBP sequences.  
335 In addition, production of CBM-containing (S)GBPs remained more or less stable after  
336 chlorophyll *a* peaks. Looking at the expression of specific laminarin PULs derived from MAGs,  
337 CBMxx-containing (S)GBPs, but also CBMyy-containing GH16\_3s were detected with a high  
338 coverage, while other PUL-encoded GHs were often not captured (e.g., GH3, GH17 or  
339 GH30\_1). Moreover, CBMxx- and CBMyy-containing proteins were highly abundant in  
340 metaproteomes, sometimes comparable to the expression of SusC/D-like 'marker' proteins  
341 (Fig. 6c).



342

343 **Fig. 6 Expression of CBMxx- and CBMyy-containing proteins correlates with the course of phytoplankton**  
 344 **blooms.** Data show chlorophyll a measurements (40) and results of semi-quantitative metaproteome analyses (0.2  
 345  $\mu\text{m}$  bacterial fractions) from three respective years, see Material and Methods. **a** Course of the bloom illustrated  
 346 using the chlorophyll a concentration. **b** Summed protein abundance values (%NWS, normalized weighted spectra)  
 347 for CBMxx- or CBMyy-containing proteins. The upper graph depicts putative (S)GBPs, containing CBMs only, and  
 348 the lower graph depicts CBM-containing catalytic proteins. Please note different scales (%NWS) for both graphs. **c**  
 349 Abundance of proteins encoded by three selected laminarin PULs. Gray tiles correspond to proteins that were not  
 350 quantified.

351

## 352 Discussion

353 Elevated photosynthetic primary production during marine phytoplankton blooms lead to  
354 pronounced increases in available substrates and nutrients to marine bacteria. These include  
355 a plethora of polysaccharides where laminarin is particularly abundant (1). The surface of many  
356 bloom-associated bacteria is therefore equipped with specific sugar-binding and -degrading  
357 proteins to ensure the recognition and docking of target substrates to the cellular surface, to  
358 allow first steps of enzymatic catalysis, but also to prevent loss of small-sized oligosaccharides  
359 that need to be shuttled into the cell. The successful and specific binding of poly- and  
360 oligosaccharides is therefore a key process in bacterial polysaccharide degradation that can  
361 only be understood by detailed analyses of the CBMs involved, as exemplified in this study by  
362 the two novel laminarin-binding families CBMxx and CBMyy.

### 363 *CBMxx and CBMyy bind laminarin, but belong to different 3D folds*

364 CBMxx and CBMyy both bind to laminarin, but differ substantially in sequence and structure.  
365 All four CBMxxs of the Cf-SGBP exhibit a  $\beta$ -sandwich fold that provides a distinct binding cleft.  
366 At the same time, differences in the surrounding loops result in variation of the wideness and  
367 depth of the cleft. While CBMxx<sub>I/II/IV</sub> display a deep and narrow cleft, it is much shallower in  
368 CBMxx<sub>III</sub>. This is consistent with observations in other multidomain proteins with multiple CBMs  
369 of the same family, e.g., the CBM22 tandem in a xylanase from *Paenibacillus barcinonensis*  
370 (41). Here, a wider groove in one of the CBM22 units was suspected to allow for binding of  
371 larger decorations. Screening of MAGs of bloom-associated bacteria for CBMxx sequences  
372 revealed similar structural variations that may affect either binding specificity or affinity. In  
373 contrast to the common  $\beta$ -sandwich fold of the CBMxx, the  $\beta$ -barrel fold of the CBMyy is rather  
374 rare in CBMs. The latter has recently been described in a CBM-containing SGBP from *B.*  
375 *fluxus*, which can now be classified as a CBMyy. Interestingly, in addition to laminarin, the Cf-  
376 GH16\_3-CBMyy was also able to bind MLG lichenan, as previously observed for the *B. fluxus*  
377 protein (21). Moreover, the Cf-GH16\_3-CBMyy lacks a distinct cleft. In the case of the *B. fluxus*  
378 protein, binding of  $\beta$ -glucan termini occurred on the top of the  $\beta$ -barrel, in a very shallow cleft

379 or platform (21), which we likewise suggest for the *C. forsetii* protein (Fig. S2). This binding  
380 site, distally positioned within the loop regions connecting the successive  $\beta$ -strands of the  
381 barrel core, is to some extent comparable to that of some CBM6 (42) or CBM32 (43) members,  
382 and of several lectin classes (<https://unilectin.unige.ch/unilectin3D/>). In summary, CBMxx  
383 classifies as a type B CBM, probably with more or less possibility to accommodate branches,  
384 while CBMyy classifies as a type C CBM (44).

385 *Shape and flexibility of CBMxx- and CBMyy-containing proteins facilitate recognition and*  
386 *processing of laminarin*

387 The laminarin-binding CBMxx and CBMyy are arranged in elongated shapes of multimodular  
388 proteins. In the case of the Cf-SGBP, with its four CBMxxs, the protein may tower out other  
389 outer membrane proteins, literally reaching out for its target substrate (Figure 4a). Elongated  
390 shapes of multimodular CBM-containing SGBPs have been determined before, e.g., for an  
391 SGBP in a heparin/heparan sulfate PUL from *B. thetaiotaomicron* (45). Like the Cf-SGBP  
392 (SAXS:  $R_g$  52 Å and  $D_{max}$  178 Å), the *B. thetaiotaomicron* protein consisted of six distinct  
393 domains and SAXS determined similar spatial expansions ( $R_g$  of 44.3 Å and  $D_{max}$  of 150 Å). In  
394 addition to the elongated shapes, analyses revealed a bent shape for CBMyy-containing Cf-  
395 GH16\_3 ( $R_g$  38 Å and  $D_{max}$  129.5 Å), seen before in another multidomain laminarinase from  
396 *Thermotoga petrophila* (CBM-GH16-CBM,  $R_g$  40 Å and  $D_{max}$  130 Å) (46). Our results also  
397 indicated that both proteins are flexible to some extent, albeit with a higher flexibility for the Cf-  
398 GH16\_3. The increased flexibility may be due to missing prolines in the Cf-GH16\_3 linkers, as  
399 central prolines are suggested to hamper flexibility (45, 47). Such prolines are present in linkers  
400 between CBMxxs in the Cf-SGBP. It was speculated that this rigidity enables distance to the  
401 membrane and thus an improved exposure of the sugar-binding site (47). This might be further  
402 supported by Ig-like fold domains that could act as spacers to the outer membrane (21). Such  
403 a projection of SGBPs into the extracellular space has recently been shown for a fructan-  
404 specific SGBP from *B. thetaiotaomicron* in a complex with a GH32 and SusC/D-like proteins,  
405 called the 'utilisome' (48). Compared to the fructan-specific SGBP, which contains one CBM,

406 the laminarin-specific Cf-SGBP from *C. forsetii* provides four CBM docking sites. However, our  
407 analyses indicated binding of two laminarin chains at maximum, possibly due to steric inhibition  
408 of binding due to the other CBMs. Nevertheless, equipping the cellular surface with SGBPs  
409 that contain multiple CBMs increases binding opportunities in 3D space and might optimize  
410 the presence of proteins on the bacterial surface as a competitive advantage compared to  
411 single CBM-containing SGBPs. Moreover, multiple binding sites in SusE and SusF are known  
412 to compensate for decreased binding due to polysaccharide capsules in gut bacteria (49). This  
413 may be transferable to the marine counterpart, which is also known to produce  
414 exopolysaccharides (50). In the fructan utilisome the sugar chain was held by an accessory  
415 tethering side of the GH32 and by the SGBP. The GH32 catalytic activity then cuts between  
416 both held ends, which was suggested to facilitate degradation (48). In general, the  
417 advantageous effect of CBM-enzyme co-occurrence has largely been demonstrated (e.g., 51,  
418 52). CBMs may also prevent the loss of released degradation products by serving as a short-  
419 term cache for cleaved sugars, which are then further digested or translocated to the SusC/D-  
420 like transporters (Fig. 4a). At the same time, the affinity of laminarin-derived oligosaccharides  
421 to the Cf-SGBP-CBM<sub>xxIV</sub> was lower, as also observed for a CBM6 from *Z. galactanivorans*  
422 Dsij<sup>T</sup> (23), which supports its major role in capturing laminarin.

#### 423 *Vivid exchange and combination of domains to increase fitness*

424 Our data from phytoplankton bloom-associated bacteria show that the *C. forsetii* proteins are  
425 representatives of diverse domain organizations of both CBMs in multimodular catalytic  
426 proteins, mostly GH16\_3s, and putative non-catalytic proteins of which a proportion might  
427 function as SGBPs. This suggests a rather dynamic intermixing of CBMs by translocations,  
428 duplications or losses, which causes also turnovers between catalytic and non-catalytic  
429 proteins. Structural variability of CBMs further contributes to this diversity, which may  
430 determine substrate specificity and/or affinity. This spectrum of domain combinations and  
431 protein structures in sugar-catching and processing proteins likely increases competitiveness  
432 to also encounter structurally similar substrates, such as different laminarins or other  $\beta$ -

433 glucans. The environmental relevance of these novel laminarin-binding CBM families CBMxx  
434 and CBMyy was further reflected by a good coverage of corresponding proteins during  
435 phytoplankton blooms, which was so far mostly observed for SusC/D-like proteins and  
436 GH16\_3s associated to laminarin degradation (4, 8). Protein expression of laminarin-binding  
437 CBMxx- and CBMyy-containing proteins correlated with the overall intensity of the investigated  
438 blooms, supporting their role in marine laminarin use.

439 In summary, marine bacteria employ CBM-containing enzymes and SGBPs on their cellular  
440 surface to scavenge laminarin from the water column. This is facilitated by elongated shapes  
441 of these multimodular proteins, where a higher flexibility of the polypeptide chain may increase  
442 catalytic activity. In comparison, multiple CBMs and the limited flexibility of SGBPs may allow  
443 the accumulation of laminarin, making SGBPs optimal sugar-trapping antenna to further  
444 enhance catalysis. Together with the wide distribution and high expression of CBMxx- and  
445 CBMyy-containing proteins during phytoplankton blooms, our findings corroborate the key role  
446 of the two novel CBMs in laminarin utilization.

## 447 **Materials and Methods**

### 448 *Carbohydrates*

449 Laminarin (*Laminaria digitata*) was acquired from Sigma-Aldrich (Merck kGaA, Darmstadt,  
450 Germany) and MLG lichenan (Icelandic Moss) from Megazyme International Ltd. (Bray,  
451 Ireland). Laminarin-derived oligosaccharides were produced following the protocol from Jam  
452 *et al.* (23) using a GH16\_3 (LamA) from *Zobellia galactanivorans* Dsij<sup>T</sup> (13).

### 453 *Nomenclature*

454 We refer to the full-length surface glycan-binding protein, which is SusE-positioned in the  
455 laminarin PUL of *Christiangramia forsetii* KT0803<sup>T</sup>, as Cf-SGBP (locus tag GFO\_RS17395).  
456 The PUL further encodes two GH16\_3s. Subject of this study was the GH16\_3 encoded  
457 downstream of the Cf-SGBP, which we refer to as Cf-GH16\_3 (locus tag GFO\_RS16360). The  
458 two proteins display a multidomain structure and specific domains were further specified, e.g.,

459 Cf-GH16\_3-Cat refers to the catalytic module of the Cf-GH16\_3. In the case of repetitive  
460 domains of the Cf-SGBP, these modules were enumerated either alphabetically for Ig-like fold  
461 domains or with Roman numerals for the CBMxxs, starting from the N-terminus and indicated  
462 as subscripts, e.g., Cf-SGBP-Ig<sub>a</sub> refers to the first of two N-terminal Ig-like fold domains of the  
463 Cf-SGBP. Whenever we refer to an SGBP in the text, we refer to CBM-containing SGBPs,  
464 excluding SusD-like proteins.

#### 465 *Cloning and site-directed mutagenesis*

466 We used Alphafold2 (28) and Phyre2 (29) to spatially define individual modules of the two  
467 multidomain proteins. LipoP (53) detected signal peptides and MeDor (54) predicted protein  
468 disorder by hydrophobic cluster analysis. In combination, these tools served to define suitable  
469 cloning sites to express full-length proteins, individual domains or combinations thereof.  
470 Corresponding genes were amplified from *C. forsetii* genomic DNA by PCR using primers listed  
471 in Table S3. The PCR products were cloned into expression vectors pRF3 (ampicillin  
472 resistance) or pET28a (kanamycin resistance) by restriction-ligation to produce recombinant  
473 N-terminal hexa-histidine tagged proteins. The vector pRF3 is a derivative of pFO4 (55). The  
474 full-length *Cf-SGBP* and *Cf-GH16\_3* were cloned into pRF3. PCR-products were digested  
475 using BamHI and AvrII, and the pRF3 vector using BamHI and NheI. In all other cases, PCR  
476 products and the pET28a vector were digested using NdeI and SacI. Digested PCR-products  
477 were ligated into expression vectors using T4 DNA ligase. Corresponding plasmids for Cf-  
478 SGBP-Ig<sub>a/b</sub>, Cf-SGBP-CBMxx<sub>i</sub>, Cf-GH16\_3-Ig and Cf-GH16\_3-CBMyy in pET28a, including  
479 NdeI and SacI restriction sites, were designed and ordered from GenScript (GenScript Biotech,  
480 Netherlands).

481 In addition, we generated two mutants of the Cf-GH16\_3-Cat, where each catalytic glutamic  
482 acid was replaced by a serine, Cf-GH16\_3-Cat<sub>E442S</sub> and Cf-GH16\_3-Cat<sub>E447S</sub>. This was  
483 achieved in two steps: first, two separate reactions of PCR, using the Cf-GH16\_Cat-encoding  
484 plasmid as a template, were carried out to produce the mutant gene as two overlapping  
485 fragments (F1 and F2). The purified PCR products F1 and F2 were then used as a template

486 for a final PCR to produce the fused product. To produce the overlapping fragments, we used  
487 a forward primer for the first fragment (F1), which annealed to the vector backbone directly  
488 upstream of *Cf-GH16\_3-Cat*, and a reverse primer for the second fragment (F2), which  
489 annealed to the vector backbone downstream of *Cf-GH16\_3-Cat* (designated as vector  
490 primers, Table S3). These primers were combined with a forward (F2) or reverse (F1) primer  
491 that contained the corresponding desired mutation and that annealed to the intended region of  
492 the mutation (Table S3). The purified PCR products F1 and F2 were digested with DpnI to  
493 remove any residual parental DNA before a final PCR fused F1 and F2 using the vector  
494 primers. PCR products were digested with NdeI and SacI and cloned into pET28a as described  
495 before.

496 Sequence identity for all constructs and successful mutation was verified by sequencing  
497 (Eurofins Genomics, Ebersberg, Germany). Plasmids were transformed into propagation  
498 strains *Escherichia coli* DH10B or DH5 $\alpha$  before being transformed into the expression strain  
499 *E. coli* BL21(DE3).

#### 500 *Protein overexpression and purification*

501 The Cf-SGBP and the Cf-GH16\_3 were produced in ZYP-5052 auto-induction medium (56),  
502 supplemented with 100  $\mu\text{g ml}^{-1}$  ampicillin, for 72 h at 20 °C and 130 rpm. For the production of  
503 the other proteins, clones were cultured in LB, supplemented with 30  $\mu\text{g ml}^{-1}$  kanamycin, at  
504 37°C and 130 rpm until IPTG induction (1 mM) and then at 20 °C and 130 rpm overnight. Cells  
505 were harvested by centrifugation and pellets were stored at -20 °C until protein purification.  
506 Cells were disrupted by chemical lysis. In brief, the cell pellet was suspended in resuspension  
507 buffer (50 mM MOPS pH 8.0, 25% sucrose, lysozyme) and agitated for 15 min at 4°C. Then,  
508 double the volume of lysis buffer (20 mM MOPS pH 7.5, 100 mM NaCl, 1% sodium DCA, 1%  
509 Triton-X) was added, followed by another incubation of 5 min at 4°C. MgCl<sub>2</sub> concentration was  
510 adjusted to 5 mM and samples were incubated with DNase at room temperature until viscosity  
511 of the sample decreased. Cell debris was removed from protein extract by centrifugation.  
512 Protein extract was loaded onto Histrap HP columns (Cytiva, Vélizy-Villacoublay, France)



513 equilibrated in buffer A (200 mM NaCl, 10 mM MOPS pH 7.8, 20 mM imidazole) and charged  
514 with 0.1 M NiSO<sub>4</sub>. In general, 1 ml columns were used for a culture volume of 250 ml and 5 ml  
515 columns for 1 l of culture. Proteins were eluted with a linear gradient of 0-100% buffer B (200  
516 mM NaCl, 10 mM MOPS pH 7.8, 500 mM imidazole) within 30 min (1 ml column) or 60 min (5  
517 ml column) at a flow rate of 1 ml min<sup>-1</sup>. For SAXS analyses, crystallography or in case of  
518 impurities, the quality of the protein sample was further increased by size exclusion  
519 chromatography (SEC) using a Superdex 200 16/60 column (Cytiva) in buffer C (200 mM NaCl,  
520 10 mM MOPS pH 7.8). Otherwise, samples were desalted on a Hiprep 26/10 desalting column  
521 (Cytiva) in buffer C. Buffers used to purify the proteins for SAXS analyses contained 100 mM  
522 NaCl. Samples were evaluated by SDS-PAGE. Protein concentration was determined on a  
523 Nanodrop Spectrophotometer using the molecular weight and the extinction coefficient. If  
524 necessary, samples were concentrated at 3,500 rpm and 4°C using Amicon Ultra centrifugal  
525 units (Merck).

#### 526 *Affinity gel electrophoresis (AGE)*

527 Prior to AGE, concentration of NaCl was decreased from 200 mM to 20 mM by diluting the  
528 protein sample with 10 mM MOPS (pH 7.8). Recombinant proteins and BSA, which served as  
529 a non-binding control, were loaded on 12% native gels without the addition of any  
530 carbohydrate, or supplemented with 0.2% laminarin or 0.1% lichenan. Runs were performed  
531 at 80 V and 4°C on ice. Proteins were stained using Coomassie Blue. Differences in length of  
532 running fronts were considered to evaluate binding.

#### 533 *Enzyme activity*

534 3,5-dinitrosalicylic acid (DNS) reducing sugar assay (57) was used to confirm activity of the Cf-  
535 GH16\_3 on laminarin and MLG lichenan. Proteins (20 mM NaCl, 10 mM MOPS pH 7.8) were  
536 mixed with 0.5% laminarin or lichenan (20 mM NaCl, 10 mM MOPS pH 7.8) in 200 µl reactions.  
537 Samples were incubated overnight at room temperature and therefore experiments correspond  
538 to endpoint measurements. The non-catalytic Cf-SGBP and samples without the addition of  
539 protein served as controls. 100 µl of DNS reagent was added to 100 µl of sample. After 5 min

540 of incubation at 95 °C, the reaction was stopped by adding 800 µl of water. Absorbance was  
541 measured at 540 nm and increased with catalytic activity. Experiments were performed in  
542 triplicates. Absorbance did not increase in the Cf-GH16\_3 incubations with MLG lichenan. In  
543 this case, samples were further investigated by fluorophore-assisted carbohydrate  
544 electrophoresis (FACE) that allows for higher sensitivity. 50 µl of the overnight incubations  
545 were dried under vacuum and mixed with 2 µl of 0.15 M 8-aminonaphtalene-1,3,6-trisulfonate  
546 (ANTS) and 2 µl of 1 M NaBH<sub>3</sub>CN. Samples were incubated at 37 °C in the dark for at least 2  
547 h and dried under vacuum. After resuspension in 20 µl of 25% glycerol, 5 µl were loaded onto  
548 27% acrylamide gels and migrated at 200 V and 4 °C. Results were visualized under UV.

#### 549 *Isothermal titration calorimetry (ITC)*

550 Proteins were dialyzed against 100 mM NaCl, 10 mM MOPS pH 7.8. The dialysis buffer was  
551 used to prepare laminarin and laminarin-derived oligosaccharide solutions, as well as for  
552 washing of the ITC cell and for controls. Prior to measurements, samples and solutions were  
553 centrifuged for degassing and removal of putative protein aggregates. In addition, protein  
554 samples were evaluated by SDS-PAGE and Nanodrop repeatedly. ITC was done on a  
555 MicroCal™ ITC 200 machine. Laminarin or laminarin-derived oligosaccharides were injected  
556 into 200 µl of protein, which was loaded into the sample cell. 2.5 mM of laminarin was injected  
557 into 147.6 µM of the Cf-SGBP, into 127.7 µM of the Cf-SGBP-CBM<sub>xxIII/III</sub> or into 94.7 µM of the  
558 Cf-SGBP-CBM<sub>xxIII/IV</sub>. 1 mM of laminarin was injected into 147.5 µM of the Cf-SGBP-CBM<sub>xxIV</sub>.  
559 For calculations with laminarin, we used the average molecular weight of laminarin from  
560 *Laminaria digitata* (2). In addition, 5 mM of laminariheptaose (DP7), 10 mM of laminaripentaose  
561 (DP5) and up to 20 mM of laminaribiose (DP2) were injected into 359.6 µM of the Cf-SGBP-  
562 CBM<sub>xxIV</sub>. Settings were as follows: cell temperature 20 °C (293.15 K), reference power 10  
563 µCal/s, stirring speed 750 rpm, filter period 1 sec, injection spacing 300 sec, 1 µl injection  
564 volume (first injection 0.3 µl), 35 injections in total. Experiments were performed in at least  
565 triplicates. 2.5 mM and 5 mM of laminarin were injected into 118.8 µM of the Cf-SGBP-Ig<sub>a</sub> as  
566 a non-interacting control. Titrations of buffer into proteins, laminarin into buffer or buffer into

567 buffer represented additional controls. MICROCAL ORIGIN v. 7 was used to analyze the data.  
568 A single-site binding model was selected.

### 569 *SEC-SAXS*

570 SAXS data of the Cf-SGBP and the Cf-GH16\_3 were collected at the Synchrotron SOLEIL on  
571 the SWING beamline. Protein samples were centrifuged prior to analyses to remove putative  
572 aggregates. SAXS measurements were coupled with prefixed SEC. The SEC column was  
573 equilibrated in buffer C (100 mM NaCl, 10 mM MOPS pH 7.8), which was used for protein  
574 purification. 70  $\mu\text{l}$  of protein were injected, the Cf-SGBP at 8.2 mg ml<sup>-1</sup> and the Cf-GH16\_3 at  
575 12.9 mg ml<sup>-1</sup>. The sample-detector-distance was 2.4 m, resulting in a scattering vector  $q$ -range  
576 of 0.012 – 0.504  $\text{\AA}^{-1}$  for the Cf-SGBP and 0.017 – 0.504  $\text{\AA}^{-1}$  for the Cf-GH16\_3. The obtained  
577 scattering data were normalized and corrected according to standard procedures. The Guinier  
578 equation was used to calculate the forward scattering  $I(0)$  and the radius of gyration  $R_g$ . The  
579 distance distribution function  $P(r)$  and the maximum particle dimension  $D_{\text{max}}$  were calculated  
580 by Fourier inversion of the scattering intensity  $I(q)$  using GNOM integrated in the PRIMUS  
581 software (ATSAS 3.1.3) (58). Models of protein envelopes were calculated from the  
582 experimental scattering curves using GASBORi (36). High  $\chi^2$  values indicated flexible proteins.  
583 Therefore, the data were analyzed using DADIMODO (37). Here, structural data of proteins  
584 are included into the analysis to identify an arrangement that best describes the experimental  
585 data. For this, rigid bodies and linkers were defined based on the structures predicted by  
586 Alphafold2 (28). Finally, EOM analysis was used to describe the flexibility of proteins in solution  
587 (38, 39), again using Alphafold2-derived coordinates. Experimental and processed data were  
588 visualized in R using the ggplot2 package.

### 589 *Crystallography*

590 Crystallization for both recombinant constructs of Cf-SGBP-CBMxx<sub>IV</sub> (at a concentration of  
591 11.7 mg ml<sup>-1</sup>) and Cf-SGBP-CBMxx<sub>III/IV</sub> (at a concentration of 9.5 mg ml<sup>-1</sup>) were tempted, in  
592 presence or absence of laminarin oligosaccharides (DP3/4), but never led to crystals in the  
593 case of Cf-SGBP-CBMxx<sub>III/IV</sub>. Crystallization screening was undertaken with the nanodrop-

594 robot Crystal Gryphon (Art Robbins instruments) starting with sparse matrix sampling kits  
595 JCSGplus, PACT, PEGs, and Wizard Classic I/II and Ligand-Friendly Screen (from Qiagen  
596 and Molecular Dimensions). The initial crystallization conditions were manually optimized and  
597 suitable crystals of Cf-SGBP-CBM<sub>xxIV</sub> in presence of laminarin-trisaccharides were obtained  
598 using the hanging drop vapor diffusion method as follows: 2 µl of protein were mixed with 1 µl  
599 of reservoir solution (500 µl) containing 2.5 M ammonium sulfate, 0.1 M BIS-TRIS propane pH  
600 7.0, 6% ethanol and 4% polyethylenglycol (PEG) 6000. Diffraction data were collected on  
601 PROXIMA1 (SOLEIL synchrotron) with a single crystal of Cf-SGBP-CBM<sub>xxIV</sub> in complex with  
602 laminarin-trisaccharide at 1.9 Å resolution. The data were integrated using XDS (59) and  
603 scaled with aimless (60). The structure was solved by molecular replacement with Phaser (61)  
604 using the AF model as starting point. The crystal structure was further enhanced with  
605 alternating cycles of refinement with REFMAC (62) and manual construction using COOT (33).  
606 All further data collection conditions and refinement statistics are given in Table S1.

#### 607 *Phytoplankton bloom data*

608 Chlorophyll *a* concentration as well as metagenomes, metagenome-assembled genomes and  
609 metaproteomes were obtained from samples taken during spring phytoplankton blooms at  
610 North Sea island of Helgoland in 2016 (9), 2018 and 2020 (10, 40). Chlorophyll *a* data were  
611 obtained from the Helgoland Roads time series (63) and have been published for the  
612 respective years (40). Metagenome sequencing, assembly and binning (MAGs) are described  
613 in (9) (2016) and (10) (2020). Preparation of metaproteome samples (0.2 µm fraction) as well  
614 as corresponding LC-MS/MS measurements and data analyses have been described in detail  
615 in (9) (2016) and in (40) (2018 and 2020).

#### 616 *Comparative genomics*

617 MAGs from 2016, 2018 and 2020 were dereplicated using dRep v1.14.6 (64) to prevent  
618 redundancy, applying a minimum completeness of 70% and contamination lower than 5% at  
619 0.95 ANI (average nucleotide identity). For each representative MAG (555), protein coding  
620 sequences for representative MAGs were predicted with Prokka v1.14.6 using default settings

621 (65). PULs, CAZymes, CBMs, SusC/D-like proteins were predicted as described previously  
622 (9), using hmmscan v3.3.2 against dbCAN-HMMdb-V1 and DIAMOND BLASTp v2.1.1.155  
623 (66) against CAZyDB.08062022 provided by dbCAN (67). CBMxx- and CBMyy-containing  
624 sequences within representative MAGs and additional draft genomes of previously published  
625 53 North Sea *Bacteroidota* (6) were identified using hmmscan as described above, adding  
626 corresponding HMM profiles to the database. Results were further filtered using the hmmscan-  
627 parser.sh script from dbCAN with an e-value cutoff of 1E-15 and a minimum coverage of 35%.  
628 The distribution of domain organization was visualized with UpSetR (68, 69). The official CAZy  
629 classification was applied to semi-manually determine the CBM boundaries, the multimodular  
630 compositions, the taxonomic distributions in CAZy and the PUL prevalence and partnerships  
631 in PULDB.

### 632 **Data Availability**

633 The metaproteome datasets for 2018 and 2020, the PDB identifier and SAXS data will be  
634 released upon acceptance of the manuscript in a peer-reviewed journal. Metagenome  
635 assemblies and MAGs from 2016, 2018 and 2020 phytoplankton blooms are available in the  
636 European Nucleotide Archive under the project accessions PRJEB28156 (2016),  
637 PRJEB38290 (2018) and PRJEB52999 (2020). Metaproteome data from the 2016 blooms are  
638 available in the PRIDE archive (70) through the identifier PXD019294. Phytoplankton data are  
639 available on request via Pangaea (<https://doi.pangaea.de/10.1594/PANGAEA.864676>).

### 640 **Acknowledgements**

641 This study was supported by the German Research Foundation (DFG) within the research unit  
642 FOR 2406 'Proteogenomics of Marine Polysaccharide Utilization' (POMPU) by grants to  
643 Hanno Teeling (TE 813/2–3), Dörte Becher (BE 3869/4-3) and Thomas Schweder (SCHW  
644 595/10-3, SCHW 595/11-3). This work also benefited from the support of the Centre National  
645 de la Recherche Scientifique (CNRS) and Sorbonne University (Paris). The authors strongly  
646 appreciated to have access to the CristalO platform (FR2424, Station Biologique de Roscoff),  
647 which is part of the core facility networks Biogenouest and EMBRC-France. We are grateful

648 for the access to the small-angle X-ray scattering beamline and MX-beamline PROXIMA1 at  
649 synchrotron SOLEIL (Saint Aubin, France) and thank the staff for their support on site,  
650 especially Aurélien Thureau and Pierre Legrand. We thank the team of the Helgoland sampling  
651 campaign, especially Lilly Franzmeyer, and the Biological Station Helgoland (Alfred-Wegener-  
652 Institut Helmholtz-Zentrum für Polar- und Meeresforschung, BAH-AWI, Germany), especially  
653 Inga Kirstein and Karen Wiltshire. The Helgoland Time Series of the Alfred Wegener Institute  
654 is supported by the Helmholtz Association as an LK-II performance category program. We  
655 thank Jana Matulla, Jasna Nikolic and Lionel Cladière for support in the lab and Stephanie  
656 Markert for helpful comments. MKZ gratefully acknowledges the DFG (SCHW 595/10-3), the  
657 PROCOPE Mobility Grant and the ERASMUS Staff Mobility for Training, all of which enabled  
658 several research stays in Roscoff.

#### 659 **Author contributions**

660 MC and MKZ directed the study with contribution from EFB and TS. MKZ wrote the first draft  
661 of the paper with writing contributions from MC and EFB. All co-authors provided their input  
662 on the manuscript and helped with editing. Figures were created by MKZ, MC and DBa. MKZ  
663 and MC performed computational sequence analyses of *C. forsetii* proteins. MKZ performed  
664 cloning, protein purification as well as experiments on catalytic activity and protein-sugar  
665 interactions. NW and RL supported cloning. EFB, MC and MJ supported experiments on  
666 protein-sugar interactions. MC and MKZ performed SAXS analyses and data treatment. AJ  
667 and MC performed protein crystallography and MC solved the 3D crystal structure. MJ and DJ  
668 produced and purified laminarin-derived oligosaccharides. MJ, TE and FT supported protein  
669 quality controls. NT classified CBMxx and CBMyy. DBa and NT performed comparative  
670 sequence analyses with support of MC and MKZ. DBa analyzed the phytoplankton bloom data  
671 with support of FW, HT, ATS and DBe. All co-authors approved the submitted manuscript.

#### 672 **Conflict of Interest**

673 The authors declare no conflict of interest.

674 **References**

675

676 1. Becker S, Tebben J, Coffinet S, Wiltshire K, Iversen MH, Harder T, et al. Laminarin is  
677 a major molecule in the marine carbon cycle. Proc Natl Acad Sci USA.

678 2020;117(12):6599-607.

679 2. Read SM, Currie G, Bacic A. Analysis of the structural heterogeneity of laminarin by  
680 electrospray-ionisation-mass spectrometry. Carbohydr Res. 1996;281(2):187-201.

681 3. Gügi B, Le Costaouec T, Burel C, Lerouge P, Helbert W, Bardor M. Diatom-Specific  
682 Oligosaccharide and Polysaccharide Structures Help to Unravel Biosynthetic

683 Capabilities in Diatoms. Mar Drugs. 2015;13(9):5993-6018.

684 4. Vidal-Melgosa S, Sichert A, Francis TB, Bartosik D, Niggemann J, Wichels A, et al.  
685 Diatom fucan polysaccharide precipitates carbon during algal blooms. Nat Commun.

686 2021;12(1):1–13.

687 5. Koch H, Dürwald A, Schweder T, Noriega-Ortega B, Vidal-Melgosa S, Hehemann JH,  
688 et al. Biphasic cellular adaptations and ecological implications of *Alteromonas*

689 *macleodii* degrading a mixture of algal polysaccharides. ISME J. 2019;13(1):92-103.

690 6. Kappelmann L, Krüger K, Hehemann JH, Harder J, Markert S, Unfried F, et al.

691 Polysaccharide utilization loci of North Sea *Flavobacteriia* as basis for using SusC/D-  
692 protein expression for predicting major phytoplankton glycans. ISME J.

693 2019;13(1):76-91.

694 7. Krüger K, Chafee M, Francis TB, del Rio TG, Becher D, Schweder T, et al. In marine  
695 *Bacteroidetes* the bulk of glycan degradation during algae blooms is mediated by few

696 clades using a restricted set of genes. ISME J. 2019;13(11):2800-16.

697 8. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennis CM, et al.

698 Substrate-controlled succession of marine bacterioplankton populations induced by a  
699 phytoplankton bloom. Science. 2012;336(6081):608-11.

700 9. Francis TB, Bartosik D, Sura T, Sichert A, Hehemann JH, Markert S, et al. Changing

701 expression patterns of TonB-dependent transporters suggest shifts in polysaccharide

- 702 consumption over the course of a spring phytoplankton bloom. ISME J.  
703 2021;15(8):2336-50.
- 704 10. Sidhu C, Kirstein IV, Meunier CL, Rick J, Fofonova V, Wiltshire KH, et al. Dissolved  
705 storage glycans shaped the community composition of abundant bacterioplankton  
706 clades during a North Sea spring phytoplankton bloom. Microbiome. 2023;11(1):77.
- 707 11. Unfried F, Becker S, Robb CS, Hehemann JH, Markert S, Heiden SE, et al. Adaptive  
708 mechanisms that provide competitive advantages to marine bacteroidetes during  
709 microalgal blooms. ISME J. 2018;12:2894–906.
- 710 12. Becker S, Scheffel A, Polz MF, Hehemann JH. Accurate quantification of laminarin in  
711 marine organic matter with enzymes from marine microbes. Appl Environ Microbiol.  
712 2017;83(9).
- 713 13. Labourel A, Jam M, Jeudy A, Hehemann JH, Czjzek M, Michel G. The  $\beta$ -glucanase  
714 ZgLamA from *Zobellia galactanivorans* evolved a bent active site adapted for efficient  
715 degradation of algal laminarin. J Biol Chem. 2014;289(4):2027-42.
- 716 14. Labourel A, Jam M, Legentil L, Sylla B, Hehemann JH, Ferrieres V, et al. Structural  
717 and biochemical characterization of the laminarinase ZgLamC(GH16) from *Zobellia*  
718 *galactanivorans* suggests preferred recognition of branched laminarin. Acta  
719 Crystallogr D. 2015;71:173-84.
- 720 15. Pollet RM, Martin LM, Koropatkin NM. TonB-dependent transporters in the  
721 Bacteroidetes: Unique domain structures and potential functions. Mol Microbiol.  
722 2021;115(3):490-501.
- 723 16. Ratliff AC, Buchanan SK, Celia H. The Ton Motor. Front Microbiol. 2022;13.
- 724 17. Glenwright AJ, Pothula KR, Bhamidimarri SP, Chorev DS, Baslé A, Firbank SJ, et al.  
725 Structural basis for nutrient acquisition by dominant members of the human gut  
726 microbiota. Nature. 2017;541(7637):407-11.
- 727 18. Anderson KL, Salyers AA. Biochemical evidence that starch breakdown by  
728 *Bacteroides thetaiotaomicron* involves outer membrane starch-binding sites and  
729 periplasmic starch-degrading enzymes. J Bacteriol. 1989;171(6):3192-8.



- 730 19. Bjursell MK, Martens EC, Gordon JI. Functional genomic and metabolic studies of the  
731 adaptations of a prominent adult human gut symbiont, *Bacteroides thetaiotaomicron*,  
732 to the suckling period. J Biol Chem. 2006;281(47):36269-79.
- 733 20. Tamura K, Foley MH, Gardill BR, Dejean G, Schnizlein M, Bahr CME, et al. Surface  
734 glycan-binding proteins are essential for cereal beta-glucan utilization by the human  
735 gut symbiont *Bacteroides ovatus*. Cell Mol Life Sci. 2019;76(21):4319-40.
- 736 21. Tamura K, Dejean G, Van Petegem F, Brumer H. Distinct protein architectures  
737 mediate species-specific beta-glucan binding and metabolism in the human gut  
738 microbiota. J Biol Chem. 2021;296:100415.
- 739 22. Mystkowska AA, Robb C, Vidal-Melgosa S, Vanni C, Fernandez-Guerra A, Höhne M,  
740 et al. Molecular recognition of the beta-glucans laminarin and pustulan by a SusD-like  
741 glycan-binding protein of a marine *Bacteroidetes*. FEBS J. 2018;285(23):4465-81.
- 742 23. Jam M, Ficko-Blean E, Labourel A, Larocque R, Czjzek M, Michel G. Unraveling the  
743 multivalent binding of a marine family 6 carbohydrate-binding module with its native  
744 laminarin ligand. FEBS J. 2016;283(10):1863-79.
- 745 24. Kumagai Y, Kishimura H, Lang W, Tagami T, Okuyama M, Kimura A.  
746 Characterization of an Unknown Region Linked to the Glycoside Hydrolase Family 17  
747  $\beta$ -1,3-Glucanase of *Vibrio vulnificus* Reveals a Novel Glucan-Binding Domain. Mar  
748 Drugs. 2022;20(4).
- 749 25. Deshmukh UB, Oren A. Proposal of *Christiangramia* gen. nov., *Neomelitea* gen. nov.  
750 and *Nicoliella* gen. nov. as replacement names for the illegitimate prokaryotic generic  
751 names *Gramella* Nedashkovskaya et al. 2005, *Melitea* Urios et al. 2008 and *Nicolia*  
752 Oliphant et al. 2022, respectively. Int J Syst Evol Microbiol. 2023;73(4).
- 753 26. Eilers H, Pernthaler J, Peplies J, Glöckner FO, Gerds G, Amann R. Isolation of novel  
754 pelagic bacteria from the German bight and their seasonal contributions to surface  
755 picoplankton. Appl Environ Microbiol. 2001;67(11):5134-42.

- 756 27. Kabisch A, Otto A, König S, Becher D, Albrecht D, Schüler M, et al. Functional  
757 characterization of polysaccharide utilization loci in the marine *Bacteroidetes*  
758 '*Gramella forsetii*' KT0803. ISME J. 2014;8(7):1492-502.
- 759 28. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold:  
760 making protein folding accessible to all. Nat Methods. 2022;19(6):679-82.
- 761 29. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal  
762 for protein modeling, prediction and analysis. Nat Protoc. 2015;10(6):845-58.
- 763 30. Holm L. Dali server: structural unification of protein families. Nucleic Acids Res.  
764 2022;50(W1):W210-W5.
- 765 31. Pluvinage B, Hehemann JH, Boraston AB. Substrate recognition and hydrolysis by a  
766 family 50 *exo-β*-agarase, Aga50D, from the marine bacterium *Saccharophagus*  
767 *degradans*. J Biol Chem. 2013;288(39):28078-88.
- 768 32. Drula E, Garron ML, Dogan S, Lombard V, Henrissat B, Terrapon N. The  
769 carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res.  
770 2022;50(D1):D571-D7.
- 771 33. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot.  
772 Acta Crystallogr D Biol Crystallogr. 2010;66(4):486-501.
- 773 34. Hudson KL, Bartlett GJ, Diehl RC, Agirre J, Gallagher T, Kiessling LL, et al.  
774 Carbohydrate-Aromatic Interactions in Proteins. J Am Chem Soc.  
775 2015;137(48):15152-60.
- 776 35. Robert X, Gouet P. Deciphering key features in protein structures with the new  
777 ENDscript server. Nucleic Acids Res. 2014;42(W1):W320-W4.
- 778 36. Svergun DI, Petoukhov MV, Koch MHJ. Determination of domain structure of proteins  
779 from X-ray solution scattering. Biophys J. 2001;80(6):2946-53.
- 780 37. Rudenko O, Thureau A, Perez J. Evolutionary refinement of the 3D structure of multi-  
781 domain protein complexes from Small Angle X-ray Scattering data. Proceedings of  
782 the 2019 Genetic and Evolutionary Computation Conference Companion (Gecco'19  
783 Companion). 2019:401-2.

- 784 38. Tria G, Mertens HDT, Kachala M, Svergun DI. Advanced ensemble modelling of  
785 flexible macromolecules using X-ray solution scattering. *IUCrJ*. 2015;2,207-17.
- 786 39. Bernadó P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI. Structural  
787 characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem*  
788 *Soc*. 2007;129(17):5656-64.
- 789 40. Schweder T, Beidler I, Steinke N, Schulze T, Sidhu C, Bartosik D, et al. Alpha-  
790 glucans from bacterial necromass indicate an intra-population loop within the marine  
791 carbon cycle. 2023. PREPRINT (Version 1) available at Research Square  
792 [<https://doi.org/10.21203/rs.3.rs-3205445/v1>]
- 793 41. Sainz-Polo MA, González B, Menéndez M, Pastor FIJ, Sanz-Aparicio J. Exploring  
794 Multimodularity in Plant Cell Wall Deconstruction: STRUCTURAL AND FUNCTIONAL  
795 ANALYSIS OF Xyn10C CONTAINING THE CBM22-1-CBM22-2 TANDEM. *J Biol*  
796 *Chem*. 2015;290(28):17116-30.
- 797 42. Henshaw JL, Bolam DN, Pires VMR, Czjzek M, Henrissat B, Ferreira LMA, et al. The  
798 family 6 carbohydrate binding module CmCBM6-2 contains two ligand-binding sites  
799 with distinct specificities. *J Biol Chem*. 2004;279(20):21552-9.
- 800 43. Ficko-Blean E, Stuart CP, Suits MD, Cid M, Tessier M, Woods RJ, et al.  
801 Carbohydrate recognition by an architecturally complex alpha-*N*-  
802 Acetylglucosaminidase from *Clostridium perfringens*. *Plos One*. 2012;7(3).
- 803 44. Boraston AB, Bolam DN, Gilbert HJ, Davies GJ. Carbohydrate-binding modules: fine-  
804 tuning polysaccharide recognition. *Biochem J*. 2004;382:769-81.
- 805 45. Cartmell A, Lowe EC, Baslé A, Firbank SJ, Ndeh DA, Murray H, et al. How members  
806 of the human gut microbiota overcome the sulfation problem posed by  
807 glycosaminoglycans. *Proc Natl Acad Sci USA*. 2017;114(27):7037-42.
- 808 46. Cota J, Alvarez TM, Citadini AP, Santos CR, Neto MD, Oliveira RR, et al. Mode of  
809 operation and low-resolution structure of a multi-domain and hyperthermophilic endo-  
810  $\beta$ -1,3-glucanase from *Thermotoga petrophila*. *Biochem Biophys Res Commun*.  
811 2011;406(4):590-4.

- 812 47. Tuzin AS, Kwiatkowski KJ, Orlovsky NI, Smith CJ, Creagh AL, Haynes CA, et al.  
813 Molecular Dissection of Xyloglucan Recognition in a Prominent Human Gut Symbiont.  
814 *Mbio*. 2016;7(2):e02134-15.
- 815 48. White JBR, Silale A, Feasey M, Heunis T, Zhu Y, Zheng H, et al. Outer membrane  
816 utilosomes mediate glycan uptake in gut Bacteroidetes. *Nature*. 2023;618(7965):583-  
817 9.
- 818 49. Cameron EA, Kwiatkowski KJ, Lee BH, Hamaker BR, Koropatkin NM, Martens EC.  
819 Multifunctional nutrient-binding proteins adapt human symbiotic bacteria for glycan  
820 competition in the gut by separately promoting enhanced sensing and catalysis. *Mbio*.  
821 2014;5(5):e01441-14.
- 822 50. Poli A, Anzelmo G, Nicolaus B. Bacterial Exopolysaccharides from Extreme Marine  
823 Habitats: Production, Characterization and Biological Activities. *Mar Drugs*.  
824 2010;8(6):1779-802.
- 825 51. Hervé C, Rogowski A, Blake AW, Marcus SE, Gilbert HJ, Knox JP. Carbohydrate-  
826 binding modules promote the enzymatic deconstruction of intact plant cell walls by  
827 targeting and proximity effects. *Proc Natl Acad Sci USA*. 2010;107(34):15293-8.
- 828 52. Najmudin S, Guerreiro CI, Carvalho AL, Prates JA, Correia MA, Alves VD, et al.  
829 Xyloglucan is recognized by carbohydrate-binding modules that interact with  $\beta$ -glucan  
830 chains. *J Biol Chem*. 2006;281(13):8815-28.
- 831 53. Juncker AS, Willenbrock H, Von Heijne G, Brunak S, Nielsen H, Krogh A. Prediction  
832 of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci*.  
833 2003;12(8):1652-62.
- 834 54. Lieutaud P, Canard B, Longhi S. MeDor: a metasever for predicting protein disorder.  
835 *BMC Genomics*. 2008;9.
- 836 55. Groisillier A, Hervé C, Jeudy A, Rebuffet E, Pluchon PF, Chevlot Y, et al. MARINE-  
837 EXPRESS: taking advantage of high throughput cloning and expression strategies for  
838 the post-genomic analysis of marine organisms. *Microb Cell Fact*. 2010;9(45).

- 839 56. Studier FW. Protein production by auto-induction in high-density shaking cultures.  
840 Protein Expr Purif. 2005;41(1):207-34.
- 841 57. Miller GL. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar.  
842 Anal Chem. 1959;31(3):426-8.
- 843 58. Manalastas-Cantos K, Konarev PV, Hajizadeh NR, Kikhney AG, Petoukhov MV,  
844 Molodenskiy DS, et al. ATSAS 3.0: expanded functionality and new tools for small-  
845 angle scattering data analysis. J Appl Crystallogr. 2021;54:343-55.
- 846 59. Kabsch W. XDS. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 2):125-32.
- 847 60. Evans PR, Murshudov GN. How good are my data and what is the resolution? Acta  
848 Crystallogr D Biol Crystallogr. 2013;69(7):1204-14.
- 849 61. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ.  
850 Phaser crystallographic software. J Appl Crystallogr. 2007;40(Pt 4):658-74.
- 851 62. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al.  
852 REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr  
853 D. 2011;67:355-67.
- 854 63. Armonies W, Asmus H, Buschbaum C, Lackschewitz D, Reise K, Rick J. Microscopic  
855 species make the diversity: a checklist of marine flora and fauna around the Island of  
856 Sylt in the North Sea. Helgol Mar Res. 2018;72.
- 857 64. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate  
858 genomic comparisons that enables improved genome recovery from metagenomes  
859 through de-replication. ISME J. 2017;11(12):2864-8.
- 860 65. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics.  
861 2014;30(14):2068-9.
- 862 66. Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale  
863 using DIAMOND. Nat Methods. 2021;18(4):366-8.
- 864 67. Zheng J, Hu B, Zhang X, Ge Q, Yan Y, Akresi J, et al. dbCAN-seq update: CAZyme  
865 gene clusters and substrates in microbiomes. Nucleic Acids Res. 2023;51(D1):D557-  
866 D63.

- 867 68. Lex A, Gehlenborg N, Strobel H, Vuillemot R, Pfister H. UpSet: Visualization of  
868 Intersecting Sets. *IEEE T Vis Comput Gr*. 2014;20(12):1983-92.
- 869 69. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of i  
870 ntersecting sets and their properties. *Bioinformatics*. 2017;33(18):2938-40.
- 871 70. Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S,  
872 Kamatchinathan S, et al. The PRIDE database resources in 2022: a hub for mass  
873 spectrometry-based proteomics evidences. *Nucleic Acids Res*. 2022;50(D1):D543-  
874 D52.
- 875