



**HAL**  
open science

## Unveiling the role of novel carbohydrate-binding modules in laminarin interaction of multimodular proteins from marine Bacteroidota during phytoplankton blooms

Marie-Katherin Zühlke, Elizabeth Ficko-Blean, Daniel Bartosik, Nicolas Terrapon, Alexandra Jeudy, Murielle Jam, Fengqing Wang, Norma Welsch, Alexandra Dürwald, Laura Torres Martin, et al.

### ► To cite this version:

Marie-Katherin Zühlke, Elizabeth Ficko-Blean, Daniel Bartosik, Nicolas Terrapon, Alexandra Jeudy, et al.. Unveiling the role of novel carbohydrate-binding modules in laminarin interaction of multimodular proteins from marine Bacteroidota during phytoplankton blooms. *Environmental Microbiology*, 2024, 26 (5), pp.e16624. 10.1111/1462-2920.16624 . hal-04579759

**HAL Id: hal-04579759**

**<https://amu.hal.science/hal-04579759v1>**

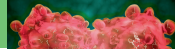
Submitted on 13 Nov 2024

**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.



Distributed under a Creative Commons Attribution 4.0 International License



# Unveiling the role of novel carbohydrate-binding modules in laminarin interaction of multimodular proteins from marine *Bacteroidota* during phytoplankton blooms

Marie-Katherin Zühlke<sup>1,2</sup> | Elizabeth Ficko-Blean<sup>3</sup> | Daniel Bartosik<sup>1,2</sup> |  
 Nicolas Terrapon<sup>4</sup> | Alexandra Jeudy<sup>3</sup> | Murielle Jam<sup>3</sup> |  
 Fengqing Wang<sup>5</sup> | Norma Welsch<sup>1,2</sup> | Alexandra Dürwald<sup>1,6</sup> |  
 Laura Torres Martin<sup>1</sup> | Robert Larocque<sup>3</sup> | Diane Jouanneau<sup>3</sup> |  
 Tom Eisenack<sup>1</sup> | François Thomas<sup>3</sup> | Anke Trautwein-Schult<sup>7</sup> |  
 Hanno Teeling<sup>5</sup> | Dörte Becher<sup>7</sup> | Thomas Schweder<sup>1,2</sup> | Mirjam Czjzek<sup>3</sup>

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

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

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

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

<sup>5</sup>Max Planck Institute for Marine Microbiology, Bremen, Germany

<sup>6</sup>Helmholtz Institute for One Health, Helmholtz Centre for Infection Research HZI, Greifswald, Germany

<sup>7</sup>Microbial Proteomics, Institute of Microbiology, University Greifswald, Greifswald, Germany

## Correspondence

Mirjam Czjzek, Integrative Biology of Marine Models (LBI2M), Station Biologique de Roscoff (SBR), Sorbonne Université, CNRS, 29680 Roscoff, Bretagne, France.  
 Email: [czjzek@sb-roscoff.fr](mailto:czjzek@sb-roscoff.fr)

Marie-Katherin Zühlke, Pharmaceutical Biotechnology, Institute of Pharmacy, University Greifswald, Greifswald, 17487, Germany.  
 Email: [marie-katherin.zuehlke@uni-greifswald.de](mailto:marie-katherin.zuehlke@uni-greifswald.de)

## Funding information

Centre National de la Recherche Scientifique; Deutsche Forschungsgemeinschaft, Grant/Award Numbers: BE 3869/4-3, SCHW 595/10-3, SCHW 595/11-3, TE 813/2-3; Sorbonne University; PROCOPE Mobility Grant; Helmholtz Association; ERASMUS+ Staff Mobility for Training

## Abstract

Laminarin, a  $\beta(1,3)$ -glucan, serves as a storage polysaccharide in marine microalgae such as diatoms. Its abundance, water solubility and simple structure make it an appealing substrate for marine bacteria. Consequently, many marine bacteria have evolved strategies to scavenge and decompose laminarin, employing carbohydrate-binding modules (CBMs) as crucial components. In this study, we characterized two previously unassigned domains as laminarin-binding CBMs in multimodular proteins from the marine bacterium *Christiangramia forsetii* KT0803<sup>T</sup>, thereby introducing the new laminarin-binding CBM families CBM102 and CBM103. We identified four CBM102s in a surface glycan-binding protein (SGBP) and a single CBM103 linked to a glycoside hydrolase module from family 16 (GH16\_3). Our analysis revealed that both modular proteins have an elongated shape, with GH16\_3 exhibiting greater flexibility than SGBP. This flexibility may aid in the recognition and/or degradation of laminarin, while the constraints in SGBP could facilitate the docking of laminarin onto the bacterial surface. Exploration of bacterial metagenome-assembled genomes (MAGs) from phytoplankton blooms in the North Sea showed that both laminarin-binding CBM families are widespread among marine *Bacteroidota*. The high protein abundance of CBM102- and CBM103-containing proteins during phytoplankton blooms further emphasizes their significance in marine laminarin utilization.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Environmental Microbiology* published by John Wiley & Sons Ltd.



## INTRODUCTION

Laminarin is an important energy and carbon storage compound in marine haptophytes and in stramenopiles, including abundant diatoms. Therefore, huge amounts of laminarin are produced during common diatom-dominated phytoplankton blooms (Becker et al., 2020). Laminarin is composed of a  $\beta(1,3)$ -linked glucose backbone with occasional  $\beta(1,6)$ -linked glucose branches that differ in length and frequency (Gügi et al., 2015; Read et al., 1996). Due to its abundance, water solubility and simple structure, marine bacteria often prefer laminarin as a substrate over less soluble and more complex polysaccharides (Koch et al., 2019; Vidal-Melgosa et al., 2021). Polysaccharide utilization loci (sg. polysaccharide utilization locus; abbr. PUL) dedicated to laminarin breakdown are thus both widely distributed (Kappelmann et al., 2019; Krüger et al., 2019) and highly expressed (Francis et al., 2021; Sidhu et al., 2023; Teeling et al., 2012; Vidal-Melgosa et al., 2021) in marine bacteria that thrive during phytoplankton blooms. PULs are distinct genetic loci that encode proteins to sense, digest and transport sugars. *Bacteroidota* are key players in marine laminarin turnover and many glycoside hydrolases (GHs) that are active on laminarin have been characterized from members of this phylum, such as GH16 and GH17 family representatives that act on the  $\beta(1,3)$ -main chain, or GH30 family representatives that remove  $\beta(1,6)$ -side chains (Becker et al., 2017; Labourel et al., 2014, 2015; Unfried et al., 2018). Some of these GHs are bound to the cell surface, where they degrade laminarin into smaller-sized oligosaccharides suitable for uptake and transport (Unfried et al., 2018). In *Bacteroidota*, the oligosaccharide shuttle to the periplasm is mediated by an accessory sugar-binding SusD-like protein, tethered to the outer membrane and a SusC-like TonB-dependent transporter (TBDT) (reviewed in Pollet et al., 2021; Ratliff et al., 2022). In complex with their target substrate, SusD-like proteins act as a hinge to relay their cargo into the associated SusC-like TBDT (Glenwright et al., 2017). In addition to SusD-like proteins, bacteria may have additional surface glycan-binding proteins (SGBPs), which are often encoded directly downstream of the *susCD*-like gene tandem and are sometimes termed ‘SusE-positioned’ proteins. The designations ‘-like’ versus ‘-positioned’ intend to distinguish between homologous versus analogous proteins, and refer to the starch utilization system (Sus) (Anderson & Salyers, 1989; Bjursell et al., 2006), which marked the beginning of the era of PUL exploration. While SusD-like proteins have a single binding site and are almost entirely  $\alpha$ -helical, SusE-positioned SGBPs may have several carbohydrate-binding modules (CBMs), which often display a  $\beta$ -sandwich fold structure and function as carbohydrate docking sites. Since SusD-like proteins also classify as SGBPs, they have sometimes been named ‘SGBP-A’ in the literature, whereas CBM-containing SGBPs are named ‘SGBP-B’ (Tamura et al., 2019). Some examples

of mixed-linkage  $\beta$ -glucan- and laminarin-related SusD-like proteins and CBM-containing SGBPs have recently been thoroughly analysed in gut strains from the *Bacteroides* genus (Tamura et al., 2021). By comparison, there is only one known structure of a marine laminarin-binding SusD-like protein, which recognizes  $\beta(1,6)$ -branches in laminarin (Mystkowska et al., 2018), and marine examples of CBM-containing SGBPs that bind laminarin are lacking. Unlike SusD-like proteins, CBM-containing SGBPs rarely show detectable similarity/homology (Tamura et al., 2019; Tamura et al., 2021) and are thus difficult to predict. Therefore, CBM-containing SGBPs are largely overlooked in gene annotations and probably underestimated. In addition to SusD-like proteins and CBM-containing SGBPs, laminarin-binding CBMs are often linked to carbohydrate-active enzymes (CAZymes) within multidomain structures (Jam et al., 2016; Kumagai et al., 2022), which further underlines the importance of CBMs in marine laminarin degradation.

The marine, laminarin-consuming model bacterium *Christiangramia forsetii* KT0803<sup>T</sup> (formerly *Gramella forsetii* (Deshmukh & Oren, 2023)), a member of the phylum *Bacteroidota*, has been isolated during a phytoplankton bloom in the North Sea (Eilers et al., 2001). A GH16 subfamily 3 (GH16\_3) (Viborg et al., 2019) and a ‘conserved hypothetical protein’, encoded in the *C. forsetii* laminarin PUL, were among the most abundantly produced proteins when growing on laminarin as a sole carbon source (Kabisch et al., 2014). Based on subproteome analyses it was suggested that the two proteins are tethered to the outer membrane and it was speculated that the ‘conserved hypothetical protein’, which is SusE-positioned, might function as an SGBP (Kabisch et al., 2014).

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

## EXPERIMENTAL PROCEDURES

### Carbohydrates

Laminarin (*Laminaria digitata*) was acquired from Sigma-Aldrich (Merck kGaA, Darmstadt, Germany), and lichenan (MLG from Icelandic Moss), laminarite-trose (DP4), laminaritriose (DP3) as well as

laminaribiose (DP2) from Megazyme International Ltd. (Bray, Ireland). Purchased laminarin-derived oligosaccharides were used for FACE, while for ITC and crystallography, oligosaccharides were produced following the protocol from Jam et al. (2016) using a GH16\_3 (LamA) from *Zobellia galactanivorans* Dsij<sup>T</sup> (Labourel et al., 2014).

## Culture conditions for *C. forsetii* KT0803<sup>T</sup>

Precultures of *C. forsetii* in Marine Broth 2216 (Difco) were used to inoculate main cultures to an OD<sub>600nm</sub> of 0.1. An artificial seawater medium (Schut et al., 1993) containing 0.2% laminarin or glucose was used for main cultures. *C. forsetii* was cultivated at 21°C and 170 rpm.

## Nomenclature

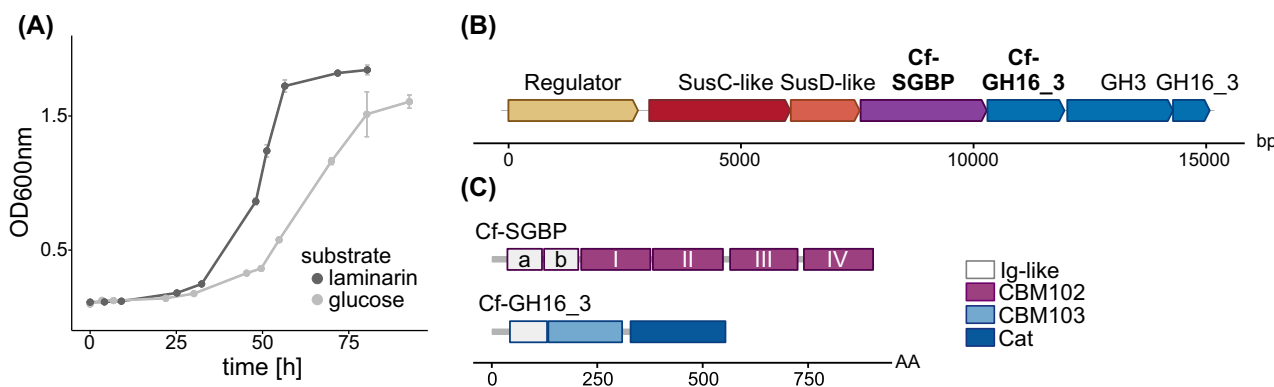
We refer to the full-length CBM-containing SGBP, which is SusE-positioned in the laminarin PUL of *C. forsetii* (see Figure 1B), as Cf-SGBP (locus tag GFO\_RS17395). The PUL further encodes two GH16\_3s. Subject of this study was the GH16\_3 encoded downstream of Cf-SGBP, which we refer to as Cf-GH16\_3 (locus tag GFO\_RS16360). The two proteins display a multidomain structure and specific domains were further specified, for example, Cf-GH16\_3-Cat refers to the catalytic module of Cf-GH16\_3. In the case of repetitive domains of Cf-SGBP, these modules were enumerated either alphabetically for Ig-like fold domains or with Roman numerals for the CBM102s, starting from the N-terminus and indicated as subscripts, for example, Cf-SGBP-Ig<sub>a</sub> refers to the first of two N-terminal Ig-like fold domains of Cf-SGBP. Whenever we refer to an SGBP in the text,

we refer to CBM-containing SGBPs, excluding SusD-like proteins.

## Cloning and site-directed mutagenesis

We used Alphafold2 within ColabFold (Mirdita et al., 2022) and Phyre2 (Kelley et al., 2015) to spatially define individual modules of the two multidomain proteins. LipoP (Juncker et al., 2003) detected signal peptides and MeDor (Lieutaud et al., 2008) predicted protein disorder by hydrophobic cluster analysis. In combination, these tools served to define suitable cloning sites to express full-length proteins, individual domains, or combinations thereof. Corresponding genes were amplified from *C. forsetii* genomic DNA by PCR using primers listed in Table S4. The PCR products were cloned into expression vectors pRF3 (ampicillin resistance) or pET28a (kanamycin resistance) by restriction-ligation to produce recombinant N-terminal hexa-histidine tagged proteins. The vector pRF3 is a derivative of pFO4 (Groisillier et al., 2010). The full-length gene sequences Cf-SGBP and Cf-GH16\_3 were cloned into pRF3. PCR products were digested using BamHI and AvrII, and the pRF3 vector using BamHI and NheI. In all other cases, PCR products and the pET28a vector were digested using NdeI and SacI. Digested PCR products were ligated into expression vectors using T4 DNA ligase. Corresponding plasmids for Cf-SGBP-Ig<sub>a/b</sub>, Cf-SGBP-CBM102<sub>i</sub>, Cf-GH16\_3-Ig and Cf-GH16\_3-CBM103 in pET28a, including NdeI and SacI restriction sites, were designed and ordered from GenScript (GenScript Biotech, Netherlands).

In addition, we generated two mutants of Cf-GH16\_3-Cat, where each catalytic glutamic acid was replaced by a serine, Cf-GH16\_3-Cat<sub>E442S</sub> and Cf-GH16\_3-Cat<sub>E447S</sub>. This was achieved in two steps:



**FIGURE 1** The laminarin PUL of *Christiangramia forsetii* encodes the multimodular Cf-SGBP and Cf-GH16\_3 containing the novel CBM102 and CBM103, respectively. (A) Growth of *C. forsetii* on laminarin (0.2%) compared to glucose (0.2%). (B) The *C. forsetii* laminarin PUL encodes multidomain proteins, including a novel CBM-containing SGBP (Cf-SGBP) and a GH16\_3 (Cf-GH16\_3), highlighted in bold. (C) Domain organization of Cf-SGBP and Cf-GH16\_3. Here, grey lines represent N-terminal lipoprotein signal peptides or linkers between modules. Cf-SGBP displays two Ig-like fold modules (a,b) followed by four CBM102s (I–IV). Cf-GH16\_3 comprises an Ig-like fold module and a CBM103 in addition to its catalytic module (Cat).





first, two separate reactions of PCR, using the Cf-GH16\_Cat-encoding plasmid as a template, were carried out to produce the mutant gene as two overlapping fragments (F1 and F2). The purified PCR products F1 and F2 were then used as a template for a final PCR to produce the fused product. To produce the overlapping fragments, we used a forward primer for the first fragment (F1), which annealed to the vector backbone directly upstream of Cf-GH16\_3-Cat, and a reverse primer for the second fragment (F2), which annealed to the vector backbone downstream of Cf-GH16\_3-Cat (designated as vector primers, Table S4). These primers were combined with a forward (F2) or reverse (F1) primer that contained the corresponding desired mutation and that annealed to the intended region of the mutation (Table S4). The purified PCR products F1 and F2 were digested with DpnI to remove any residual parental DNA before a final PCR fused F1 and F2 using the vector primers. PCR products were digested with NdeI and SacI and cloned into pET28a as described before.

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

## Protein overexpression and purification

Cf-SGBP and Cf-GH16\_3 were produced in ZYP-5052 auto-induction medium (Studier, 2005), supplemented with 100  $\mu\text{g mL}^{-1}$  ampicillin, for 72 h at 20°C and 130 rpm. For the production of the other proteins, clones were cultured in LB, supplemented with 30  $\mu\text{g mL}^{-1}$  kanamycin, at 37°C and 130 rpm until IPTG induction (1 mM) and then at 20°C and 130 rpm overnight. Cells were harvested by centrifugation and pellets were stored at -20°C until protein purification. Cells were disrupted by chemical lysis. In brief, the cell pellet was suspended in resuspension buffer (50 mM MOPS pH 8.0, 25% sucrose, lysozyme) and agitated for 15 min at 4°C. Then, double the volume of lysis buffer (20 mM MOPS pH 7.5, 100 mM NaCl, 1% sodium DCA, 1% Triton-X) was added, followed by another incubation of 5 min at 4°C. MgCl<sub>2</sub> concentration was adjusted to 5 mM and samples were incubated with DNase at room temperature until viscosity of the sample decreased. Cell debris was removed from protein extract by centrifugation. Protein extract was loaded onto Histrap HP columns (Cytiva, Vélizy-Villacoublay, France) equilibrated in buffer A (200 mM NaCl, 10 mM MOPS pH 7.8, 20 mM imidazole) and charged with 0.1 M NiSO<sub>4</sub> using an ÄKTA Prime system. In general, 1 mL columns were used for a culture volume of 250 mL and 5 mL columns for 1 L of culture.

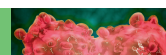
Proteins were eluted with a linear gradient of 0%–100% buffer B (200 mM NaCl, 10 mM MOPS pH 7.8, 500 mM imidazole) within 30 min (1 mL column) or 60 min (5 mL column) at a flow rate of 1 mL min<sup>-1</sup>. For SAXS analyses, crystallography or in case of impurities, the quality of the protein sample was further increased by size exclusion chromatography (SEC) using a Superdex 200 16/60 column (Cytiva) in buffer C (200 mM NaCl, 10 mM MOPS pH 7.8). Otherwise, samples were desalted on a HiPrep 26/10 desalting column (Cytiva) in buffer C. Buffers used to purify the proteins for SAXS analyses contained 100 mM NaCl. Samples were evaluated by SDS-PAGE. Protein concentration was determined on a Nanodrop Spectrophotometer using the molecular weight and the extinction coefficient. If necessary, samples were concentrated at 3500 rpm and 4°C using Amicon Ultra centrifugal filter units (3, 10, or 30 kDa cut-off, Merck).

## Affinity gel electrophoresis (AGE)

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

## Enzyme activity

To interpret our results from AGE, 3,5-dinitrosalicylic acid (DNS) reducing sugar assay (Miller, 1959) was used to confirm activity of Cf-GH16\_3 on laminarin and lichenan. Proteins (20 mM NaCl, 10 mM MOPS pH 7.8) were mixed with 0.5% laminarin or lichenan (20 mM NaCl, 10 mM MOPS pH 7.8) in 200  $\mu\text{L}$  reactions. Samples were incubated overnight at room temperature and therefore experiments correspond to endpoint measurements. The non-catalytic Cf-SGBP and samples without the addition of protein served as controls. One hundred microliter of DNS reagent was added to 100  $\mu\text{L}$  of sample. After 5 min of incubation at 95°C, the reaction was stopped by adding 800  $\mu\text{L}$  of water. Absorbance was measured at 540 nm and increased with catalytic activity. Experiments were performed in triplicates. Absorbance did not increase in Cf-GH16\_3 incubations with lichenan. In this case, samples were further investigated by fluorophore-assisted carbohydrate electrophoresis (FACE). Fifty microliter of the overnight incubations were dried under vacuum and mixed with 2  $\mu\text{L}$  of 0.15 M 8-aminonaphthalene-



1,3,6-trisulfonate (ANTS) and 2  $\mu\text{L}$  of 1 M  $\text{NaBH}_3\text{CN}$ . Samples were incubated at 37°C in the dark for at least 2 h and dried under vacuum. After resuspension in 20  $\mu\text{L}$  of 25% glycerol, 5  $\mu\text{L}$  were loaded onto 27% acrylamide gels and migrated at 200 V and 4°C. Results were visualized under UV.

In small-scale screening, the activity of all PUL-encoded CAZymes on laminarin and laminarin-derived oligosaccharides was investigated by FACE. For this, protein extract of *E. coli* BL21(DE3) producing the respective CAZyme was used. Protein extract of *E. coli* BL21(DE3) carrying the empty vector served as an additional control.

## Isothermal titration calorimetry (ITC)

Proteins were dialyzed against 100 mM NaCl, 10 mM MOPS pH 7.8. The dialysis buffer was used to prepare laminarin and laminarin-derived oligosaccharide solutions, as well as for washing of the ITC cell and for controls. Prior to measurements, samples and solutions were centrifuged for degassing and removal of putative protein aggregates. In addition, protein samples were evaluated by SDS-PAGE and Nanodrop repeatedly. ITC was done on a MicroCal™ ITC 200 machine. Laminarin or laminarin-derived oligosaccharides were injected into 200  $\mu\text{L}$  of protein, which was loaded into the sample cell. 2.5 mM of laminarin was injected into 147.6  $\mu\text{M}$  of Cf-SGBP, into 127.7  $\mu\text{M}$  of Cf-SGBP-CBM102<sub>III/III</sub> or into 94.7  $\mu\text{M}$  of Cf-SGBP-CBM102<sub>III/IV</sub>. 1 mM of laminarin was injected into 147.5  $\mu\text{M}$  of Cf-SGBP-CBM102<sub>IV</sub>. For calculations with laminarin, we used the average molecular weight of laminarin from *Laminaria digitata* (Read et al., 1996). In addition, 5 mM of laminariheptaose (DP7), 5 or 10 mM of laminaripentaose (DP5) and up to 20 mM of laminaribiose (DP2, non-binding, data not shown) were injected into 359.6  $\mu\text{M}$  of Cf-SGBP-CBM102<sub>IV</sub>. Settings were as follows: cell temperature 20°C (293.15 K), reference power 10  $\mu\text{Cal/s}$ , stirring speed 750 rpm, filter period 1 s, injection spacing 300 sec, 1  $\mu\text{L}$  injection volume (first injection 0.3  $\mu\text{L}$ ), 35 injections in total. Experiments were performed in at least triplicates. 2.5 mM and 5 mM of laminarin were injected into 118.8  $\mu\text{M}$  of Cf-SGBP-Ig<sub>a</sub> as a non-interacting control. Titrations of buffer into proteins, laminarin into buffer or buffer into buffer represented additional controls. MICROCAL ORIGIN v. 7 was used to analyse the data. A single-site binding model was selected.

## SEC-SAXS

SAXS data of Cf-SGBP and Cf-GH16\_3 were collected at the Synchrotron SOLEIL on the SWING beamline. Protein samples were centrifuged prior to analyses to

remove putative aggregates. SAXS measurements were coupled with prefixed SEC. The SEC column was equilibrated in buffer C (100 mM NaCl, 10 mM MOPS pH 7.8), which was used for protein purification. Seventy microliter of protein were injected, Cf-SGBP at 8.2 mg mL<sup>-1</sup> and Cf-GH16\_3 at 12.9 mg mL<sup>-1</sup>. The sample-detector-distance was 2.4 m, resulting in a scattering vector  $q$ -range of 0.012–0.504 Å<sup>-1</sup> for Cf-SGBP and 0.017–0.504 Å<sup>-1</sup> for Cf-GH16\_3. The obtained scattering data were normalized and corrected according to standard procedures. The Guinier equation was used to calculate the forward scattering  $I(0)$  and the radius of gyration  $R_g$ . The distance distribution function  $P(r)$  and the maximum particle dimension  $D_{\text{max}}$  were calculated by Fourier inversion of the scattering intensity  $I(q)$  using GNOM integrated in the PRIMUS software (ATSAS 3.1.3) (Manalastas-Cantos et al., 2021). Models of protein envelopes were calculated from the experimental scattering curves using GASBORi (Svergun et al., 2001). High  $\chi^2$  values indicated flexible proteins. Therefore, the data were analysed using DADIMODO (Rudenko et al., 2019). Here, structural data of proteins are included into the analysis to identify an arrangement that best describes the experimental data. For this, rigid bodies and linkers were defined based on the structures predicted by AlphaFold2 within ColabFold (Mirdita et al., 2022). Finally, EOM analysis was used to describe the flexibility of proteins in solution (Bernadó et al., 2007; Tria et al., 2015), again using AlphaFold2-derived coordinates. Experimental and processed data were visualized in R using the ggplot2 package. Figures of models were created using the PyMOL Molecular Graphics System v2.5.4 (Schrödinger, LLC.).

## Crystallography

Crystallization for both recombinant constructs of Cf-SGBP-CBM102<sub>IV</sub> (at a concentration of 11.7 mg mL<sup>-1</sup>) and Cf-SGBP-CBM102<sub>III/IV</sub> (at a concentration of 9.5 mg mL<sup>-1</sup>) was attempted, in the presence or absence of laminarin and laminarin-derived oligosaccharides (DP3/4), but never led to crystals in the case of Cf-SGBP-CBM102<sub>IV</sub> with laminarin or Cf-SGBP-CBM102<sub>III/IV</sub>. Crystallization screening was undertaken with the nanodrop-robot Crystal Gryphon (Art Robbins instruments) starting with sparse matrix sampling kits JCSGplus, PACT, PEGs and Wizard Classic I/II and Ligand-Friendly Screen (from Qiagen and Molecular Dimensions). The initial crystallization conditions were manually optimized and suitable crystals of Cf-SGBP-CBM102<sub>IV</sub> in the presence of laminarin-oligosaccharides were obtained using the hanging drop vapour diffusion method as follows: 2  $\mu\text{L}$  of protein were mixed with 1  $\mu\text{L}$  of reservoir solution (500  $\mu\text{L}$ ) containing 2.5 M ammonium sulphate, 0.1 M BIS-TRIS propane pH 7.0, 6% ethanol and 4% polyethylene glycol (PEG)



6000. Diffraction data were collected on PROXIMA1 (SOLEIL synchrotron) with a single crystal of Cf-SGBP-CBM102<sub>IV</sub> in complex with laminarin-trisaccharide at 1.9 Å resolution. The data were integrated using XDS (Kabsch, 2010) and scaled with aimless (Evans & Murshudov, 2013). The structure was solved by molecular replacement with Phaser (McCoy et al., 2007) using the AF model as a starting point. The crystal structure was further enhanced with alternating cycles of refinement with REFMAC (Murshudov et al., 2011) and manual construction using COOT (Emsley et al., 2010). All further data collection conditions and refinement statistics are given in Table S1. Figures of structures were created using the PyMOL Molecular Graphics System v2.5.4 (Schrödinger, LLC.).

## Phytoplankton bloom data

Chlorophyll *a* concentration as well as metagenomes, metagenome-assembled genomes and metaproteomes were obtained from samples taken during spring phytoplankton blooms at the North Sea island of Helgoland in 2016 (Francis et al., 2021), 2018 (Wang et al., 2024) and 2020 (Beidler et al., 2023; Sidhu et al., 2023). Chlorophyll *a* data were obtained from the Helgoland Roads time series (Armonies et al., 2018) and have been published for the respective years (Beidler et al., 2023). Metagenome sequencing, assembly and binning (MAGs) are described in Francis et al. (2021) (2016), in Wang et al. (2024) (2018) and in Sidhu et al. (2023) (2020). The preparation of metaproteome samples (0.2 µm fraction) as well as corresponding LC-MS/MS measurements and data analyses have been described in detail in Francis et al. (2021) (2016) and in Beidler et al. (2023) (2018 and 2020). Metaproteome data are available in the PRIDE archive (Perez-Riverol et al., 2022) (see data availability statement).

## Comparative genomics

MAGs from 2016, 2018 and 2020 were dereplicated using dRep v1.14.6 (Olm et al., 2017) to prevent redundancy, applying a minimum completeness of 70% and contamination lower than 5% at 0.95 ANI (average nucleotide identity). For each representative MAG ( $n = 555$ ), protein-coding sequences for representative MAGs were predicted with Prokka v1.14.6 using default settings (Seemann, 2014). PULs, CAZymes, CBMs and SusC/D-like proteins were predicted as described previously (Francis et al., 2021), using hmmscan v3.3.2 against dbCAN-HMMdb-V1 and DIAMOND BLASTp v2.1.1.155 (Buchfink et al., 2021) against CAZyDB.08062022 provided by dbCAN (Zheng et al., 2023). CBM102- and CBM103-containing sequences within representative MAGs and additional draft

genomes of previously published 53 North Sea *Bacteroidota* (Kappelmann et al., 2019) were identified using hmmscan as described above, adding corresponding HMM profiles to the database. Results were further filtered using the hmmscan-parser.sh script from dbCAN with an e-value cutoff of  $1E-15$  and a minimum coverage of 35%. The distribution of domain organization was visualized with UpSetR (Conway et al., 2017; Lex et al., 2014). The official CAZy DB classification was applied to semi-manually determine the CBM boundaries, the multimodular compositions, the taxonomic distributions in CAZy and the PUL prevalence and partnerships in PULDB.

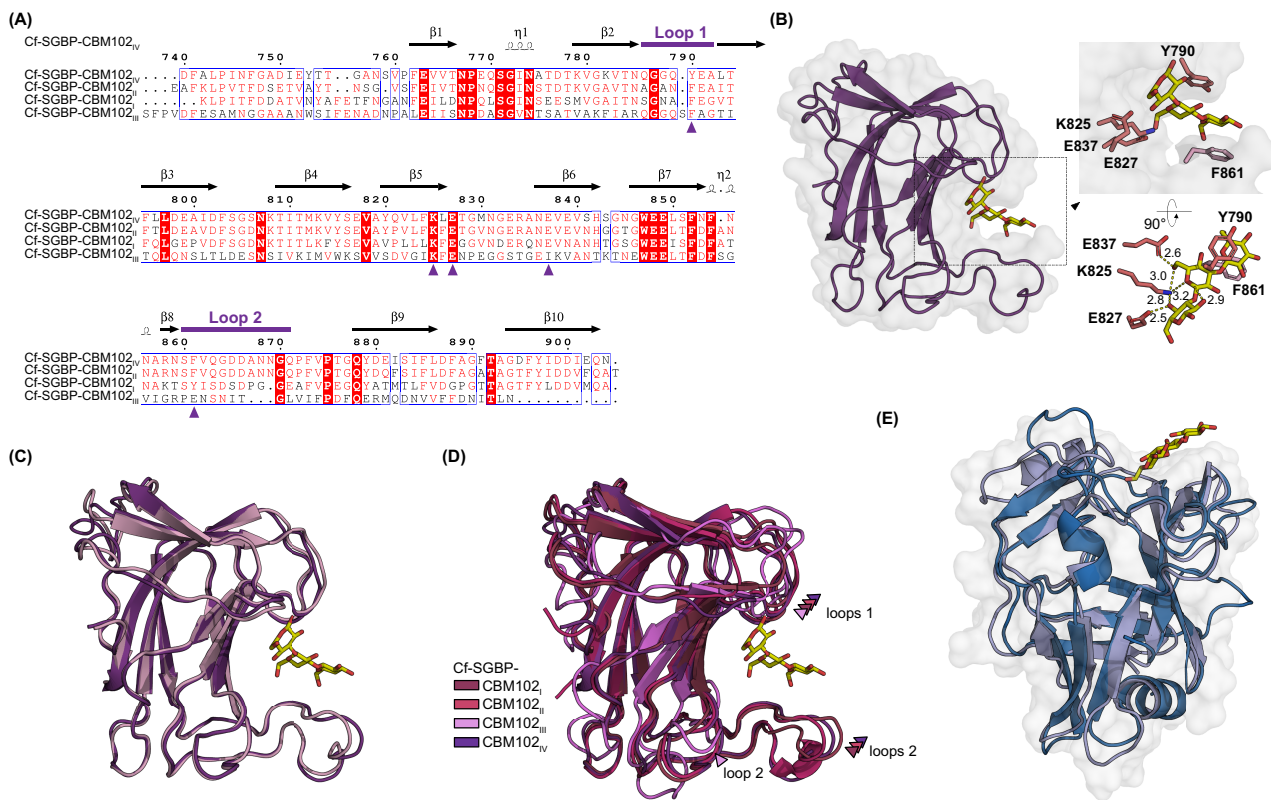
## RESULTS

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

*C. forsetii* grows to similar optical densities on laminarin relative to glucose, though with a shorter lag-phase (Figure 1A). In-depth subproteome analyses had already shown that the production of laminarin PUL proteins was increased in laminarin-grown *C. forsetii* compared to alginate-grown *C. forsetii* (Kabisch et al., 2014). The *C. forsetii* laminarin PUL encodes a regulator, a SusD-like protein/SusC-like TBDT transport complex, a putative CBM-containing SGBP, two GH16\_3s and a GH3 (Figure 1B). We investigated the SGBP (Cf-SGBP, locus tag GFO\_RS17395) and a protein containing a putative CBM and a GH16\_3 catalytic module (Cf-GH16\_3, locus tag GFO\_RS16360) (Figure 1C, further information on domain nomenclature is provided in Section 2.3), which are encoded downstream of the SusC/D-like pair (Figure 1B). Subproteome analyses (Kabisch et al., 2014) and the presence of a lipoprotein signal peptide support that Cf-SGBP and Cf-GH16\_3 are tethered to the outer membrane. Structural prediction using AlphaFold2 (AF) within ColabFold (Mirdita et al., 2022) and Phyre2 (Kelley et al., 2015) uncovered that Cf-SGBP displays two Ig-like fold ( $\beta$ -sandwich) modules at its N-terminus followed by four putative CBMs (I–IV) (Figure 1C) with high sequence similarity (47%–84%), suggesting internal duplication events (Figure 2A). Cf-GH16\_3 on the other hand is predicted to feature a single N-terminal Ig-like folded module followed by a putative CBM, with no significant similarity to CBMs of Cf-SGBP, and the C-terminal GH16\_3 catalytic domain.

All putative CBM sequences were submitted to Dali searches (Holm, 2022), which did not reveal any close relatives with solved 3D structures. The four CBMs from Cf-SGBP were only remotely similar to a module attached to a GH50 from *Saccharophagus degradans* (PDB ID 4BQ2) (Pluvinage et al., 2013), with a z-score >12 and 10%–15% identity. To date, this module is not





**FIGURE 2** Multiple CBM102 domains of Cf-SGBP and CBM103 of Cf-GH16\_3: binding clefts versus platform. (A) Alignment of Cf-SGBP-CBM102<sub>s</sub> visualized using ESPrnt 3.0 (<https://esprnt.ibcp.fr>) (Robert & Gouet, 2014). Identical residues are highlighted in red boxes, and similar residues using red letters. Residues involved in substrate binding in Cf-SGBP-CBM102<sub>IV</sub> are indicated by violet arrowheads; the two loops flanking the substrate binding groove are indicated by violet lines above the alignment. (B) 3D crystal structure of Cf-SGBP-CBM102<sub>IV</sub> with laminaritriose, highlighting interacting residues. (C) Overlay of the crystal structure of Cf-SGBP-CBM102<sub>IV</sub> (violet) and the corresponding AF prediction (light pink), supporting the accuracy of AF. (D) Overlay of AF-predicted CBM102<sub>s</sub> (I–III) with the crystal structure of Cf-SGBP-CBM102<sub>IV</sub>. In particular, Cf-SGBP-CBM102<sub>III</sub> differs in one of the loops delimiting the binding groove, which results in a flat binding region flanked only by one loop (see also Figure S1). (E) CBM103 of Cf-GH16\_3 displays a  $\beta$ -barrel structure, missing a distinct binding cleft compared to CBM102. The AF-predicted model (blue) superimposes very closely to a CBM of an SGBP from *Bacteroidetes fluxus* (Tamura et al., 2021) binding laminarin and mixed-linkage  $\beta(1,3)/\beta(1,4)$ -glucan on a platform (light blue, here laminaritriose, PDB ID 7KV7, see also Figure S2).

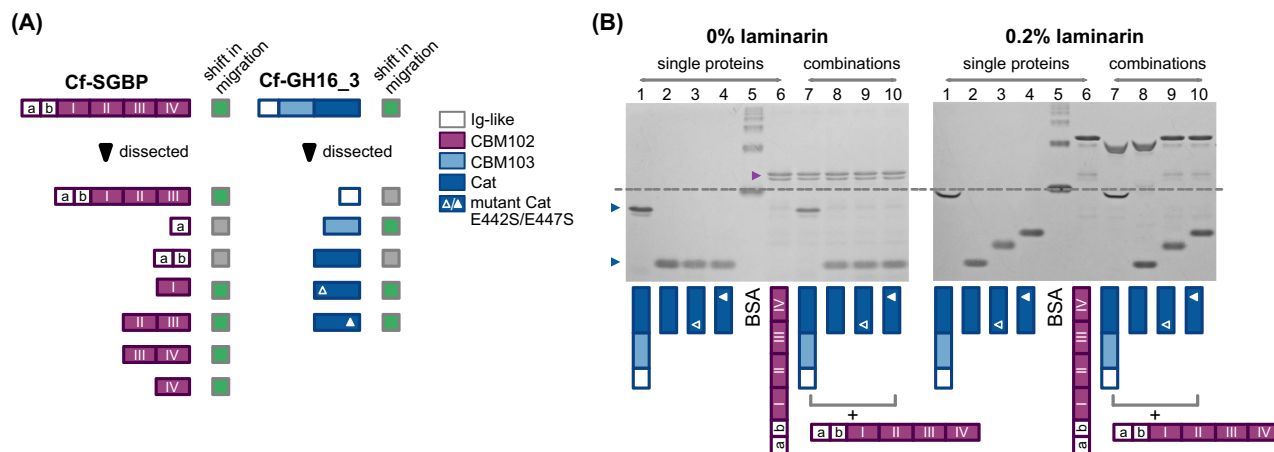
classified in any CBM family in the CAZy DB classification, due to the absence of a characterized close homologue. Similarly, the CBM of Cf-GH16\_3 returned uncharacterized proteins, including hypothetical CBMs, as best hits with z-scores  $\sim 8$  and 9%–12% sequence identity (PDB ID 4L3R and 4QRL).

Sequence analysis confirmed that the putative *C. forsetii* CBMs did not exhibit any domains belonging to a known CBM family. Hence, the functional characterization of the CBMs of Cf-SGBP and Cf-GH16\_3, described herein, provides the basis for founding the novel CAZy DB (Drula et al., 2022) CBM families CBM102 and CBM103, respectively. Using X-ray crystallography, we determined the 3D structure at 2.0 Å resolution for the isolated C-terminal CBM102, Cf-SGBP-CBM102<sub>IV</sub>, in complex with laminarin-derived oligosaccharides (Figure 2B, Table S1). Least square superimposition of the corresponding AF predicted structure to the 3D crystal structure using COOT (Emsley et al., 2010) and matching all atoms leads to an RMSD of 1.765. This high agreement (Figure 2C)

supported reliability for the AF predictions for the other CBM102s, which further indicated two things: the four CBM102s have a highly similar fold and CBM103 differs structurally from CBM102.

The crystal structure of Cf-SGBP-CBM102<sub>IV</sub> reveals a typical CBM  $\beta$ -sandwich fold, consisting of two antiparallel  $\beta$ -sheets composed of four and five  $\beta$ -strands that form a concave and convex face, respectively (Figure 2B). The concave face forms a deep and narrow binding cleft, also established by two expansive loops. While the concave face contains residues mediating polar interactions with the substrate, distal loops contain aromatic residues (Figure 2B) to establish CH- $\pi$  interactions (Hudson et al., 2015). AF predicted a similar fold for all four CBM102s (Figure 2D) and residues that mediate interaction with laminaritriose in the Cf-SGBP-CBM102<sub>IV</sub> crystal structure align (Figure 2A). Remarkably, differences are seen in Cf-SGBP-CBM102<sub>III</sub> loop 2, which results in a wider and very shallow binding cleft (Figure 2A,D, Figure S1).

Instead of the  $\beta$ -sandwich fold of CBM102, CBM103 displays a  $\beta$ -barrel structure with two additional



**FIGURE 3** Identification of laminarin-binding domains in Cf-SGBP and Cf-GH16<sub>3</sub>. (A) Summary of AGE runs, where green boxes indicate binding, which was concluded from shifts in corresponding gels in Figure 3B, Figures S3 and S4. (B) AGE of Cf-GH16<sub>3</sub> revealed catalytic activity during runs. Cf-GH16<sub>3</sub>, the dissected catalytic module Cf-GH16<sub>3</sub>-Cat and two mutants thereof (Cf-GH16<sub>3</sub>-Cat<sub>E442S</sub> and Cf-GH16<sub>3</sub>-Cat<sub>E447S</sub>) were loaded individually (lanes 1–4) or in combination with the non-catalytic Cf-SGBP (lanes 7–10), compared to Cf-SGBP alone (lane 6). While a shift in migration in laminarin gels indicated binding (using BSA as a control, lane 5), an altered ‘moustache’-like band shape of CBM-containing proteins in laminarin gels (lanes 1, 7 and 8) revealed the presence of GH16<sub>3</sub> activity within the lane (this activity masks the migration shift for the dissected catalytic module Cf-GH16<sub>3</sub>-Cat, which lacks CBM103 compared to the full-length protein, lanes 2 and 8). Detailed information is provided in the main text.

$\alpha$ -helices and one additional  $\beta$ -strand pair. Together with a binding platform instead of a groove, the *C. forsetii* CBM103 is structurally related to a CBM present in a *Bacteroidetes fluxus* SGBP (Figure 2E, Figure S2; Tamura et al., 2021) devoid of a catalytic GH16<sub>3</sub>. Our analyses reveal that the *B. fluxus* CBM should be classified as a CBM103.

### CBM102 and CBM103 bind laminarin

Affinity gel electrophoresis (AGE) demonstrated that CBM102 and CBM103 bind laminarin (Figure 3A, Figures S3 and S4). By comparison, laminarin-binding was not detected for the N-terminal Ig-like modules (Figure 3A, Figures S3 and S4), which was confirmed by isothermal titration calorimetry (ITC) in the case of Cf-SGBP-Ig<sub>a</sub> (data not shown). In addition to laminarin-binding by CBM102 and CBM103, both PUL-encoded GH16<sub>3</sub>s (Figure 1B) were active on laminarin (Figure S5). The CBM103-containing Cf-GH16<sub>3</sub> hydrolyzed laminaritriose (DP3) releasing laminaribiose (DP2) and glucose as final products (Figure S5).

In our AGE experiments, this catalytic activity degrades the laminarin in the gel lanes while migrating, leaving residual laminarin in between the affinity gel lanes. The CBM-containing polypeptides (Figure 3B, lanes 1, 7 and 8) can bind to the residual laminarin in between the gel lanes resulting in a ‘moustache-like’ band appearance. Importantly, we here distinguish the ‘moustache effect’ from the more typical ‘smiling’ effect caused by uneven temperatures within a gel. Smiling can generally be resolved by decreasing the voltage/

temperature of a gel. In our case, the moustache effect is localized to specific gel lanes where there is a combination of substrate, catalytic activity and ligand binding. Consequently, laminarin degradation by the dissected catalytic module Cf-GH16<sub>3</sub>-Cat, which lacks CBM103 compared to the full-length protein Cf-GH16<sub>3</sub>, masked the migration shift (Figure 3B, lanes 2 and 8). By comparison, inactivated mutants of the catalytic module were retained by laminarin (Figure 3B, lanes 3 and 4) and did not induce the altered band-shape in the co-loaded Cf-SGBP (Figure 3B, lanes 9 and 10), as they do not degrade laminarin in these lanes.

In addition to laminarin binding, AGE also demonstrated binding of lichenan, a mixed-linkage  $\beta(1,3)/\beta(1,4)$ -glucan (MLG) from Icelandic Moss, by Cf-GH16<sub>3</sub>-CBM103 (Figure S6). This was in agreement with the activity of Cf-GH16<sub>3</sub> on lichenan; however, our analysis indicated that the activity on lichenan was much lower compared to laminarin (Figure S7a,b).

### Multiple CBM102 domains in Cf-SGBP: Is ‘more’ really ‘better’?

While the *C. forsetii* Cf-GH16<sub>3</sub> contains a single CBM103 to bind laminarin, Cf-SGBP contains four consecutive CBM102s. AGE did not provide indications on whether the higher number of CBMs in Cf-SGBP translates into a higher binding efficiency (Figure S3). To explore the putative benefits of multiple CBM102 domains in more detail, we investigated laminarin binding by ITC. We selected the full-length Cf-SGBP, the tandem constructs Cf-SGBP-CBM102<sub>I/III</sub> and Cf-SGBP-CBM102<sub>III/IV</sub>,




**TABLE 1** Affinity of laminarin and its oligosaccharides (DP7 and DP5) to Cf-SGBP and/or dissected modules thereof.

<b>Cf-SGBP and selected modules—laminarin</b>					
	$K_a$ ( $M^{-1}$ )	$\Delta H$ (kcal mol $^{-1}$ )	$\Delta S$ (cal mol $^{-1}$ K $^{-1}$ )	$N$ (sites)	$\Delta G$ (kcal mol $^{-1}$ )
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-CBM102 <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-CBM102 <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-CBM102 <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-CBM102<sub>IV</sub>—laminarin-derived oligosaccharides (DP7 and DP5)</b>					
	$K_a$ ( $M^{-1}$ )	$\Delta H$ (kcal mol $^{-1}$ )	$\Delta S$ (cal mol $^{-1}$ K $^{-1}$ )	$N$ (sites)	$\Delta G$ (kcal mol $^{-1}$ )
Cf-SGBP-CBM102 <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-CBM102 <sub>IV</sub> DP5	$1.39 (\pm 0.07) \times 10^4$	$-9.14 \pm 0.18$	$-12.20 \pm 0.70$	$0.61 \pm 0.04$	$-5.56 \pm 0.04$

Note: Mean values and standard deviation ( $\pm$ ) of ITC experiments ( $n \geq 3$ ). Measurements were performed at 293.15 K. For each run,  $\Delta G$  was calculated (Gibbs–Helmholtz equation,  $\Delta G = \Delta H - T\Delta S$ ).

as well as the single module Cf-SGBP-CBM102<sub>IV</sub>, for which we had a 3D crystal structure, to cover a range of CBM102 repetitions for comparisons. Other constructs were either insufficiently stable or insufficiently produced and thus unfortunately not available for ITC measurements.

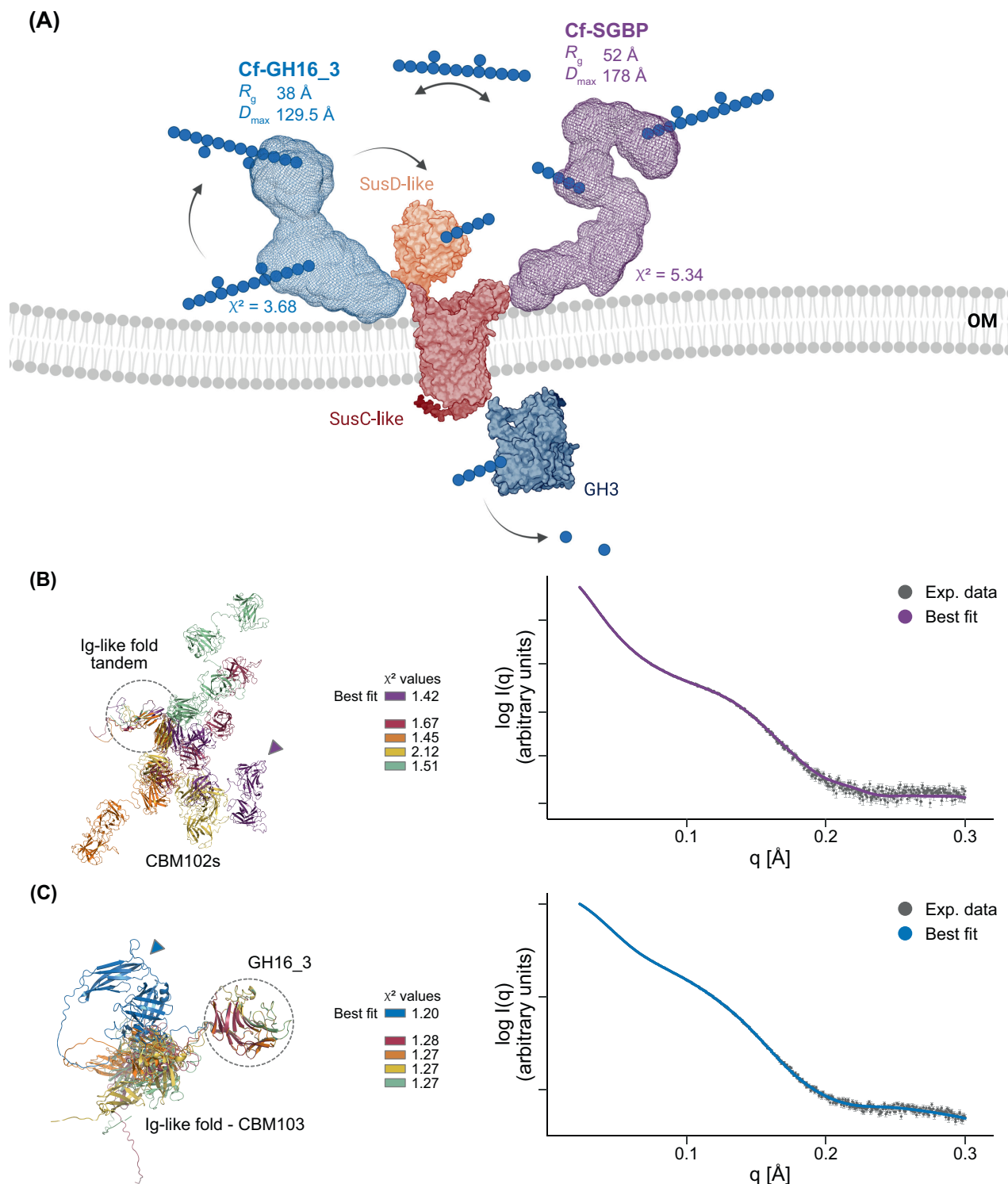
ITC confirmed laminarin-binding for the selected proteins (Table 1, Figure S8a). Unexpectedly, the highest affinity for laminarin was determined for the individual Cf-SGBP-CBM102<sub>IV</sub> ( $1.04 \times 10^5 M^{-1}$ ), which exceeded affinity of the other constructs by an order of magnitude (Table 1) and is in the same range as described for the N-terminal CBM6 of a GH16\_3 (LamC) in *Zobellia galactanivorans* Dsij<sup>T</sup> (Jam et al., 2016). Using non-linear curve fitting and a single-site binding model, the stoichiometry ( $n$ ) was determined from the molar ratio at the inflection point of the binding isotherm. A binding stoichiometry of close to 0.5 indicates two CBM102<sub>IV</sub>s are able to interact with one ligand molecule. The binding stoichiometry is increased to  $\sim 0.75$  for the double module constructs and the full-length Cf-SGBP binds 1.6 laminarin molecules per polypeptide chain, though no significantly increased affinity was seen. Instead, affinities of the full-length Cf-SGBP and the two CBM102 tandems were relatively similar (Cf-SGBP  $3.09 \times 10^4 M^{-1}$ , Cf-SGBP-CBM102<sub>II/III</sub>  $2.76 \times 10^4 M^{-1}$  and Cf-SGBP-CBM102<sub>III/IV</sub>  $3.63 \times 10^4 M^{-1}$ ). We speculate that the increased affinity of the individual CBM102<sub>IV</sub> for laminarin relative to the tandem CBMs is provided by the absence of steric/spatial inhibition of binding from the other CBM102. Thus, in the absence of the other modules, two individual CBM102<sub>IV</sub> are better accommodated on the same laminarin chain. Conversely, the steric hindrance of successive CBMs in Cf-SGBP may contribute to partial binding site inaccessibility. Furthermore, the longer constructs pay more entropic penalties than the single Cf-SGBP-CBM102<sub>IV</sub>, probably related to the loss of conformational freedom upon ligand binding. Notably, Cf-SGBP-CBM102<sub>IV</sub> also bound laminarin-derived

oligosaccharides DP7 with a higher affinity than DP5, although with a lower affinity than laminarin (Table 1, Figure S8b). This is coherent with increased affinity of the individual Cf-SGBP-CBM102<sub>IV</sub> on laminarin, as it suggests that this CBM can accommodate more than DP7 in the ligand binding site. For all interactions, binding was enthalpy-driven, with an entropic penalty (Table 1).

## Two elongated multidomain proteins allow for laminarin recognition and scavenging

The multidomain structure of the *C. forsetii* proteins and corresponding ITC experiments for the Cf-SGBP constructs raised the question about the spatial organization of their polypeptide chains. AF predictions were highly uncertain for the linkers connecting the individual modules. Indeed, in Cf-SGBP the linkers between the CBMs consisted of 5, 14 and 15 amino acids among which 3, 7 and 6 are glycine residues, respectively, while the linker between the catalytic domain and the CBM in Cf-GH16\_3 is 19 residues long (see Table S2 for linker compositions). Therefore, we analysed Cf-SGBP and Cf-GH16\_3 by small-angle X-ray scattering (SAXS), which revealed that both multidomain proteins feature an elongated structural organization (Figure 4A, Figure S9a–f, Table S3).

The values for the radius of gyration ( $R_g$ ) and the maximum particle distance ( $D_{max}$ ) derived from the scattering curve are  $R_g$  52 Å and  $D_{max}$  178 Å for Cf-SGBP and  $R_g$  38 Å and  $D_{max}$  129.5 Å for Cf-GH16\_3 (Table S3, Figure S9a,c,d,f). The elongated structures were confirmed by the GASBORi (Svergun et al., 2001) *ab initio* calculated envelopes (Figure 4A, Figure S10a,b). All calculations for Cf-GH16\_3 returned a bent shape, while much more variable shapes were calculated for Cf-SGBP (Figure S10a,b). However, the relatively high  $\chi^2$  values (all >3) of calculated individual shapes for both proteins potentially indicated that the



**FIGURE 4** Two elongated multidomain proteins to sense and capture laminarin. (A)  $R_g$  and  $D_{max}$  values calculated from the experimental SAXS data revealed an elongated structure for Cf-SGBP and Cf-GH16\_3, which is also reflected in two calculated models for protein envelopes using GASBORi. The figure was created with BioRender. GH3 activity has been previously suggested (Kabisch et al., 2014). The spatial organization of domains in (B) Cf-SGBP and (C) Cf-GH16\_3 was investigated using DADIMODO. Shown are five models (left graphs) in which domains used for superimposition are circled. The best fit (lowest  $\chi^2$ ) is highlighted with an arrowhead and was plotted against the experimental data (right panels). Further information on domains is provided in Figure 1C.

experimental curves were not well fitted by a single envelope, which implicated possible conformational variability. Nevertheless, using DADIMODO, spatial

arrangements of the successive domains based on those predicted by AF could be determined for both Cf-SGBP and Cf-GH16\_3 that individually fitted the


**TABLE 2** Summary of the EOM analyses to determine the flexibility of Cf-SGBP and Cf-GH16\_3 in solution.

	$R_g$ (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

Note: Given are the final  $R_g$  and  $D_{max}$  derived from the unique models for the selected ensembles, averaged  $R_g$  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% or  $R_{sigma}$  of 1 corresponds to a fully flexible system.

experimental curve (Figure 4B,C, right panels). Superimposition of the individual models, however, covered variable arrangements (Figure 4B,C, left panels). To further analyse the potential conformational flexibility of these proteins in solution, we used the ensemble optimization method (EOM) (Bernadó et al., 2007; Tria et al., 2015). Here, an ensemble is selected from a large pool of theoretical conformers, for each of which a corresponding theoretical scattering curve is generated. The selected ensembles for both proteins were able to fit the experimental data, but in contrast to Cf-GH16\_3, the  $\chi^2$  value of the EOM fit of Cf-SGBP (10.85 versus 4.57 for Cf-GH16\_3, Figure S11a,b) was much higher than  $\chi^2$  values of the GASBORi (Figure 4A, Figure S10a,b) or DADIMODO single fits (Figure 4B,C), indicative of lower conformational flexibility. Despite the presence of several glycine residues, the shorter linkers in Cf-SGBP apparently induce less flexibility or variable conformers. In addition, the calculated  $R_{flex}$  for the selected ensemble of Cf-SGBP (77%) was lower compared to that of Cf-GH16\_3 (81%), although the  $R_{flex}$  of the pool and the final  $R_{sigma}$  were almost the same (Table 2).  $R_{flex}$  and  $R_{sigma}$  are values to describe the flexibility in the EOM approach, where 100% or 1 corresponds to a fully flexible system, respectively. Finally, a higher flexibility of Cf-GH16\_3 was supported by shifts to higher  $R_g$  and  $D_{max}$  values for the selected ensemble compared to the pool, whereas they were biased to lower values for Cf-SGBP (Figure S12a–d).

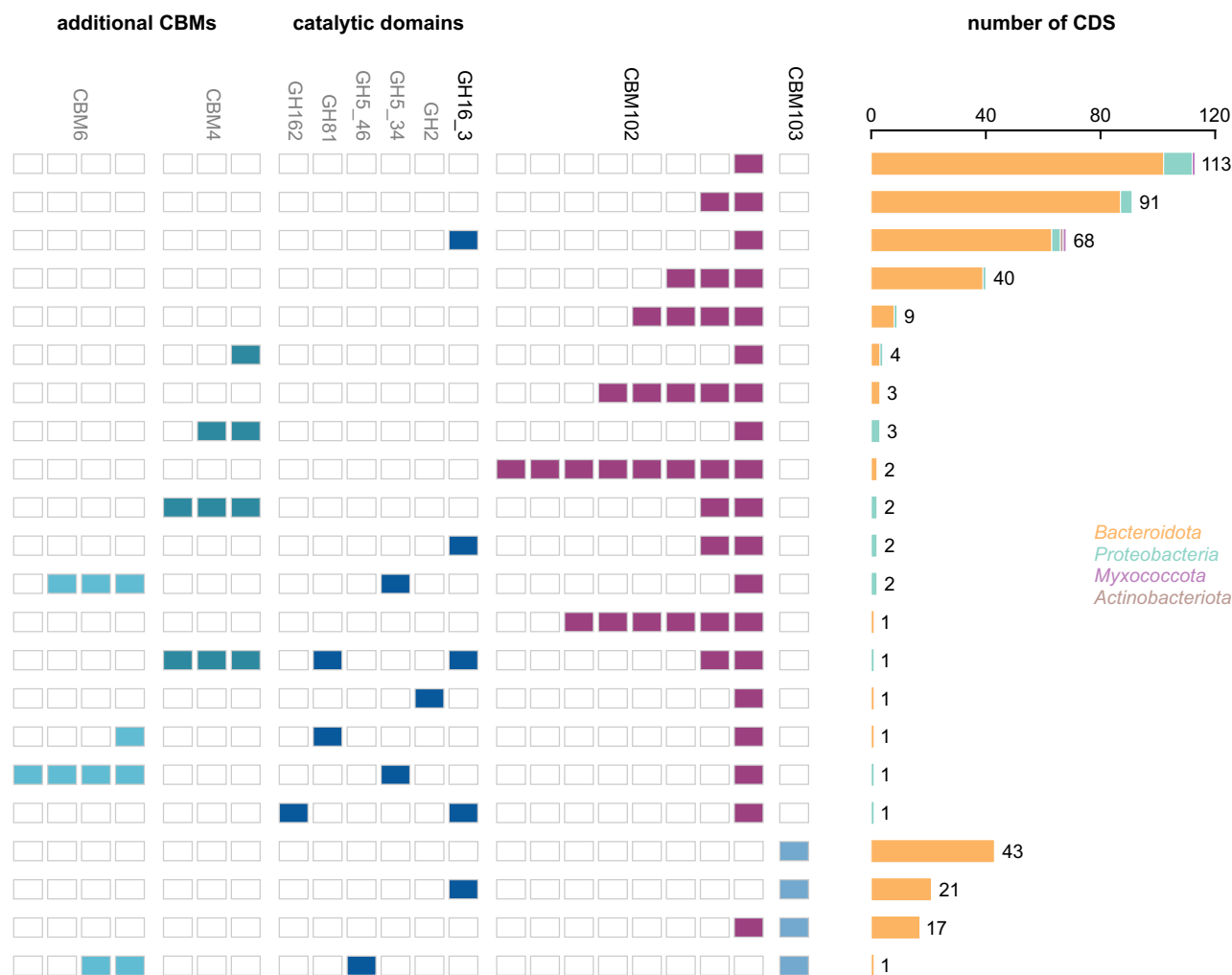
### CBM102 and CBM103 are widely distributed in marine *Bacteroidota* that thrive during phytoplankton blooms

Screening of bacterial metagenome-assembled genomes (MAGs) obtained from North Sea phytoplankton bloom-associated microbial biomass samples at Helgoland Roads in 2016, 2018 and 2020, yielded 427 sequences coding for CBM102- and CBM103-containing proteins (555 representative MAGs of which 201 belonged to *Bacteroidota*). The proportion of CBM102-containing proteins was higher (362) compared to CBM103 (82), whereas only 17 sequences contained both modules (Figure 5). We found a total of 22 domain combinations, in which CBM102 and/or CBM103 were detected in protein-coding sequences with or without known catalytic modules, for example, linked to CBM4, CBM6, GH16\_3 or rare examples of GH2, GH5\_34, GH5\_46,

GH81 and GH162. All these modules have known  $\beta$ -glucosidase or  $\beta$ -glucanase activity or bind  $\beta$ -glucans, with the exception of the xylanase GH5\_34. The most frequent organizations included CBM102- or CBM103-only-proteins (72% for CBM102 with up to eight repeats; 52% for CBM103, always alone) followed by CBM102 or CBM103 associated with a GH16\_3 (20% for CBM102, 26% for CBM103). In addition, corresponding sequences were frequently co-localized with genes coding for known laminarin-active CAZymes, for example, GH16\_3, GH17 or GH30 family members, but also with *susCD*-like genes. Similar pictures of domain organization and genomic context for both CBMs emerged from screening draft genomes of 53 published North Sea *Bacteroidota* (Kappelmann et al., 2019), including *C. forsetii*, and using the CAZy PULDB for both CBMs (as of May 2023). A total of 92 CBM102- and/or CBM103-containing sequences were detected in 38 of the strains from the North Sea (Figure S13) and >400 in over ~2000 *Bacteroidota* genomes with additional relevant partners such as GH3, GH149 and GH5\_46 using the CAZy PULDB. These GH families are frequently encoded in laminarin PULs (Kappelmann et al., 2019; Mystkowska et al., 2018; Unfried et al., 2018). While GH3 and GH5\_46 have  $\beta$ -glucosidase and/or  $\beta$ -glucanase activity, GH149 is a  $\beta$ -1,3-glucan phosphorylase (Kuhaudomlarp et al., 2018). In the bloom data set, CBM102 was predominant in *Bacteroidota*, but also occurred in *Proteobacteria* or as rare exceptions in *Actinobacteriota* and *Myxococcota*, whereas CBM103 was exclusive to *Bacteroidota* (Figure 5). This was confirmed by analysis of the CAZy DB, which additionally showed that CBM102 and CBM103 occur also in environments other than marine. Structural deviations that we have already discovered in CBM102s of Cf-SGBP are also reflected in sequences from bacterial MAGs (Figures S14 and S15). Nevertheless, residues mediating polar interactions with the substrate in Cf-SGBP-CBM102<sub>IV</sub>, and also some structural residues, were highly conserved (Figure S15). Since not all putative non-catalytic proteins may be attached to the outer membrane to function as CBM-containing SGBPs, we will henceforth use the term (S)GBPs.

### Protein abundance of the novel CBMs correlated with the course of the bloom

Metaproteome analyses showed that the overall production of CBM102- and CBM103-containing proteins

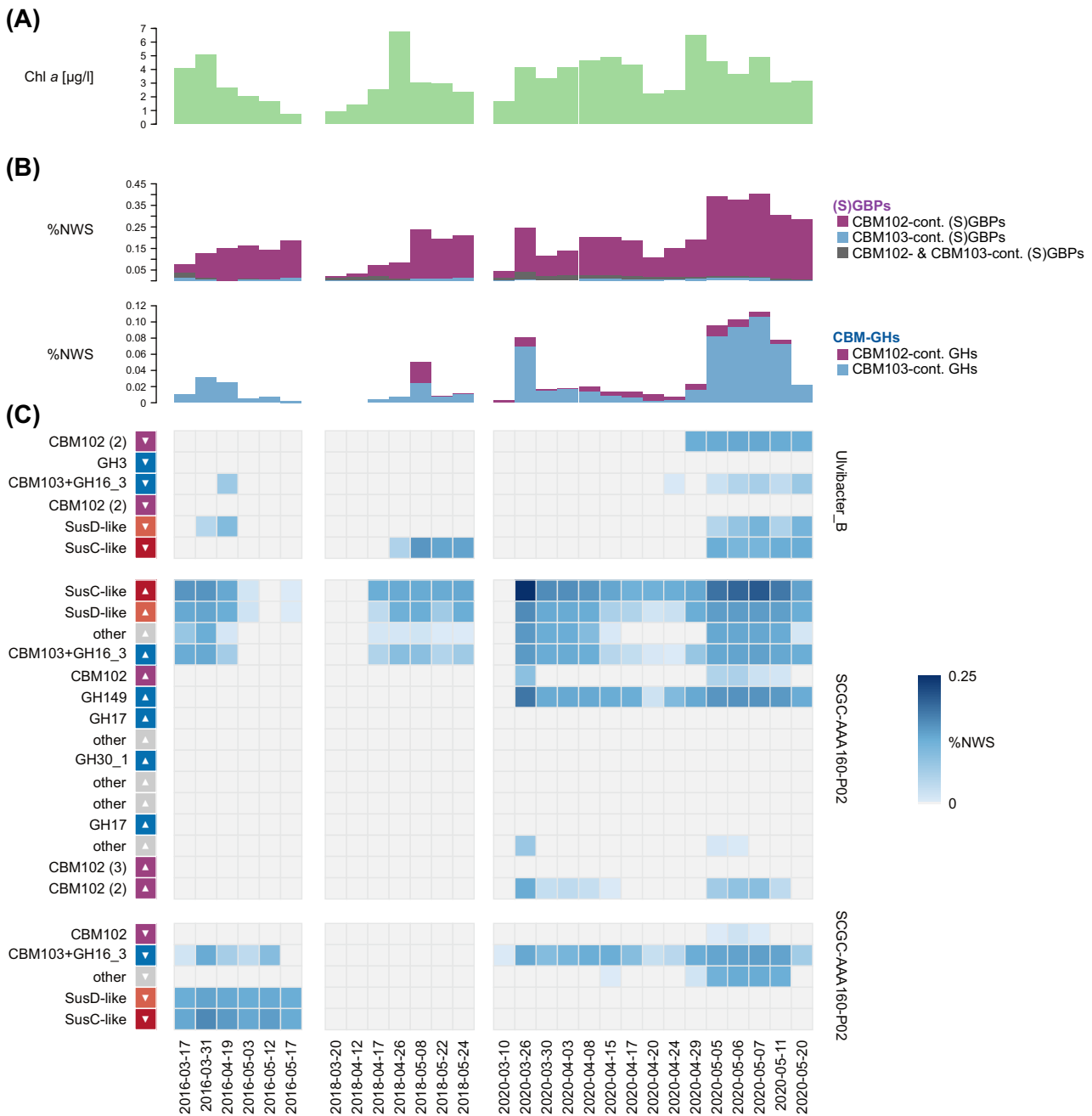


**FIGURE 5** Distribution of CBM102 and CBM103 in proteins encoded in bacterial MAGs from algae blooms in the North Sea. Overview of detected CBM102- and CBM103-coding sequences and their co-occurrence with CBMs and catalytic modules in multimodular proteins as well as the frequency of occurrence of detected domain combinations and the assigned taxonomy. The graph also depicts the multiplicity of certain modules, where each tile represents a detected domain. Note that this graph does not represent the organization of domains within the polypeptide chain. While CBM102 crossed phyla, CBM103 was restricted to *Bacteroidota* in this data set.

correlated with bloom intensities (Figure 6A,B), as assessed by chlorophyll *a* measurements (Beidler et al., 2023). CBM-containing (S)GBPs were more prevalent than CBM-containing CAZymes (mainly GH16\_3s), possibly due to the higher number of putative (S)GBP sequences. Altogether, CBM102- and CBM103-containing proteins accounted for 26%–50% of all detected CAZymes and (S)GBPs in metaproteomes, and many of them were detected in high quantities in this fraction (Figure S16). In addition, the production of CBM-containing (S)GBPs remained more or less stable after chlorophyll *a* peaks. Looking at specific laminarin PULs derived from MAGs, CBM102-containing (S)GBPs, but also CBM103-containing GH16\_3s were detected with high coverage, while other PUL-encoded GHs were often not captured (e.g., GH3, GH17 or GH30\_1) (Figure 6C).

## DISCUSSION

Elevated photosynthetic primary production during marine phytoplankton blooms leads to pronounced increases in available substrates and nutrients for marine bacteria. These include a plethora of polysaccharides where laminarin is particularly abundant (Becker et al., 2020). The surface of many bloom-associated bacteria is therefore equipped with specific sugar-binding and sugar-degrading proteins to ensure the recognition and docking of target substrates to the cellular surface. This allows the first steps of enzymatic catalysis but also prevents the loss of small-sized oligosaccharides that need to be shuttled into the cell. The successful and specific binding of poly- and oligosaccharides is therefore a key process in bacterial polysaccharide degradation that can only be understood by detailed analyses of the CBMs



**FIGURE 6** Abundance of CBM102- and CBM103-containing proteins correlates with the course of phytoplankton blooms. Data show chlorophyll *a* measurements (Beidler et al., 2023) and results of semi-quantitative metaproteome analyses (0.2 µm bacterial fractions) from three respective years, see Section 2.11. (A) Course of the bloom illustrated using the chlorophyll *a* concentration. (B) Summed relative protein abundance values (%NWS, normalized weighted spectra) for CBM102- or CBM103-containing proteins. The upper graph depicts putative (S)GBPs, containing CBMs only, and the lower graph depicts CBM-containing catalytic proteins. Please note different scales (%NWS) for both graphs. (C) Relative abundance of proteins (%NWS) encoded by three selected laminarin PULs (arrowhead upwards: sense strand, arrowhead downwards: antisense strand). The number of respective domains is indicated in brackets. Grey tiles in protein abundance plots correspond to proteins that were not quantified.

involved, as exemplified in this study of the two novel laminarin-binding families CBM102 and CBM103.

### CBM102 and CBM103 bind laminarin but belong to different 3D folds

CBM102 and CBM103 both bind to laminarin but differ substantially in sequence and structure. All four

CBM102s of Cf-SGBP exhibit a  $\beta$ -sandwich fold that provides a distinct binding cleft. At the same time, differences in the surrounding loops result in variations in the wideness and depth of the cleft. While CBM102<sub>I/III/IV</sub> display a deep and narrow cleft, it is much shallower in CBM102<sub>III</sub>. This is consistent with observations in other multidomain proteins with multiple CBMs of the same family, for example, the CBM22 tandem in a xylanase from *Paenibacillus barcinonensis* (Sainz-Polo et al.,





2015). Here, a wider groove in one of the CBM22 units was suspected to allow for the binding of larger decorations. Screening of MAGs of bloom-associated bacteria for CBM102 sequences revealed similar structural variations that may affect either binding specificity or affinity, which will be addressed in the future.

In contrast to the common  $\beta$ -sandwich fold of CBM102, the  $\beta$ -barrel fold of CBM103 is rather rare in CBMs. The latter has recently been described in a CBM-containing SGBP from *B. fluxus*. At that time, this CBM could not be classified as a novel family because it was not attached to a catalytic module. It can now be classified as CBM103, as this module is attached to a GH16\_3 catalytic module encoded in a laminarin-specific PUL of *C. forsetii* and in metagenome sequences from our environmental samples. Interestingly, in addition to laminarin, Cf-GH16\_3-CBM103 was also able to bind lichenan, as previously observed for the *B. fluxus* SGBP (MLG from Barley) (Tamura et al., 2021). Moreover, Cf-GH16\_3-CBM103 lacks a distinct cleft. In the case of the CBM103-containing SGBP of *B. fluxus*, binding of  $\beta$ -glucan termini—including laminarin and MLG-derived oligosaccharides—occurs on the top of the  $\beta$ -barrel, in a very shallow cleft or platform (Tamura et al., 2021), which we likewise suggest for the *C. forsetii* protein (Figure 2E, Figure S2). This binding site, distally positioned within the loop regions connecting the successive  $\beta$ -strands of the barrel core, is to some extent comparable to that of some CBM6 (Henshaw et al., 2004) or CBM32 (Ficko-Blean et al., 2012) members, and of several lectin classes (<https://unilectin.unige.ch/unilectin3D/>). In summary, CBM102 classifies as a type B CBM, probably with more or less possibility to accommodate branches, while CBM103 classifies as a type C CBM (Boraston et al., 2004).

### The shape and flexibility of CBM102- and CBM103-containing proteins facilitate the recognition and processing of laminarin

The laminarin-binding CBM102 and CBM103 are arranged in elongated shapes of multimodular proteins. Elongated shapes of multimodular CBM-containing SGBPs have been determined before (Cartmell et al., 2017; Stevenson et al., 2021), for example, for an SGBP encoded in a heparin/heparan sulphate PUL from *B. thetaiotaomicron* (Cartmell et al., 2017). Like Cf-SGBP (SAXS:  $R_g$  52 Å and  $D_{max}$  178 Å), the *B. thetaiotaomicron* protein consisted of six distinct domains and SAXS determined similar spatial expansions ( $R_g$  44.3 Å and  $D_{max}$  150 Å). In addition to the elongated shapes, analyses revealed a bent shape for the CBM103-containing Cf-GH16\_3 ( $R_g$  38 Å and  $D_{max}$  129.5 Å), seen before in another multidomain laminarinase from *Thermotoga petrophila* (CBM-GH16-CBM,  $R_g$  40 Å and  $D_{max}$  130 Å) (Cota et al., 2011). Our

results also indicated that both proteins are flexible to some extent, albeit with a higher flexibility for Cf-GH16\_3, possibly due to the length (19 aa) and composition of the linker. The increased flexibility may be due to missing prolines in the Cf-GH16\_3 linker, as central prolines are suggested to hamper flexibility (Cartmell et al., 2017; Tauzin et al., 2016). Although such prolines are mostly absent in linkers between CBM102s in Cf-SGBP (except for the first linker that contains a central proline between the Ig-like domain and CBM102), the much shorter linker regions apparently rather act as spacers than as flexible linkers. It was speculated that this rigidity of SGBPs enables distance to the membrane and thus an improved exposure of the sugar-binding site (Tauzin et al., 2016). This might be further supported by the presence of Ig-like fold domains that could act as spacers to the outer membrane (Tamura et al., 2021). Such a projection of SGBPs into the extracellular space has recently been shown for a fructan-specific SGBP from *B. thetaiotaomicron* in complex with a GH32 and SusC/D-like proteins, called the ‘utilisome’ (White et al., 2023). Compared to the fructan-specific SGBP, which contains one CBM, the laminarin-specific Cf-SGBP from *C. forsetii* provides four CBM docking sites. Together with the elongated shape, one could speculate that the CBMs of the protein protrude beyond the other outer membrane proteins, literally reaching out for its target substrate (Figure 4A). However, to date we can only speculate on this spatial arrangement as stated before (Cartmell et al., 2017), and in addition, our analyses indicate binding of two laminarin chains at maximum, possibly due to steric inhibition of binding due to the presence of the other CBMs. Nevertheless, equipping the cellular surface with SGBPs that contain multiple CBMs increases binding opportunities in 3D space and might optimize the presence of proteins on the bacterial surface as a competitive advantage compared to single CBM-containing SGBPs. Moreover, multiple binding sites in SusE and SusF are known to compensate for decreased binding due to polysaccharide capsules in gut bacteria (Cameron et al., 2014). This may be transferable to the marine counterpart, which is also known to produce exopolysaccharides (Poli et al., 2010). In the fructan utilisome the sugar chain is held by an accessory tethering side of the GH32 and by the SGBP. The GH32 catalytic activity then cuts between the two held ends, which was suggested to facilitate degradation (White et al., 2023). In general, the advantageous effect of CBM-CAZyme co-occurrence has largely been demonstrated e.g., Hervé et al., 2010; Najmudin et al., 2006. CBMs may also prevent the loss of released degradation products by serving as a short-term cache for cleaved sugars, which are then further digested or translocated to the SusC/D-like transporters (Figure 4A). At the same time, the affinity of laminarin-derived oligosaccharides to Cf-SGBP-CBM102<sub>IV</sub> was



lower relative to laminarin, as also observed for a CBM6 from *Z. galactanivorans* Dsij<sup>T</sup> (Jam et al., 2016), which supports its major role in capturing laminarin.

## Dynamic combination of domains to increase fitness

Our data from phytoplankton bloom-associated bacteria show that the *C. forsetii* proteins are representatives of diverse domain organizations of both CBMs in multimodular catalytic proteins, mostly GH16\_3s, and putative non-catalytic proteins of which a proportion might function as SGBPs. This suggests a rather dynamic intermixing of CBMs by translocations, duplications, or losses, which also causes exchanges between catalytic and non-catalytic proteins. Structural variability of CBMs further contributes to this diversity, which may determine substrate specificity and/or affinity. This spectrum of domain combinations and protein structures in sugar-catching and sugar-processing proteins likely increases competitiveness to also encounter structurally similar substrates, such as different laminarins or other  $\beta$ -glucans. The environmental relevance of these novel laminarin-binding CBM families CBM102 and CBM103 was further reflected by a good coverage of corresponding proteins during phytoplankton blooms, which was so far mostly observed for SusC/D-like proteins and GH16\_3s associated with laminarin degradation (Teeling et al., 2012; Vidal-Melgosa et al., 2021). Protein abundance of laminarin-binding CBM102- and CBM103-containing proteins correlated with the overall intensity of the investigated blooms, supporting their role in marine laminarin use.

In summary, marine bacteria employ CBM-containing CAZymes and SGBPs on their cellular surface to scavenge laminarin from the water column. This is facilitated by the elongated shapes of these multimodular proteins, where a higher flexibility of the polypeptide chain may increase catalytic activity of CAZymes. In comparison, multiple CBMs and the limited flexibility of SGBPs may allow the accumulation of laminarin, making SGBPs an optimal sugar-trapping antenna to further enhance catalysis. Together with the wide distribution and high expression of CBM102- and CBM103-containing proteins during phytoplankton blooms, our findings corroborate the key role of the two novel CBMs in marine laminarin utilization.

## AUTHOR CONTRIBUTIONS

**Marie-Katherin Zühlke:** Writing – original draft; conceptualization; investigation; visualization; funding acquisition; methodology; project administration; formal analysis. **Elizabeth Ficko-Blean:** Writing – review and editing; project administration; investigation; methodology. **Daniel Bartosik:** Visualization; writing – review and

editing; formal analysis; software. **Nicolas Terrapon:** Writing – review and editing; formal analysis; software. **Alexandra Jeudy:** Investigation; methodology; writing – review and editing. **Murielle Jam:** Writing – review and editing; validation; methodology. **Fengqing Wang:** Software; formal analysis. **Norma Welsch:** Writing – review and editing; methodology. **Alexandra Dürwald:** Investigation; writing – review and editing. **Laura Torres Martin:** Investigation. **Robert Larocque:** Methodology; writing – review and editing. **Diane Jouanneau:** Investigation; writing – review and editing. **Tom Eisenack:** Validation. **François Thomas:** Validation; writing – review and editing. **Anke Trautwein-Schult:** Formal analysis; writing – review and editing. **Hanno Teeling:** Formal analysis; funding acquisition; resources; writing – review and editing. **Dörte Becher:** Resources; formal analysis; funding acquisition. **Thomas Schweder:** Resources; conceptualization; writing – review and editing; funding acquisition. **Mirjam Czjzek:** Conceptualization; funding acquisition; investigation; writing – review and editing; visualization; methodology; formal analysis; resources; project administration.

## ACKNOWLEDGEMENTS

The authors strongly appreciated having access to the CristalO platform (FR2424, Station Biologique de Roscoff), which is part of the core facility networks Biogenouest and EMBRC-France. We are grateful for the access to the small-angle X-ray scattering beamline and MX-beamline PROXIMA1 at synchrotron SOLEIL (Saint Aubin, France) and thank the staff for their support on-site, especially Aurélien Thureau and Pierre Legrand. We thank the team of the Helgoland sampling campaign, especially Lilly Franzmeyer, and the Biological Station Helgoland (Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung, BAH-AWI, Germany), especially Inga V. Kirstein and Karen H. Wiltshire. The Helgoland Time Series of the Alfred Wegener Institute is supported by the Helmholtz Association as an LK-II performance category program. We thank Jana Matulla, Jasna Nikolic and Lionel Cladière for support in the lab as well as Stephanie Markert and Michelle Teune for helpful comments. This study was supported by the German Research Foundation (DFG) within the research unit FOR 2406 ‘Proteogenomics of Marine Polysaccharide Utilization’ (POMPU) by grants to Hanno Teeling (TE 813/2–3), Dörte Becher (BE 3869/4-3) and Thomas Schweder (SCHW 595/10-3, SCHW 595/11-3). This work also benefited from the support of the Centre National de la Recherche Scientifique (CNRS) and Sorbonne University. MKZ gratefully acknowledges the DFG (SCHW 595/10-3), PROCOPE Mobility Grant and the ERASMUS+ Staff Mobility for Training, all of which enabled several research stays in Roscoff. Open Access funding enabled and organized by Projekt DEAL.



## CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

The X-ray crystallography data (Cf-SGBP-CBM102<sub>IV</sub> in complex with laminaritriose) are available in the Protein Data Bank (PDB) under the accession code 8QX6, and the SAXS data are available in the Small Angle Scattering Biological Data Bank (SASBDB) via the accession codes SASDU93 (Cf-GH16\_3) and SASDUA3 (Cf-SGBP). Metagenome assemblies and MAGs from 2016, 2018 and 2020 phytoplankton blooms are available in the European Nucleotide Archive under the project accessions PRJEB28156 (2016), PRJEB38290 (2018) and PRJEB52999 (2020). Metaproteome data are available in the PRIDE archive through the identifiers PXD019294 (2016), PXD042676 (2018) and PXD042805 (2020). Phytoplankton data are available via Pangaea (<https://doi.pangaea.de/10.1594/PANGAEA.864676>).

## ORCID

Marie-Katherin Zühlke  <https://orcid.org/0000-0002-8081-9969>


Elizabeth Ficko-Blean  <https://orcid.org/0000-0002-3915-4715>


Daniel Bartosik  <https://orcid.org/0000-0002-5103-4564>

Nicolas Terrapon  <https://orcid.org/0000-0002-3693-6017>


Alexandra Jeudy  <https://orcid.org/0009-0008-8633-3952>

Murielle Jam  <https://orcid.org/0000-0002-9357-5676>

Alexandra Dürwald  <https://orcid.org/0000-0001-9503-485X>

Diane Jouanneau  <https://orcid.org/0000-0003-1964-292X>


Tom Eisenack  <https://orcid.org/0000-0003-4829-4696>

François Thomas  <https://orcid.org/0000-0003-1896-0774>

Anke Trautwein-Schult  <https://orcid.org/0000-0003-3272-5746>

Hanno Teeling  <https://orcid.org/0000-0003-1889-7859>

Dörte Becher  <https://orcid.org/0000-0002-9630-5735>

Thomas Schweder  <https://orcid.org/0000-0002-7213-3596>

Mirjam Czjzek  <https://orcid.org/0000-0002-7483-2841>

## REFERENCES

- Anderson, K.L. & Salyers, A.A. (1989) Biochemical evidence that starch breakdown by *Bacteroides thetaiotaomicron* involves outer membrane starch-binding sites and periplasmic starch-degrading enzymes. *Journal of Bacteriology*, 171(6), 3192–3198.
- Armonies, W., Asmus, H., Buschbaum, C., Lackschewitz, D., Reise, K. & Rick, J. (2018) Microscopic species make the diversity: a

- checklist of marine flora and fauna around the Island of Sylt in the North Sea. *Helgoland Marine Research*, 72, 11.
- Becker, S., Scheffel, A., Polz, M.F. & Hehemann, J.H. (2017) Accurate quantification of laminarin in marine organic matter with enzymes from marine microbes. *Applied and Environmental Microbiology*, 83(9), e03389-16.
- Becker, S., Tebben, J., Coffinet, S., Wiltshire, K., Iversen, M.H., Harder, T. et al. (2020) Laminarin is a major molecule in the marine carbon cycle. *Proceedings of the National Academy of Sciences of the United States of America*, 117(12), 6599–6607.
- Beidler, I., Steinke, N., Schulze, T., Sidhu, C., Bartosik, D., Krull, J., et al. (2023) *Alpha-glucans from bacterial necromass indicate an intra-population loop within the marine carbon cycle*. PREPRINT (Version 1) available at Research Square. <https://doi.org/10.121203/rs3rs-3205445/v1>
- Bernadó, P., Mylonas, E., Petoukhov, M.V., Blackledge, M. & Svergun, D.I. (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *Journal of the American Chemical Society*, 129(17), 5656–5664.
- Bjursell, M.K., Martens, E.C. & Gordon, J.I. (2006) Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, *Bacteroides thetaiotaomicron*, to the suckling period. *The Journal of Biological Chemistry*, 281(47), 36269–36279.
- Boraston, A.B., Bolam, D.N., Gilbert, H.J. & Davies, G.J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *The Biochemical Journal*, 382, 769–781.
- Buchfink, B., Reuter, K. & Drost, H.G. (2021) Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, 18(4), 366–368.
- Cameron, E.A., Kwiatkowski, K.J., Lee, B.H., Hamaker, B.R., Koropatkin, N.M. & Martens, E.C. (2014) Multifunctional nutrient-binding proteins adapt human symbiotic bacteria for glycan competition in the gut by separately promoting enhanced sensing and catalysis. *MBio*, 5(5), e01441-14.
- Cartmell, A., Lowe, E.C., Baslé, A., Firbank, S.J., Ndeh, D.A., Murray, H. et al. (2017) How members of the human gut microbiota overcome the sulfation problem posed by glycosaminoglycans. *Proceedings of the National Academy of Sciences of the United States of America*, 114(27), 7037–7042.
- Conway, J.R., Lex, A. & Gehlenborg, N. (2017) UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*, 33(18), 2938–2940.
- Cota, J., Alvarez, T.M., Citadini, A.P., Santos, C.R., Neto, M.D., Oliveira, R.R. et al. (2011) Mode of operation and low-resolution structure of a multi-domain and hyperthermophilic endo- $\beta$ -1,3-glucanase from *Thermotoga petrophila*. *Biochemical and Biophysical Research Communications*, 406(4), 590–594.
- Deshmukh, U.B. & Oren, A. (2023) Proposal of *Christiangramia* gen. nov., *Neomelitea* gen. nov. and *Nicolliella* gen. nov. as replacement names for the illegitimate prokaryotic generic names *Gramella* Nedashkovskaya et al. 2005, *Melitea* Urios et al., 2008 and *Nicolia* Oliphant et al. 2022, respectively. *International Journal of Systematic and Evolutionary Microbiology*, 73(4), 5806.
- Drula, E., Garron, M.L., Dogan, S., Lombard, V., Henrissat, B. & Terrapon, N. (2022) The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Research*, 50(D1), D571–D577.
- Eilers, H., Pemthaler, J., Peplies, J., Glöckner, F.O., Gerdt, G. & Amann, R. (2001) Isolation of novel pelagic bacteria from the German bight and their seasonal contributions to surface picoplankton. *Applied and Environmental Microbiology*, 67(11), 5134–5142.
- Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. (2010) Features and development of Coot. *Acta Crystallographica. Section D, Biological Crystallography*, 66(4), 486–501.
- Evans, P.R. & Murshudov, G.N. (2013) How good are my data and what is the resolution? *Acta Crystallographica. Section D, Biological Crystallography*, 69(7), 1204–1214.





- Ficko-Blean, E., Stuart, C.P., Suits, M.D., Cid, M., Tessier, M., Woods, R.J. et al. (2012) Carbohydrate recognition by an architecturally complex alpha-N-acetylglucosaminidase from *Clostridium perfringens*. *PLoS One*, 7(3), e33524.
- Francis, T.B., Bartosik, D., Sura, T., Sichert, A., Hehemann, J.H., Markert, S. et al. (2021) Changing expression patterns of TonB-dependent transporters suggest shifts in polysaccharide consumption over the course of a spring phytoplankton bloom. *The ISME Journal*, 15(8), 2336–2350.
- Glenwright, A.J., Pothula, K.R., Bhamidimarri, S.P., Chorem, D.S., Baslé, A., Firbank, S.J. et al. (2017) Structural basis for nutrient acquisition by dominant members of the human gut microbiota. *Nature*, 541(7637), 407–411.
- Groissillier, A., Hervé, C., Jeudy, A., Rebuffet, E., Pluchon, P.F., Chevolut, Y. et al. (2010) MARINE-EXPRESS: taking advantage of high throughput cloning and expression strategies for the post-genomic analysis of marine organisms. *Microbial Cell Factories*, 9, 45.
- Gügi, B., Le Costaouec, T., Burel, C., Lerouge, P., Helbert, W. & Bardor, M. (2015) Diatom-specific oligosaccharide and polysaccharide structures help to unravel biosynthetic capabilities in diatoms. *Marine Drugs*, 13(9), 5993–6018.
- Henshaw, J.L., Bolam, D.N., Pires, V.M.R., Czjzek, M., Henrissat, B., Ferreira, L.M.A. et al. (2004) The family 6 carbohydrate binding module CmCBM6-2 contains two ligand-binding sites with distinct specificities. *The Journal of Biological Chemistry*, 279(20), 21552–21559.
- Hervé, C., Rogowski, A., Blake, A.W., Marcus, S.E., Gilbert, H.J. & Knox, J.P. (2010) Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. *Proceedings of the National Academy of Sciences of the United States of America*, 107(34), 15293–15298.
- Holm, L. (2022) Dali server: structural unification of protein families. *Nucleic Acids Research*, 50(W1), W210–W215.
- Hudson, K.L., Bartlett, G.J., Diehl, R.C., Agirre, J., Gallagher, T., Kiessling, L.L. et al. (2015) Carbohydrate-aromatic interactions in proteins. *Journal of the American Chemical Society*, 137(48), 15152–15160.
- Jam, M., Ficko-Blean, E., Labourel, A., Larocque, R., Czjzek, M. & Michel, G. (2016) Unraveling the multivalent binding of a marine family 6 carbohydrate-binding module with its native laminarin ligand. *The FEBS Journal*, 283(10), 1863–1879.
- Juncker, A.S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H. & Krogh, A. (2003) Prediction of lipoprotein signal peptides in gram-negative bacteria. *Protein Science*, 12(8), 1652–1662.
- Kabisch, A., Otto, A., König, S., Becher, D., Albrecht, D., Schüler, M. et al. (2014) Functional characterization of polysaccharide utilization loci in the marine *Bacteroidetes* ‘*Gramella forsetii*’ KT0803. *The ISME Journal*, 8(7), 1492–1502.
- Kabsch, W.X.D.S. (2010) XDS. *Acta Crystallographica. Section D, Biological Crystallography*, 66(Pt 2), 125–132.
- Kappelman, L., Krüger, K., Hehemann, J.H., Harder, J., Markert, S., Unfried, F. et al. (2019) Polysaccharide utilization loci of North Sea *Flavobacteriia* as basis for using SusC/D-protein expression for predicting major phytoplankton glycans. *The ISME Journal*, 13(1), 76–91.
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. & Sternberg, M.J.E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10(6), 845–858.
- Koch, H., Dürwald, A., Schweder, T., Noriega-Ortega, B., Vidal-Melgosa, S., Hehemann, J.H. et al. (2019) Biphasic cellular adaptations and ecological implications of *Alteromonas macleodii* degrading a mixture of algal polysaccharides. *The ISME Journal*, 13(1), 92–103.
- Krüger, K., Chafee, M., Francis, T.B., del Rio, T.G., Becher, D., Schweder, T. et al. (2019) In marine *Bacteroidetes* the bulk of glycan degradation during algae blooms is mediated by few clades using a restricted set of genes. *The ISME Journal*, 13(11), 2800–2816.
- Kuhaulomlarp, S., Patron, N.J., Henrissat, B., Rejzek, M., Saalbach, G. & Field, R.A. (2018) Identification of *Euglena gracilis*  $\beta$ -1,3-glucan phosphorylase and establishment of a new glycoside hydrolase (GH) family GH149. *The Journal of Biological Chemistry*, 293(8), 2865–2876.
- Kumagai, Y., Kishimura, H., Lang, W., Tagami, T., Okuyama, M. & Kimura, A. (2022) Characterization of an unknown region linked to the glycoside hydrolase family 17  $\beta$ -1,3-glucanase of *Vibrio vulnificus* reveals a novel glucan-binding domain. *Marine Drugs*, 20(4), 250.
- Labourel, A., Jam, M., Jeudy, A., Hehemann, J.H., Czjzek, M. & Michel, G. (2014) The  $\beta$ -glucanase ZgLamA from *Zobellia galactanivorans* evolved a bent active site adapted for efficient degradation of algal laminarin. *The Journal of Biological Chemistry*, 289(4), 2027–2042.
- Labourel, A., Jam, M., Legentil, L., Sylla, B., Hehemann, J.H., Ferrieres, V. et al. (2015) Structural and biochemical characterization of the laminarinase ZgLamC(GH16) from *Zobellia galactanivorans* suggests preferred recognition of branched laminarin. *Acta Crystallographica Section D*, 71, 173–184.
- Lex, A., Gehlenborg, N., Strobel, H., Vuillemot, R. & Pfister, H. (2014) UpSet: visualization of intersecting sets. *IEEE Transactions on Visualization and Computer Graphics*, 20(12), 1983–1992.
- Lieutaud, P., Canard, B. & Longhi, S. (2008) MeDor: a metasever for predicting protein disorder. *BMC Genomics*, 9(Suppl 2), S25.
- Manalastas-Cantos, K., Konarev, P.V., Hajizadeh, N.R., Kikhney, A.G., Petoukhov, M.V., Molodenskiy, D.S. et al. (2021) ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *Journal of Applied Crystallography*, 54, 343–355.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. & Read, R.J. (2007) Phaser crystallographic software. *Journal of Applied Crystallography*, 40(Pt 4), 658–674.
- Miller, G.L. (1959) Use of Dinitrosalicylic acid reagent for determination of reducing sugar. *Analytical Chemistry*, 31(3), 426–428.
- Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S. & Steinegger, M. (2022) ColabFold: making protein folding accessible to all. *Nature Methods*, 19(6), 679–682.
- Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A. et al. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D*, 67, 355–367.
- Mystkowska, A.A., Robb, C., Vidal-Melgosa, S., Vanni, C., Fernandez-Guerra, A., Höhne, M. et al. (2018) Molecular recognition of the beta-glucans laminarin and pustulan by a SusD-like glycan-binding protein of a marine *Bacteroidetes*. *The FEBS Journal*, 285(23), 4465–4481.
- Najmudin, S., Guerreiro, C.I., Carvalho, A.L., Prates, J.A., Correia, M.A., Alves, V.D. et al. (2006) Xyloglucan is recognized by carbohydrate-binding modules that interact with  $\beta$ -glucan chains. *The Journal of Biological Chemistry*, 281(13), 8815–8828.
- Olm, M.R., Brown, C.T., Brooks, B. & Banfield, J.F. (2017) dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME Journal*, 11(12), 2864–2868.
- Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S. et al. (2022) The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Research*, 50(D1), D543–D552.
- Pluinage, B., Hehemann, J.H. & Boraston, A.B. (2013) Substrate recognition and hydrolysis by a family 50 *exo*- $\beta$ -agarase, Aga50D, from the marine bacterium *Saccharophagus degradans*. *The Journal of Biological Chemistry*, 288(39), 28078–28088.



- Poli, A., Anzelmo, G. & Nicolaus, B. (2010) Bacterial exopolysaccharides from extreme marine habitats: production, characterization and biological activities. *Marine Drugs*, 8(6), 1779–1802.
- Pollet, R.M., Martin, L.M. & Koropatkin, N.M. (2021) TonB-dependent transporters in the Bacteroidetes: unique domain structures and potential functions. *Molecular Microbiology*, 115(3), 490–501.
- Ratliff, A.C., Buchanan, S.K. & Celia, H. (2022) The ton motor. *Frontiers in Microbiology*, 13, 13.
- Read, S.M., Currie, G. & Bacic, A. (1996) Analysis of the structural heterogeneity of laminarin by electrospray-ionization-mass spectrometry. *Carbohydrate Research*, 281(2), 187–201.
- Robert, X. & Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, 42(W1), W320–W324.
- Rudenko, O., Thureau, A. & Perez, J. (2019) Evolutionary refinement of the 3D structure of multi-domain protein complexes from small angle X-ray scattering data. In M. López-Ibáñez, (Ed.), *Proceedings of the 2019 Genetic and Evolutionary Computation Conference Companion (Gecco'19 Companion)*. New York, NY: ACM, pp. 401–402.
- Sainz-Polo, M.A., González, B., Menéndez, M., Pastor, F.I.J. & Sanz-Aparicio, J. (2015) Exploring multimodularity in plant Cell Wall deconstruction: Structural and functional analysis of Xyn10C containing the CBM22-1-CBM22-2 tandem. *The Journal of Biological Chemistry*, 290(28), 17116–17130.
- Schut, F., de Vries, E.J., Gottschal, J.C., Robertson, B.R., Harder, W., Prins, R.A. et al. (1993) Isolation of typical marine bacteria by dilution culture: growth, maintenance, and characteristics of isolates under laboratory conditions. *Applied and Environmental Microbiology*, 59(7), 2150–2160.
- Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14), 2068–2069.
- Sidhu, C., Kirstein, I.V., Meunier, C.L., Rick, J., Fofonova, V., Wiltshire, K.H. et al. (2023) Dissolved storage glycans shaped the community composition of abundant bacterioplankton clades during a North Sea spring phytoplankton bloom. *Microbiome*, 11(1), 77.
- Stevenson, J., Ngo, M., Brandt, A., Weadge, J.T. & Suits, M.D.L. (2021) Analysis of two SusE-like enzymes from *Bacteroides thetaiotaomicron* reveals a potential degradative capacity for this protein family. *Frontiers in Microbiology*, 12, 12.
- Studier, F.W. (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification*, 41(1), 207–234.
- Svergun, D.I., Petoukhov, M.V. & Koch, M.H.J. (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophysical Journal*, 80(6), 2946–2953.
- Tamura, K., Dejean, G., Van Petegem, F. & Brumer, H. (2021) Distinct protein architectures mediate species-specific beta-glucan binding and metabolism in the human gut microbiota. *The Journal of Biological Chemistry*, 296, 100415.
- Tamura, K., Foley, M.H., Gardill, B.R., Dejean, G., Schnizlein, M., Bahr, C.M.E. et al. (2019) Surface glycan-binding proteins are essential for cereal beta-glucan utilization by the human gut symbiont *Bacteroides ovatus*. *Cellular and Molecular Life Sciences*, 76(21), 4319–4340.
- Tauzin, A.S., Kwiatkowski, K.J., Orlovsky, N.I., Smith, C.J., Creagh, A.L., Haynes, C.A. et al. (2016) Molecular dissection of xyloglucan recognition in a prominent human gut symbiont. *MBio*, 7(2), e02134-15.
- Teeling, H., Fuchs, B.M., Becher, D., Klockow, C., Gardebrecht, A., Bennke, C.M. et al. (2012) Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science*, 336(6081), 608–611.
- Tria, G., Mertens, H.D.T., Kachala, M. & Svergun, D.I. (2015) Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ*, 2, 207–217.
- Unfried, F., Becker, S., Robb, C.S., Hehemann, J.H., Markert, S., Heiden, S.E. et al. (2018) Adaptive mechanisms that provide competitive advantages to marine bacteroidetes during microalgal blooms. *The ISME Journal*, 12, 2894–2906.
- Viborg, A.H., Terrapon, N., Lombard, V., Michel, G., Czejek, M., Henrissat, B. et al. (2019) A subfamily roadmap of the evolutionarily diverse glycoside hydrolase family 16 (GH16). *The Journal of Biological Chemistry*, 294(44), 15973–15986.
- Vidal-Melgosa, S., Sichert, A., Francis, T.B., Bartosik, D., Niggemann, J., Wichels, A. et al. (2021) Diatom fucan polysaccharide precipitates carbon during algal blooms. *Nature Communications*, 12(1), 1–13.
- Wang, F.Q., Bartosik, D., Sidhu, C., Siebers, R., Lu, D.C., Trautwein-Schult, A. et al. (2024) Particle attached bacteria act as gatekeepers in the decomposition of complex phytoplankton polysaccharides. *Microbiome*, 12(32), 32.
- White, J.B.R., Silale, A., Feasey, M., Heunis, T., Zhu, Y., Zheng, H. et al. (2023) Outer membrane utilisomes mediate glycan uptake in gut bacteroidetes. *Nature*, 618(7965), 583–589.
- Zheng, J., Hu, B., Zhang, X., Ge, Q., Yan, Y., Akresi, J. et al. (2023) dbCAN-seq update: CAZyme gene clusters and substrates in microbiomes. *Nucleic Acids Research*, 51(D1), D557–D563.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Zühlke, M.-K., Ficko-Blean, E., Bartosik, D., Terrapon, N., Jeudy, A., Jam, M. et al. (2024) Unveiling the role of novel carbohydrate-binding modules in laminarin interaction of multimodular proteins from marine *Bacteroidota* during phytoplankton blooms. *Environmental Microbiology*, 26(5), e16624. Available from: <https://doi.org/10.1111/1462-2920.16624>