PROTEIN STRUCTURE REPORT

Crystal structure of a marine glycoside hydrolase family 99-related protein lacking catalytic machinery

Craig S. Robb,[1,2] Agata Anna Mystkowska,[1,2] and Jan-Hendrik Hehemann [1,2]*

[1]Center for Marine Environmental Sciences, University of Bremen (MARUM), Bremen 28359, Germany
[2]Max Planck-Institute for Marine Microbiology, Bremen 28359, Germany

Received 19 July 2017; Accepted 6 September 2017
DOI: 10.1002/pro.3291
Published online 8 September 2017 proteinscience.org

Abstract: Algal polysaccharides of diverse structures are one of the most abundant carbon resources for heterotrophic, marine bacteria with coevolved digestive enzymes. A putative sulfo-mannan polysaccharide utilization locus, which is conserved in marine flavobacteria, contains an unusual GH99-like protein that lacks the conserved catalytic residues of glycoside hydrolase family 99. Using X-ray crystallography, we structurally characterized this protein from the marine flavobacterium *Ochrovirga pacifica* to help elucidate its molecular function. The structure reveals the absence of potential catalytic residues for polysaccharide hydrolysis, which—together with additional structural features—suggests this protein may be noncatalytic and involved in carbohydrate binding.

Keywords: marine polysaccharides; polysaccharide utilization locus; glycoside hydrolases; GH99; X-ray crystal structure

Introduction

Microalgae, which form algal blooms, synthesize substantial amounts of polysaccharides in nutrient rich marine surface waters. Polysaccharides serve diverse biological functions including energy storage, maintenance of cell wall structure, and as secreted exudates. They also provide carbon and energy for heterotrophic organisms including bacteria maintaining energy and carbon flow through marine food webs and they have been recently suggested as promising resource for the production of biofuels and useful chemicals.[1] Unlike plant polysaccharides, algal polysaccharides often include monosaccharides with different covalent modifications including sulfate groups, which require molecular pathways for enzymatic conversion that have not been explored. Algal bloom-associated bacteria are enriched in putative carbohydrate active enzymes, which likely target the polysaccharides of the algae and therefore represent a promising source of new enzymes.[2] Many bacteria, including marine flavobacteria, contain functionally co-evolved carbohydrate active enzymes within genomic clusters that target a single type of polysaccharide and these clusters are commonly referred to as polysaccharide utilization loci (PUL).[3] In addition to enzymes, PULs also frequently encode for proteins that function as cell surface glycan-binding proteins typified by SusD and its homologs[4,5] but also including ancillary binding proteins such as SusE and SusF as a large complex on the outer surface

Grant sponsor: Deutsche Forschungsgemeinschaft (DFG); Grant number: HE 7217/1-1; Grant sponsor: Max-Planck Society; Grant sponsor: Beamline EMBO; Grant number: MX-454.

*Correspondence to: Jan-Hendrik Hehemann, Center for Marine Environmental Sciences, University of Bremen (MARUM), Bremen 28359, Germany. E-mail: jhhehemann@marum.de

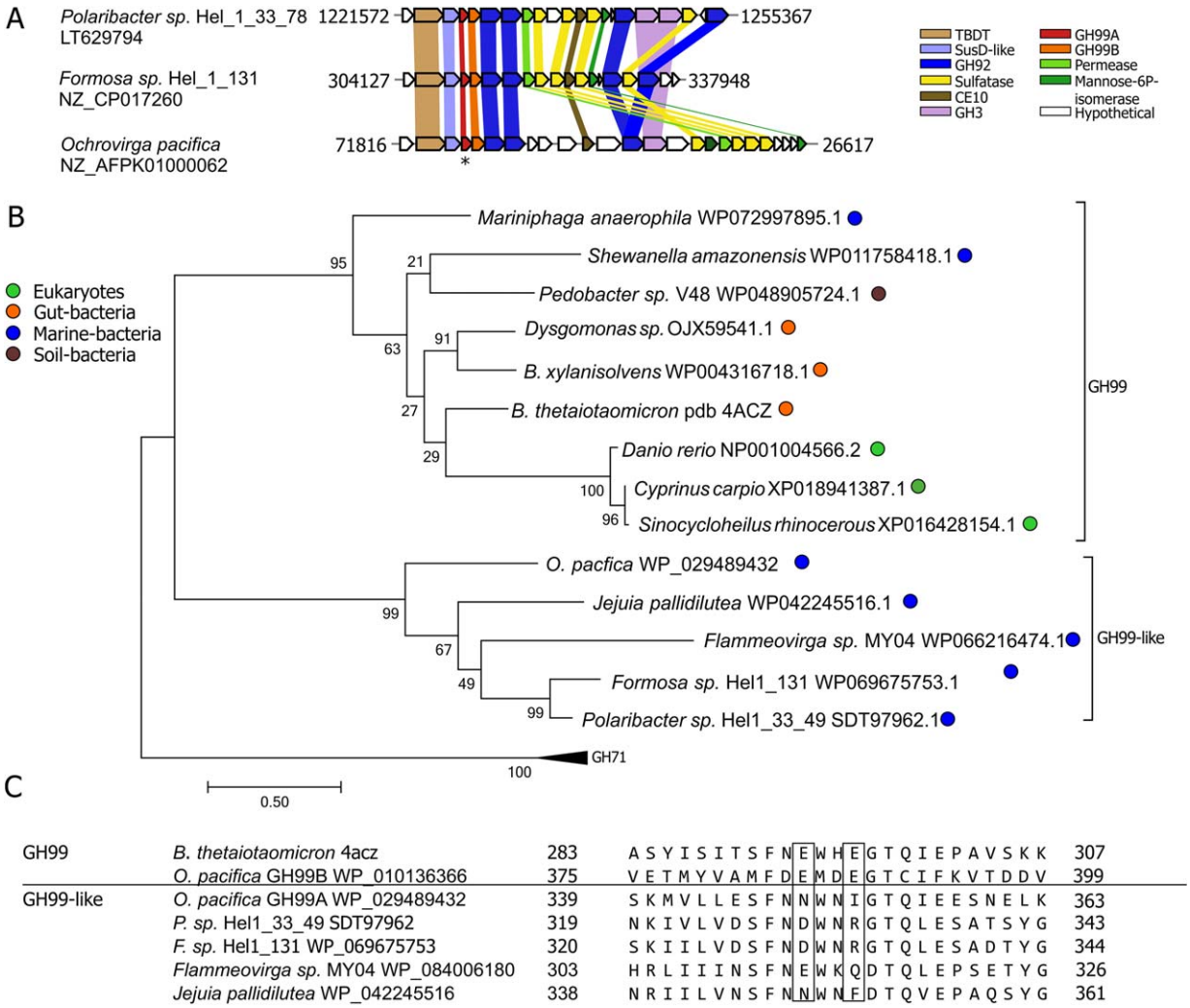

Figure 1. Phylogenetic comparison of OpGH99A with related proteins from marine and terrestrial organisms. (A) The putative sulfo-mannan PULs of the three marine isolates have 15 genes in common. The colors correspond to the function of the protein and proteins with high sequence identity have been linked with colored ribbons. The target protein of this study is marked with a star. (B) A rooted maximum likelihood phylogenetic protein tree shows OpGH99A is distantly related to glycoside hydrolase family 99. The numbers at the nodes present bootstrap values. The tree was calculated with amino acid sequences and the scale bar represents the number of amino acid substitutions per site. Five sequences from glycoside hydrolase family GH71 were used as outgroup. Displayed on the tree are the species names and the accession number or the PDBid of the corresponding sequence. Colored dots indicate origin or habitat (green, eukaryote; orange, intestine; blue, marine; brown, soil). (C) A 3D structure-guided sequence alignment shows that the catalytic residues, two glutamates of signature **E**xx**E**, which are conserved in the family GH99, are not conserved in sequences belonging to the GH99-like proteins. Also shown, OpGH99B also found in the PUL does bear the catalytic residues of the glycoside hydrolase family.

of the bacterial cell.[6] The metagenomic analyses of a series of recently monitored North Sea Spring microalgal blooms revealed a recurrent putative sulfo-mannan PUL that includes mannanases of family GH92, GH99 and sulfatases[2,7,8] suggesting that they play an important role in depolymerization and turnover of sulfated mannans produced by the algae. Specifically, the family GH99—found in the bacterial metagenomes—is composed of endo-mannanases that are thought to operate with an unusual, substrate assisted retaining mechanism involving an 1,2-anhydro-sugar intermediate.[9,10] The catalytic residues involved in hydrolysis of the glycosidic linkage are highly conserved within this enzyme family; however, we show that these catalytic residues are absent within a subgroup of related sequences that include one of the GH99 from the marine sulfo-mannan PUL. We investigated the 3D crystal structure of a member of the GH99-like subfamily and exposed a lack of alternative catalytic residues suggesting that this protein may function as a glycan-binding protein rather than an enzyme. This protein has a predicted lipoprotein signal peptide that would result in secretion and possible display on the cell surface as seen in other systems.[11,12]

Table I. *Data collection and refinement statistics*

Data collection	
Data set	OpGH99A
X-ray source	EMBL P13
Wavelength (Å)	0.9763
Space group	$P3_121$
Unit cell *a*, *b*, *c* (Å)	102.05, 102.05, 80.11
Resolution range, (Å)	88.38–2.40 (2.53–2.40)
R_{merge}	0.112 (0.741)
Completeness (%)	100.0 (100.0)
Redundancy	9.6 (9.8)
$<I/\sigma(I)>$	13.6 (3.0)
No. of reflections	184,087 (27,039)
No. unique	19,223 (2761)
Mosaicity	0.23
Refinement	
R_{work}/R_{free} (%)	0.166/0.215
No. of atoms	3005
Protein	2846
Bromide	5
Water	164
B factors	
Overall	48.32
Protein	48.41
Bromide	57.95
Water	46.64
R.m.s. deviations	
Bond lengths (Å)	0.015
Bond angles (°)	1.653
Ramachandran statistics (%)	
Favored	97.5
Allowed	2.5
Outliers	0
PDB accession code	5NGW

Results and Discussion

OpGH99A phylogenetic results

The sulfo-mannan PUL is found in marine bacteria including isolates from the North Sea algal bloom *Polaribacter* spp.[8] and *Formosa* sp. Hel1_33_131.[13] The sulfo-mannan PUL consists of enzymes from the glycoside hydrolase families GH92 and GH99 and numerous sulfatases [Fig. 1(A)]. Based on the functional prediction of the proteins encoded for in the PUL, the putative substrate may be a sulfated-α-mannan, the source of which remains unknown; however, the phylogenetic analysis shows the proteins are conserved in marine bacteria suggesting the substrate is produced in the sea. *Ochrovirga pacifica*, a marine flavobacterium isolated from the Pacific[14] also has a highly similar PUL consisting of highly similar sulfatases and glycoside hydrolases suggesting that this organism targets the same or a related substrate.

Phylogenetic analysis reveals OpGH99A is member of a monophyletic group of proteins including structurally and biochemically characterized homologs that are archetypical members of family GH99. However, these enzymes form two, distantly related clades and the one occupied by OpGH99A is composed exclusively of sequences from marine bacteria [Fig. 1(B)]. A sequence alignment confirms absence of the conserved catalytic residues among the members of the GH99-like clade [Fig. 1(C)]. The other gene clusters with homologs of OpGH99A include transporters and glycoside hydrolases suggesting they are involved in polysaccharide degradation.

Structural characterization of OpGH99A

To shed light on the structure and help understand the lack of conserved active site residues we used X-ray crystallography. We obtained crystals of OpGH99A that diffracted to 2.4 Å in space group $P3_121$ with 58% solvent. Given the low sequence identity to other GH99 structures (17% identity), we initially aimed for experimental phasing by collecting a bromide derivative dataset. However, no anomalous signal was observed during the three-wavelength MAD experiment. Instead, the structure was solved by molecular replacement using Phyre model generated from the structure of BtGH99 (pdb: 4acy) and the sequence of OpGH99A.[15] The final 3D structure comprised a single protein molecule in the asymmetric unit with residues 11–370, 164 water molecules, and 5 partially occupied bromide atoms (Table I).

The OpGH99A structure adopts an $(\alpha/\beta)_8$ fold, which consists of an 8-stranded barrel at the center of the protein and with 8 alpha helices around the outside arranged in an alternating alpha/beta pattern [Fig. 2(A)]. The alpha helices and connecting loops extend above the central opening of the barrel, where the active site residues are found in the hydrolase members of this fold.[16] These extensions shape the walls of a potential functional site and form a cleft that extends along the length of the protein ~50 Å that is between ~9 Å and ~13 Å wide. The cleft is lined with 18 asparagine or glutamine residues out of 63 such residues in the protein sequence [Fig. 2(B)]. The electrostatic potential of the putative functional site is close to neutral given an almost equal number of positively and negatively charged residues. There are two carboxyl-bearing residues located in the putative binding cleft E329 and D81 located 8.4 Å apart in the center toward the bottom of the cleft, which is in the range of values for catalytic residues of glycoside hydrolases.[17] However, in OpGH99A, these two residues are divided by K38 that would hinder a functional interaction and thus would not likely participate in a hydrolysis reaction due to steric hindrance [Fig. 2(C)]. Furthermore, these residues are not conserved in the close homologs of OpGH99A. If this subclade has a conserved function, then these residues could be dispensable. Given the lack of apparent catalytic residues, we propose that this protein may not be a glycoside hydrolase and may function as a binding protein. We tested binding of OpGH99A to yeast

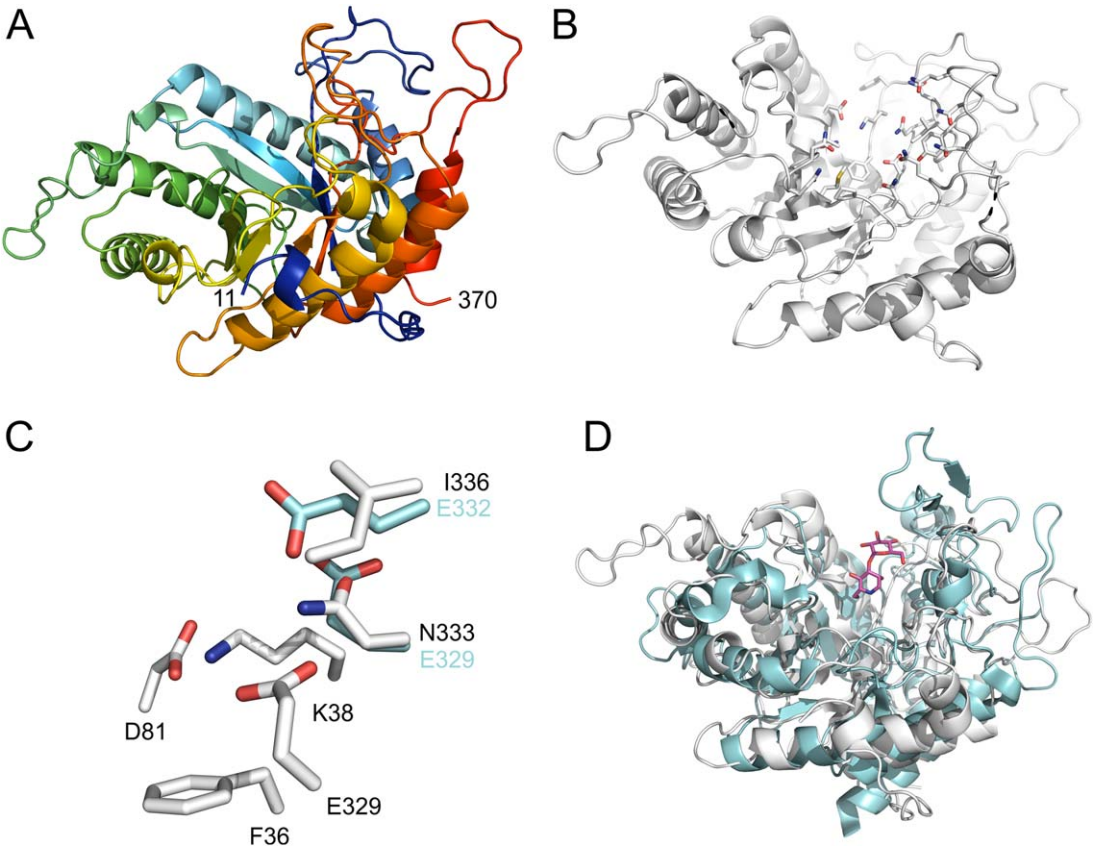

Figure 2. Structure of OpGH99A. (A) The structure of OpGH99A shown as a cartoon representation color ramped from the N-terminus (blue) to the C-terminus (red). OpGH99A adopts a $(\alpha/\beta)_8$ fold and is shown centering on the putative functional site located in the C-terminal end of the barrel. (B) The putative functional site of the protein is rich in asparagine and glutamine. (C) A close-up of two acidic residues in the putative functional site, D81 and E329 separated by K38. The catalytic residues of BtGH99 (cyan) and the corresponding residues from OpGH99A (white) are shown for reference. (D) The structure of OpGH99A (white) aligns with an RMSD of 2.5 along 275 of 341 residues to the structure of BtGH99 (cyan), its closest structural homolog sharing a pairwise sequence identity of 17% shown in complex with the inhibitor glucose-1,3-deoxymannojirimycin (red).

mannan using a gel-retardation assay but unlike the control, the protein showed no binding activity (data not shown).

Several features distinguish the structure of OpGH99A from its closest structurally characterized homolog, BtGH99 (PDB: 4acy).[10] The two proteins share 17% pairwise sequence identity and a backbone root mean square deviation of 2.5 Å along 275 aligned residues [Fig. 2(D)].[18] In addition to the aforementioned lack of conserved catalytic residues, the largest difference is the lack of hydrophobic residues in the cleft of OpGH99A. Classical protein–carbohydrate interactions are largely mediated by hydrophobic residues.[19] In BtGH99, hydrophobic side chains mediate specific interactions with the substrate in this way. As a result, the cleft of OpGH99A is deeper, wider, and longer. Aside from overall fold, these two proteins have little in common.

Conclusion

Metagenomic, genomic, and proteomic information previously provided insights into putative sulfo-mannan degradation by marine heterotrophic bacteria.[5] This putative pathway may be a critical agent involved in the recycling of marine mannan. Here, OpGH99A one protein of the PUL, which is distantly related to characterized endo-mannanases of gut bacteria, likely operates in a fundamentally different way in the recycling of mannans. The absence of catalytic residues within a cleft that would be conducive for glycan interaction and its presence within mannose degrading PULs suggests that this protein may bind a mannose rich glycan. Overall, these results underline the fact that simple association to GH families does not allow for functional annotation of genes in metagenome data. Moreover, the structure and phylogenetic analysis provide the first basis for a molecular level understanding of this group of GH99-like proteins and their possible role in the degradation of mannan in the marine environment.

Materials and Methods

Cloning, protein production, and purification

The gene encoding for GH99A lacking its signal peptide (residues 6–371) from *O. pacifica* was amplified from genomic DNA by PCR using gene specific

primers (Biomers) and cloned using Gibson assembly in *E. coli* DH5α (New England Biolabs). Clones were sequenced in house by Sanger sequencing using Bigdye (Thermofisher). For protein production, plasmid DNA was transformed into *E. coli* BL21 (DE3) and the proteins were produced in 1 L batches of autoinduction media incubated at 20°C for 4 days. Cells were harvested by centrifugation and stored at −20°C until processing. Cell lysis was conducted by chemical lysis as previously described.[20] Frozen cell pellets were resuspended in 20 mL of sucrose solution (25% sucrose, 50 mM Tris pH 8.0). Lysozyme was added at a concentration of 1 mg/mL and the lysis was incubated 10 min at room temperature with spinning. Forty milliliters of deoxycholate solution (1% deoxycholate, 1% Triton X-10, and 100 mM NaCl) was added followed by $MgCl_2$ to a final concentration of 1 mM and DNase to a concentration of 1 mg/mL. The resulting lysate was centrifuged at 16,000*g* for 45 min at 4°C. For purification, clarified lysate was applied to a 5 mL prepacked IMAC column (GE) previously equilibrated in Buffer A (20 mM Tris pH 8 and 500 mM NaCl) using an AKTA start FPLC. The column was washed extensively with Buffer A and His-tagged protein was eluted using a gradient of imidazole to 500 mM in Buffer A. Purified protein was concentrated using a stirred cell ultrafiltration device with a 10 kDa membrane. The concentrated protein was polished using size exclusion chromatography S200 (GE) in 20 mM Tris pH 8 with 200 mM NaCl. Finally, protein was concentrated to 20 mg/mL prior to further experiments as determined by absorbance at 280 nm using the extinction coefficient of 1.40 for OpGH99A.[21]

Crystallization, data collection, structure solving, and refinement

Pure concentrated protein was screened for crystallization conditions in a 96-well sitting drop format in a vapor diffusion experiment. Initial crystal conditions were optimized in 24-well hanging drops and the final crystallization condition was 0.1 M Tris pH 8, 20% PEG 3350. Crystals were soaked in mother liquor supplemented with 1 M KBr prior to cryo protection with 25% glycerol and then flash frozen in liquid nitrogen. Diffraction data were collected at EMBL Beamline P13[22]; 1200 images of 0.1° width were collected (Table I).

The data were processed using XDS, Pointless and Scala.[23,24] Molecular replacement was conducted using PHASER[25] in CCP4[26] and a Phyre[15] model based on BtGH99[10] (pdb id: 4acy) modified by chainsaw[27] to polyalanine and trimmed of spurious surface loops in Coot.[28] The top placed model with a TFZ-score of 6.7 was further trimmed of regions outside of density in Coot. The final trimmed polyalanine model was resubmitted to PHASER for a single solution with a TFZ-score of 10.2. The OpGH99A model was built using Autobuild in the Phenix package and finalized using Coot and Refmac5.[28–30] Molprobity was used for structure validation.[31]

Phylogenetic analysis

Sequences were extracted from NCBI Genbank using BlastP: five with OpGH99A (ZP_09498382), nine with BtGH99 (WP_011108998), and five with GH71 (GAA88202.1). The sequences were aligned using PROMALS3D with default values and the structures of OpGH99A, BxGH99 (pdb: 4ad1) and BtGH99 (pdb: 4acy) to structurally guide the alignment. A maximum likelihood reconstruction of phylogeny was conducted in MEGA7 using the Jones Taylor Thornton model and 1000 bootstrap replications.

Acknowledgment

The authors thank Thomas R. Schneider for beamtime support and Isabel Bento for assisting with the data collection.

REFERENCES

1. Wargacki AJ, Leonard E, Win MN, Regitsky DD, Santos CNS, Kim PB, Cooper SR, Raisner RM, Herman A, Sivitz AB, Lakshmanaswamy A, Kashiyama Y, Baker D, Yoshikuni Y (2012) An engineered microbial platform for direct biofuel production from brown macroalgae. Science 335:308–313.
2. Teeling H, Fuchs BM, Bennke CM, Krüger K, Chafee M, Kappelmann L, Reintjes G, Waldmann J, Quast C, Glöckner FO, Lucas J, Wichels A, Gerdts G, Wiltshire KH, Amann RI (2016) Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. Elife 5:1–29.
3. Martens EC, Koropatkin NM, Smith TJ, Gordon JI (2009) Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. J Biol Chem 284:24673–24677.
4. Sonnenburg ED, Zheng H, Joglekar P, Higginbottom SK, Firbank SJ, Bolam DN, Sonnenburg JL (2010) Specificity of polysaccharide use in intestinal *Bacteroides* species determines diet-induced microbiota alterations. Cell 141:1241–1252.
5. Koropatkin NM, Martens EC, Gordon JI, Smith TJ (2008) Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. Structure 16:1105–1115.
6. Cameron EA, Maynard MA, Smith CJ, Smith TJ, Koropatkin NM, Martens EC (2012) Multidomain carbohydrate-binding proteins involved in *Bacteroides thetaiotaomicron* starch metabolism. J Biol Chem 287: 34614–34625.
7. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, Kassabgy M, Huang S, Mann AJ, Waldmann J, Weber M, Klindworth A, Otto A, Lange J, Bernhardt J, Reinsch C, Hecker M, Peplies J, Bockelmann FD, Callies U, Gerdts G, Wichels A, Wiltshire KH, Glöckner FO, Schweder T, Amann R (2012) Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. Science 336:608–611.
8. Xing P, Hahnke RL, Unfried F, Markert S, Huang S, Barbeyron T, Harder J, Becher D, Schweder T, Glöckner FO, Amann RI, Teeling H (2015) Niches of two polysaccharide-degrading *Polaribacter* isolates

from the North Sea during a spring diatom bloom. ISME J 9:1410–1422.
9. Hakki Z, Thompson AJ, Bellmaine S, Speciale G, Davies GJ, Williams SJ (2015) Structural and kinetic dissection of the endo-α-1,2-mannanase activity of bacterial GH99 glycoside hydrolases from *Bacteroides* spp. Chemistry 21:1966–1977.
10. Thompson AJ, Williams RJ, Hakki Z, Alonzi DS, Wennekes T, Gloster TM, Songsrirote K, Thomas-Oates JE, Wrodnigg TM, Spreitz J, Stütz AE, Butters TD, Williams SJ, Davies GJ (2012) Structural and mechanistic insight into N-glycan processing by endo-α-mannosidase. Proc Natl Acad Sci USA 109:781–786.
11. Larsbrink J, Rogers TE, Hemsworth GR, McKee LS, Tauzin AS, Spadiut O, Klinter S, Pudlo NA, Urs K, Koropatkin NM, Creagh AL, Haynes CA, Kelly AG, Cederholm SN, Davies GJ, Martens EC, Brumer H (2014) A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. Nature 506:498–502.
12. Juncker AS, Willenbrock H, von Heijne G, Brunak S, Nielsen H, Krogh A (2003) Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Sci 12: 1652–1662.
13. Hahnke RL, Bennke CM, Fuchs BM, Mann AJ, Rhiel E, Teeling H, Amann R, Harder J (2015) Dilution cultivation of marine heterotrophic bacteria abundant after a spring phytoplankton bloom in the North Sea. Environ Microbiol 17:3515–3526.
14. Kwon Y-K, Kim JH, Kim JJ, Yang S-H, Ye B-R, Heo S-J, Hyun J-H, Qian Z-J, Park H-S, Kang D-H, Oh C (2014) *Ochrovirga pacifica* gen. nov., sp. nov., a novel agar-lytic marine bacterium of the family Flavobacteriaceae isolated from a seaweed. Curr Microbiol 69:445–450.
15. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10:845–858.
16. Reardon D, Farber GK (1995) The structure and evolution of α/β barrel proteins. FASEB J 9:497–503.
17. Davies G, Henrissat B (1995) Structures and mechanisms of glycosyl hydrolases. Structure 3:853–859.
18. Holm L, Sander C (1998) Touring protein fold space with Dali/FSSP. Nucleic Acids Res 26:316–319.
19. Boraston AB, Bolam DN, Gilbert HJ, Davies GJ (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem J 382:769–781.
20. Hehemann J-H, Smyth L, Yadav A, Vocadlo DJ, Boraston AB (2012) Analysis of keystone enzyme in agar hydrolysis provides insight into the degradation (of a polysaccharide from) red seaweeds. J Biol Chem 287:13985–13995.
21. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein Identification and Analysis Tools on the ExPASy Server. In: The Proteomics Protocols Handbook. pp. 571–607.
22. Cianci M, Bourenkov G, Pompidor G, Karpics I, Kallio J, Bento I, Roessle M, Cipriani F, Fiedler S, Schneider TR (2017) P13, the EMBL macromolecular crystallography beamline at the low-emittance PETRA III ring for high- and low-energy phasing with variable beam focusing. J Synchr Radiat 24:323–332.
23. Kabsch W (2010) XDS. Acta Cryst D66:125–132.
24. Evans PR (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. Acta Cryst D67:282–292.
25. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. J Appl Cryst 40:658–674.
26. Collaborative Computational Project N 4 (1994) The CCP4 suite: programs for protein crystallography. Acta Cryst D50:760–763.
27. Stein N (2008) CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. J Appl Cryst 41:641–643.
28. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of *Coot*. Acta Cryst D66: 486–501.
29. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA (2011) REFMAC 5 for the refinement of macromolecular crystal structures. Acta Cryst D67:355–367.
30. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Cryst D66:213–221.
31. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity : all-atom structure validation for macromolecular crystallography. Acta Cryst D66:12–21.