
Carbohydrate-binding proteins from marine bacteria

Ph.D. thesis by
Agata Anna Mystkowska

Dissertation zur Erlangung
des Doktorgrades der Naturwissenschaften
-Dr. rer. nat.-
im Fachbereich Geowissenschaften der Universität Bremen

April 2019



Bremen International Graduate School
for Marine Sciences



Max Planck Institute
for Marine Microbiology

The present work was developed in the MARUM-MPI Bridge Group for Marine Glycobiology at the MARUM Center for Marine Environmental Sciences of the University of Bremen and the Max Planck Institute for Marine Microbiology.

Title: Carbohydrate-binding proteins from marine bacteria

Autor: Agata Anna Mystkowska, M.Sc.
MARUM – Center for Marine Environmental Sciences
Max Planck Institute for Marine Microbiology
University of Bremen
Institute of Biochemistry, Greifswald University

Date of colloquium: 05.07.2019

Reviewers: Dr. Jan-Hendrik Hehemann
MARUM – Center for Marine Environmental Sciences
Max Planck Institute for Marine Microbiology
University of Bremen

Prof. Dr. Nicole M. Koropatkin
University of Michigan

Study motivation - in brief:

Polysaccharides present outstanding structural diversity and are difficult to analyze with current methods. They are a major carbon source in the ocean but little is known about polysaccharide structures in the sea; thus, we need novel tools to identify and quantify these polymers. Marine bacteria are highly specialized in polysaccharide utilization. Specifically, they contain carbohydrate-binding proteins to detect and bind polysaccharides from the environment. Particular carbohydrate-binding proteins from marine bacteria could potentially be exploited as polysaccharide-specific probes. In my thesis, I studied in detail the function of several carbohydrate-binding proteins and investigated their potential as glycan probes aiding marine polysaccharide analyses.

Summary

Marine phytoplankton is responsible for about half of the CO₂ fixation on Earth. Algal photosynthesis results in the production of organic carbon, a substantial fraction of which is in the form of polysaccharides. Bacterial utilization of polysaccharides is a highly relevant process, which constitutes a major carbon and energy transfer in the ocean. *Bacteroidetes*, key marine polysaccharide degraders, are known to employ a complex protein system to bind, transport and digest polysaccharides. It was first characterized for the starch utilization system in *Bacteroides thetaiotaomicron*. In contrast to terrestrial or human gut microbes, polysaccharide degradation by marine bacteria remains largely unexplored.

The main aim of my thesis was to investigate the molecular details of microbial polysaccharide utilization in the ocean. More specifically, during my Ph.D. thesis, I investigated carbohydrate-binding proteins involved in the recognition of substrates at the surface of bacterial cells. Based on the knowledge from homologous systems and bioinformatic predictions, we assumed that marine *Bacteroidetes* assemble outer membrane complexes composed of surface-glycan binding proteins to acquire polysaccharides. Previous experiments using fluorescently-labeled polysaccharides showed that marine *Bacteroidetes* take up polysaccharides in a selfish manner. These bacteria use surface-associated enzymes and binding proteins to partially degrade polysaccharides and minimize production of freely diffused hydrolysis products. For this approach, bacteria must have evolved a highly efficient and selective binding apparatus, which I studied in detail in this thesis. These in-depth analyses were necessary to evaluate the potential of carbohydrate-binding proteins as novel glycan probes.

In the first manuscript, I present the characterization of GMSusD protein. We focused our analyses on proteins putatively specific to the highly abundant marine polysaccharide - laminarin. Biochemical and structural analyses on the GMSusD from *Gramella* sp. MAR_2010_102 revealed the predicted laminarin binding. Surprisingly, the protein was specific to a particular type of laminarin structure. There is a big discrepancy between vast “omics” sequence data and functionally or structurally characterized proteins. To provide more accurate support for bioinformatic predictions we performed structure-guided alignment of metagenomes of global surface water datasets using the structure of GMSusD as a guide. We found SusD-like proteins with structurally conserved residues of the binding site in different locations in the ocean, suggesting a similar manner of laminarin recognition by these proteins.

In the second manuscript, we identified two additional laminarin-binding proteins from the same planktonic bacteria. Upstream to the GMSusD, in the gene cluster called a Polysaccharide utilization locus, there were two genes encoded with previously unknown

function. After excluding their enzymatic activity, we analyzed binding abilities. Based on the identified laminarin binding activity and predicted three-dimensional structures, we propose that these two proteins, GMSusE and GMSusF, belong to a highly unexplored group of SusEF-like binding proteins.

Finally, I applied medium throughput expression of recombinant putative carbohydrate-binding proteins to investigate their potential as glycan probes. The field of marine glycobiology needs to be extended, since little is known about structures of polysaccharides present in the ocean. Thus, we proposed taking advantage of bacterial proteins, which are expressed in response to algal blooms in the North Sea. We investigated a library of forty-seven constructs resulting in the production of twelve soluble recombinant proteins, the binding of which was tested with environmental algal extracts and well-defined polysaccharide controls. These analyses allowed us to discover four novel carbohydrate-binding proteins specific to laminarin, α -mannan and β -mannan. However, we encountered some limitations of this approach, which are discussed in the third manuscript.

The research performed in this thesis contributes to our greater quest to understand algal carbohydrate binding by marine microbes, which is a crucial mechanism for bacterial polysaccharide utilization and therefore key in the marine carbon cycle.

Zusammenfassung

Marines Phytoplankton ist für etwa die Hälfte der globalen CO₂-Fixierung verantwortlich. Die Photosynthese von Algen führt zur Produktion von organischem Kohlenstoff, der zu einem erheblichen Teil in Form von Polysacchariden vorliegt. Die bakterielle Nutzung von Polysacchariden ist ein hochgradig relevanter Prozess, der den Haupttransfer von Kohlenstoff und Energie im Ozean darstellt. Bekanntermaßen verwenden *Bacteroidetes*, wichtige marine Polysaccharid degradierenden Bakterien, ein komplexes Proteinsystem um Polysaccharide zu binden, transportieren und zu verdauen. Ein solches System wurde erstmals für das Verwertungssystem der Stärke in *Bacteroides thetaiotaomicron* charakterisiert. Im Gegensatz zu terrestrischen Bakterien oder Mikroben des menschlichen Darms bleibt der Abbau von Polysacchariden durch marine Bakterien weitgehend unerforscht.

Der Fokus meiner Dissertation liegt in der Erforschung der molekularen Details der mikrobiellen Polysaccharidnutzung im Ozean. Während meiner Promotion untersuchte ich kohlenhydratbindende Proteine die an der Erkennung des Substrats an der Oberfläche der bakteriellen Zelle beteiligt sind. Basierend auf den Erkenntnissen homologer Systeme und bioinformatischer Vorhersagen nahmen wir an, dass marine *Bacteroidetes* zur Aufnahme von Polysacchariden einen äußeren Membrankomplex bilden, der aus Oberflächen-Glycan-Bindungsproteinen besteht. Frühere Versuche mit fluoreszenzmarkierten Polysacchariden zeigten, dass marine *Bacteroidetes* auf eine eigennützige Art und Weise Polysaccharide aufnehmen, wobei Zellen oberflächenassoziierte Enzyme und Bindungsproteine verwenden, um die Polysaccharide teilweise abzubauen und die Freisetzung von frei diffundierenden Hydrolyseprodukten zu minimieren. Für einen solchen Ansatz evolvierten Bakterien einen hocheffizienten und selektiven Bindungsapparat, den ich in dieser Arbeit ausführlich untersuche. Diese detaillierten Analysen waren notwendig, um das Potential von CBPs als neuartige Glycan-Sonden zu beurteilen.

Im ersten Manuskript präsentiere ich die Charakterisierung des GMSusD-Proteins. Wir fokussierten unsere Analyse auf Proteine, die spezifisch für das sehr abundante marine Polysaccharid Laminarin sind. Biochemische und strukturelle Untersuchungen an GMSusD von *Gramella* sp. MAR_2010_102 zeigten die vorhergesagte Laminarinbindung. Überraschenderweise ist das Protein spezifisch für einen bestimmten Typ von Laminarinstruktur. Es gibt eine große Diskrepanz zwischen großen "omics"-Sequenzdaten und funktionell oder strukturell charakterisierten Proteinen. Um eine genauere Unterstützung bioinformatischer Vorhersagen zu bieten, haben wir ein strukturgesteuertes Alignment von Metagenomen globaler Oberflächenwasser-Datensätze vorgenommen, wobei die Struktur von GMSusD als Leitfaden diente. Wir fanden SusD-ähnliche Proteine mit

strukturell konservierten Aminosäureresten an verschiedenen Stellen im Ozean, was auf eine ähnliche Art der Erkennung von Laminarinen durch diese Proteine hinweist.

Im zweiten Manuskript haben wir zwei weitere laminarinbindende Proteine aus denselben Planktonbakterien identifiziert. Im Gencluster, dem sogenannten Polysaccharide utilization locus (PUL), dem GMSusD vorgelagert, sind zwei Gene mit bisher unbekannter Funktion. Nachdem ihre enzymatische Aktivität ausgeschlossen wurde, analysierten wir ihre Bindungsfähigkeiten. Anhand der Bindung von Laminarin und der vorhergesagten dreidimensionalen Strukturen schließen wir darauf, dass die beiden Proteine, GMSusE und GMSusF, zu einer sehr wenig erforschten Gruppe von SusEF-ähnlichen Bindungsproteinen gehören.

Schlussendlich habe ich die Expression von rekombinanten kohlenhydratbindenden Proteinen mit mittlerem Durchsatz angewendet, um deren Potential als Glycan-Sonden zu untersuchen. Das Gebiet der marinen Glykobiologie bedarf einer Erweiterung angesichts der vielen unerforschten Polysaccharidstrukturen im Ozean. Wir schlagen daher vor, bakterielle Proteine zu nutzen, die als Reaktion auf Algenblüten in der Nordsee exprimiert werden. Wir untersuchten eine Zusammenstellung von etwa fünfzig Konstrukten, die zur Produktion von elf löslichen, rekombinanten Proteinen führten und mit Algenextrakten aus der Umwelt und definierten Polysaccharidkontrollen getestet wurden. Auf diese Weise konnten wir vier neuartige kohlenhydratbindende Proteine finden, die für Laminarin, α - und β -Mannan spezifisch sind. Jedoch gibt es Einschränkungen dieses Ansatzes, die im dritten Manuskript behandelt werden.

Die gesamte Forschungsarbeit trägt zum größeren Bestreben bei, die Bindung von Kohlenhydraten aus Algen durch Mikroben besser zu verstehen. Dieser Mechanismus ist entscheidend für die bakterielle Verwertung von Polysacchariden und spielt folglich eine ebenso fundamentale Rolle im Kohlenstoffkreislauf der Ozeane.

CONTENTS

1. INTRODUCTION	15
Polysaccharides as a component of marine organic matter	17
Outstanding complexity of polysaccharide molecules	19
Marine polysaccharides	20
Bacterial polysaccharide utilization	22
Surface-glycan binding proteins	25
Selfish uptake of marine polysaccharides	29
Applications of carbohydrate-binding proteins	30
Research aims and thesis outline	34
Contribution to manuscripts and further publications	35
2. MANUSCRIPTS	39
Manuscript 1: Molecular recognition of the beta-glucans laminarin and pustulan by a SusD-like glycan-binding protein of a marine <i>Bacteroidetes</i>	41
Manuscript 2: Marine SusEF-like proteins from <i>Gramella</i> sp. Mar_2010_102 bind laminarin	73
Manuscript 3: Exploring the potential of bacterial carbohydrate-binding proteins as novel probes to detect marine polysaccharides	93
3. GENERAL DISCUSSION	115
4. CONCLUDING REMARKS	125
POPULAR SUMMARY	129
ACKNOWLEDGMENTS	135
BIBLIOGRAPHY	137
APPENDIX	153

Abbreviation list

BSA	Bovine serum albumin
CAZymes	Carbohydrate-active enzymes
CBMs	Carbohydrate-binding modules
CBPs	Carbohydrate-binding proteins
CE	Carbohydrate esterase
DLS	Dynamic light scattering
ELISA	Enzyme-linked immunosorbent assay
FLA-PS	Fluorescently labeled polysaccharides
GH	Glycoside hydrolases
GOS	Global Ocean Sampling
HPAEC-PAD	High Performance Anion Exchange Chromatography with Pulsed Amperometric Detection
HPLC	High Performance Liquid Chromatography
IMAC	Immobilized Metal Ion Affinity Chromatography
Ig	Immunoglobulin
IO	Indian Ocean
ITC	Isothermal titration calorimetry
MMseqs2	<i>Many-against-Many sequence searching 2</i>
MPBS	Milk phosphate buffered saline
MS	Mediterranean Sea
NAO	North Atlantic Ocean
NPO	North Pacific Ocean
ORFs	Open Reading Frames
OSD	Ocean Sampling Day
PAHBAH	p-hydroxybenzoic acid hydrazide
PDB	Protein Data Bank

PKD	Polycystic kidney disease
PL	Polysaccharide lyases
PUL	Polysaccharide utilization locus
SAO	South Pacific Ocean
SEC	Size Exclusion Chromatography
SGBPs	Surface glycan binding proteins
SO	Southern Ocean
SP	Signal peptide
SPO	South Pacific Ocean
SR-SIM	Super-resolution structured illumination microscopy
Sus	Starch utilization system
TARA	TARA Ocean Expedition
TBDTs	TonB-dependent transporters
TEP	Transparent exopolymer particles

Introduction

Most biological processes are driven by molecular interactions. Protein-carbohydrate binding has profound meaning for bacterial utilization of marine polysaccharides, which is an important part of the global carbon cycle. In detail studies on protein-carbohydrate interactions will elucidate the microbial role in the marine biogeochemical cycles. Exploring biological function of carbohydrate-binding proteins is necessary to investigate their potential as glycan-binding probes to detect marine polysaccharides. Growing research on marine polysaccharides is stimulated by scientific curiosity and also by the potential for pharmaceutical, cosmetic and food industries.

Polysaccharides as a component of marine organic matter

Atmospheric carbon dioxide is fixed into organic carbon by photosynthetic organisms in the upper layer of the ocean. This great reservoir of organic material is divided into DOM - dissolved organic matter and POM - particulate organic matter. POM and DOM convert into each other; therefore, there is a fine line between dissolved and insoluble fractions. The operational difference between POM and DOM is defined by the pore size used by researchers to separate the two fractions, where DOM passes through the 0.2 μm cut-off filter and POM retains on this filter (Druffel *et al.*, 1992). Phytoplankton contributes to DOM by releasing extracellular exudates into the seawater, which stimulates bacterial production leading to multiple mutual benefits for both groups of organisms. For example, iron is a limiting factor for phytoplankton growth. This element can be bound by bacteria and as a consequence iron concentration in the environment decreases, which is in favor of phytoplankton (Emerson & Hedges, 2008). DOM release can be either the active exudate or passive leakage by diffusion across the cell. It is hypothesized that in response to increasing CO_2 concentration in the atmosphere, DOM exudation undergoes the overflow mechanism. It states that photosynthesis is more rapid than required for growth, which results in the primary products being released directly to the water (Thornton, 2014). Phytoplankton biomass enters the food web during viral lysis (Bettarel *et al.*, 2005), sloppy feeding by predators (Møller, 2007) or cell death (Veldhuis *et al.*, 2001), which makes it available as DOM and POM to heterotrophic bacteria (Figure 1.1). POM is transported by sinking in the form of “marine snow”. These sinking particles create different niches for diverse microorganisms since conditions like temperature and light vary on different depths. Gradually with increasing depth, the concentration of an organic matter is lower as heterotrophic bacteria degrade it. These processes constitute a basic pillar of the biological carbon pump (Figure 1.1).

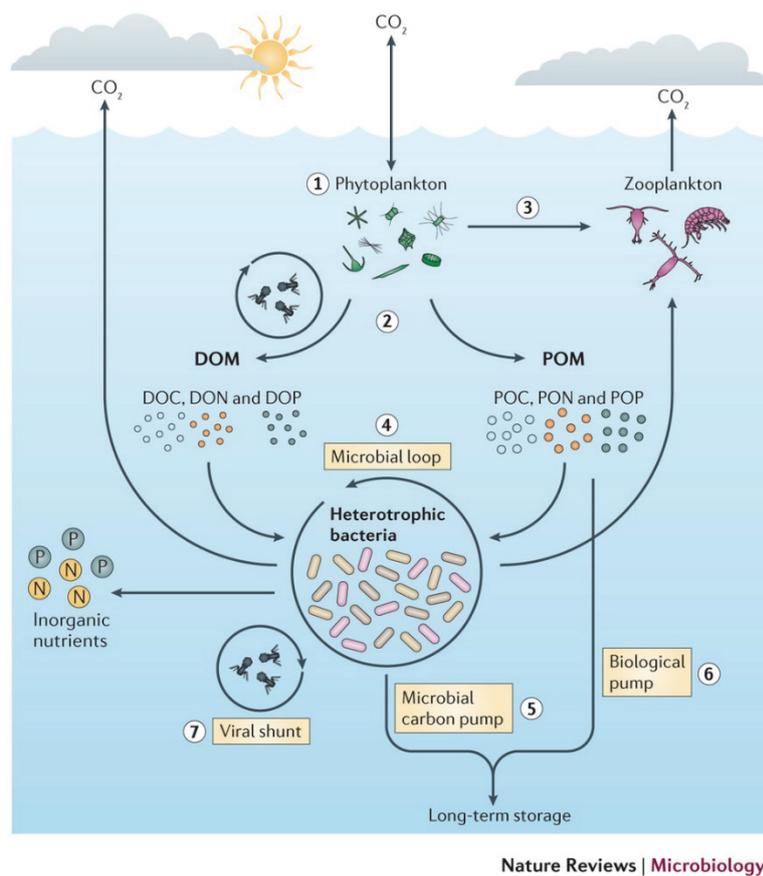


Figure 1.1 Transformation of algal-derived organic matter. Phytoplankton fixes CO_2 to organic carbon (1); dissolved organic matter (DOM) and particulate organic matter (POM) are released from phytoplankton (2); phytoplankton biomass is consumed by zooplankton grazers (3) and the organic matter undergoes mineralization and recycling by diverse heterotrophic bacteria (4). A fraction of the heterotrophic bacteria is consumed by zooplankton, and the carbon is further transferred to the food web. Part of the organic carbon is transferred into recalcitrant DOC (5) and POM (6). A viral-mediated cell lysis contributes to the release of dissolved and particulate matter from both the phytoplanktonic and bacterial pools (7). The figure was taken from (Buchan *et al.*, 2014).

Marine organic material contains substances such as amino acids, lipids, nucleic acids, and carbohydrates (Lee *et al.*, 2004). The latter makes up a large fraction of the organic carbon in the sea. Carbohydrates (including mono-, oligo- and polysaccharides) comprised 20-40% of the phytoplankton biomass (Parsons *et al.*, 1961) and constitute up to 50% of DOM (Benner *et al.*, 1992; Borch & Kirchman, 1997; Aluwihare & Repeta, 1999). However, the exact composition of carbohydrates, especially polysaccharides remains largely uncharacterized (Arnosti, 1995, 2003; Lee *et al.*, 2004)

Outstanding complexity of polysaccharide molecules

Carbohydrates serve as storage and structural components in living organisms. In comparison to DNA, RNA or proteins, structures of polysaccharides (polymeric carbohydrates) are much more complex. In contrast to a linear peptide or phosphodiester bonds, glycosidic linkages between monosaccharides create extended and branched structures. Each building block possesses multiple hydroxyl groups, which potentially form a bond within polysaccharide chains (regiochemistry). Moreover, monosaccharides can be linked by α - or β -glycosidic bonds (stereochemistry). In polysaccharide chains, non-reducing and reducing ends can be distinguished, where the second retains its chemical reactivity as it has a free aldehyde or ketone group in an open-chain form. Additionally, polysaccharides are in some cases chemically modified with carboxyl, phosphate and/or sulfuric ester groups. For instance, sulfation is a very common modification in the marine environment (Bertozzi & Rabuka, 2009). To illustrate the diversity of these molecules, the number of theoretically unique structures of polysaccharide and nucleic acid polymers was compared. It has been calculated that hexasaccharides can form 1.05×10^{12} theoretically unique structures, whereas a hexanucleotide can form 4.1×10^3 (Laine, 1994).

Specific glycan probes are powerful tools to overcome this high complexity of polysaccharides. Note, glycan defined by International Union of Pure and Applied Chemistry as “compounds consisting of a large number of monosaccharides linked glycosidically” are used as synonyms of polysaccharides. In figure 1.2A an example of two similar glucan structures have been presented. Both have the same monosaccharide composition, but different linkages, which makes them difficult to differentiate. However, there are proteins which specifically recognize these fine variations (Figure 1.2B). CBM3a is an example of commercially available probe (Blake *et al.*, 2006), which is extensively used for the detection of plant-derived polysaccharides. Glycan probes are either in form of antibodies or carbohydrate-binding proteins (CBPs), but we miss probes specific to marine polysaccharides. To extend the collection of glycan probes and apply them to detect algal-derived polysaccharides, I propose in detail characterization of CBPs from marine bacteria. To judge the potential of CBPs as glycan probes to identify polysaccharides in the environment, I will first review what is known about marine polysaccharides, their bacterial uptake and finally the applications.

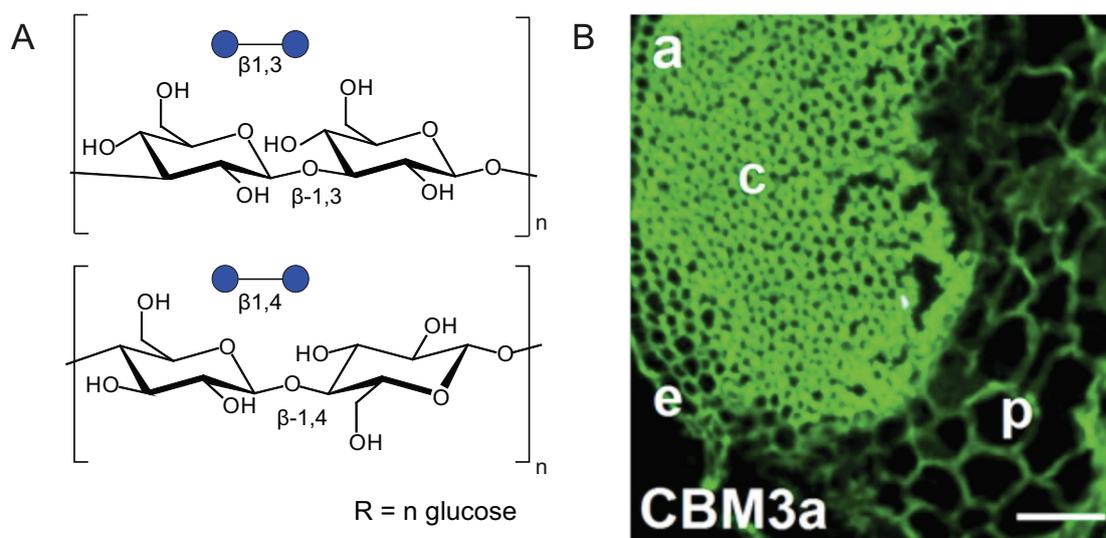


Figure 1.2 Diversity of polysaccharides can be overcome with the specific glycan probes. **A:** Example of two glucose polymers with a different type of glycosidic bond: β -1,3-linked and β -1,4-linked glucans. Above the schematic chemical structure, the cartoon representation is shown in blue. **B:** Image of immunofluorescent detection of glycan probe CBM3a binding to sections of celery petioles. Letters stand for epidermis (e), collenchyma (c), and parenchyma (p). The part B of this figure was taken from (Blake *et al.*, 2006).

Marine polysaccharides

There is a considerable discrepancy in the knowledge about algal and plant polysaccharide composition, where the second is extensively characterized. The following monosaccharides can be found in the two environments: arabinose, fucose, galactose, galacturonic acid, glucose, glucuronic acid, mannose, rhamnose, xylose (Painter, 1983; Gao *et al.*, 2011). Several structures of plant polysaccharides have been observed in the ocean, such as cellulose, mannan, and xylan (Painter, 1983). However, algae also possess different polysaccharides than land plants, since these two groups of organisms have evolved independently (Michel *et al.*, 2010). A common modification present in algal and absent in plant polysaccharides is sulfate-esterification. Sulfated galactans such as agarose and carrageenan (Painter, 1983) contain D-galactose based units alternatively linked by β -1,4 and α -1,3 linkages but differ in the number and position of the ester sulfate groups (Maciel *et al.*, 2007). Sulfated galactans are widespread in the cell wall of brown algae, whereas alginate, a polymer of the uronic acids and fucose-containing sulfated polysaccharides (FCSP) are present in the cell wall of the brown algae (Ale & Meyer, 2013). Another sulfated polysaccharide ulvan, found in green algae is mainly composed of glucose, glucuronic acid

rhamnose, and xylose (Lahaye & Robic, 2007). The algal storage polysaccharide is laminarin, a glucose polymer (Gobet *et al.*, 2018).

Furthermore, structures of polysaccharides from unicellular algae – diatoms are even more unexplored. Diatom polysaccharide structures have been studied for years (Rees, 1963), but the complexity of polysaccharides is highly challenging and therefore, there are still many new polysaccharides to explore and analyze in detail. For example, the sulfated α -mannan has been discovered in the cell wall of diatom *Phaeodactylum tricornutum* (Le Costaouëc *et al.*, 2017). The relatively well-studied marine polysaccharide is so-called chrysolaminarin, the storage component of microalgae (Gügi *et al.*, 2015). Macroalgal laminarin typically terminates by mannitol end-groups, whereas chrysolaminarin from diatoms lacks this group (Myklestad & Haug, 1972). In the following thesis, it will be only referred as laminarin. This degradation of widespread laminarin is an objective of this dissertation and will be further described in the following subchapter.

Laminarin is one of the most abundant polysaccharides in the ocean and it serves as a carbon source for bacteria. An annual algal laminarin production is estimated to 18±9 gigatons (Becker, 2018). This polysaccharide is defined as a β -1,3-linked glucose polymer with β -1,6-linked glucose side chains (Figure 1.3). The molecular weight of laminarin is assumed to be approximately 5 kDa (Myklestad & Haug, 1972; Alderkamp *et al.*, 2007; Kadam *et al.*, 2015). However, laminarins vary notably between algal species (Rioux & Turgeon, 2015). For example, the molecular weight of laminarin isolated from *Eisenia bicyclis* was reported as notably high (19-27 kDa) (Menshova *et al.*, 2014). Another clear difference concerns the amount of β -1,6-linked glucose branching. For instance, the diatom *Phaeodactylum tricornutum* was characterized as an essentially linear β -1,3-linked glucan with only a small degree of branching (Ford & Percival, 1965). More precisely, it has been characterized by NMR spectroscopy for example for *Chaetoceros debilis*, which contains even up to 37% of β -1,6-glucan content and it is called laminarin with pustulan-like branches (Note: pustulan is a β -1,6-glucose polymer) (Størseth *et al.*, 2006). Moreover, β -1,2-linked glucose branching has been reported (Rioux *et al.*, 2010). Laminarin accumulated in the presence of light is stored in the vacuole inside the chloroplast and its concentration decreases in the darkness (Chiovitti *et al.*, 2004). In the literature, laminarin is also referred to as laminaran.

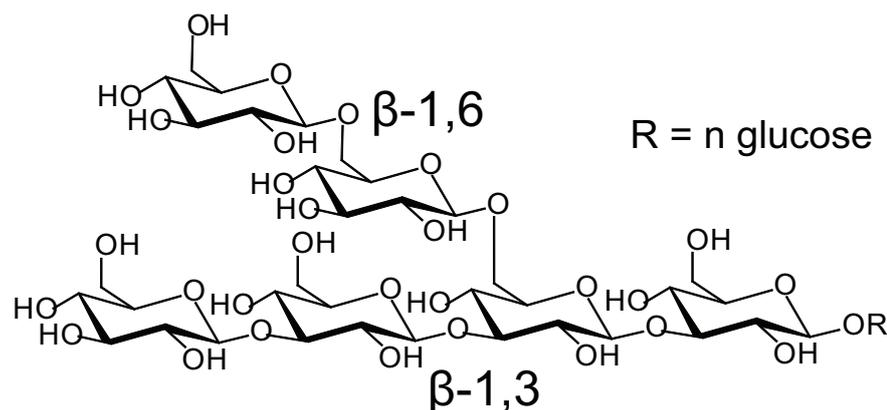


Figure 1.3 Structure of laminarin molecule. The scheme represents a structure of laminarin, which is a linear β -1,3-glucan backbone with β -1,6-glucose branches.

Marine polysaccharides containing various modifications and frequent branching represent an outstanding heterogeneity. Additionally, their structures vary depending on conditions such as nutrient sources, diel or life cycles of the phytoplankton (Chiovitti *et al.*, 2004; Gügi *et al.*, 2015). However, particular bacterial species are highly specialized in the recognition and degradation of these complex molecules. In the following subchapter, I will describe the current state of knowledge about bacterial polysaccharide utilization and the details of the proteins responsible for recognition and binding of polysaccharides.

Bacterial polysaccharide utilization

Polysaccharides, one of the main components of algae (Kraan, 2012) are known to be a carbon source for marine bacteria. Given the great variety of algae in the ocean, diverse niches have been provided for microbes (Martin *et al.*, 2014). Algal blooms are particularly interesting for monitoring bacterial-algal interactions and dynamics of the substrate succession. One of the major algal groups are diatoms, which are single-cell organisms with a silica-based cell wall. They contribute to 40% of primary production in the marine ecosystem (Field *et al.*, 1998). Extensive, environmental studies on annual diatom bloom were conducted at the coastal station Helgoland, located in the German Bight, North Sea (Teeling *et al.*, 2012; Xing *et al.*, 2014; Teeling *et al.*, 2016). Bacterial community composition changes in response to the algal-derived organic matter release. Pronounced response of *Bacteroidetes* was observed upon the beginning of the bloom. Further *Gammaproteobacteria* and the *Roseobacter* also increased their abundance, albeit with less extent, whereas other groups such as free-living planktonic bacterial clade SAR11 decreased (Teeling *et al.*, 2012).

Reoccurrence of *Bacteroidetes* in bloom-associated communities has been observed in various areas throughout the ocean (Kirchman, 2002; Pinhassi *et al.*, 2004; Buchan *et al.*, 2014). This phylum, which exists in the literature also as *Cytophaga-Flavobacterium*, encompasses four classes: *Bacteroidia*, *Cytophagia*, *Flavobacteria*, and *Sphingobacteria* (Krieg *et al.*, 2010). Members of *Bacteroidetes* are globally distributed, Gram-negative bacteria, which are present in marine environments such as coastal waters (Kirchman *et al.*, 2003), open ocean (Schattenhofer *et al.*, 2009), sediments (Wang *et al.*, 2018), and hydrothermal vents (Kormas *et al.*, 2006). Marine *Bacteroidetes* have been considered since a long time to be the key players in mineralization of complex organic matter, in particular, polysaccharides (Reichenbach & Weeks, 1981). However, Teeling *et al.* demonstrated for the first time that marine *Bacteroidetes* express particular proteins involved in polysaccharide degradation in response to polysaccharide release during algal blooms (Teeling *et al.*, 2012).

Metagenomic and metaproteomic analysis of *Bacteroidetes* during Helgoland bloom revealed expression of carbohydrate-active enzymes (CAZymes), TonB-dependent transporters (TBDTs) and CBPs. Genome sequencing combined with functional studies confirmed *Bacteroidetes* specialization in glycan degradation. Various molecular adaptations in these bacteria have been observed. For example, *Gramella forsetii* KT0803^T possesses an enlarged set of polysaccharide utilization enzymes and therefore is a generalist degrader (Kabisch *et al.*, 2014), whereas *Formosa* strains A and B are specialized as they have genomic ability to process an only small set of sugars (Unfried *et al.*, 2018). Genes that encode CAZymes, TBDTs and CBPs are clustered together in the genomes of bacteria from this phylum.

Bacteroidetes contain genomic islands dedicated to polysaccharide degradation called PUL- Polysaccharide utilization locus. According to the classical definition, they have to contain homologs of at least two of the genes that encode SusC: TBDTs and SusD: main CBP (Martens, Chiang, *et al.*, 2009). However, PULs without SusCD pairs have also been described. For example, a functional carrageenan PUL does not include a SusCD pair, but it has been showed that a distal acting SusCD pair is responsible for the oligo-carrageenan uptake (Ficko-Blean *et al.*, 2017). PULs encode CBPs and a set of CAZymes such as carbohydrate esterases (CEs), GHs and polysaccharide lyases (PLs). Besides, there are regulators, proteases, and sulfatases in case of marine polysaccharides (Kappelmann *et al.*, 2019). Due to the high diversity of glycans, there are various PULs containing different CAZymes families, which allows bacteria to utilize several polysaccharides. The total numbers of PULs within one species vary and is connected with polysaccharides-degrading capability. Based on the current studies they range from few to more than a hundred, for example, *Alistipes putredinis* DSM 17216 possesses only 1 PUL and *Bacteroides ovatus* ATCC 8483 has 106 PULs (Terrapon

Introduction

et al., 2015). Kappelmann *et al.* performed an analysis of 400 PULs from the 53 genomes of North Sea *Flavobacteria*. In this environment, on average 7.5 PULs per bacterial genome have been identified. Around 78% of surface water isolates contained at least one laminarin PUL. They distinguished four variants based on the encoded GH, which reflects the structural differences of laminarin (Kappelmann *et al.*, 2019).

Genes encoded in the PUL are expressed to proteins which assemble into an outer membrane multiprotein system. Salyers' group identified the first cluster encoding proteins involved in the starch binding, transport, and degradation. The system was composed of susR, A, B, C, D, E, F and G proteins and named the starch utilization system (Sus) (Figure 1.4) (Tancula *et al.*, 1992; D'Elia & Salyers, 1996). The Sus system is assembled in the outer membrane and periplasm of Gram-negative bacteria. Polysaccharides are specifically recognized and bound by the proteins SusE, SusF, and SusD; and digested to oligosaccharides by the outer membrane glycoside hydrolases (GH) SusG. Shorter chains are transported across the outer membrane by SusCD proteins and afterward degraded by the periplasmic GH-like SusA and SusB. When the disaccharide of glucose is present in the periplasm, SusR triggers the upregulation of the PUL gene cluster (Martens, Koropatkin, *et al.*, 2009). PULs and Sus-like systems are common features of *Bacteroidetes*. Research on *Bacteroides thetaiotaomicron*, a model polysaccharide degrader and a prominent member of the human gut microbiota, provided valuable insight into the mechanism of polysaccharide utilization.

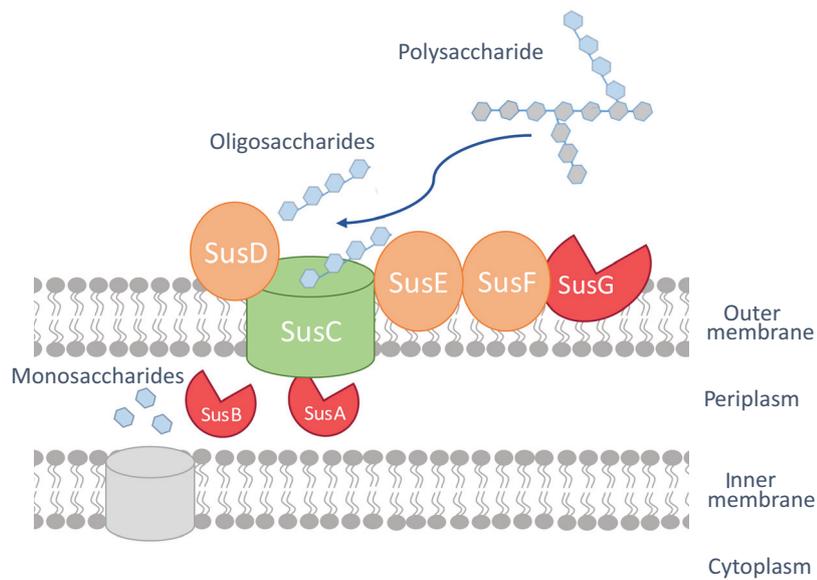


Figure 1.4 Scheme of the starch utilization system (Sus). It is assembled in the outer membrane and periplasm of the *Bacteroidetes*. A polysaccharide is recognized and bound by surface-glycan binding proteins and processed by outer membrane-associated glycoside hydrolases. Generated oligosaccharides are transported across the outer membrane by SusC-like protein. Saccharides digested into mono- or disaccharides by periplasmic glycan-degrading enzymes are transported across inner membrane. Figure adapted from (Bolam & Koropatkin, 2012).

Surface glycan-binding proteins

There is a multiplicity of polysaccharides in the marine environment. Members of the phylum *Bacteroidetes* evolved CBPs called surface glycan-binding proteins (SGBPs) to sense appropriate polysaccharides from the surrounding environment. Seminal work of Abigail Salyers on gut *Bacteroidetes* laid the foundation for studying the initial steps involved in the polysaccharide binding (Anderson & Salyers, 1989; Shipman *et al.*, 2000) and this research continues as so many questions remain or continue to appear.

The most characterized CBP from Sus-like system is called SusD-like. The first SusD protein was described for starch, but it was also identified for other polysaccharides present in the gut environment (Koropatkin *et al.*, 2008; Tauzin *et al.*, 2016; Glenwright *et al.*, 2017) as well as in soil (Larsbrink *et al.*, 2016). The presence of SusD-like genes has been observed in multiple PULs (Bjursell *et al.*, 2006) showing the conserved role of this protein. The three-dimensional structure of SusD revealed α -helical composition (Koropatkin *et al.*, 2008), which was uncommon for characterized carbohydrate-binding modules (CBMs) that comprised β -sandwich fold (Boraston *et al.*, 2004). Also, the molecular weight for SusD-like proteins (around 50-60 kDa) is higher than for typical CBMs. Although the sequence similarity between SusD-like proteins is often relatively low, the architecture of the structure is alike

Introduction

between known SusD-like proteins (Koropatkin *et al.*, 2009). SusD-like protein structures have highly conserved tetratricopeptide repeats (TPRs) motif and the single binding site, which shape differs depending on the substrate. TPRs are known to be involved in protein-protein interaction, but so far this function has not been observed for SusD-like proteins. The starch binding SusD shows a higher affinity towards cyclic than linear forms, which indicates its preference for shape recognition rather than the monomeric composition of the substrate. This type of protein shows relatively low affinity to the carbohydrate compared to CBMs. Other experiments revealed that SusD is essential for the growth of *B. thetaiotaomicron* on a substrate longer than six sugar units (Koropatkin *et al.*, 2008). However, it seems that the most important function of SusD-like proteins aside from binding is to create an outer membrane complex with other SGBPs.

Glenwright *et al.* reported the first structure of SusCD-like proteins. Since the complex of SusCD-like proteins could not be produced as soluble recombinant proteins in *E. coli*, they extracted it directly from *B. thetaiotaomicron*. This approach was successful as the complex is highly expressed (approximately 1 mg l⁻¹). Other than SusCD-like proteins, the isolated complex contained two additional small lipoproteins. All four proteins are encoded together (BT2261–64), but not in a PUL or in the vicinity of any other polysaccharide utilizing proteins. The two additional lipoproteins tightly connect with SusCD-like proteins after purification and have a structure akin to the characterized starch specific SusEF proteins. This result stresses the importance of the outer membrane complex formation. Based on the molecular dynamics simulations and SusCD structure obtained via X-ray crystallography, Glenwright *et al.* proposed a “pedal bin” mechanism, where SusD is located on top of SusC and they interact in a hinge-like manner. In the absence of the ligand, SusD is mobile and dissociates from SusC to expose the carbohydrate-binding site. Upon ligand binding, SusD behaves as a lid and enters a “closed” state in relation to SusC. Additionally, SusC changes the conformation of the plug domain, which allows the substrate to go through the outer membrane (Figure 1.5) (Glenwright *et al.*, 2017).

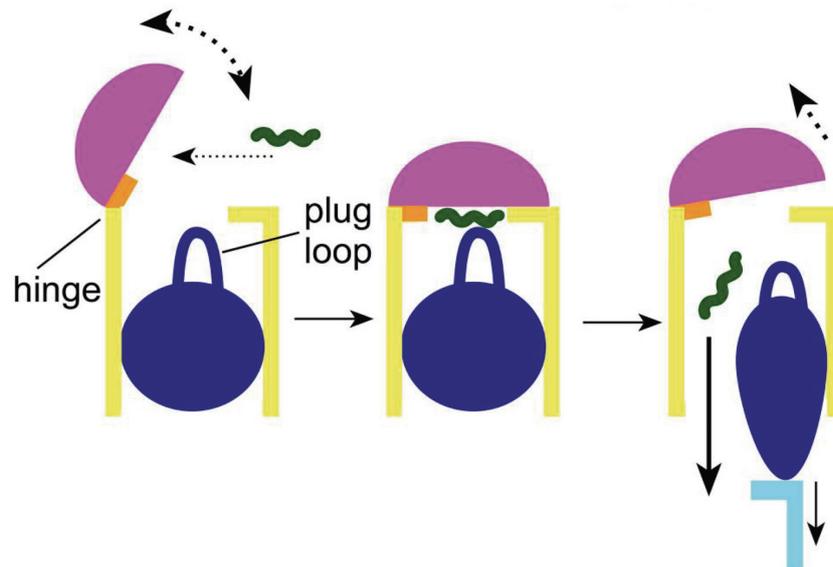


Figure 1.5 Cartoon representation of the substrate transport by SusCD complex. Substrate is transported in to the cell by the proposed pedal bin mechanism, where SusD acts as a lid. SusC is coloured yellow, the plug domain is dark blue; SusD is shown in magenta. The figure was taken from (Glenwright *et al.*, 2017).

Polysaccharide binding to the bacterial cell is also coordinated by SusEF-like proteins. Unlike SusD-like proteins, SusEF are not extensively characterized. Low sequence identities within these group of proteins makes them difficult to annotate. The SusEF-like proteins are encoded in the vicinity of SusCD genes, sometimes named as SusEF-positioned genes. The structures of starch-specific SusE and SusF proteins revealed a multidomain architecture comprising CBMs - three for SusE and two for SusF. Each of the domains binds with a slightly different affinity towards various starch oligosaccharides. This allows them to accommodate distinct regions of the polysaccharide. The N-terminus of SusF consists of a β -barrel fold similar to immunoglobulin-like (Ig-like) domains, whereas the N-terminal domain for SusE was not resolved in the crystal structure, but modeled to have also an Ig-like fold (Cameron *et al.*, 2012). Another characterized SusE-positioned protein specific for xyloglucan - SGBP-B is composed of three tandem immunoglobulin Ig-like domains followed by a xyloglucan-binding CBM (Tauzin *et al.*, 2016). Interestingly, SusE-like proteins specific for starch and xyloglucan have no similarity at the level of amino acid sequence.

Besides the binding ability, SGBPs are involved in the outer membrane complex formation. Studies on the model starch binding system show that presence of SusD is more important than its binding ability for the bacterial growth on starch. The *B. thetaiotaomicron* mutant with the deletion of SusD (Δ susD) is not able to grow on amylopectin, one of the two components of starch. To test if the loss of growth ability is due to the function or the presence of SusD, another mutant was created, where SusD protein was inactivated as the binding

residues were replaced to alanine (SusD*). This also resulted in a lack of growth but it could be restored by the addition of a low concentrated maltose, an α -1,4-glucose disaccharide. In the case of a strain with complete loss of SusD expression, addition of maltose was not efficient enough to restore the growth on amylopectin (Figure 1.6) suggesting that functional SusD is crucial for optimal binding of available oligosaccharides. Further analysis of the *susC* transcripts expression consistently revealed that SusD binding is critical for the sensing of starch and activates the optimal transcriptional response. A mutant lacking SusD binding activity needs a much higher concentration of available oligosaccharides to activate the *sus* operon (Cameron *et al.*, 2014). Further analyses of particular SGBPs revealed that SusE can compensate for the lack of binding during SusD* growth on oligosaccharides. Transport of the starch oligosaccharides through the SusC requires the presence of SusD and SusE, but does not depend on their binding affinity. The lack of SusD binding residues possibly favor a firm complex of SusD with SusC, which impedes the oligosaccharide carrier. It has been speculated that SusCDE complex formation is more important than their glycan binding roles (Foley *et al.*, 2018). Binding sites of SusEFG are important for the growth on high-molecular weight starch as their deletion resulted in an extended lag phase of *B. thetaiotaomicron* on high-molecular weight amylopectin. Many gut *Bacteroidetes* possess a thick polysaccharide capsule, which may create a barrier difficult to penetrate, especially for extended polysaccharides. The SusEFG have been suggested to improve the binding of high molecular polysaccharides by overcoming this capsular layer of polysaccharides (Cameron *et al.*, 2014). Further studies on xyloglucan uptake also suggest the importance of the SGBPs complex assembly, where xyloglucan binding may promote outer membrane orientation of xyloglucan (Foley *et al.*, 2019). Compared to the knowledge on carbohydrate uptake in intestinal *Bacteroidetes*, only few studies have investigated the nature of this process in the aquatic environment.

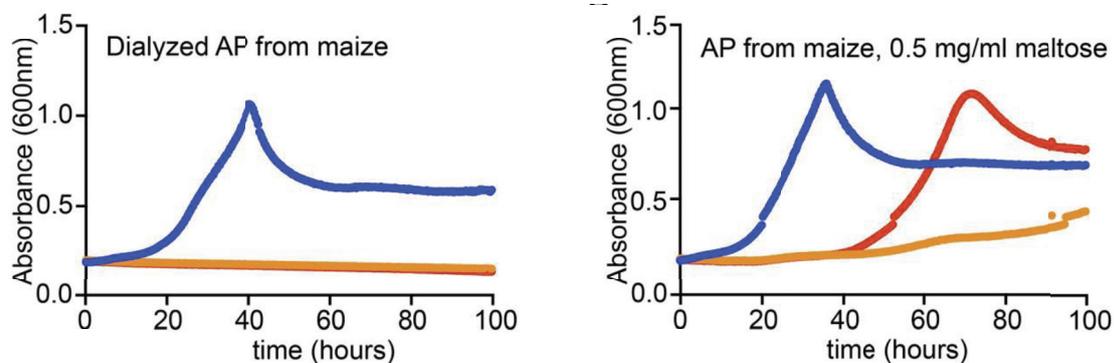


Figure 1.6 Influence of maltose, glucose disaccharide, on the SusD mutant strains. Growth curves of *B. thetaiotaomicron* showing that requirement for SusD binding during growth on starches is overcome by adding a low concentration of maltose. Cameron *et al.* used the following strains: wild

type (blue), strain with *susD* deletion (orange), strain with binding-deficient *SusD* (red). The figure was taken from (Cameron *et al.*, 2014).

Selfish uptake of marine polysaccharides

Currently, it is assumed that marine bacteria degrade carbohydrates in three different ways. Heterotrophs carrying extracellular enzymes represent a group of sharing bacteria (Figure 1.7, middle panel). These bacteria initiate polysaccharide degradation, which results in the release of small hydrolysis products available in the environment as so-called “public goods”. They represent mono- and oligosaccharides and serves as a carbon source for other bacterial communities, which do not produce external enzymes for polysaccharides digestion, so-called scavenging (Figure 1.7, right panel) (Arnosti, 2004). This two-player model has been extended with alternative uptake (Reintjes *et al.*, 2017) where marine microbes acquire substrate with the selfish mechanism, similar to gut bacteria (Figure 1.7, left panel)

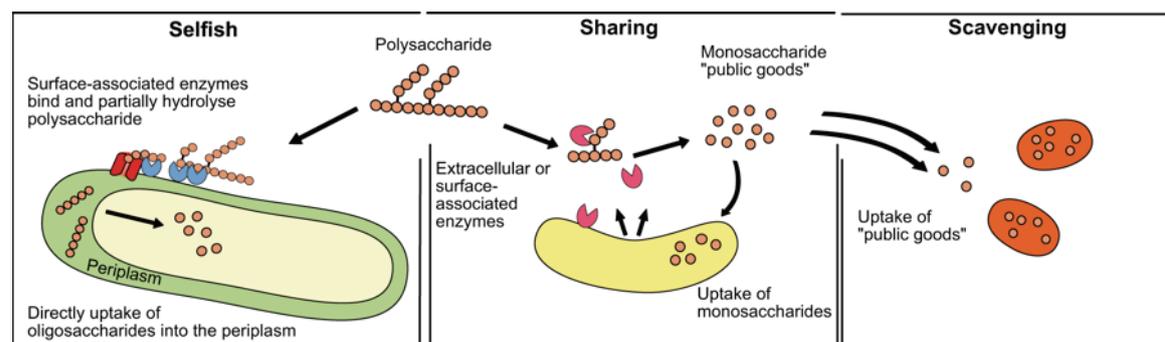


Figure 1.7 Three-player model of polysaccharide uptake in marine bacteria. Selfish behavior: bacteria use surface-located enzymes and binding proteins (left panel); sharing: cells employ extracellular and surface-attached enzymes, which leads to release of extracellular hydrolysis products: “public goods” (middle panel); scavenging: cells take up extracellular hydrolysis products produced by other organisms (right panel). The figure was taken from (Reintjes *et al.*, 2018).

Selfish substrate uptake relies on a highly efficient system for retaining the substrate at the cell surface. Reintjes *et al.* developed a technique, where fluorescently labeled polysaccharides (FLA-PS) (Arnosti, 2003) are monitored with super-resolution structured illumination microscopy (SR-SIM). In association with fluorescence in situ hybridization (FISH) it allows identification of species taking up particular polysaccharides. The ability to carry the substrate into the cell was observed for *Bacteroidetes* in environmental samples and further tested with pure cultures of a marine member of this phylum - *Gramella forsetii*. Examination of individual cells revealed uptake of FLA-PS (green) into the periplasmic space (Figure 1.8). They analyzed laminarin, xylan, and chondroitin sulfate uptake, where laminarin stain was acquired rapidly and at the highest abundance by the cells in sea water. These experiments supported by genomic analyses that revealed the potential for polysaccharide

utilization, suggested a selfish-like substrate uptake in the marine bacteria (Reintjes *et al.*, 2017). Cuskin *et al.* previously demonstrated the selfish mechanism for human gut *Bacteroidetes* with extended biochemical analysis and targeted gene disruption experiments (Cuskin *et al.*, 2015). This mechanism implies cell-surface located initial digestion of large oligosaccharides that are subsequently transferred to the periplasmic space. However, molecular information about proteins involved in the polysaccharides acquisition on the cell surface is needed to fully understand this process in marine *Bacteroidetes*. It is most likely that the evolution of the cell attached enzymes and SGBPs enabled *Bacteroidetes* to efficiently take up polysaccharides and reduce the diffusion of hydrolysis products. The work of this thesis provides the detailed analyses of the marine SGBPs: GMSusD, GMSusE, and GMSusF from planktonic bacteria *Gramella* sp. MAR_2010_102. In detail characterization of these proteins is important for understanding biological function, but also necessary to investigate their potential as glycan binding probes. Potential applications of CBPs for polysaccharide detection are described in the next subchapter.

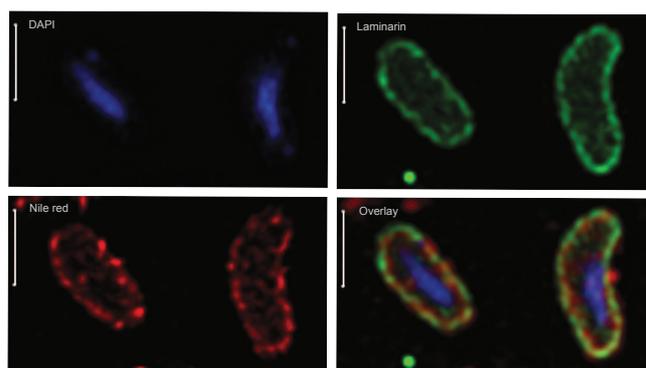


Figure 1.8 Selfish uptake of FLA-laminarin (green) by *Bacteroidetes*. Microscopy images showing the accumulation of laminarin in the periplasm of bacteria. Cell stained using Nile red (red) and DAPI (blue) to visualize the membranes and DNA, respectively. Scale bar = 0.5 μm . The images were taken from (Reintjes *et al.*, 2017).

Applications of carbohydrate binding proteins

The outstanding complexity of polysaccharides challenges their accurate analysis, but there are numerous valuable methods. Acid hydrolysis coupled with reducing sugar assay or chromatographic analyses provide information about sugar composition. However, the structure of polysaccharide is not only defined by the type and order of monosaccharides, but also by the configuration of linkages (see also figure 1.2). Multiple chromatographic methods such as thin-layer chromatography, size-exclusion chromatography (SEC) or high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) have been developed to analyze oligosaccharides (Xu *et al.*, 2017), but they require

characterized standards. More advanced information about linkages and anomeric configuration can be provided by mass spectrometry and nuclear magnetic resonance. Relatively high sample concentration and purity of samples is needed for these techniques. The current methods are improving over the years and enhance overall information about the carbohydrates. However, we are lacking a method for high-throughput profiling of complex polysaccharide mixtures.

As outlined above bacteria possess efficient systems to bind and utilize polysaccharides. The idea is to use these proteins as a tool for glycan analysis. Becker *et al.* established the laminarin quantification assay, where enzymes from marine microbes were produced as recombinant proteins and optimized for selective laminarin degradation. The method is more specific than classical acid hydrolysis, provides information about linkages and can be applied to environmental samples (Becker *et al.*, 2017; Becker & Hehemann, 2018). Another approach includes characterization of CBPs from marine bacteria and their application as probes for high-throughput microarrays.

Microarrays are a very precise, high throughput method for glycan analysis, mainly used for plant cell wall components. Hundreds of different polysaccharides can be screened rapidly. The polysaccharides are printed on a membrane or slide. They can be immobilized either in a covalent or non-covalent manner. As long molecules, polysaccharides are usually immobilized on the membrane by passive adsorption, whereas oligosaccharides require the coupling to a protein or lipid, which results in a larger glycoconjugate. Another option would be the attachment of oligosaccharides with covalent chemical linkers. The arrays are based on two main approaches: extracted glycan arrays and defined arrays. The first refers to an experiment, where we can monitor the composition of extracted polysaccharides by using well-established antibodies. In contrast, defined arrays include the immobilization of already known carbohydrates and are used for testing CBPs or enzymes activity (Fangel *et al.*, 2012). The developing procedure is analogous to the enzyme-linked immunosorbent assay (ELISA), but carbohydrate arrays are a high throughput method, which require less samples and reagents. According to the procedure, printed carbohydrates are detected by the molecular probes (monoclonal antibodies or CBPs) and linked through a secondary antibody. The latter are specific either based on the source of the probe, for example, anti-mouse antibodies or they target specific tags such as a hexahistidine tag (His-tag). Binding is detected based on the enzymatic reaction, where the specific substrate interacts with enzyme conjugated to the secondary antibody, mostly alkaline phosphatase. The signal is quantified and presented in the form of a heatmap (Figure 1.9). Probes for plant cell wall polysaccharides are well-established in contrast to probes for algal-related polysaccharides.

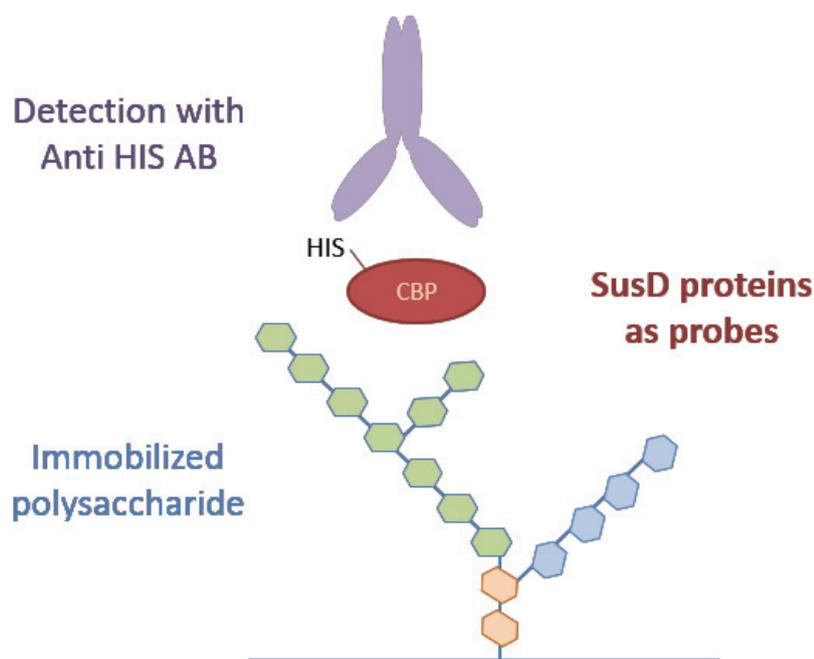


Figure 1.9 Basis of the polysaccharide recognition in the microarray technique. Polysaccharide spotted on a membrane is probed with CBPs containing His-tags, which is further detected with secondary antibodies.

The specificity displayed by CBPs enables them to assess their target among a variety of substrates, which prompted their development as probes detecting polysaccharides. Marine glycobiology is relatively unexplored thus exploitation of novel CBPs as probes is important to extend our knowledge in this field. There is a wide collection of monoclonal antibodies directed toward specific polysaccharides from the plant cell wall (www.plantprobes.net). Production of monoclonal antibodies against polysaccharides with limited immunogenicity requires sufficient purity (Rydahl *et al.*, 2017), which is a challenge for marine polysaccharides. However, the collection of monoclonal antibodies to brown algae cell wall polysaccharides is expanding (www.sb-roscoff.fr/en/seaprobes). Currently, monoclonal antibodies against fucoidan (Torode *et al.*, 2015) and alginate (Torode *et al.*, 2016) are available. Additionally, ulvan antibodies (Rydahl *et al.*, 2017) were developed. Genetic engineering provides a complementary tool for the probe's formulation. As mentioned earlier, diatom material is rapidly decomposed by microbes. During the bloom in Helgoland, marine *Bacteroidetes* upregulate CBPs such as SusD-like proteins. A single bacterium has a specific set of CBPs targeting particular polysaccharides in the overall organic material released by phytoplankton. This makes these proteins promising candidates to develop them as carbohydrate probes. The exact composition of the polysaccharides in the ocean is not yet

known. We cannot exclude the scenario, where marine polysaccharides contain different structures than plant cell wall-related polysaccharides. In this case, proteins

extracted from the environment have potential as valuable probes. Additionally, the production of probes in the form of recombinant proteins has multiple advantages such as lower costs, faster and easier procedure. Once probes are developed, they can be used not only for high-throughput screening but also for immunolabeling experiments showing the localization of target polysaccharides or they may be exploited for polysaccharide purification. There is also high interest in the use of CBPs in biotechnological applications (Shoseyov *et al.*, 2006). Exploring novel CBPs and their potential as probes helps to advance the field of glycobiology.

Research aims and thesis outline

The overall aim of this thesis was to explore novel CBPs from marine bacteria. Our research focuses on the surface located proteins of the Sus-like system from marine *Bacteroidetes*. SGBPs from gut *Bacteroidetes* create an outer membrane complex and bind polysaccharides on the bacterial surface. The knowledge about these proteins in bacteria from the ocean is limited, therefore I present the first biochemical characterization of SGBPs from marine bacteria. We believe studying these proteins contribute to a better understanding of the way how *Bacteroidetes* take up algal-derived polysaccharides. Furthermore, we screened multiple marine SusD-like proteins for their potential as novel probes for glycan analysis.

In the first manuscript, I investigated a SusD-like protein from the planktonic bacteria *Gramella* sp. MAR_2010_102. I studied the functions of this protein by performing binding assays and solving the three-dimensional structure. To support the bioinformatic predictions and better understand SusD-like proteins from the marine environment, we applied the protein structure obtained with X-ray crystallography to the alignment with the metagenomic sequences. The results were published in FEBS Journal.

In the second manuscript, I set out to analyze two novel proteins. Their genes encoded in the vicinity of the SusCD in the marine PULs were annotated as unknown or PKD domains. We tested the hypothesis that they represent members of a relatively unexplored group of SusEF-like carbohydrate-binding proteins. The manuscript is in the preparation to be submitted to FEBS letters Journal.

In the third manuscript, the major objective was to test the potential of marine SusD-like proteins as probes for polysaccharides analysis. We extracted CBPs from bacteria, which respond to the algal bloom in Helgoland, North Sea. The forty-seven CBPs were produced as recombinant proteins and tested for binding ability. Studying these proteins contributes to a better understanding of the protein-carbohydrate interaction and can potentially further develop the field of polysaccharide analysis.

In the appendix, a putative glycosyl hydrolase was studied in detail. The structural analysis revealed the absence of catalytic residues suggesting that this protein may bind a mannose-rich glycan. The paper was published in Protein Science Journal.

Contribution to manuscripts and further publications

Manuscript 1: Molecular recognition of the beta-glucans laminarin and pustulan by a SusD-like glycan binding protein of a marine *Bacteroidetes*

Agata Anna Mystkowska, Craig Robb, Silvia Vidal-Melgosa, Chiara Vanni, Antonio Fernandez-Guerra, Matthias Höhne, Jan-Hendrik Hehemann

Published in **The FEBS Journal**, 285(23), DOI:10.1111/febs.14674

JHH, MH and AM planned and designed the research. AM produced, purified, crystallized GMSusD and performed binding assays. AM and CR collected the X-ray diffraction data, determined and refined the structure of GMSusD. AM and SVM conducted the carbohydrate microarrays experiment. AM performed phylogenetic analysis. JHH, AM and CR analyzed the data. AM and JHH wrote the manuscript, all authors contributed to the editing of the manuscript.

Manuscript 2: Marine SusEF-like proteins from *Gramella* sp. MAR_2010_102 bind laminarin

Agata Anna Mystkowska, Andrew Hettle, Silvia Vidal-Melgosa, Alisdair Boraston, Matthias Höhne, Jan-Hendrik Hehemann

This manuscript is in preparation for a submission to The FEBS Letters Journal.

JHH, AB, MH, and AM planned and designed the research. AM produced and purified GMSusE and GMSusF proteins. AM and SVM conducted the carbohydrate microarrays experiment. AM and AH performed ITC analysis. AM and JHH wrote the manuscript, all authors contributed to the editing of the manuscript.

Introduction

Manuscript 3: Exploring the potential of bacterial carbohydrate-binding proteins as novel probes to detect marine polysaccharides

Agata Anna Mystkowska, Silvia Vidal-Melgosa, Jan-Hendrik Hehemann

JHH and AM planned and designed the research. AM produced and purified protein library. AM and SVM conducted the carbohydrate microarrays experiment. AM took part in the Helgoland sampling campaign 2016. AM wrote the manuscript, SVM and JH edited the manuscript.

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

Craig S. Robb, Agata Anna Mystkowska, Jan-Hendrik Hehemann

Published in **Protein Science Journal**, 26(12), DOI: 10.1002/pro.3291

AM analyzed the binding ability of GH99-like protein.

Manuscripts

Manuscript 1:

**Molecular recognition of the β -glucans laminarin and pustulan by
a SusD-like glycan-binding protein of a marine *Bacteroidetes***

Molecular recognition of the beta-glucans laminarin and pustulan by a SusD-like glycan-binding protein of a marine *Bacteroidetes*

Agata Anna Mystkowska^{1,2}, Craig Robb^{1,2}, Silvia Vidal-Melgosa^{1,2}, Chiara Vanni^{2,3}, Antonio Fernandez-Guerra^{2,3}, Matthias Höhne⁴ and Jan-Hendrik Hehemann^{1,2}

¹ Center for Marine Environmental Sciences (MARUM), University of Bremen, Germany

² Max Planck Institute for Marine Microbiology, Bremen, Germany

³ Jacobs University Bremen gGmbH, Bremen, Germany

⁴ Institute of Biochemistry, Greifswald University, Germany

Keywords

Bacteroidetes; carbohydrate-binding proteins; laminarin; microalgae; polysaccharides

Correspondence

J.-H. Hehemann, MARUM MPG Joint Research Group Marine Glycobiology, University of Bremen, Center for Marine Environmental Sciences (MARUM), Bremen 28359, Germany

Fax: +49 421 2028690

Tel: +49 421 218 65775

E-mails: jhhehemann@marum.de,

jhehemann@mpi-bremen.de

and

M. Höhne, Institute of Biochemistry, University of Greifswald, Felix-Hausdorff-Str. 4, 17487 Greifswald, Germany

Tel: +49 3834 420 4417

Fax: +49 (0)3834 420 4373

E-mail: Matthias.Hoehne@uni-greifswald.de

(Received 27 May 2018, revised 19 September 2018, accepted 5 October 2018)

doi:10.1111/febs.14674

Marine bacteria catabolize carbohydrate polymers of algae, which synthesize these structurally diverse molecules in ocean surface waters. Although algal glycans are an abundant carbon and energy source in the ocean, the molecular details that enable specific recognition between algal glycans and bacterial degraders remain largely unknown. Here we characterized a surface protein, GMSusD from the planktonic *Bacteroidetes*-*Gramella* sp. MAR_2010_102 that thrives during algal blooms. Our biochemical and structural analyses show that GMSusD binds glucose polysaccharides such as branched laminarin and linear pustulan. The 1.8 Å crystal structure of GMSusD indicates that three tryptophan residues form the putative glycan-binding site. Mutagenesis studies confirmed that these residues are crucial for laminarin recognition. We queried metagenomes of global surface water datasets for the occurrence of SusD-like proteins and found sequences with the three structurally conserved residues in different locations in the ocean. The molecular selectivity of GMSusD underscores that specific interactions are required for laminarin recognition. In conclusion, our findings provide insight into the molecular details of β -glucan binding by GMSusD and our bioinformatic analysis reveals that this molecular interaction may contribute to glucan cycling in the surface ocean.

Abbreviations

CAZymes, carbohydrate-active enzymes; CBMs, carbohydrate-binding modules; DLS, dynamic light scattering; GH, glycoside hydrolases; GOS, global ocean sampling; HPAEC-PAD, high-performance anion exchange chromatography with pulsed amperometric detection; HPLC, high-performance liquid chromatography; IMAC, immobilized metal ion affinity chromatography; IO, Indian Ocean; MMseqs2, Many-against-Many sequence searching 2; MPBS, milk phosphate-buffered saline; MS, Mediterranean Sea; NAO, North Atlantic Ocean; NPO, North Pacific Ocean; ORFs, open reading frames; OSD, Ocean sampling day; PDB, Protein Data Bank; PUL, polysaccharide utilization loci; SAO, South Pacific Ocean; SEC, size exclusion chromatography; SGBPs, surface glycan-binding proteins; SO, Southern Ocean; SPO, South Pacific Ocean; Sus, starch utilization system; TARA, TARA Ocean Expedition.

Abstract:

Marine bacteria catabolize carbohydrate polymers of algae, which synthesize these structurally diverse molecules in ocean surface waters. Although algal glycans are an abundant carbon and energy source in the ocean, the molecular details that enable specific recognition between algal glycans and bacterial degraders remain largely unknown. Here we characterized a surface protein, GMSusD from the planktonic *Bacteroidetes-Gramella* sp. MAR_2010_102 that thrives during algal blooms. Our biochemical and structural analyses show that GMSusD binds glucose polysaccharides such as branched laminarin and linear pustulan. The 1.8 Å crystal structure of GMSusD indicates that three tryptophan residues form the putative glycan binding site. Mutagenesis studies confirmed that these residues are crucial for laminarin recognition. We queried metagenomes of global surface water datasets for the occurrence of SusD-like proteins and found sequences with the three structurally conserved residues in different locations in the ocean. The molecular selectivity of GMSusD underscores that specific interactions are required for laminarin recognition. In conclusion, our findings provide insight into the molecular details of β -glucan binding by GMSusD and our bioinformatic analysis reveals that this molecular interaction may contribute to glucan cycling in the surface ocean.

Introduction:

Approximately one half of the global primary production of organic carbon occurs in the surface ocean, where carbon dioxide is converted by photosynthetic algae into biomass (Field *et al.*, 1998). More than 50% of the phytoplankton primary production is present in the form of polysaccharides (Baines & Pace, 1991). One of the most abundant algal polysaccharides is laminarin, a β -1,3-glucan with β -1,6-linked glucose branches (Painter, 1983). Polysaccharides serve as main carbon source for many marine bacteria (Alderkamp *et al.*, 2007; Kabisch *et al.*, 2014) such as *Bacteroidetes*, which consume a wide variety of polysaccharides using enzymes, associated binding proteins and transporters that are encoded in operons known as a Polysaccharide utilization locus (PULs) (Bjursell *et al.*, 2006). Bacteroidetal PULs, which are specific for different polysaccharides, have in common genes encoding homologs of two protein families - SusC and SusD. These SusCD pairs are surrounded by variable genes including glycoside hydrolases (GH), surface glycan-binding proteins (SGBPs) and accessory functions frequently belonging to different protein families. The high diversity of glycans is mirrored by the genetic and functional heterogeneity of PULs containing different families of Carbohydrate-Active Enzymes (CAZymes), which depolymerize polysaccharides (Lombard *et al.*, 2014). The proteins encoded in PULs assemble into a multi-protein system, first described for the starch utilization system (Sus) (Anderson & Salyers, 1989), which is located in the inner and outer membrane passing through the periplasm of Gram-negative bacteria.

The Sus-like system is widely spread across distinct ecological niches. In terrestrial and gut bacteria this system enables specific interactions between microbes and their glycan substrates, suggesting that similar synergy takes place in the sea. Compared to the advanced knowledge on carbohydrate uptake in intestinal *Bacteroidetes*, only few studies have investigated the nature of this process in the aquatic environment. The analysis of the isolate *Gramella forsetii* KT0803 revealed a specific protein response of the strain to the presence of algal polysaccharides, particular PULs were induced during the growth on laminarin and alginate (Kabisch *et al.*, 2014). Several studies have provided important biochemical information on the enzymes and carbohydrate-binding modules (CBMs) from the aquatic bacteria involved in the degradation of polysaccharides (Jam *et al.*, 2005; Hehemann *et al.*, 2010; Hehemann *et al.*, 2014; Labourel *et al.*, 2015; Becker *et al.*, 2017). The first PUL for the degradation of an algal glycan has been recently extensively described in a marine bacterium. This PUL encodes the carrageenan degradation pathway of the marine bacterium *Zobelia galactanivorans* (Ficko-Blean *et al.*, 2017). Large scale screening of marine carbohydrate binding proteins was exploited in so called Double Blind Comprehensive Microarray Polymer Profiling method (Salmeán *et al.*, 2018). However, direct studies of

SGBPs such as the SusD-like proteins from marine *Bacteroidetes*, with in-depth biochemical insight remain rare.

It has been shown that the genes of SusD-like proteins are among the most highly expressed bacterial proteins during algal blooms in the North Sea (Mann *et al.*, 2013). The growth response and CAZyme expression of *Bacteroidetes* during algal blooms, when high amounts of polysaccharides are released by algae, indicate the important role of *Bacteroidetes* as well as their SusD-like proteins and PULs in polysaccharide degradation (Teeling *et al.*, 2012).

The deficit of empirical studies leaves a lack of knowledge about the specificity of marine SGBPs towards their substrate. In gut *Bacteroidetes* the role of SGBPs such as protein homologs of SusD, SusE or SusF is to bind and concentrate the glycan on the cell surface, where outer membrane enzymes initially degrade them (Martens, Koropatkin, *et al.*, 2009). The SusD-like proteins are known to be outer membrane carbohydrate binding proteins. The structure of SusCD complex from *Bacteroides thetaiotaomicron* – a model gut bacterium, shows additional involvement of SusD-like proteins in the transport of polysaccharide into the cell. Based on the proposed pedal bin mechanism, SusD-like protein interacts with SusC-like TonB-dependent transporter and might guide the polysaccharide to the periplasm helping to move the polysaccharide through the membrane (Glenwright *et al.*, 2017). Biophysical conditions of the ocean as compared to the intestines or land diverge in many ways. For instance, nutrient concentrations, oxygen level or salinity, as well as flow variations and glycan substrates differ between both systems. Consequently, the recognition of polysaccharides by SusD-like proteins from marine microbes could be different. This study provides new insights into this relevant interaction for carbon cycling by and energy acquisition in marine bacteria.

We characterized a SusD-like protein named GMSusD from a predicted laminarin PUL of *Gramella* sp. MAR_2010_102. The protein was expressed during an algal bloom in the North Sea near Helgoland and may play an important role in this environment (Teeling *et al.*, 2012). This structural and biochemical analysis of a SusD-like protein from a marine *Bacteroidetes* shows that GMSusD binds selectively to branched laminarin and pustulan. Our investigation provides a framework to elucidate and annotate the function of globally distributed SusD-like proteins and other marine glycan-binding proteins that are of biogeochemical relevance.

Results:**SusD-like genes are present in surface waters across the oceans**

Previous studies found locally enriched expression of SusD-like (and SusC) proteins in response to algal blooms suggesting that they are involved in the turnover of algal glycans (Teeling *et al.*, 2012). Expanding on these observations, we queried metagenomes of global surface water datasets for the occurrence of SusD-like proteins, which might be involved in glycan cycling. A comprehensive Open Reading Frames (ORFs) dataset (Table 2.1) was analyzed by collecting the ORFs from 3 of the major metagenomic environmental surveys of the ocean microbiome: Global Ocean Sampling survey (Williamson *et al.*, 2008), TARA Ocean expedition (Karsenti *et al.*, 2011) and the Ocean Sampling Day (Kopf *et al.*, 2015), which indicate presence of the SusD-like genes in marine habitats. We were able to annotate to SusD-like domains 25,394 ORFs, from 422 samples. From them, we selected 14,306 ORFs coming from 301 surface water samples. The distribution of SusD-like genes in the oceans is extensive, as shown in figure 2.1.1 which displays the proportion of the ORFs annotated as SusD-like domains in terms of percentage of the total amount of ORFs in each sample. The average proportion of SusD-like domains is 42% of that of RecA, which is present as a single gene in most bacterial genomes (Wu *et al.*, 2011). The results of our analyses indicate SusD-like genes are globally present in the ocean.

Table 2.1.1 Open Reading Frames (ORFs) originated from 472 samples collected by three major mega-sequencing projects of the ocean microbiome

Project	Samples	ORFs
TARA	242	111,903,261 (70.1%)
OSD	150	7,015,383 (4.4%)
GOS	80	20,068,580 (12.6%)

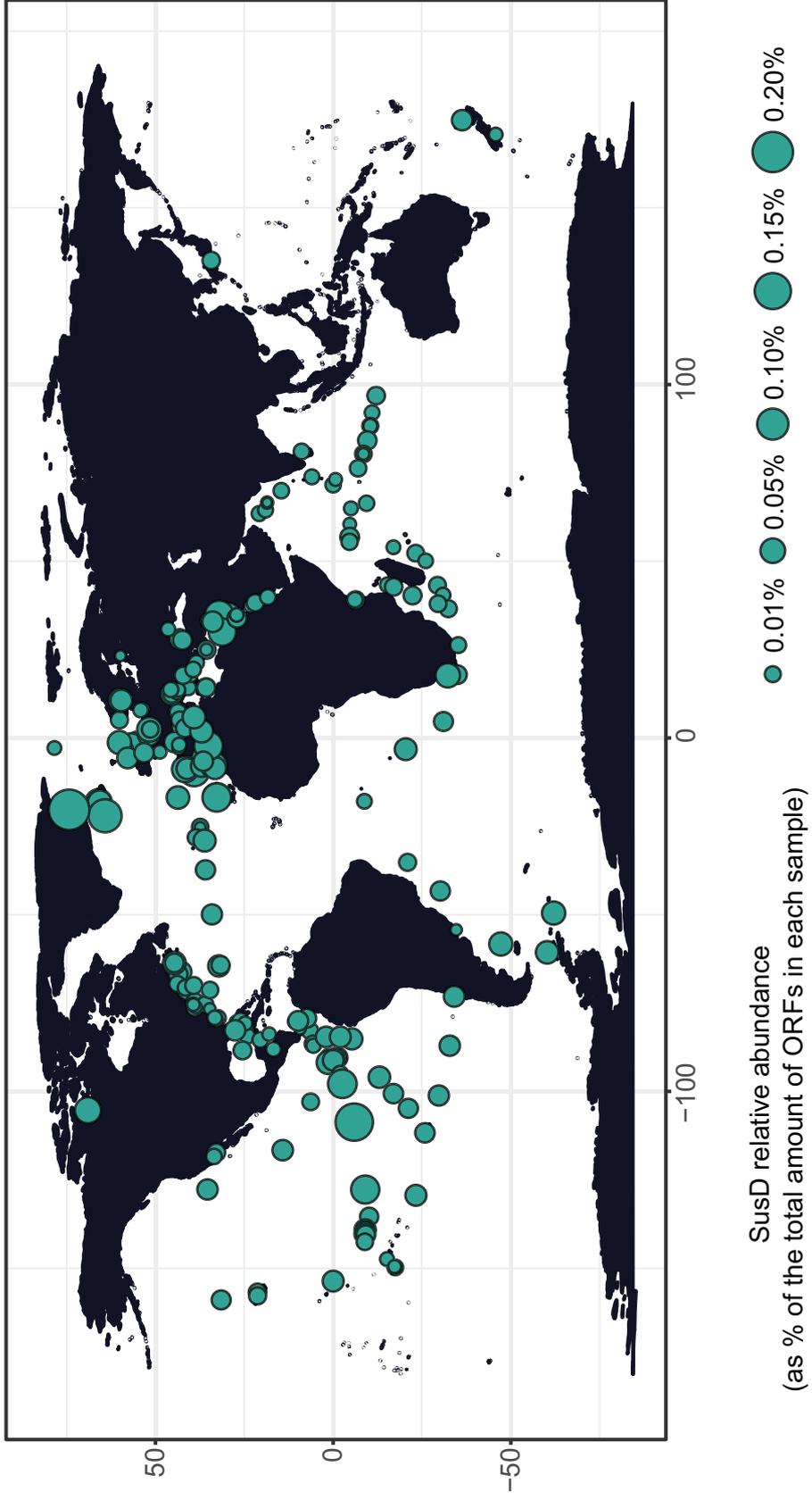


Figure 2.1.1 Occurrence of SusD-like genes in surface water metagenomes. Data are based on the metagenomic environmental surveys: the Ocean Expedition (TARA) (Williamson *et al.*, 2008), the Global Ocean Sampling (GOS) (Karsenti *et al.*, 2011) and the Ocean Sampling Day (OSD) (Kopf *et al.*, 2015).

GMSusD is a laminarin binding protein

SusD-like proteins from the gut and terrestrial microbes have been found to bind different plant polysaccharides (Table 2.1.2). Moreover, the specificity of a SusD-like protein usually coevolves with the specificity of the CAZymes in the same PUL (Grondin *et al.*, 2017). Recently a series of aquatic microbial PULs with putative specificity for algal glycans have been described. We focused our efforts on a PUL with putative laminarin or related β -glucan degrading function from the marine microbe *Gramella* sp. MAR_2010_102 (Figure 2.1.2). The CAZymes: GH3, GH5, GH30 and two GH16 present in this PUL suggest GMSusD may be binding to β -glucan polysaccharides based on specificity of the described enzymes from the homolog PUL belonging to GH families with laminarinases (Labourel *et al.*, 2014; Becker *et al.*, 2017; Nelson *et al.*, 2017). In order to test this hypothesis, we cloned and produced the protein in recombinant form in *E. coli* and purified it to high homogeneity and purity (Figure 2.1.3A). The molecular weight of GMSusD predicted based on the amino acid sequence is 50 kDa. Dynamic Light Scattering (DLS) measurements established a molecular weight of 50 kDa \pm 1 (mean \pm standard deviation) as well as a hydrodynamic radius of 3.2 nm \pm 0.6, percent polydispersity 9.5 \pm 1.1. Separation of GMSusD on a SEC analytical column indicated the size of protein to be 46 kDa in solution when compared to the proteins that were used as standards (Figure 2.1.3A, B). The migration of GMSusD on the SDS-PAGE gel showed the molecular weight of approximately 50 kDa (Figure 2.1.3C). No aggregation was observed in the collected fractions after IMAC and SEC purification, together all data indicate the protein is monomeric in solution.

Table 2.1.2 Specificity of SusD-like proteins from gut and terrestrial *Bacteroidetes*

Bacteria	Ecosystem	Substrate	Reference
<i>B. thetaiotaomicron</i>	Human gut	Starch	(Koropatkin <i>et al.</i> , 2008)
<i>B. ovatus</i>	Human gut	Xyloglucan	(Tauzin <i>et al.</i> , 2016)
<i>B. thetaiotaomicron</i>	Human gut	Levan	(Glenwright <i>et al.</i> , 2017)
<i>F. johnsoniae</i>	Soil	Chitin	(Larsbrink <i>et al.</i> , 2016)
<i>B. fragilis</i>	Human gut	Sialic acid	(Phansopa <i>et al.</i> , 2014)

To elucidate the glycan specificity of GMSusD we performed carbohydrate-binding assays with affinity gel electrophoresis. In brief, the folded protein is migrating by electrophoresis in a native polyacrylamide gel that is supplemented with a polysaccharide, in this case, laminarin or other glucans. Delayed movement of the protein, compared to a not interacting protein control, indicates GMSusD bind specifically to the tested polysaccharides. Specificity of GMSusD was tested with commercially available substrates containing different β -linked glucans. This experiment revealed varied binding ability of GMSusD to the laminarins from different species (Figure 2.1.4A). GMSusD bound laminarin from *Eisenia bicyclis*, whereas almost no shift was observed on the gel with laminarin from *Laminaria. digitata* (Figure 2.1.4A). We further verified the binding of GMSusD towards the *E. bicyclis* laminarin by digesting it with a laminarinase mixture (Becker *et al.*, 2017). The binding of GMSusD on the native gel with digested polysaccharide was completely abolished, confirming the specificity of GMSusD to laminarin from *E. bicyclis*.

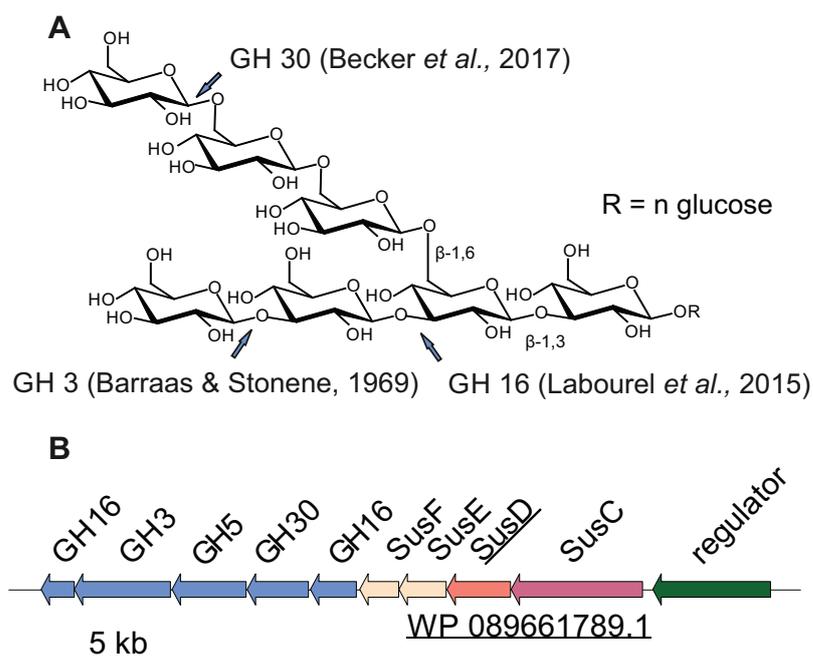


Figure 2.1.2 Enzyme composition indicates a branched laminarin specific polysaccharide utilization locus in marine *Gramella MAR*. **A:** Laminarin is a glucan consisting of a β -1,3-D-glucose polysaccharide with β -1,6-linked side chains. **B:** The CAZy families – GH16, GH30, GH3 and GH5 are involved in β -glucan degradation. The GH3 has been described as β -1,3-glucan exohydrolases (Barraas & Stone, 1969) whereas GH16 acts as endo β -1,3- laminarinase (Labourel *et al.*, 2015). Interestingly, there is GH30 encoded in the PUL, which indicates specificity for branched laminarin, since GH30 is known as exo β -1,6- laminarinase (Becker *et al.*, 2017). The SusD-like protein GMSusD, which is investigated in this study, is underlined (WP 089661789.1). SusC-like proteins=TBDR; SusD=SusD-family protein; GH=glycoside hydrolase.

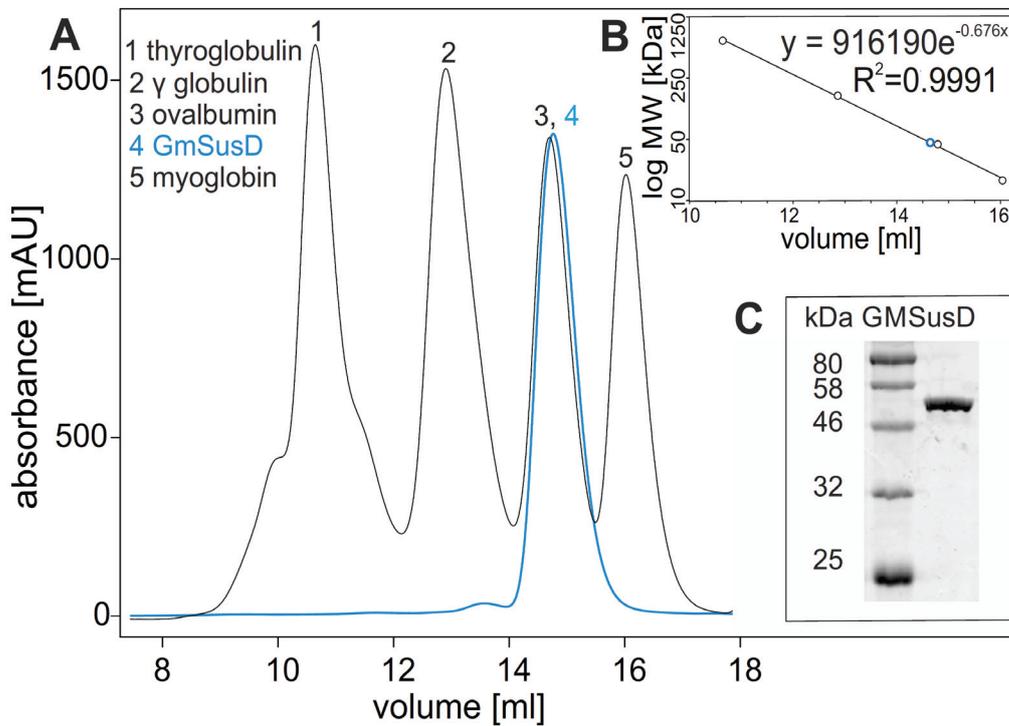


Figure 2.1.3 The recombinant GMSusD from *Gramella MAR* is a monomeric protein. **A:** Analytical SEC profile of purified GMSusD and size exclusion standards: thyroglobulin (670 kDa), γ -globulin (158 kDa), ovalbumin (44 kDa), myoglobin (17 kDa) (1511901, Bio Rad). **B:** The molecular weight of GMSusD was calculated with the calibration curve as 46 kDa. **C:** SDS-PAGE analysis of purified protein, approximately 10 μ g of protein was loaded.

GMSusD requires β -1,6-configured glucose for binding

Laminarin is a β -1,3-glucan with β -1,6-linked glucose side chains, where the degree of branching differs between species (Read *et al.*, 1996; Menshova *et al.*, 2014). Variability of the laminarin structures suggests that binding of proteins towards these polysaccharides could depend on the amount of side chains. *E. bicyclis* laminarin contains more β -1,6 linked side chains compared to other β -glucans (Pang *et al.*, 2005), such as the one from *L. digitata*. In the binding assay, we showed that GMSusD binds stronger to laminarin from *E. bicyclis* in comparison to laminarin from *L. digitata* (Figure 2.1.4A). We experimentally confirmed the higher amount of β -1,6 configured glucose side chains present on *E. bicyclis* laminarin. The same mass of both laminarin types diluted in water were digested independently by a specific exo- β -1,6-glucanase from *Formosa* sp. nov strain Hel1_33_131 (Becker *et al.*, 2017) and the amount of released glucose was compared. The glucose peak measured by HPAEC-PAD after 60 min of the enzymatic

reactions was around five times higher for *E. bicyclis* laminarin than for *L. digitata* laminarin (Figure 2.1.4B, C). The results clearly indicate higher amounts of β -1,6-linked glucose in the *E. bicyclis* laminarin, which is in agreement with previous studies (Menshova *et al.*, 2014). Next, we tested binding to the glucose polymer pustulan and found interaction with this polysaccharide on the affinity gel (Figure 2.1.4A). As pustulan is a linear β -1,6-glucan, this result further suggested that GMSusD requires β -1,6-linkages for binding. To validate this proposed substrate specificity, we debranched *E. bicyclis* laminarin with a GH30 enzyme. After enzymatic treatment, where we assume an absence of β -1,6-side chains, we observed that almost no GMSusD retention occurred on the gel (Figure 2.1.4A), suggesting that GMSusD does indeed require the β -1,6-linkages of the glucose side chains for binding.

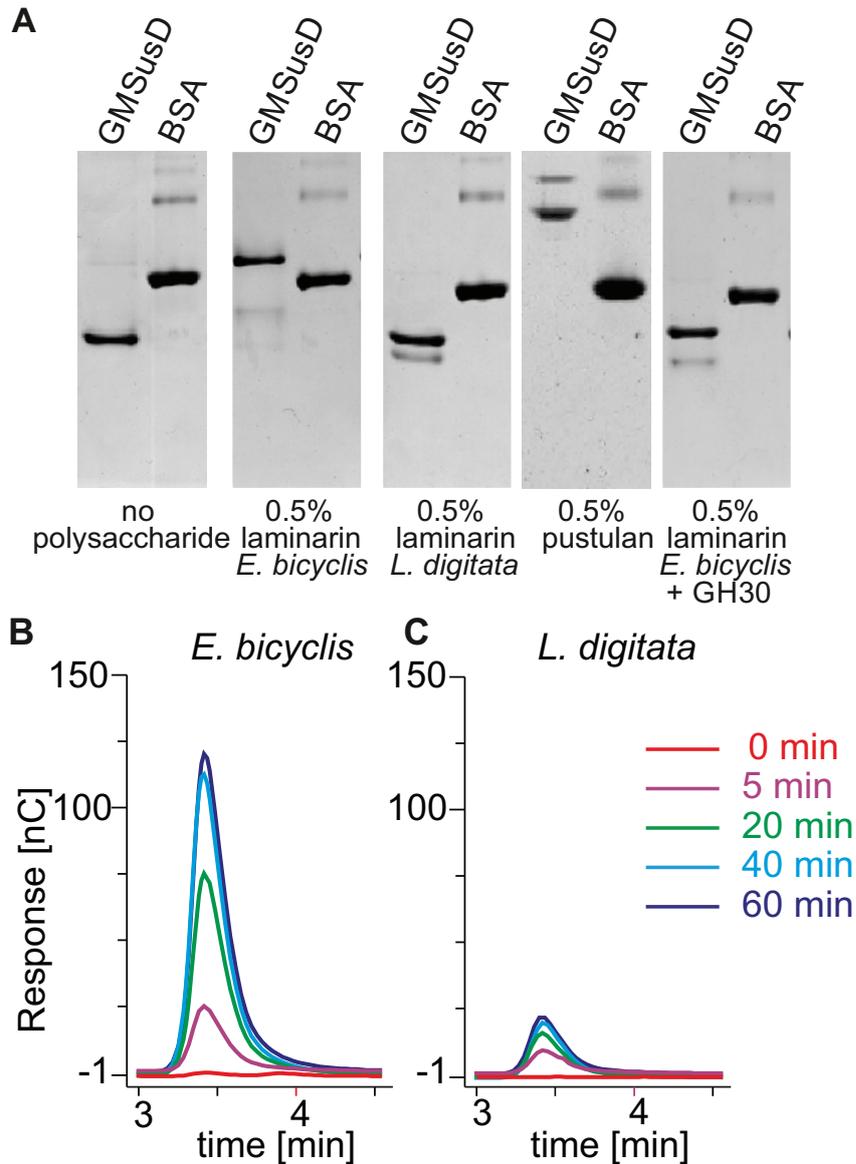


Figure 2.1.4 GMSusD binds to branched laminarin and pustulan polysaccharides. **A:** Affinity gel electrophoresis of the native GMSusD on a gel containing no polysaccharide (negative control) and on gels with 0.5% (w/v) polysaccharides: laminarin from *E. bicyclis*, laminarin from *L. digitata* and pustulan. Moreover, binding of GMSusD towards 0.5% of laminarin from *E. bicyclis* digested with GH30 exo- β -1,6-glucanase was analyzed. About 10 μ g of each protein was loaded on the gel: GMSusD (SusD, line 1) and control - BSA (Bovine Serum Albumin) (BSA, line 2). **B and C:** Different amounts of branching on *E. bicyclis* laminarin compared to *L. digitata* was analyzed by using the GH30 exo- β -1,6-glucanase from the marine bacterium *Formosa* sp. nov strain Hel1_33_131 to hydrolyze β -1,6-linked glucose in the polysaccharide. The released glucose concentration is proportional to the amount of β -1,6 glucose present in the laminarin side chains. HPAEC-PAD measurements after digestion of two different laminarins (*E. bicyclis* and *L. digitata*) with the GH30 showing the release of glucose.

GMSusD binds to pustulan on carbohydrate microarrays

The binding profile of GMSusD was further evaluated by using carbohydrate microarrays (Moller *et al.*, 2008). A total of twelve polysaccharide structures were tested. Three monoclonal antibodies with previously established specificities were included as positive controls, which showed the expected binding to their cognate epitopes: BS-4002 bound to β -1,3-glucan (Meikle *et al.*, 1991), BS-4003 to β -1,3;1,4-glucan (Meikle *et al.*, 1994) and LM11 to arabinoxylan (McCartney *et al.*, 2005) (Figure 2.1.5). A recombinant SusD-like protein (SusD1) from *Salegentibacter* specific to α -mannan was included as negative control where no binding signal was detected. Binding of GMSusD to pustulan was detected (Figure 2.1.5). These results are in agreement with the affinity gel assay. All together these independent techniques confirm the ability of GMSusD to bind β -1,6-linked glucose chains.

	GMSusD	BS-4002	BS-4003	LM11	SusD1
Glucomannan (β mannan)	0	0	0	0	0
Arabinoxylan (β xylan)	0	0	0	100	0
Pullulan (α -1,4;1,6 glucan)	0	0	0	0	0
Carboxymethyl Cellulose (β -1,4 glucan)	0	0	0	0	0
Laminarin <i>Eisenia bicyclis</i>	0	0	0	0	0
Laminarin <i>Laminaria digitata</i>	0	0	0	0	0
Pachyman (β -1,3 glucan)	0	64	0	0	0
Scleroglucan (β -1,3;1,6 glucan)	0	0	0	0	0
β -glucan (β -1,3;1,4 glucan)	0	0	62	0	0
Lichenan (β -1,3;1,4 glucan)	0	60	56	0	0
Pustulan (β -1,6 glucan)	32	71	0	0	0
Xyloglucan (β -1,4 glucan with 1,6 xylose)	0	0	0	0	0

Figure 2.1.5 Heat map showing the analysis of the binding of GMSusD to different polysaccharides using carbohydrate microarrays. Different polysaccharides (listed at the left) were printed on microarrays and single arrays were probed with different protein probes: GMSusD protein and the controls: BS-4002, BS-4003, LM11 and SusD1 (listed at the top). Values on the heat map correspond to the mean binding signal intensities derived from the triplicates, where the highest value was set to 100 and all other values were normalized accordingly. Grey color represents no binding detected. SusD1: recombinant SusD-like protein from *Salegentibacter* specific to α -mannan. Note: Glucomannan has a β -mannan structure.

3D structure analysis of GMSusD and identification of the binding site

To explore the structural details including the key residues that mediate the interaction between protein and glycan we solved the 3D crystal structure of GMSusD (Table 2.1.3) at a resolution of 1.8 Å via molecular replacement using a homologous structure PDB code 3QNK (23% identity). Monomeric protein with residues 41 to 501 visible as continuous electron density, revealed a SusD-like structure comprising alpha helices with four tetratricopeptide repeat (TPR) units on the one side of the protein, which are proposed to serve as a structural scaffold (Koropatkin *et al.*, 2008; Glenwright *et al.*, 2017). The TPR1 unit is composed of a pair of helices: $\alpha 1$ (residues 4–21) and $\alpha 4$ (residues 67–89), TPR2: $\alpha 5$ (residues 94–119) and $\alpha 6$ (residues 142–156), TPR3: $\alpha 8$ (residues 174–188) and $\alpha 9$ (residues 191–202), TPR4: $\alpha 15$ (residues 361–372) and $\alpha 16$ (residues 377–390). The structure contains twenty α -helices and four β -sheets (Figure 2.1.6A). In order to predict binding residues a structural alignment was performed. GMSusD overlay with archetypal *B. thetaiotaomicron* SusD in complex with β -cyclodextrin, PDB code 3CK8 (Koropatkin *et al.*, 2008) with RMSD of 3.7 Å (Figure 2.1.6F, B). The location of the TPR based scaffold on the opposite site of the binding surface is a conserved feature of the SusD-like proteins. Therefore, a relative position of the TPR face together with the knowledge from available structures in complex with their native substrates enable to predict that the aromatic residues W287, W290 and W323 (Figure 2.1.6C, D) are likely involved in glycan recognition.

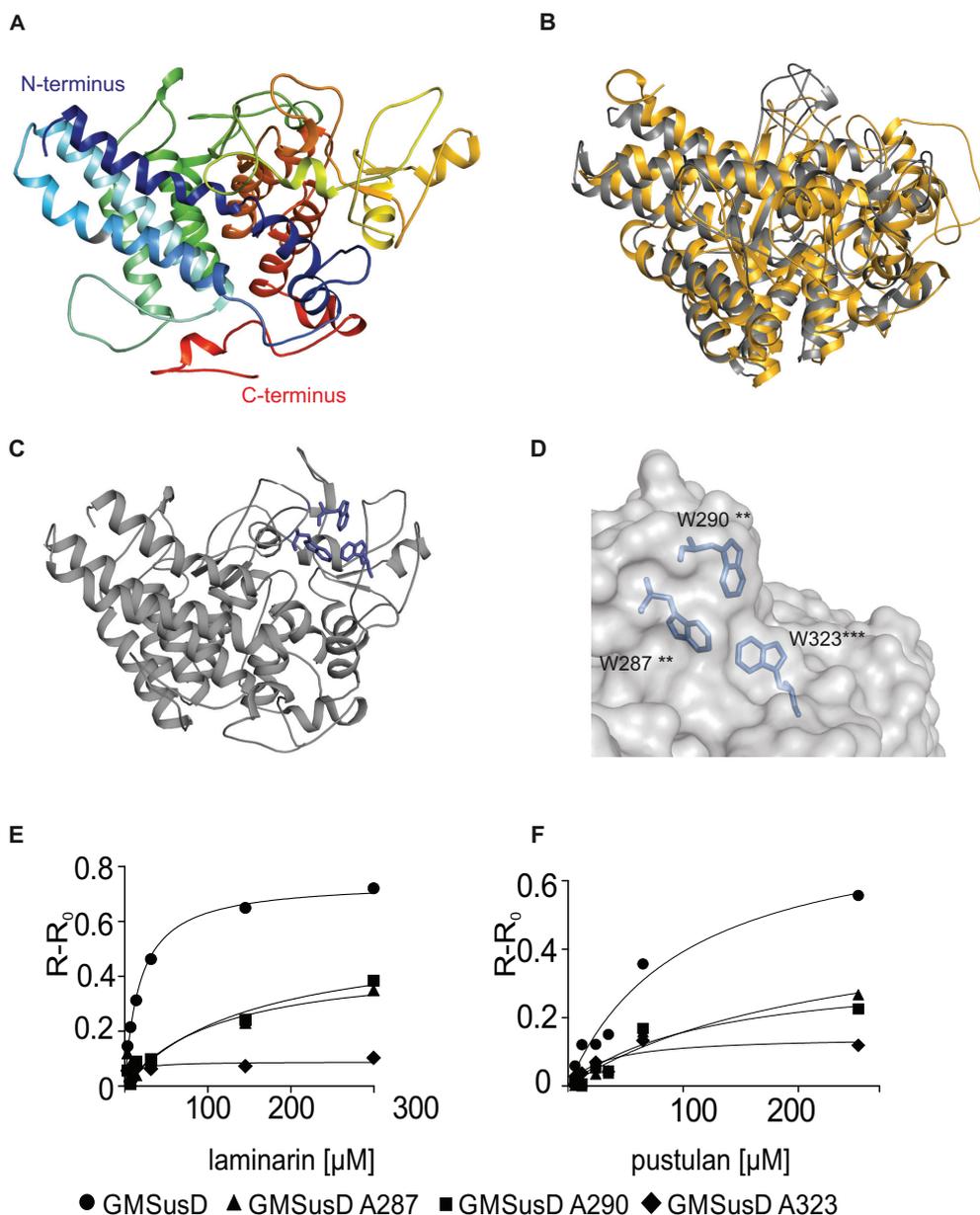


Figure 2.1.6 The three-dimensional structure of GMSusD. **A:** Overall view of GMSusD protein structure represented as a cartoon color ramped from blue (N-terminus) to red (C-terminus). **B:** Comparison of GMSusD structure in grey, with the 3D structure of the original starch SusD (PDB code 3CK8) binding protein. **C:** The binding site of GMSusD contains three tryptophan residues in dark blue, which are located on the surface of the protein. **D:** Zoom of the surface display of the glycan binding site and the positions of tryptophan residues of GMSusD (dark blue). Site-directed mutagenesis experiment confirmed the importance of each of the three tryptophan residues, three stars indicates the highest impact for the polysaccharide binding. **E and F:** The graphs displaying relative mobility ($R-R_0$) of GMSusD and mutants as a function of laminarin (E) and pustulan (F) concentration calculated using Prism 7 with a one-site specific binding equation. The symbols are as follows: GMSusD WT closed circle (●), GMSusD 287A closed triangle (▲), GMSusD 290 closed squares (■), and GMSusD 323A diamond (◆).

Table 2.1.3 Data collection and refinement statistics for the crystal structure of GMSusD.

PDB ID no.	6GCZ
Beamline	EMBL P13
Wavelength (Å)	1.0332
Space group	P 2 ₁ 2 ₁
Unit cell (Å)	a=67.39 b=69.42 c=115.59
Resolution (Å)	67.38-1.80 (1.90-1.80)
No. of reflections	652323 (95960)
No. Of unique reflections	50993 (7347)
Completeness (%)	99.9 (99.9)
Redundancy	12.8 (13.1)
Mean I/σ(I)	11.1 (2.5)
R _{sym} (%)	0.152 (1.22)
Refinement statistics	
R _{work}	0.1717
R _{free}	0.2072
No. of non-hydrogen atoms	
All	4041
Macromolecules	3545
Glycerol	2
Water	484
Avg B-factors	
All	28.60
Macromolecules	27.8
Glycerol	26.6
Water	38.0
RMSD	
Bond length (Å)	0.0189
Bond angle (°)	1.676
Ramachandran statistics	
Favored (%)	98.47
Outliers (%)	0.00

Carbohydrates are known to interact with proteins via their aromatic rings thus these aromatic protein residues likely create a necessary contact surface for binding. In protein glycan interactions the hydrophobic flat face of aromatic side chains aligns to the pyranose sugar ring as seen in carbohydrate binding proteins (Boraston *et al.*, 2004) and enzymes. Additionally, the residues F65, V379 and F325 might also be involved in the interaction by providing hydrogen-bonding interactions. Likely due to tightly packed crystal lattice GMSusD *holo* structure was not solved, since soaked oligosaccharides could not enter the water channels of the crystal packing, therefore the binding capabilities of aromatic residues in the predicted binding site were tested via mutagenesis instead. Residues W287, W290 and W323 were each substituted by alanine and the three variant proteins each carrying a single tryptophan to alanine substitution were produced in *E. coli*. The AGE experiment with mutated GMSusD (287A, 290A, 323A) revealed that the relative mobility of each of three mutants was lower in comparison to native GMSusD, indicating the importance of these residues for binding (Figure 2.1.6E, F, see also: Table 2.1.4)

Table 2.1.4 Relative motilities of proteins with alanine point mutations (three different proteins with alanine instead of W287, W290 and W323 respectively) were compared with GMSusD based on the affinity gel electrophoresis.

WT	R-R ₀ ± SD	R-R ₀ ± SD
	laminarin	Pustulan
GMSusD WT	0.70 ± 0.01	0.56 ± 0.01
GMSusD 287A	0.23 ± 0.01	0.27 ± 0.01
GMSusD 290A	0.24 ± 0.02	0.23 ± 0.01
GMSusD 323A	0.07 ± 0.01	0.12 ± 0.03

Structural insights enable the annotation of homolog proteins in environmental datasets

The obtained biochemical and structural information about GMSusD from *Gramella* sp. MAR_2010_102 might be used for a better understanding of β -glucan binding by SusD-like proteins from environmental datasets. From the set of 14,306 ORFs (Table 2.1.1), found in surface samples and annotated as SusD-like proteins, we retrieved 405 ORFs matching specifically to GMSusD, with coverage of 0.999. The deduplication step led then to 220 non-redundant ORFs. The sequence alignment guided by GMSusD structure (PDB code 6GCZ) showed the presence of twenty-five ORFs. Adding the duplicates of these ORFs we retrieved twenty-nine ORFs with the conserved tryptophan residues at the binding site (TARA project). The resulting ORFs are all complete (the gene is fully contained within the contig) and have average entropy of 0.02 per ORF, indicating a low base variability and consequently validating the hypothesis that the conserved residues in these ORFs code for tryptophan and are not metagenomic artifacts. The distribution of this set of ORFs cover almost all ocean regions (with samples from the North Atlantic Ocean (NAO), the South Pacific Ocean (SAO), the North Pacific Ocean (NPO), the South Pacific Ocean (SPO), the Southern Ocean (SO), the Indian Ocean (IO) and from the Mediterranean Sea (MS); for a total of twenty-nine different samples. TARA sequences with conserved tryptophans at the binding site belong with GMSusD to a monophyletic group on the phylogenetic tree (Figure 2.1.7A). Other SusD-like protein sequences from predicted or previously analyzed laminarin PULs do not belong to the same branch as GMSusD and do not possess all three tryptophans at the conserved positions. For simplicity, a reduced structure guided alignment of eight out of twenty-nine TARA sequences and three other SusD-like proteins was shown (Figure 2.1.7B). This result indicates a different mode of laminarin binding, perhaps they do not require β -1,6-linked glucose.

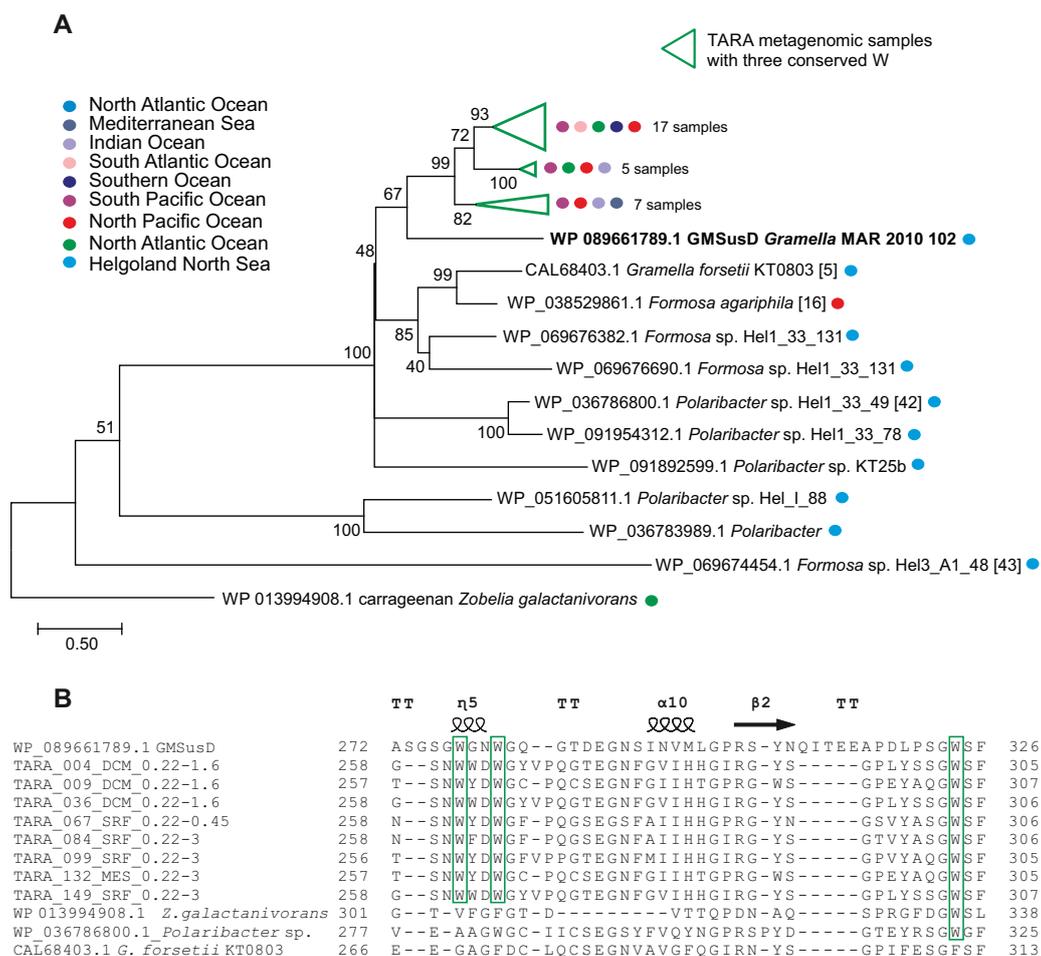


Figure 2.1.7 Phylogenetic analysis of predicted laminarin binding proteins and conservation of binding site residues of GMSusD among homologs from marine bacteria. A: The phylogenetic protein tree based on a structure guided alignment using the structure of GMSusD as guide shows that GMSusD and twenty-nine identified metagenomic homologs, which have about 40% sequence identity and contain the three tryptophan residues that bind branched laminarin. The SusD-like protein sequences from previously analyzed laminarin PULs: were included: CAL68403.1 from *G. forsetii* KT0803 [5], WP_038529861.1 from *F. agariphila* [16], WP_036786800.1 from *Polaribacter* sp. Hel1_33_49 [42] and WP_069674454.1 from *Formosa* sp. Hel3_A1_48 [43]. A carrageenan-specific SusD-like protein was used as outgroup. Colored dots indicate different environmental origin of the sequences. The scale bar describes substitutions per site and the numbers at nodes represent bootstrap values. **B:** A structure guided alignment shows conserved tryptophan residues within SusD-like protein sequences obtained from TARA metagenomes, whereas other SusD-like sequences do not possess all three tryptophan residues that are predicted to be structurally conserved. To simplify, only eight out of twenty-nine TARA sequences with conserved three tryptophan residues and three other SusD-like proteins are depicted on the figure. On the top, secondary structure elements of GMSusD are shown. Following representation was applied: α helices with squiggles, β strands with arrows, TT letters for turns and η denotes a helix with 3 residues per turn (Robert & Gouet, 2014).

Discussion:

In the present study, we functionally characterized and solved the 3D structure of the SusD-like protein GMSusD from the marine bacterium *Gramella* sp. MAR_2010_102. GMSusD interacts with highly β -1,6-branched laminarin and pure β -1,6-linked pustulan. Laminarin is produced by many algal species in the ocean, thus the mechanisms through which it is acquired by bacteria needs to be examined in detail. We evaluated the presence of marine SusD-like genes by analyzing metagenomic datasets, which revealed their broad distribution in different regions of the ocean (Figure 2.1.1). By coupling these results with biochemical characterization of GMSusD we gain a better understanding of laminarin recognition and binding and provide a strategy for the functional annotation of other SGBPs in metagenome datasets.

The substrate specificity of GMSusD was determined by affinity gel electrophoresis (Figure 2.1.4) This experiment revealed binding of GMSusD protein from *Gramella* sp. MAR_2010_102 towards laminarin, which is in agreement with bioinformatic predictions. The unanticipated finding was that GMSusD recognizes laminarin derived from *E. bicyclis*, whereas it binds only weakly the laminarin from *L. digitata* (Figure 2.1.4). It has been shown that laminarin from *E. bicyclis* possesses branches with a degree of polymerization up to three β -1,6-glucoses (Menshova *et al.*, 2014). Our results show higher glucose release from *E. bicyclis* laminarin, when compared to *L. digitata*, after β -1,6-exo-acting laminarinase treatment, demonstrating indeed a different β -1,6- glucose content. The basic structure of laminarin is a β -1,3-glucan with β -1,6-glucose side chains. However, the laminarin isolated from *E. bicyclis* has a higher content of β -1,6-linked glucose and a higher molecular weight. The ratio of *E. bicyclis* laminarin linkages (1 \rightarrow 3; 1 \rightarrow 6)- β is 3:2, whereas for *L. digitata* laminarin bonds (1 \rightarrow 3; 1 \rightarrow 6)- β ratio is 7:1 (Pang *et al.*, 2005). This phenomenon influences the binding of GMSusD on the retardation gel, what was additionally demonstrated with the gel assay where the binding to laminarin from *E. bicyclis* was diminished after enzymatic removal of β -1,6-linked glucoses. Moreover, GMSusD recognizes linear β -1,6-glucan pustulan (Figure 2.1.4).

The differences of binding by GMSusD on affinity gel electrophoresis could lead to two different conclusions. The most striking explanation would be that GMSusD requires β -1,6-configured glucose for selective binding. Besides the specificity of GMSusD for branched laminarin and pustulan observed on the retardation gels, two other aspects support this hypothesis. There are many enzymes thought to be involved in laminarin degradation: endo-acting GH16, GH17, GH55, GH64, and GH81, and exo-acting laminarinases belonging to the GH3 family, all are responsible for cleavage of β -1,3

linkage. Additionally, the β -1,3-1,6-endoglucanase from GH5 family (Wang *et al.*, 2016) and the β -1,6 exo-acting laminarinases from GH30 family have been characterized (Becker *et al.*, 2017). In light of the known specificities of these families, GMSusD is encoded in a PUL that suggests laminarin specificity. Besides the β -1,3 laminarinases, there are two genes: GH5 and GH30 indicating capability to recognize β -1,6 linkages. The GH30 from GMSusD PUL shares 29% sequence identity with the characterized GH30 β -1,6 exo-acting laminarinase homolog from *Formosa* sp. nov strain Hel1_33_131. Based on bioinformatics of the genome, this is the only PUL predicted as a laminarin specific PUL encoded in *Gramella* sp. MAR_2010_102. GMSusD was upregulated during spring diatom bloom in Helgoland, Germany. Interestingly, the marine planktonic diatoms *Chaetoceros debilis* which is known to produce pustulan-like laminarin with high content of β -1,6 glucose (Størseth *et al.*, 2006) appears frequently during spring blooms in Helgoland (Kraberg *et al.*, 2010). Assuming that GMSusD is specific to highly branched laminarin and pustulan, it makes this diatom a possible candidate for producing polysaccharides available for *Gramella* sp. MAR_2010_102 to metabolize. Taken together, we propose a possibility of specific ecological relationship between *Gramella* sp. MAR_2010_102 and *C. debilis*, which is governed by the production and subsequent metabolism of β -1,6 glucose rich laminarin. The presence of GMSusD accommodates the utilization of such a specialized substrate and could provide a selective advantage over other bacteria in competing for β -glucans such as laminarin.

Given the low affinity of the protein, the here presented results are based on one method, which has limited ability to compare binding between different polysaccharides. Retardation gels report binding or its absence, but the stoichiometry of the reaction is not included in the calculations. Possible scenarios could be that GMSusD has a comparable affinity towards both commercially available laminarins, but laminarin from *E. bicyclis* as longer polymer possesses a higher binding density, what makes the retention stronger. Previously characterized SusD-like proteins are known to recognize a spectrum of structurally similar polysaccharides. Growth experiments reveal the importance of archetypal SusD for maltooligosaccharides, including amylopectin and pullulan, where the latter contains α -1,6-branches. Absence or presence of the branches does not change drastically binding abilities in that case (Koropatkin *et al.*, 2008). According to the recent proposed “pedal bin” mechanism of SusD-SusC complex, the SusD acts as a lid during nutrient import (Glenwright *et al.*, 2017), which suggests low affinity of the protein and limits the binding assays. Hence, the possibility that low affinity is the reason why binding is not observed for less branched laminarin from *L. digitata* cannot be excluded. Our

inability to detect significant retention of GMSusD on the gel with *L. digitata* laminarin does not necessarily indicate a complete lack of binding.

In order to verify the binding abilities with another method we used carbohydrate microarrays, which were populated with commercially available glucans as well as with β -mannan and arabinoxylan. GMSusD showed only binding to pustulan, what was also observed with AGE. However, there was no signal detected towards laminarin. In the microarrays used for our analysis, polysaccharides are non-covalently bound onto the array surface. Small polymers, like oligosaccharides, bind less on the array, they require coupling before immobilization (Pedersen *et al.*, 2012). The lack of signal detected against laminarin might be due to the fact that this molecule has a relatively low degree of polymerization and probably did not immobilize onto the array surface. This hypothesis is supported by the absence of signal against laminarin detected for BS-4002 - a commercially available antibody known to be specific for β -1,3-linked glucan. The BS-4002 recognized pachyman, but no binding was observed for any of the two laminarins (Figure 2.1.5).

The crystal structure of the recombinant protein GMSusD has been determined at 1.8 Å resolution (PDB code 6GCZ), which exposed a fold typical for a SusD-like protein family. The SusD-like scaffold composed of multiple TPRs, which is a conserved feature within the mentioned family, with loops forming the binding site to assure flexibility, allowing a protein to recognize different polysaccharides. Our inability to solve the protein structure with the appropriate ligand might be caused by crystal packing. The crystals were soaked with oligosaccharides, but tight packing might have blocked access to the binding sites. However, based on the comparative structural and mutagenesis analysis we postulate that three tryptophan residues are responsible for polysaccharide binding W287, W290 and W323. An analysis of metagenomic data revealed conservation of these residues within the phylogenetic tree. We can assume with high confidence that proteins with structurally conserved three tryptophan in the binding site recognize laminarin in the same way as GMSusD. The variety of suggested laminarin structures must therefore be considered when annotating these proteins in environmental datasets. The picture of laminarin as a β -1,3-linkage glucan has emerged. However, various degree of branches and length of the backbone for laminarin have been reported (McConville *et al.*, 1986; Størseth *et al.*, 2005; Gügi *et al.*, 2015), which most likely reflects on laminarin recognition by binding proteins and enzymes. Differences in target polysaccharide may result in varied composition and arrangement of genes encoded in the PUL. For example, the laminarin-induced PUL from *Polaribacter* sp. Hel1_33_49 encodes enzymes from the families: GH16, GH3, GH17, GH30 (Xing *et al.*, 2014), and a SusD-like protein that does not

contain the three structurally conserved binding site tryptophan that defines the GMSusD binding site. The *Polaribacter* sp. Hel1_33_49 SusD-like protein clusters separately on the phylogenetic tree, indicating differences in the laminarin recognition. Comparing the above *Polaribacter* sp. and *Gramella* sp. MAR_2010_102, one imposing property would be the presence of both during diatom-dominated spring phytoplankton blooms nearshore the North Sea island Helgoland. Hence, it could conceivably be hypothesized that bacteria need to compete for similar laminarin substrate and fine differences allow bacteria to adapt to different polysaccharides.

In conclusion, the biochemical and structural analysis together with mutagenesis experiments provides a molecular insight into GMSusD binding, which can be used to identify SusD-like proteins with higher molecular resolution. Hence to validate the accuracy of the *in-silico* predictions of protein binding in the context of aquatic ecology more biochemical studies, such as GMSusD characterization, reported here, are necessary. There is a considerable imbalance between bioinformatics models and biochemical and structural studies. Thus, future research on the molecular basis of the carbohydrate recognition and degradation should be undertaken to develop a better understanding of laminarin acquisition, additional studies including characterization of enzymes and other carbohydrate binding proteins will be needed.

Material and Methods:**Distribution of SusD-like proteins in the ocean**

For the following analyses we considered a data set of ~140M Open Reading Frames (ORFs) originated from 472 samples collected by three major mega-sequencing projects of the ocean microbiome: the TARA Ocean Expedition (TARA) (Williamson *et al.*, 2008) the Global Ocean Sampling (GOS) (Karsenti *et al.*, 2011) and the Ocean Sampling Day (OSD) (Kopf *et al.*, 2015) (Table 2.1.1). The whole ORF set was searched against the SusD profile HMMs, found in the *Pfam* database of protein domain families (version 30) (Finn *et al.*, 2016) using the *hmmsearch* program of the *HMMER* package (version 3.1b2) (Finn, *et al.* 2011). From this set of ORFs we selected those from surface water samples (defined by a maximum depth of 5 meters). Additionally, we screened the same data set against the RecA profile HMMs of the *Pfam* database, in the same way as for the SusD domain, and we selected the ORFs from the surface samples. We then retrieved the abundances for the annotated ORFs in the selected samples. For TARA and OSD samples we used the read coverage to each ORF as a proxy for abundance. The coverage was calculated by mapping the reads against the assemblies using the *bwa-mem* algorithm from BWA (Li & Durbin, 2009). The coverage for each ORF was calculated using BEDTOOLS (Quinlan & Hall, 2010) performing the intersection of the ORF coordinates to the assemblies, then the per-base coverage was normalised by the length of the ORF. Each ORF from GOS had, instead, a count of 1.

Finally, we compared the proportion of SusD-like genes against those of RecA, which is usually present as a single copy gene in most microbial genomes.

Supplementary Table 2.1.1: Primers that were used for site-directed mutagenesis.

287W/A F	AGCGGTGCGGGTAACTGGGGTCAGGGTACCGACG
287W/A R	GTTACCCGCACCGCTACCGCTCGCGTTGGTGT
290W/A F	GGTAACGCGGGTCAGGGTACCGACGAGGGTAACA
290W/A R	CTGACCCGCGTTACCCCAACCGCTACCGCTCGCG
323W/A F	GCGGTGCGAGCTTCAACCCGGTTCTGCCG
323W/A R	AAGCTCGCACCGCTCGGCAGATCCGGCG

Heterologous protein expression

The recombinant plasmid pET28 (a+) containing GMSusD (NCBI Reference Sequence: WP_089661789, residues from 41-501) from *Gramella* sp. MAR 2010 102 and SusD-like protein from *Salegentibacter* were obtained by gene synthesis (GenScript). The first forty amino acids of GMSusD were not included in the final construct as they correspond to the signal peptide, which is likely removed when the SusD-like protein is assembled in the outer membrane. Competent *Escherichia coli* BL21 (DE3) cells (New England Biolabs Ipswich, Massachusetts, United States) were transformed with this plasmid. The recombinant proteins were designed to contain His-tags at the N terminus. Two ml of LB medium were inoculated with a single bacterial colony and incubated overnight at 37°C. 0.5 ml of the preculture was used to inoculate 2 L of ZYP-5052 expression medium containing 150 µg/ml kanamycin (Studier, 2005). Cells were grown in autoinduction medium at 20°C with shaking (150 rpm) for 4 days. To avoid foaming, 1 ml of Antifoam (Silicone Antifoam, Sigma-Aldrich, Saint Louis, Missouri, USA) was added. Cells were harvested by centrifugation (45 min, 4500 g) and suspended in 25 ml of sucrose solution [(25% w/v), 50 mM Tris-HCl, pH 8.0]. Ten milligrams of lysozyme were added followed by mixing for 15 min with a magnetic stir bar at 20°C. 50 ml of a deoxycholate solution [1% deoxycholate (w/v), 1% Triton X-100 (v/v), 20 mM Tris-HCl, pH 7.5, 100 mM NaCl] was added followed by stirring for 10 min. One ml of 1 M MgCl₂ and DNase were added and stirred for 10 minutes. Cell debris was removed by centrifugation at 16,000 g for 45 min at 4°C.

Mutagenesis

In order to mutate residues at the binding site: W287, W290 and W323 predicted based on the structure analysis; the primers present in the Supplementary Table 2.1.1, were designed. The experiment was performed in the same manner as previously described (Liu & Naismith, 2008). Concisely, the PCR was used to amplify pET28 – GMSusD plasmid with point mutations, which then was transformed to *E. coli* DH5α. The three single mutants have been constructed and for each of them alterations of tryptophan to alanine were confirmed by sequencing.

Purification by Immobilized Metal Ion Affinity Chromatography (IMAC) and Size Exclusion Chromatography (SEC)

Recombinant GMSusD protein was purified using a 5 mL Hi-Trap metal affinity column (GE Healthcare, Chicago, Illinois, USA). The cell lysate was loaded in buffer (20 mM Tris, 500 mM NaCl, pH 8.0) and was eluted with an imidazole gradient (15-500 mM). Fractions were analyzed by the SDS-PAGE, those containing the protein of interest were dialyzed against 20 mM Tris, 500 mM NaCl, 10 mM CaCl₂ buffer in parallel to the thrombin digestion in the dialysis tube, 100 µl of thrombin was added (Thrombin Clean Cleave Kit, Sigma-Aldrich). Dialysis and digestion were accomplished at 4°C, overnight. The protein was concentrated in a stirred ultrafiltration unit (Amicon, Sigma-Aldrich) into 1.5 ml of 40 mg/ml, the concentration value was determined by the absorbance at A280 using the extinction coefficient of 1.25 for GMSusD. Size exclusion chromatography – SEC, with a 120 ml HiPrep 16/60 Sephacryl S-200 HR column, elution was performed with 20 mM Tris-HCl, 150 mM NaCl pH 8, flow rate 0.5 ml/min. Additionally, the analytical SEC ENrich SEC 650 10 x 300 column was calibrated with five proteins with known molecular weight [thyroglobulin (bovine) 670 kDa, γ-globulin (bovine) 158 kDa, ovalbumin (chicken) 44 kDa, myoglobin (horse) 17 kDa] (BioRad, Hercules, California, United States), at the same conditions in order to determine the size of purified protein. The recombinant protein after SEC purification was concentrated again in a stirred ultrafiltration unit (Amicon, Sigma-Aldrich) to 26 mg/ml. The concentrated protein was centrifuged for 20 min at 12000 g in order to remove aggregates.

Dynamic Light Scattering (DLS)

20 µl of filtered proteins at the concentration of 1 mg/ml, was applied into the plate. It was centrifuged 5 min to remove bubbles and measured in a DLS DynaPro Plate Reader. The temperature was maintained at constant 25°C during the analysis.

Polysaccharide substrates

The following polysaccharides were used for binding studies: glucomannan (konjac), xylan (beechwood), pullulan (1,4;1,6- α -D-glucan), Carboxymethyl Cellulose 4M (1,4- β -D-glucan), pachyman, (1,3- β -D-Glucan), β -glucan (barley), lichenan (icelandic moss), xyloglucan (tamarind) from Megazyme (Wicklow, Leinster, Ireland); laminarin (*E. bicyclis*) and scleroglucan (1,3- β -D-glucan with 1,6- β -glucose side chains) from Carbosynth; laminarin (*L. digitata*) from Sigma-Aldrich and pustulan (1,6- β -D-glucan) from Elicityl, Crolles, Auvergne-Rhône-Alpes, France.

Affinity gel electrophoresis (AGE)

12% native acrylamide gels were prepared, gels without polysaccharide as control and others containing different polysaccharide concentrations from 0,01% up to 1%. We tested AGE with laminarin (*E. bicyclis* and *L. digitata*), pustulan, mixed-linkage glucan, lichenan and scleroglucan. Ten micrograms of GMSusD, Bovine Serum Albumin BSA and a α -mannan specific SusD-like protein from *Salegentibacter* sp. Hel 1_6 were loaded onto each gel. The motilities of the proteins were quantified between the native gel and gel containing one of the respective polysaccharides. Electrophoresis was performed for 2 h at a constant voltage of 80 V on ice. The proteins were visualized by staining with Coomassie Blue. The graphs were prepared in Prism 7 assuming one specific binding site (GraphPad Software).

Carbohydrate microarray analysis

The polysaccharides were dissolved in water to 5 mg/ml, except pachyman, which was dissolved in 4 M NaOH with 0.1% (w/v) NaBH₄. Subsequently they were diluted in printing buffer (55.2% glycerol, 44% water, 0.8% Triton X-100) by a 2-fold dilution and added in wells of a 384-microwell plate. The plate was centrifuged for 10 min at 4000 rpm and the content was printed at 22°C with 55% humidity onto nitrocellulose membrane with a pore size of 0.45 μ m (Whatman, Sigma-Aldrich) using a microarray robot (Sprint, Arrayjet, Roslin, UK). Each polysaccharide was printed in triplicate.

The microarrays were analyzed as previously described (Vidal-Melgosa *et al.*, 2015). Briefly, microarrays that contained the printed polysaccharides were blocked for 1 h in PBS (140 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.7 mM KH₂PO₄, pH 7.5) with 5% (w/v) low fat milk powder (MPBS). Next, the arrays were incubated for 2 h with probes: his tagged proteins produced in the *E. coli* expression system: GMSusD (50 μ g/ml) and SusD1 from *Salegentibacter Hel 1_6* (50 μ g/ml); and monoclonal antibodies: BS-4002 (1:1000) (Meikle *et al.*, 1991), BS-4003 (1:1000) (Meikle *et al.*, 1994) and LM11 (1:10) (McCartney *et al.*, 2005), all diluted in MPBS. Then, microarrays were washed in PBS and incubated for 2 h with anti-His tag (for recombinant proteins produced in *E. coli*), anti-rat or anti-mouse secondary antibodies (for antibodies produced in different host animals) conjugated to alkaline phosphatase (Sigma-Aldrich) diluted 1:1500 (for his tag) or 1:5000 (anti-rat and anti-mouse) in MPBS. After washing in PBS and deionized water, microarrays were developed in a solution containing 5-bromo-4-chloro-3 indolylphosphate and nitro blue tetrazolium in alkaline phosphatase buffer (100 mM NaCl, 5 mM MgCl₂, 100 mM diethanolamine, pH 9.5), which allows detection of the

antigens bound to the membrane. These substrates develop insoluble purple product on the membrane when they are exposed to alkaline phosphatase conjugated antibodies. Once developed, the arrays were scanned at 2400 dots/inch and probe binding signals were quantified using pertinent software (Array-Pro Analyzer 6.3, Media Cybernetics, Rockville, Maryland, United States). The highest mean spot signal intensity was set to 100 and the rest of values normalized consequently. Data are shown in a heat map where color intensity is correlated to mean spot signal value. A cut-off of 5 arbitrary units was applied.

Crystallization, data collection and structure solving

Crystallization conditions were screened via the sitting drop method of vapor diffusion in 96-well plates. The protein at a concentration of 26 mg/ml was screened with the Jena Bioscience buffers (Jena Bioscience, Jena, Thuringia, Germany) and stored at 20°C in a vibration free incubator for crystallization. The optimization of the well B11 from the screen MCSG1 (0.2 M magnesium chloride; 0.1 M Tris: HCl, pH 8.5; 20 % PEG 8000) (Anatrace, Maumee City, Ohio, United States) via 1 µl hanging drop was performed, resulting in the final crystals at 0.2 M magnesium chloride pH 7.8, 16% PEG 8000. Crystals were transferred into mother liquor supplemented to a concentration of 30% of glycerol followed by flash freezing in liquid nitrogen. Diffraction data were collected at the EMBL Hamburg, Beamline P11 at 80 K. Indexing and integration were carried out using XDS autoprocess (Kabsch, 2010) and the integrated data were scaled and merged in Aimless (Evans, 2011). The structure was solved by molecular replacement using the SusD-like protein (BF3747) from *Bacteroides fragilis* as a search model (PDB code 3QNK) in Phaser3 (Winn *et al.*, 2011)(McCoy *et al.*, 2007). The structure was built and refined using Buccaneer (Cowtan, 2006), Coot (Emsley *et al.*, 2010) and Refmac (Murshudov *et al.*, 2011) and validated with MolProbity (Davis *et al.*, 2007) before deposition. Single crystals were soaked with laminarin oligosaccharides, but did not lead to a complex structure. Co-crystallisation of GMSusD with laminarin hexose and 1% laminarin from *E. bicyclis* digested for 1h at 37°C with the endo-laminarinase (GH16) or a set of enzymes (GH16, GH17 and GH30) (Becker *et al.*, 2017) did not produce protein crystals in screens or with the condition that gave the apo-structure.

Enzyme assay of laminarin with GH30

The amount of β -1,6-linked glucose released after GH30 digestion was measured with a High-Performance Anion Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD). The GH30 enzyme specific for the β -1,6- side chain (Becker *et al.*, 2017) was incubated with 0.1% laminarin: from *E. bicyclis* and *L. digitata*. The time

points (5 min, 20 min, 40 min and 60 min) at 37°C incubation of the polysaccharide with the GH30 were collected and the enzymatic reaction was stopped by heating (90°C, 5 min). A 25 µl of each sample was diluted in 475 µl of Milli-Q water and analyzed for the glucose release with the Dionex ICS-5000+ HPAEC-PAD System (Thermo Fischer Scientific Inc., Waltham, Massachusetts, USA). The separation was performed with the eluent 1: 0.15 M NaOH (HPLC grade, VWR) and eluent 2: 0.15 M NaOH and 1 M sodium acetate (HPLC grade, Sigma), which were dissolved in Milli-Q water and degassed with helium for 10 min. Eluent 2 was filtered through a 0.2 µm nylon filter membrane and then degassed. Each time point sample was analyzed separately by a linear gradient from 100% eluent 1 to 50% eluent 1 and 50% eluent 2 for 19.5 min, afterward an increase of the eluent 2 concentration to 100% over 120 s was performed. At the end the concentration of eluent 1 was returned to 100% over 30 s. The electrochemical detection on the gold working electrode and a pH reference electrode (Ag/AgCl) were used. Samples were separated at 25°C on a Dionex CarboPac PA100 analytical column (2×250 mm) coupled with a Dionex CarboPac PA100 guard column (2×50mm). The reference sugars at concentration of 1 µg/mL each were used: glucose (Sigma-Aldrich), laminaribiose, laminaritriose, laminaritetraose, laminaripentaose and laminarihexaose (Megazyme, Wicklow, Leinster, Ireland). The entire separation was conducted by isocratic flow rate of 0.25 mL/min.

Structure guided alignment and phylogenetic analysis

To test if annotated ORFs contain the same location of predicted, structurally conserved binding residues as GMSusD and in consequence a similar manner of binding, we performed a structure-guided multiple sequence alignment. The conserved residues are specific to the GMSusD; therefore, we searched the set of ORFs from surface samples, annotated to SusD-like protein, against the *Gramella* sp. MAR_2010_102 SusD-like protein sequence, using the MMseqs2 (*Many-against-Many sequence searching 2*) software (Steinegger & Söding, 2017) with a coverage threshold of 0.999 and a sensitivity of 8. The subset of matching ORFs was then deduplicated using again the MMSeqs2 software (Steinegger & Söding, 2017). We created a template file referring to the structure of GMSusD (PDB code 6GCZ) to structurally guide the alignment. Hence, we applied the *T-COFFEE* multiple sequence alignment program (Notredame *et al.*, 2000), using the *TAlign_pair* method, to perform the structure guided alignment with the PDB file of GMSusD as a 3D template. From the alignment we selected the ORFs with conserved residues in the binding site, we added the respective duplicated ORFs, and we checked both the degree of completeness and the average frequency of bases mismatching for each

ORFs These two evaluation steps are critical, since the considered ORFs came from metagenomic assemblies and can be fragmented or contain a high base variability, which is a measure of how many bases, in the reads building the assembly consensus, result different from those of the consensus at each position. Usually the consensus is based on the most frequently encountered nucleotide at each position. We calculated the average entropy, as measure of the base variability, for each ORF using the processing script *diversiutils* from the *DiversiTools* package[<http://josephhughes.github.io/DiversiTools/>]. Based on the metagenomics analysis we have chosen sequences with predicted structurally conserved binding site analogous to GMSusD. Additionally, the SusD-like protein sequences from predicted (WP_069676382.1, WP_069676690.1, WP_091954312.1, WP_091892599.1, WP_051605811.1, WP_036783989.1) or previously analyzed laminarin PULs: (CAL68403.1 (Kabisch *et al.*, 2014), WP_038529861.1 (Mann *et al.*, 2013), WP_036786800.1 (Xing *et al.*, 2014), WP_069674454.1 (Unfried *et al.*, 2016)) were selected. Carrageenan specific SusD-like protein (Ficko-Blean *et al.*, 2017) was added to the analyzed sequences. Quoted collection was objected with the structure-based alignment mentioned above. A maximum likelihood reconstruction of phylogeny was conducted in MEGA7 using the LG model and 250 bootstrap replications.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) grant HE 7217/1-1 to Jan-Hendrik Hehemann and by the Max Planck Society. Matthias Höhne was supported by DFG grant HO 4754/5-1 and we acknowledge support by the DFG Forschergruppe (FOR 2406/1). We thank Nicole Koropatkin, Alisdair Boraston, Carol Arnosti and Thorsten Dittmar for scientific advice. We thank Hanno Teeling and Rudolf Amann for access to genomic information *Gramella* sp. MAR_2010_102, which was sequenced by the JGI within the COGITO project. This work was supported by beamline EMBO grant MX-454. We thank Olga Lorbeer for beamtime support.

Manuscript 2:

Marine SusEF-like proteins from *Gramella* sp. MAR_2010_102

bind laminarin

Title: Marine SusEF-like proteins from *Gramella* sp. MAR_2010_102 bind laminarin.

Agata Anna Mystkowska^{1,2,3}, Andrew Hettle⁴, Silvia Vidal-Melgosa^{1,2}, Alisdair Boraston⁴, Matthias Höhne^{3*}, Jan-Hendrik Hehemann^{1,2*}

¹University of Bremen, Center for Marine Environmental Sciences, (MARUM), Bremen, 28359, Germany.

²Max Planck Institute for Marine Microbiology, Bremen, 28359, Germany.

³Institute of Biochemistry, Greifswald University, Greifswald 17487, Germany

⁴Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia V8W 3P6, Canada

*To whom correspondence should be addressed:

Jan-Hendrik Hehemann, MARUM MPG Joint Research Group Marine Glycobiology, University of Bremen, Center for Marine Environmental Sciences, (MARUM), Bremen, 28359, Germany; Max Planck Institute for Marine Microbiology, Bremen, 28359, Germany, Phone: +49 421 218-65775, Email: jhehemann@marum.de, jheheman@mpi-bremen.de

Matthias Höhne, Institute of Biochemistry, University of Greifswald, Felix-Hausdorff-Str. 4, 17487 Greifswald, Germany, Phone: +49 3834 420 4417
Email: Matthias.Hoehne@uni-greifswald.de

Keywords: surface-glycan binding proteins, laminarin, *Bacteroidetes*, polysaccharides, PKD domains

Abstract

Marine *Bacteroidetes* play a key role in the degradation of diverse types of polysaccharides produced by diatoms during algal blooms. Members of this phylum have adapted to decompose polysaccharides with complexes of surface localized Sus-like proteins encoded within a Polysaccharide utilization locus. However, the precise molecular characterization of marine surface glycan binding proteins involved in molecular recognition of algal polysaccharides remains unknown. We discovered carbohydrate binding abilities of two surface glycan binding proteins from *Gramella* sp. MAR_2010_102, initially annotated as hypothetical. We found subtle differences in laminarin recognition between the two proteins, which are discussed. We postulate that these proteins belong to a relatively unexplored glycan-binding family similar to the SusEF- proteins described in starch utilization.

Introduction

Degradation of polysaccharide by marine bacteria is a specific process and a key step in the carbon cycle. Polysaccharides appear as transparent exopolymer particles (TEP) generated by phytoplankton (Passow & Alldredge, 1995), algal cell walls (Michel *et al.*, 2010) or as sinking particles called marine snow (Engel *et al.*, 2004). Bacteria that use polysaccharides as a carbon source need an efficient system to utilize complex polysaccharides present throughout the water column. *Bacteroidetes* are known to be the key player in glycan degradation (Teeling *et al.*, 2016). To overcome the high diversity of polysaccharides (Turnbull & Field, 2007), *Bacteroidetes* evolved various genomic islands called polysaccharide utilization locus (PUL) (Martens, Koropatkin, *et al.*, 2009).

Although the proteins that compose the starch utilization system (Sus) have been extensively studied especially in gut bacteria (Anderson & Salyers, 1989), the functions of several genes encoded in the PULs of marine bacteria remain unknown. In the case of marine bacteria there are only a few functional studies on the Sus-like proteins (for example Chen *et al.*, 2018; Unfried *et al.*, 2018). A few components of the Sus-like system of marine bacteria show low amino-acid sequence identities to characterized proteins and in consequence, they are often wrongly annotated and/or omitted in bioinformatics analysis. This seems to be the case for homologs of SusEF proteins, whose role is to bind and recognize starch (Shipman *et al.*, 2000; Cameron *et al.*, 2012). When SusEF-like genes are present in PULs, they are often encoded near to the SusCD genes, which products are responsible for binding and transport of oligomeric fragments into the periplasm. By analogy, marine SusEF-like proteins might be important for substrate binding to the outer membrane — the first crucial step of polysaccharide acquisition, which is highly unexplored in the context of marine bacteria.

The carbohydrate-binding SusD-like protein and SusC-like transporter mediating polysaccharide acquisition are assembled on the surface of the bacterial cell. A study in Helgoland, Germany, showed that homologs of both proteins were highly expressed in response to diatom blooms, when polysaccharides are produced (Teeling *et al.*, 2012). Based on the observation of fluorescently labeled polysaccharides being rapidly ingested by bacteria leading to fluorescent cells, Reintjes *et al.* proposed selfish uptake mechanisms for marine bacteria (Reintjes *et al.*, 2017). This mechanism was also demonstrated previously for mannan in the human gut *Bacteroidetes thetaiotaomicron* (Cuskin *et al.*, 2015). The selfish mechanism implies that bacteria do not share or try to minimize sharing of digested polysaccharide with neighboring communities by transporting relatively long oligosaccharides into the cell. The experiments with fluorescent polysaccharides strongly indicate that marine *Bacteroidetes* must have an

elaborate system to recognize and bind polysaccharides.

The best-studied example of polysaccharide utilization in *B. thetaiotaomicron* shows that starch binding is mediated by four surface glycan-binding proteins (SGBPs): SusG is a glycoside hydrolase containing a carbohydrate-binding module (CBM); and SusD, SusE and SusF are carbohydrate-binding proteins (CBPs) without enzymatic activity (Koropatkin *et al.*, 2008; Koropatkin & Smith, 2010; Cameron *et al.*, 2012). The latter two are multidomain proteins with three or two glycan binding sites, respectively. Each binding site of SusE and SusF binds with a slightly different affinity towards various starch oligosaccharides, allowing them to accommodate distinct regions of the polysaccharide (Cameron *et al.*, 2012).

Aside from the binding ability, the assembly of the SusCDE complex on the cell surface is crucial for uptake of starch (Cameron *et al.*, 2014; Foley *et al.*, 2018) and xyloglucan (Tauzin *et al.*, 2016; Foley *et al.*, 2019). Glenwright *et al.* presented the structure of the functionally distinct SusCD-like complex from *B. thetaiotaomicron*, where SusD-like protein might push the substrate through the SusC-like protein based on the proposed pedal bin mechanism (Glenwright *et al.*, 2017). Interestingly, the SusCD-like complex was co-purified with BT2261 and BT2262 proteins encoded upstream to SusCD. These two proteins, which were tightly connected with SusCD-like proteins after purification, had a structure akin to the characterized starch-specific SusEF. Nevertheless, defining the specificity of SusEF-like proteins for various polysaccharides from different environments is important to fully understand their role.

In this study, we identified laminarin-specific GMSusE and GMSusF proteins from the planktonic *Gramella* sp. MAR_2010_102. Laminarin is a highly abundant substrate in the ocean (Becker, 2018) and studying its microbial acquisition is of the great importance to understand the marine carbon cycle at molecular resolution. Here, we provide the biochemical characterization of two marine SusEF-like proteins. Our results expand the pool of knowledge about these proteins.

Results

Sequence screening and recombinant protein production

The genome of *Gramella* sp. MAR_2010_102 contains two genes of unknown function encoded upstream of the previously characterized GMSusD (Figure 2.2.1A) (Mystkowska *et al.*, 2018). To elucidate the domain architecture and to compare to the other proteins we used BLAST (Altschul *et al.*, 1997). One has a “PKD domain” (cd00146), while the other does not contain any conserved domains. In both cases, most of the proteins that came up in BLAST searches were annotated as hypothetical proteins, cell-surface proteins, and PKD domain-containing proteins. While screening genes in the vicinity of SusCD in other predicted laminarin PULs from bacteria present during Helgoland bloom, we systematically observed annotations as either polycystic kidney disease (PKD) domains or hypothetical proteins. The predicted three-dimensional structures of tested proteins from *Gramella* sp. MAR_2010_102 obtained from the PHYRE2 server (Sternberg & Kelley, 2009) shows a multimodular organization consisting of an immunoglobulin-like β -sandwich. However, the structural models are based on the very low identity (around 20%) and have many unsolved regions. The multiple PSI-BLAST alignments and PHYRE2 models suggest potential protein domains as shown in figure 2.2.1B.C. However, at this stage this results remains highly hypothetical.

To test for a potential enzymatic function of the proteins, we expressed the genes recombinantly in *E. coli* and obtained the 37 kDa GMSusE and slightly smaller 30 kDa GMSusF. Ion metal affinity chromatography (IMAC) followed by size exclusion chromatography (SEC) resulted in pure proteins that were used for further analysis. Initially, we tested them for catalytic activity. However, p-hydroxybenzoic acid hydrazide (PAHBAH) and high-performance ion chromatography with pulsed amperometric detection (HPAEC-PAD) analyses did not reveal any digestion products. The other group of proteins found in gut microbiome PULs includes SusEF-like proteins (Cameron *et al.*, 2012; Tauzin *et al.*, 2016). This is a relatively unexplored group of SGBPs with low sequence identity, which makes them difficult to identify. As the PULs are conserved within diverse environments (Grondin *et al.*, 2017), we decided to further investigate these two hypothetical proteins from *Gramella* sp. MAR_2010_102 for their potential carbohydrate-binding abilities.

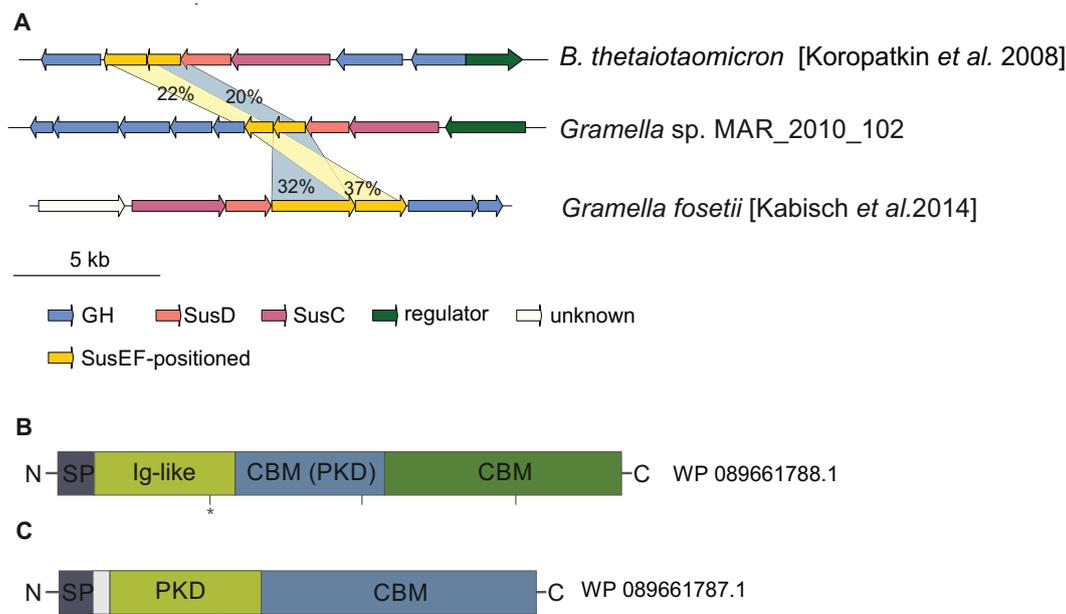


Figure 2.2.1 Two novel carbohydrate-binding proteins encoded upstream to SusD. The genomic representation of the PUL encoding GMSusE and GMSusF **(A)** (WP_089661787-089661788) and potential scheme of the protein domains for GMSusE **(B)** and GMSusF **(C)**. The grid lines indicated with asterisk represent the length of 100 amino-acids, N and C stand for N and C-terminus, SP- signal peptide. Of note, in case of SusE model the last residues (225-307) were not modeled as the identity to any known structures was too low.

Qualitative polysaccharides binding analyses

Both analyzed proteins migrated more slowly through the gels containing laminarin from different sources, whereas the control protein BSA (bovine serum albumin) behaved in the same way under all conditions. The retention of the investigated proteins in the gel with laminarin from *Eisenia bicyclis* appears to be stronger than in the gel containing *Laminaria digitata* laminarin (Figure 2.2.2). Binding was not detected in negative controls in the form of a native gel with galactomannan.

In addition to the gel-shift assays, we performed microarray analysis to detect possible interactions with other polysaccharides. The microarray analysis depicts the specificity of GMSusF towards pachyman, which is a linear β -1,3-glucan polysaccharide (Figure 2.2.3). Additionally, GMSusF shows a low affinity towards β -1,3;1,4-glucan, but the signal intensity is close to the cutoff value. We converted all values ≤ 5 to 0. In this experiment, we focused on GMSusF, because GMSusE did not show binding on the microarrays.

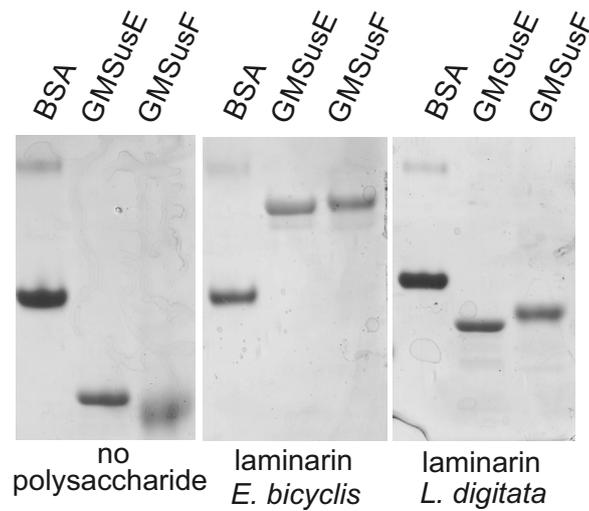


Figure 2.2.2 Affinity gel electrophoresis analyses. Proteins were electrophoresed in the native gels containing 0.5% (w/v) polysaccharides: laminarin from *E. bicyclis* or laminarin from *L. digitata*; and the gel without polysaccharide (control). The behavior of investigated proteins on the gels was compared to BSA as a control, which migrates in the same manner independently of the presence or lack of the polysaccharide.

To be comprehensive, we broadened the microarray analysis with epitope deletion. For this approach, an immobilized polysaccharide on the microarray membrane is digested by a particular enzyme. The cleaved epitopes are washed away from the membrane and thus not available for the investigated carbohydrate-binding protein. We performed pachyman epitope deletion by an β -1,3-glucanase treatment and included as a control a β -1,3;1,4-glucanase that does not digest β -1,3-glucans. Importantly, the pachyman signal for GMSusF was completely reduced after β -1,3-glucanase treatment (Figure 2.2.3). In the control with the β -1,3;1,4-glucanase, GMSusF binding was not changed, compared to the buffer-treated arrays that had no added enzyme. Positive controls with well-established antibodies were prepared: signals of BS-4002 (Meikle *et al.*, 1991) specific for β -1,3-glucan and BS-4003 (Meikle *et al.*, 1994) which recognizes β -1,3:1,4-glucan were reduced with the corresponding enzymes (Figure 2.2.3).

	GMSusE	GMSusF	GMSusF + β -1,3-glucanase	GMSusF + β -1,3;1,4-glucanase	BS-4002	BS-4002 + β -1,3-glucanase	BS-4003	BS-4003 + β -1,3;1,4-glucanase
Glucomannan (β mannan)	0	0	0	0	0	0	0	0
Arabinoxylan (β xylan)	0	0	0	0	12	0	0	0
Pullulan (α -1,4;1,6-glucan)	0	0	0	0	0	0	0	0
Carboxymethyl Cellulose (β -1,4-glucan)	0	0	0	0	0	0	0	0
Laminarin <i>Eisenia bicyclis</i>	0	0	0	0	0	0	0	0
Laminarin <i>Laminaria digitata</i>	0	0	0	0	0	0	0	0
Pachyman (β -1,3-glucan)	0	100	0	77	85	13	0	0
Scleroglucan (β -1,3;1,6-glucan)	0	0	0	0	0	0	0	0
β -glucan (β -1,3;1,4-glucan)	0	9	0	0	0	0	59	0
Xyloglucan (β -1,4-glucan with 1,6-xylose)	0	0	0	0	0	0	0	0

Figure 2.2.3 Carbohydrate microarray. Analysis including printed polysaccharides that were probed with His-tagged proteins: GMSusE and GMSusF or antibodies as controls followed by detection with secondary antibodies. Particular arrays were preincubated with the polysaccharide degrading enzymes as indicated on the top. Values on the heat map correspond to the mean binding signal intensities derived from the triplicates, where the highest value was set to 100 (yellow) and all other values were normalized accordingly. Positive controls: BS-4002 specific for β -1,3-glucan and BS-4003 for β -1,3;1,4-glucan.

Quantitative carbohydrate-binding analyses

The native gel and microarrays analyses were followed with quantitative isothermal titration calorimetry (ITC) studies (Figure 2.2.4). We evaluated the binding ability of GMSusE and GMSusF for β -1,3-glucooligosaccharides and laminarin by titrating the carbohydrates into the protein. The experiment revealed GMSusE and GMSusF specificity towards both oligo- and polysaccharides consisting of a β -1,3-glucan (Table 1). Values of enthalpically favorable and entropically unfavorable thermodynamics were relatively consistent for all measurements, and typical of protein-carbohydrate interactions (PMID: 15214846). Both GMSusF and GMSusE bound to laminaripentaose and laminarihexaose. Neither protein displayed a strong dependence of the affinity on the size of the oligosaccharide; however, the overall affinity of GMSusF for oligosaccharides was approximately 5-10-fold higher than GMSusE (Table 2.2.1).

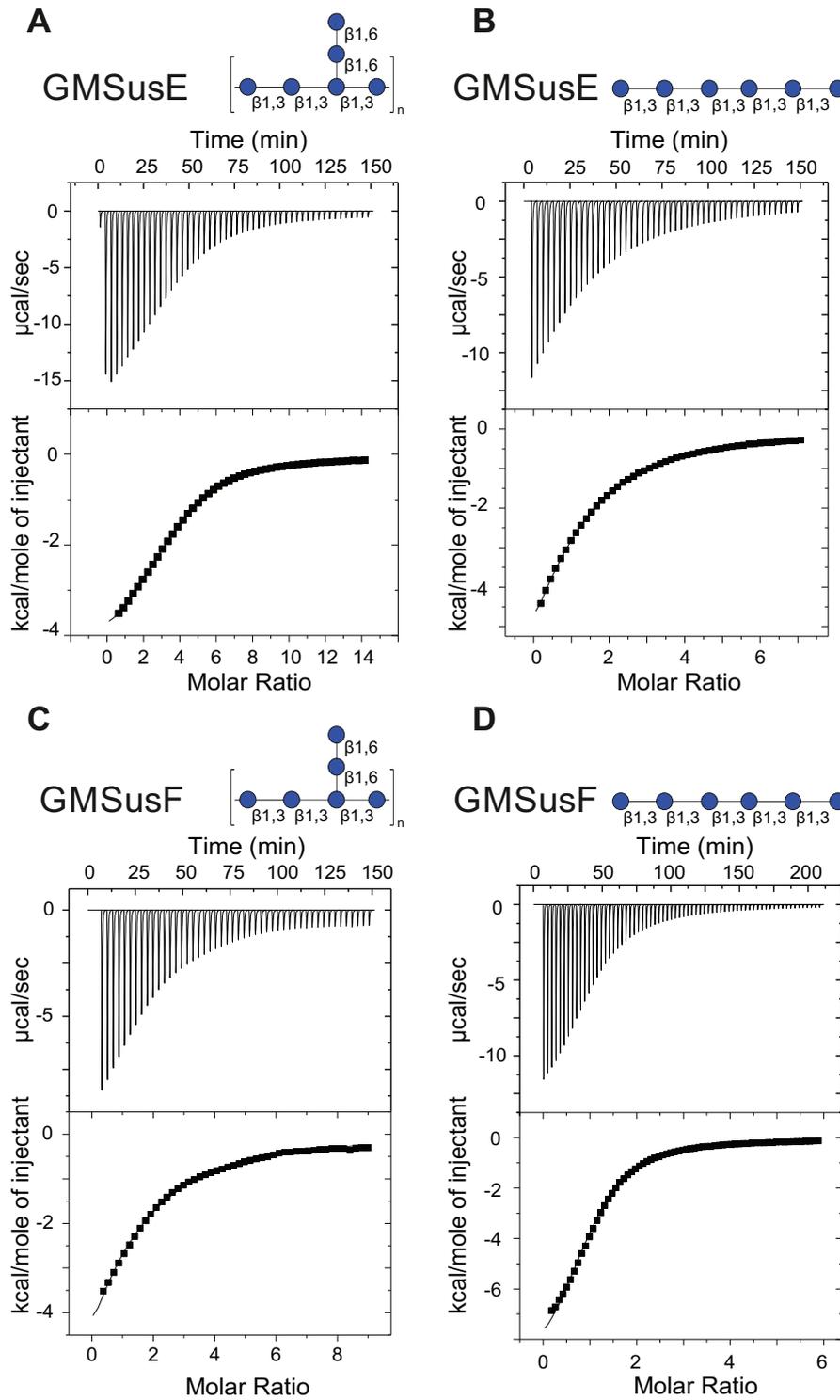


Figure 2.2.4 Qualitative isothermal titration calorimetry analyses. Representative analyses of 0.3 mM *E. bicyclis* (A, B) laminarin or 15 mM laminarin hexose (C, D) in the syringe titrated to GMSusE and GMSusF in the cell. The top panels show raw injection heats, the lower half, the integrated heats as solid squares fitted using a single-site model. The symbolic representation of tested carbohydrate is showed on the top of each ITC graph.

Using ITC, we also investigated the ability of GMSusF and GMSusE to bind laminarin from either *E. bicyclis* or *L. digitata*, which contain different degrees of β -1,6-glucose branches. Both GMSusF and GMSusE were able to bind these preparations of laminarin, though with opposite preferences (Figure 2.2.4). GMSusF showed a higher affinity for laminarin from *L. digitata* than *E. bicyclis*, while GMSusE preferred laminarin from *E. bicyclis* over laminarin from *L. digitata* (Table 2.2.1). The GMSusF data yielded a stoichiometry of 1.90 (\pm 0.19) molecules of L10/SusF-binding site for *E. bicyclis* laminarin and 1.23 (\pm 0.03) molecules of L10/SusF-binding site for *L. digitata* laminarin. This equates to one SusF-binding site per \sim 5 glucose subunits and one SusF-binding site per \sim 8 glucose subunits of *E. bicyclis* laminarin and *L. digitata* laminarin, respectively. In contrast, the GMSusE data gave a stoichiometry 0.29 (\pm 0.00) molecules of L10/SusE-binding site and 1.99 (\pm 0.05) molecules of L10/SusE-binding site, which equates to a footprint of one SusE-binding site per \sim 33 glucose subunits and one SusE-binding site per \sim 5 glucose subunits on laminarin from *E. bicyclis* and *L. digitata*, respectively. Additionally, GMSusF was tested against β -1,3;1,4-glucan and pustulan (β -1,6-glucan), but no interaction was observed.

Table 2.2.1 Affinity of GMSusF and GMSusE for oligosaccharides and polysaccharides determined by isothermal titration calorimetry at 25 °C in phosphate-buffered saline, pH 8.0

Protein	Ligand	K_a $\times 10^3 M^{-1}$	N	ΔH kcal/mol	ΔS cal/mol/K	ΔG kcal/mol
GMSusE	laminaripentaose	0.94 ± 0.02	1	-11.54 ± 0.11	-25.06 ± 0.39	-4.06 ± 0.01
	laminarihexaose	1.70 ± 0.01	1	-11.36 ± 0.02	-24.17 ± 0.08	-4.41 ± 0.00
	Laminarin <i>L. digitata</i>	4.30 ± 0.33	1.99 ± 0.05	-4.62 ± 0.15	0.49 ± 0.52	-4.96 ± 0.05
	Laminarin <i>E. bicyclis</i>	13.03 ± 0.61	0.29 ± 0.00	-15.55 ± 0.12	-34.82 ± 0.4	-5.61 ± 0.03
GMSusF	laminaripentaose	7.48 ± 0.21 ^a	1	-9.93 ± 0.08	-15.57 ± 0.26	-5.28 ± 0.02
	laminarihexaose	10.39 ± 0.26 ^b	1	-9.91 ± 0.06	-14.85 ± 0.22	-5.48 ± 0.01
	Laminarin <i>L. digitata</i>	17.10 ± 1.64	1.23 ± 0.03	-7.16 ± 0.11	-4.63 ± 0.41	-5.77 ± 0.06
	Laminarin <i>E. bicyclis</i>	2.72 ± 0.41 ^b	1.90 ± 0.19	-12.32 ± 0.21	-25.61 ± 0.75	-4.69 ± 0.09
	β -1,3;1,4-glucan pustulan	NB NB	NB NB	NB NB	NB NB	NB NB

^a These values were determined by a single experiment. Errors were determined during the data fitting process.

^b The errors given are the S.D. of experiments performed in duplicates. Otherwise, experiments were performed in triplicate.

^c NB, no binding.

Discussion

Bacteroidetes rely on the surface-located proteins of the Sus-like system to recognize polysaccharide substrates in their environment. This system has not been extensively explored in aquatic habitats. For example, we do not find any functional studies on the marine homologs of SusEF-like proteins, which has been studied for gut polysaccharides such as starch (Cameron *et al.*, 2012, 2014; Foley *et al.*, 2018), xyloglucan (Tauzin *et al.*, 2016; Foley *et al.*, 2019) and heparin sulfate (Cartmell *et al.*, 2017). In this study, we determined the laminarin-binding ability of the two proteins, GMSusE and GMSusF, from the marine bacteria *Gramella* sp. MAR_2010_102 of which their function was previously unknown. Characterized proteins despite low sequence identity to other described SusEF-like proteins, show similar carbohydrate-binding function and are encoded at the same genomic position. We propose that GMSusE and GMSusF might belong to the family of SusEF-positioned proteins, which play important roles in glycan binding and the assembly of the outer membrane protein complex.

We have noted that SusEF-positioned genes are annotated as hypothetical proteins or PKD domains in the marine PULs (Unfried *et al.*, 2016; Kappelmann *et al.*, 2019). As the two genes encoded upstream of GMSusD (Figure 2.2.1) did not show enzymatic activity in laminarin degradation, we analyzed the function of the PKD domains in greater detail. PKD domains are usually found in extracellular segments and are involved in protein–protein or protein–carbohydrate interactions across various species. The structure prediction suggested a β -sandwich Ig-like fold, which is a common feature shared with the starch-specific SusF (PDB 4FE9). The N-terminal domain of protein from the starch-specific PUL SusE (PDB 4FEM) was not resolved in the crystal structure, but modeled also as an Ig-like fold (Cameron *et al.*, 2012). Beyond this N-terminal domain, SusE and SusF consist of multiple β -sandwich CBMs. Another characterized carbohydrate-binding protein SGBP-B (PDB 5E7G) encoded in the xyloglucan PUL extracted from *Bacteroides ovatus*, is composed of three tandem Ig-like domains followed by a xyloglucan-binding domain at the C terminus (Tauzin *et al.*, 2016). Although there was no sequence homology of SGBP-B to SusE or SusF, growth experiments on xyloglucan carbohydrates revealed a similar function. Our interpretation of the PKD domain annotations for many SusEF-like genes in marine PULs could be due to a common Ig-like fold. Bioinformatics studies generate a large amount of data, which are crucial for understanding ecologically relevant substrates and how bacteria respond to them, however, some of the proteins show low sequence identity. This could lead to misannotation but simultaneously provides an opportunity to discover novel characteristics of unknown proteins. Neither of the SusEF-positioned genes from *Gramella* sp. MAR_2010_102 showed homology to the characterized carbohydrate-binding proteins from gut *Bacteroidetes*.

However, the respective PHYRE2 models (Sternberg & Kelley, 2009) suggest multimodular architecture composed of an Ig-like fold and potential carbohydrate-binding sites as has been shown for SusEF-like proteins specific for starch (Cameron *et al.*, 2012) or xyloglucan (Tauzin *et al.*, 2016). These shared structural features along with the location in the genomic island let us hypothesize that GMSusE and GMSusF from *Gramella* sp. MAR_2010_102 belong to a relatively unexplored family of SusEF-positioned glycan-binding proteins.

To provide insight into the function of the putative laminarin-binding complex, we investigated the ability of GMSusE and GMSusF to bind polysaccharides. We used laminarins from distinct algae, *E. bicyclis* and *L. digitata*, as these polysaccharides vary in the length of their β -1,3-glucan chain and the amount of β -1,6-glucose branches. Native affinity gel electrophoresis qualitatively revealed binding to both types of laminarin.

To extend the investigation of potential ligands, various polysaccharides were screened using defined carbohydrate microarrays. The polysaccharides immobilized on the microarray membrane were detected by specific binding proteins. GMSusF distinctively recognizes pachyman, which shares the β -1,3-glucan motif with laminarin. A weak signal for β -1,3;1,4-glucan may indicate tolerance of GMSusF for β -1,4-linked glucose residues, though β -1,3-glucan as a contaminant in the β -1,3;1,4-glucan may also explain this result; however, ITC measurements did not show binding of GMSusF to β -1,3;1,4-glucan. Treatment of the arrays with β -1,3-glucanase dramatically reduced the binding confirming the specificity of GMSusF towards epitopes containing β -1,3-linked glucose. Laminarin is likely not attached to the nitrocellulose membrane, as it is a relatively short molecule that would require coupling prior to immobilization (Pedersen *et al.*, 2012). The lack of any GMSusE-binding signal on the array may be due to the tertiary structure of GMSusE. The His-tag needs to be accessible for the anti-his antibody recognition. The terminus containing the His-tag may be sequestered in some manner, and therefore inaccessible for the anti-his antibodies.

Our ITC analysis suggests another explanation for the lack of GMSusE binding of pachyman on the microarray. In the ITC experiment, GMSusE exhibited an affinity for both laminarins tested with a similar affinity range, but showed a preference for the more branched *E. bicyclis* laminarin (Table 1). The amount of β -1,6-glucose branches is much higher in *E. bicyclis* than in *L. digitata* laminarin (Pang *et al.*, 2005), whereas pachyman is a long linear β -1,3-glucan. There have been studies reporting the presence of a few β -1,6-glucose branches in pachyman, but the ratio of (1,3 \rightarrow 1,6)- β is negligible when considering the length of the polymer consisting a degree of polymerization of 255 (Saito *et al.*, 1968). The lack of the GMSusE binding signal for pachyman on the microarray may be due to the low abundance of β -1,6-branch points in this polysaccharide. Interestingly, GMSusF shows the opposite tendency. When the binding abilities of these two proteins are compared, the biggest

difference is in the affinity for *E. bicyclis* laminarin (Figure 2.2.3). GMSusF shows a much lower affinity for branched *E. bicyclis* laminarin and for pachyman on the microarray, which indicates that GMSusF does not need β -1,6-glucose to bind.

Fine structural variations in polysaccharides cause differences in the binding manner of particular SGBPs located on the surface of *Gramella* sp. MAR_2010_102 cells. This relates to the roles of SusEF-like proteins. Members of this family are responsible not only for the sensing and binding polysaccharides from the environment but also for keeping the substrate at the surface during hydrolysis. It is essential to consider the function of these proteins as a cooperative system. They may show slightly different preferences regarding the substrate length and number of branches, because they accommodate distinct parts of the polysaccharide. From a biological perspective, the mentioned variations could also reflect the specialization of *Gramella* sp. MAR_2010_102 towards a particular type of laminarin. The structure of this highly abundant β -1,3-glucan seems to vary between algae species (Gügi *et al.*, 2015). Bacterial adaptation of SGBPs to one type of polysaccharide structure could be a way to overcome competition for the glycan resource. However, this must be further tested, for example with growth experiments using targeted genes disruption. Additionally, one should remember that the binding abilities of particular SGBPs may possibly change depending on protein-protein interactions. This potential protein complex may behave differently when SGBPs interact with each other. In summary, in this report we explore the function of GMSusE and GMSusF proteins extending the knowledge about marine bacterial utilization of an abundant carbon source - laminarin.

Materials and methods

Expression and purification of recombinant proteins

The GMSusE and GMSusF genes from *Gramella* sp. MAR 2010 102 (NCBI Reference Sequence: WP_089661788.1 and WP_089661787.1 respectively) were encoded on the obtained by gene synthesis (GenScript, Piscataway, NJ, USA). Constructs correspond to the amino acid residues 21-372 from GMSusE and 25-307 for GMSusF, both designed to contain His-tags at the N-terminal end. Recombinant proteins were derived from pET28 plasmid, expressed in competent *E. coli* BL21 (DE3) cells (New England Biolabs, New England Biolabs, Ipswich, MA, USA), purified by IMAC, and polished by SEC using standard methodology (Mystkowska *et al.*, 2018).

Affinity gel electrophoresis

Determination of the GMSusF and GMSusE binding towards soluble polysaccharides was performed by affinity gel electrophoresis (AGE). The 12% native acrylamide gels supplemented with 0.5% laminarin from *E. bicyclis* (Carbosynth, Compton, UK), laminarin *L. digitata* (Sigma-Aldrich, Saint Louis, MO, USA), galactomannan (Megazyme, Wicklow, Leinster, Ireland) (negative control) and a gel without the polysaccharide were prepared. Ten micrograms of GMSusE, GMSusF, and BSA onto each gel and electrophoresis was carried out for 2 h at a constant voltage of 80 V on ice. The proteins were visualized by staining with Coomassie Blue.

Carbohydrate microarray analysis

Defined microarrays were prepared and analyzed as previously described (Mystkowska *et al.*, 2018). In parallel to regular microarray screening, we performed epitope deletion experiment, where the targets for the binding proteins were removed by enzymatic treatment. The microarrays with printed polysaccharides were blocked for 1 h in PBS (140 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.7 mM KH₂PO₄, pH 7.5) with 5% (w/v) low fat milk powder (MPBS). Then, they were incubated for 2h at 37 °C with endo- β -1,3-glucanase, endo- β -1,3;1,4-glucanase (1 U mL⁻¹) (Megazyme, Wicklow, Leinster, Ireland) or with buffer to keep all conditions identical. Afterward, the arrays were incubated for 2 h with probes: His-tagged proteins produced in the *E. coli* expression system: GMSusE, GMSusF (10 ug/ml); and monoclonal antibodies: BS-4002 (1:1000) BS-4003 (1:1000) all diluted in MPBS. Next, microarrays were washed in PBS and incubated for 2 h with anti-His tag or anti-mouse secondary antibodies conjugated to alkaline phosphatase (Sigma-Aldrich) diluted 1:1500 (for His tag) or 1:5000 (anti-rat and anti-mouse) in MPBS. After washing in PBS and deionized

water, microarrays were developed in a solution containing 5-bromo-4-chloro-3 indolyl phosphate and nitro blue tetrazolium in alkaline phosphatase buffer (100 mM NaCl, 5 mM MgCl₂, 100 mM diethanolamine, pH 9.5). Once developed, the arrays were scanned using pertinent software (Array-Pro Analyzer 6.3, Media Cybernetics, Rockville, Maryland, United States). The highest mean spot signal intensity was set to 100, the rest of values normalized consequently. Data are presented in a heat map form where color intensity is correlated to mean spot signal value.

Isothermal titration calorimetry

The proteins were extensively dialyzed against PBS buffer pH 8, which was kept for dilution of polysaccharides: laminarin *E. bicyclis*, laminarin *L. digitata*, pustulan (Elicityl, Crolles, Auvergne-Rhone-Alpes, France) and β -1,3;1,4-glucan; or oligosaccharides: laminaripentaose and laminarihexaose (Megazyme). ITC of GMSusF, GMSusE, oligosaccharides, and polysaccharides were conducted in a VP-ITC titration calorimeter (MicroCal, Northampton, MA, USA) at 25 °C. All solutions of protein and ligand were filtered and degassed immediately prior to the ITC experiment. Protein concentrations between 0.28 and 0.5 mM and oligosaccharides were at 15 mM. Due to the ambiguous molecular weight of laminarin, concentrations of laminarin from *L. digitata* and *E. bicyclis* were calculated based on equivalents of β -1,3-glucodecaose, L10. Titrations were performed by titrating oligosaccharide or polysaccharide into protein, Oligosaccharide titrations were analyzed using a standard one-site binding model. However, because of the unknown number of binding sites present on the laminarin preparations, the laminarin titrations were analyzed using a one site binding model but where the protein present in the sample cell was treated as the ligand and the oligosaccharides or polysaccharides present in the syringe were treated as the macromolecule. All data show the average and S.D. of three independent titrations unless otherwise indicated in Table 1.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) grant HE 7217/1-1 to Jan-Hendrik Hehemann and by the Max Planck Society and by a Natural Sciences and Engineering Research Council of Canada Discovery Grant to Alisdair Boraston (FRN_2014-04355). Matthias Höhne was supported by DFG grant HO 4754/5-1 and we acknowledge support by the DFG Forschergruppe (FOR 2406/1). We would like to thank GLOMAR graduate school for supporting the research stay of Agata Mystkowska at University of Victoria, Canada, where the experiments contributing to this paper were performed.

Manuscript 3:

**Exploring the potential of bacterial carbohydrate-binding proteins
as novel probes to detect marine polysaccharides**

Title: Exploring the potential of bacterial carbohydrate-binding proteins as novel probes to detect marine polysaccharides

Agata Anna Mystkowska^{1,2,3}, Silvia Vidal-Melgosa^{1,2}, Jan-Hendrik Hehemann^{1,2*}

¹University of Bremen, Center for Marine Environmental Sciences, (MARUM), Bremen, 28359, Germany.

²Max Planck Institute for Marine Microbiology, Bremen, 28359, Germany.

³Institute of Biochemistry, Greifswald University, Greifswald 17487, Germany

*To whom correspondence should be addressed:

Jan-Hendrik Hehemann, MARUM MPG Joint Research Group Marine Glycobiology, University of Bremen, Center for Marine Environmental Sciences, (MARUM), Bremen, 28359, Germany; Max Planck Institute for Marine Microbiology, Bremen, 28359, Germany, Phone: +49 421 218-65775, Email: jhehemann@marum.de, jheheman@mpi-bremen.de

Keywords: glycan probes, SusE, SusF, surface-glycan binding proteins, *Bacteroidetes*, marine polysaccharides,

Abstract

Microalgae are responsible for about half of the global primary production. Although polysaccharides are one of the major products of photosynthesis, their structures remain highly unexplored. A suitable method to investigate polysaccharides is the application of molecular probes, which represent highly specific proteins distinctive for recognizing a particular epitope. To detect polysaccharides, molecular probes in the form of antibodies and carbohydrate-binding proteins have been developed, but microalgae-specific probes are rather scarce, while most current probes bind to land plant polysaccharides. Here, we propose to exploit the vast diversity of carbohydrate-binding proteins available in nature with hundreds of thousands of marine microorganisms that utilize microalgal carbohydrates. Subsequent “omics” studies during diatom blooms at the North Sea, close to the island of Helgoland, showed that putative proteins associated with polysaccharide uptake were highly expressed. Among them, SusD-like proteins, which are specific carbohydrate-binding proteins located on the bacterial surface in members of the *Bacteroidetes* phylum, were detected. As their proposed biological function is to bind polysaccharides derived from the same niche, we decided to investigate their potential as microalgal specific probes. A protein library was created starting with forty-seven constructs expressed in *Escherichia coli*, from which we successfully obtained twelve soluble recombinant proteins. Their binding activity was evaluated with affinity gel electrophoresis and carbohydrate microarrays. The arrays included well-characterized polysaccharides but also polysaccharide extracts from algal lab cultures and environmental samples. We found four novel carbohydrate-binding proteins: SusD 1 binding towards α -mannan, SusD 19 to β -mannan, SusD 46 and CBM 1 both recognizing laminarin. However, most of the produced SusD-like proteins were not mono-specific as they bound negatively charged polysaccharides independently of the proteins predicted substrate. In conclusion, we propose testing further marine carbohydrate-binding proteins considering that negatively charged polysaccharides might create nonspecific binding for the recombinantly expressed SusD-like proteins.

Keywords: SusD, carbohydrate-binding proteins, carbohydrate microarrays, marine polysaccharides, carbohydrate-specific probes.

Introduction

Bacteria utilization of polysaccharides in the ocean has a great impact on the global carbon cycle. The phylum *Bacteroidetes* is one of the main marine polysaccharide-degrading bacterial groups (Teeling *et al.*, 2012; Buchan *et al.*, 2014). Their genome encodes Polysaccharide utilization locus (PULs) including carbohydrate-active enzymes (CAZymes), carbohydrate-binding proteins (CBPs) and transporters allowing for efficient acquisition of polysaccharides from the environment (Kappelman *et al.*, 2019). Interestingly, it was shown that these genes were upregulated in response to a microalgae bloom at the North Sea, at Helgoland, Germany, when high amounts of microalgal polysaccharides would be present (Teeling *et al.*, 2012). Aside from the CAZymes and transporters, a pronounced abundance of SusD-like proteins was observed (Teeling *et al.*, 2012). SusD-like proteins are main carbohydrate-binding proteins in the so-called starch utilization system (Sus), initially discovered for starch utilization (Koropatkin *et al.*, 2008). A recent model of the system proposes that the SusD-like protein interacts with the SusC-like transporter, where SusD is located on the top of SusC and acts as a lid pushing the substrate through the SusC. In the open form of the SusCD complex, the binding site of SusD is exposed to the outside of the bacterial cell (Glenwright *et al.*, 2017). The Sus-like system was discovered and characterized for *Bacteroidetes* present in the gut microbiome, but this system seems to be conserved as it is encoded within bacteria from diverse environments (Bjursell *et al.*, 2006) such as marine (Kappelman *et al.*, 2019) and soil (Xie *et al.*, 2007). We currently lack accurate knowledge about Sus-like system from marine bacteria, particularly about CBPs. Studying these proteins is of great interest not only because of their biological relevance, but also due to their potential as tools for glycobiology. Additionally, bacteria possess another group of CBPs named carbohydrate-binding modules (CBMs), which are non-catalytic domains of glycoside hydrolases (GH) (Boraston *et al.*, 2004). Investigating and characterizing the specificities of CBMs from marine bacteria could improve our existing collection of algal-specific probes.

In contrast to the plant cell wall, an accurate characterization of algal polysaccharides structure is not established (Popper *et al.*, 2011). Regarding the known algal polysaccharides, some of the structures are present in both algal- and plant-derived polysaccharides, for example cellulose (Painter, 1983). However, most of them differ substantially in the linkage type between particular monomers and the modifications, for instance marine polysaccharides are highly sulfated (Popper *et al.*, 2011). The current collection of probes specific for marine polysaccharides is limited. Therefore, there is a need for the development of probes specific towards marine polysaccharides to aid the identification and characterization of algal glycan structures. Salmeán *et al.* proposed the double-blind microarray-based polysaccharide profiling, a strategy to explore simultaneously

uncharacterized polysaccharides in biological samples and carbohydrate-binding proteins with unknown specificities. They found four CBPs specific towards carrageenans, arabinoxylan, and xyloglucan. Interestingly, the presence of arabinoxylan and xyloglucan in macroalgal cell wall had not been previously observed. Additionally, they reported carrageenan specific probes for the first time (Salmeán *et al.*, 2018). Bennke *et al.* suggested another approach, where they used lectins to map glycoconjugates involved in the specific interactions of *Bacteroidetes* and diatoms sampled from the North Sea, Helgoland. The study revealed binding of fucose-, galactose-, and mannose-specific lectins towards diatom exudates. However, only 12 out of 77 tested lectins resulted in a binding signal, where 5 lectins were retrieved from the marine environment while the others were isolated from plants. This emphasizes the necessity for more marine specific probes (Bennke *et al.*, 2013). Monoclonal antibodies are challenging to obtain as polysaccharides show low immunogenicity and their isolation in the required amounts is difficult (Rydahl *et al.*, 2017). To our knowledge, probes for marine polysaccharides in the form of monoclonal antibodies have been developed and characterized for alginate (Torode *et al.*, 2016), fucoïdan (Torode *et al.*, 2015), and ulvan (Rydahl *et al.*, 2017). The current research on the development of algal-polysaccharide probes focuses mainly on macroalgae rather than on single-celled algae - microalgae. The detailed investigation of SusD-like proteins and their target substrates, proposed in this study, may contribute to the current collection of probes specific for marine polysaccharides.

In this research, we prepared a protein library using the advantages of genetic engineering for relatively rapid production of potential probes in the form of recombinant proteins. We tested the expression of forty-seven CBPs constructs in *E. coli*; twelve of these were soluble and further investigated with binding assays. The binding of twelve proteins towards plant polysaccharides and algal-derived extracts was analyzed. Furthermore, carbohydrate microarrays with polysaccharide extracts from environmental samples of Helgoland were examined. Our screening provided novel insights into how CBPs interact with marine polysaccharides.

Results

Library of recombinant CBPs

We selected proteins, which were present in the metagenomes or showed high expression in metaproteomes of bacteria during spring algal blooms at the North Sea (54°11.3'N, 7°54.0'E) near the island of Helgoland, Germany. Sequences from relatively evenly separated, distant branches of a maximum likelihood phylogenetic tree (data not shown) were selected for in-depth analysis. They were from varied marine *Bacteroidetes* targeting different polysaccharides. The library also included one CBM (Table 2.3.1). Predicted polysaccharide targets were determined based on the CAZymes encoded in the PULs, which SusD-like proteins belonged to. Specificity of CBM was predicted based on the activity of connected GH (Becker *et al.*, 2017).

A total of forty-seven sequences were expressed in the *Escherichia coli* BL21. All proteins were designed to contain a double hexahistidine (his) tag. Initially, protein expression was performed on a small-scale in deep-well blocks to screen high soluble expression (Studier, 2005). Both fractions of lysates were directly loaded on SDS-PAGE gels: supernatant representing soluble fractions and pellet with insoluble proteins resuspended in urea. We observed high soluble expression for twelve proteins, while twelve constructs were expressed but formed inclusion bodies (Figure 2.3.1AB, Table 2.3.1). We scaled-up the expression volume for the constructs showing soluble protein production. A combination of metal affinity and size exclusion chromatography resulted in the library of ten SusD-like proteins and one CBM from bacterioplankton (Figure 2.3.1B). Based on the bioinformatic analyses of the genomic context of recombinantly produced proteins, we predicted the following substrates: α -mannan (SusD 1), β -mannan (SusD 6, SusD 14, SusD 16, SusD 19, SusD 20, SusD 26, SusD 42), xylan (SusD 9), fucose-containing polysaccharide (SusD 25) and laminarin (SusD 46, CBM 1). The library of proteins was created, where aliquots at 1 mg ml⁻¹ concentrations were flash-frozen and kept in -80°C. These proteins were used for the binding activity experiments.

Table 2.3.1. List of the genes encoding SusD-like proteins and CBM used in the present study. The constructs resulting in soluble protein expression are indicated. NCBI reference number for sequences extracted from metagenomes indicate the most similar sequence and are marked with the asterisk (*).

Name	Bacteria	Expression	Reference Number
SusD 1	<i>Salegentibacter Hel 1 6</i>	soluble	WP_037317849.1
SusD 2	<i>Maribacter</i> sp. HTCC2170		EAR00462.1*
SusD 3	<i>Maribacter</i> MAR 2009 72		WP_089259059.1
SusD 4	<i>Maribacter</i> Hel1 7		WP_027068111.1
SusD 5	<i>Gramella forsetii</i> KT0803		WP_011709047.1
SusD 6	<i>Leeuwenhoekiella</i> MAR 2009 132	soluble	WP_052108500.1
SusD 7	<i>Polaribacter</i> Hel1 85 G22 134		WP_052108500.1
SusD 8	<i>Maribacter</i> MAR 2009 72		WP_100869917.1
SusD 9	<i>Zobellia uliginosa</i> MAR 2009 138	soluble	WP_038231891.1
SusD 10	<i>Maribacter</i> sp._HTCC2170		WP_013305629.1 *
SusD 11	<i>Leeuwenhoekiella</i> MAR 2009 132		WP_031427507.1
SusD 12	<i>Flavobacteriaceae bacterium</i> PRS1		WP_122514876.1 *
SusD 13	<i>Gramella forsetii</i> KT0803		WP_011709590.1
SusD 14	<i>Muricauda</i> MAR 2010 75	soluble	WP_036379614.1
SusD 15	<i>Flavobacteriaceae bacterium</i>		PCH76841.1*
SusD 16	<i>Arenibacter palladensis</i> MAR 2009 79	soluble	WP_084532430.1
SusD 17	<i>Polaribacter tangerinus</i>		WP_088323384.1*
SusD 18	<i>Salegentibacter</i> Hel1 6		WP_081912658.1
SusD 19	<i>Leeuwenhoekiella</i> Hel1 48	soluble	WP_028377317.1
SusD 20	<i>Salegentibacter</i> Hel1 6	soluble	WP_051935792.1
SusD 21	<i>Leeuwenhoekiella blandensis</i>		WP_038231962.1 *
SusD 22	<i>Zobellia uliginosa</i> MAR 2009 138		WP_038238410.1
SusD 23	<i>Zobellia uliginosa</i> MAR 2009 138		WP_038238410.1
SusD 24	<i>Flavobacteriaceae bacterium</i> TMED145		OUV85246.1*
SusD 25	<i>Gramella forsetii</i> KT0803	soluble	WP_011709610.1
SusD 26	<i>Flavimarina</i> sp. Hel_I_48	soluble	WP_031428275.1
SusD 27	<i>Zobellia uliginosa</i> MAR 2009 138		WP_081889078.1
SusD 28	<i>Polaribacter</i> sp. MED152		WP_015480796.1
SusD 29	<i>Maribacter dokdonensis</i>		WP_083365713.1
SusD 30	<i>Polaribacter irgensii</i>		WP_004569336.1 *

SusD 31	<i>Maribacter</i> Hel1 7		WP_027064618.1 *
SusD 32	<i>Maribacter</i> sp. HTCC2170		WP_013305629.1 *
SusD 33	<i>Polaribacter irgensii</i>		EAR12146.1 *
SusD 34	<i>Sediminibacter</i> Hel1 10		WP_026753464.1
SusD 35	<i>Sediminibacter</i> Hel1 10		WP_026754909.1
SusD 36	<i>Zobellia uliginosa</i>		WP_038231970.1
SusD 37	<i>Sediminibacter</i> Hel1 10		WP_081820764.1
SusD 38	<i>Gramella forsetii</i>		WP_069676382.1 *
SusD 39	<i>Polaribacter irgensii</i>		WP_018942494.1 *
SusD 40	<i>Polaribacter irgensii</i>		WP_004568757.1 *
SusD 41	<i>Polaribacter</i> sp. KT25b		WP_091892599.1 *
SusD 42	<i>Polaribacter</i> Hel1 33 49	soluble	WP_036785225.1
SusD 43	<i>Dokdonia donghaensis</i>		WP_035326951.1 *
SusD 44	<i>Maribacter sedimenticola</i>		WP_089259059.1
SusD 45	<i>Salegentibacter</i> Hel1 6		WP_037316924.1
SusD 46	<i>Gramella</i> sp. MAR_2010_102	soluble	WP_089661789.1
CBM 1	<i>Formosa agariphila</i>	soluble	WP_038529859.1

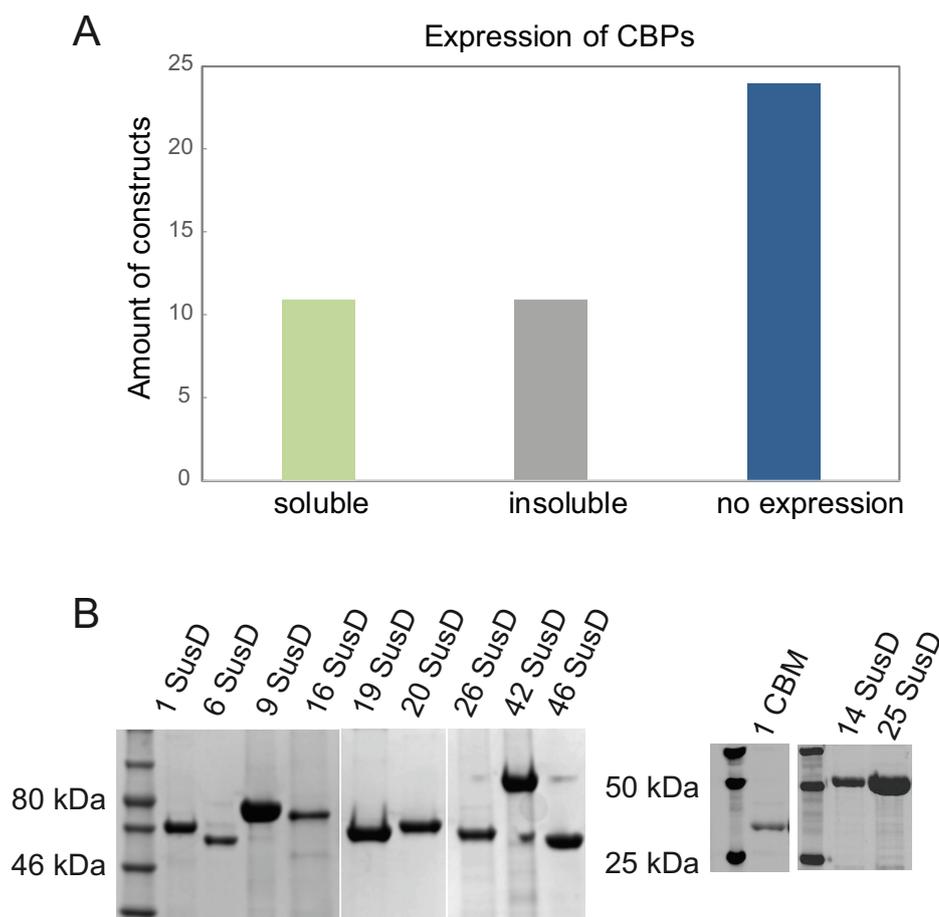


Figure 2.3.1. Production of CBPs as recombinant proteins in *E. coli* expression system. **A:** Plot representing the amount of proteins expressed as soluble (green), insoluble (grey) and constructs, for which did not observe expression of recombinant proteins (blue). **B:** SDS-PAGE of purified SusD-like proteins and CBM expressed in *E. coli* as soluble proteins. Three gels are shown. All soluble proteins except: 1 CBM, 14 SusD and 25 SusD were loaded on one gel, but the unimportant lanes were intentionally omitted, as indicated by vertical lines. A protein marker is shown in the first well from the left of each gel.

Binding ability analyses and optimization of the protocol

To explore the carbohydrate-binding ability of our recombinantly expressed and purified proteins we used affinity gel electrophoresis. The following polysaccharides were tested: galactomannan, glucomannan, α -mannan, and laminarin. We observed retention on the affinity gel for some of the proteins. Our results indicated that SusD 1 is specific towards α -mannan, SusD 19 recognizes galacto- and glucomannan, while SusD 46 and CBM 1 bind to laminarin (Figure 2.3.2). The substrates correlate with the genomic context of the binding protein i.e. the putative target polysaccharides.

Further, we tested SusD-like proteins with carbohydrate microarrays, which allows for high-throughput analysis of multiple polysaccharides. Our study included defined arrays and comprehensive microarray polymer profiling (CoMPP) (Moller *et al.*, 2007). By defined

arrays, we screened the binding activity of our proteins towards polysaccharides of defined structures. A total of twenty-five polysaccharides were printed on microarrays and then individual arrays were incubated with each of our proteins. Binding of the probes to particular polysaccharides was detected by using an anti-his secondary antibody. Our results showed that most of the proteins tend to bind carrageenan and fucoidan except for SusD 19 and SusD 20 that bound only to fucoidans (Figure 2.3.3A). As the availability of pure marine polysaccharides is limited, especially for microalgae polysaccharides, we extended our analysis with CoMPP arrays. Polysaccharides were sequentially extracted (using three different solvents) from our microalgae lab cultures as well as from red and brown macroalgae. The extracts were printed onto microarrays and we tested the binding of our proteins to the extracted polysaccharides. The most intense protein binding signal was observed towards polysaccharides extracted with water from brown and red algae (Figure 2.3.3B). While most of the tested proteins bound to the extracted polymers of brown and red algae, only binding to red algae extracts was detected for SusD 1 and SusD 42. SusD 9 was the only protein that bound to microalgae extracts in addition to the red and brown algae extracts. SusD 9 signal was not detected in microalgae H₂O extracts but in all EDTA extracts of *Thalassiosira weissflogii*, *Emiliania huxleyi* and *T. pseudonana*. It suggests that this protein bound negatively charged polysaccharides that were extracted with this chelating agent.

Additionally, we used another set of CoMPP arrays that contained samples obtained during a spring diatom bloom in the North Sea (near Helgoland) for our analysis. The array contained material including both particulate organic matter (POM) and high molecular weight-dissolved organic matter (HMW-DOM). Polysaccharides were extracted from POM and HMW-DOM biomass also by sequential extraction using three solvents and all extracts were printed on microarrays. The CoMPP arrays, populated with all the extracts, were incubated with our proteins and the detected binding is presented in figure 2.3.4. Only protein SusD 9 showed binding towards polysaccharides extracted from the Helgoland bloom, both in POM and in HMW-DOM (Figure 2.3.4). The highest intensity was detected in HMW-DOM samples. However, POM and DOM samples were obtained by different technical approaches and therefore the probe signal (corresponding to polysaccharide relative abundance) should be considered separately by POM and by DOM. SusD 9 signal intensity showed fluctuations during time. There was a slight increase in signal by time in the 3 μ m POM and intensity peaks were observed at the beginning of the bloom for HMW-DOM extracts. However, as the results from the defined array indicated this SusD bound to two of our defined polysaccharides (galactofucan and carrageenan), it is not possible to conclude if the detected signal corresponds to binding to one polysaccharide or to more than one structure present during the bloom.

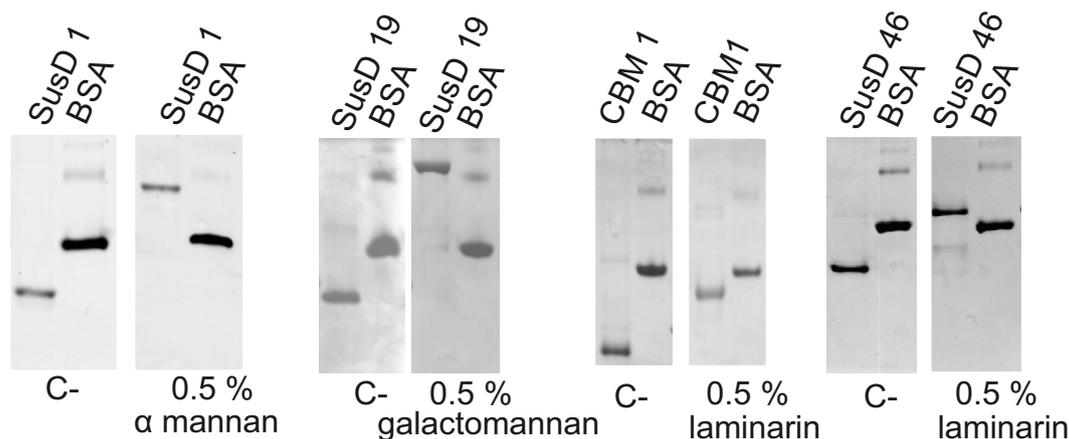


Figure 2.3.2. Affinity gel electrophoresis binding analyses. SusD-like proteins (SusD 1, SusD 19, SusD 46) and CBM 1 were analyzed in the native conditions on the gel containing no polysaccharide (C-, negative control) and on gels with 0.5% (w/v) polysaccharides: α -mannan, galactomannan or laminarin. Bovine Serum Albumin (BSA) was used as a non-binding control.

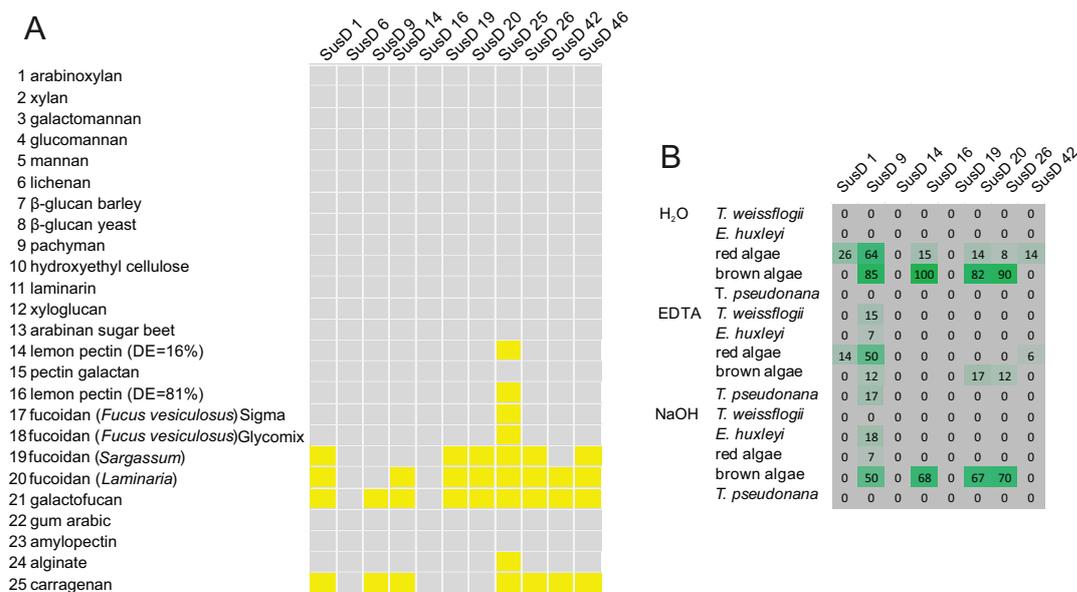


Figure 2.3.3. Microarray analysis of defined and extracted polysaccharides. Heat maps showing the results of binding analyzing of SusD-like proteins to different polysaccharides using defined arrays **(A)** and CoMPP **(B)**. Polysaccharides or algal extracts (listed at the left) were printed on microarrays and each single array was probed with a particular SusD-like protein (listed at the top). For the defined arrays yellow indicates binding **(A)**, whereas for the CoMPP arrays binding is presented in green **(B)**. Gray color represents no binding detected. Values on the CoMPP heat map correspond to the mean binding signal intensities derived from the triplicates, where the highest value was set to 100 and all other values were normalized accordingly. DE, degree of esterification.

To reduce non-specific binding on the microarrays we tried multiple optimization protocols. We adjusted the salt concentration, performed all steps in the cold room and extended the blocking with milk powder PBS, but none of the modifications changed the result. Considering that only the binding toward negatively charged polysaccharides had been observed, we also tested various protocols to capture the signal to the predicted substrate. For instance, we performed the blocking step with BSA instead of milk powder PBS (O’Riordan *et al.*, 2014), but no changes in the binding signals were observed.

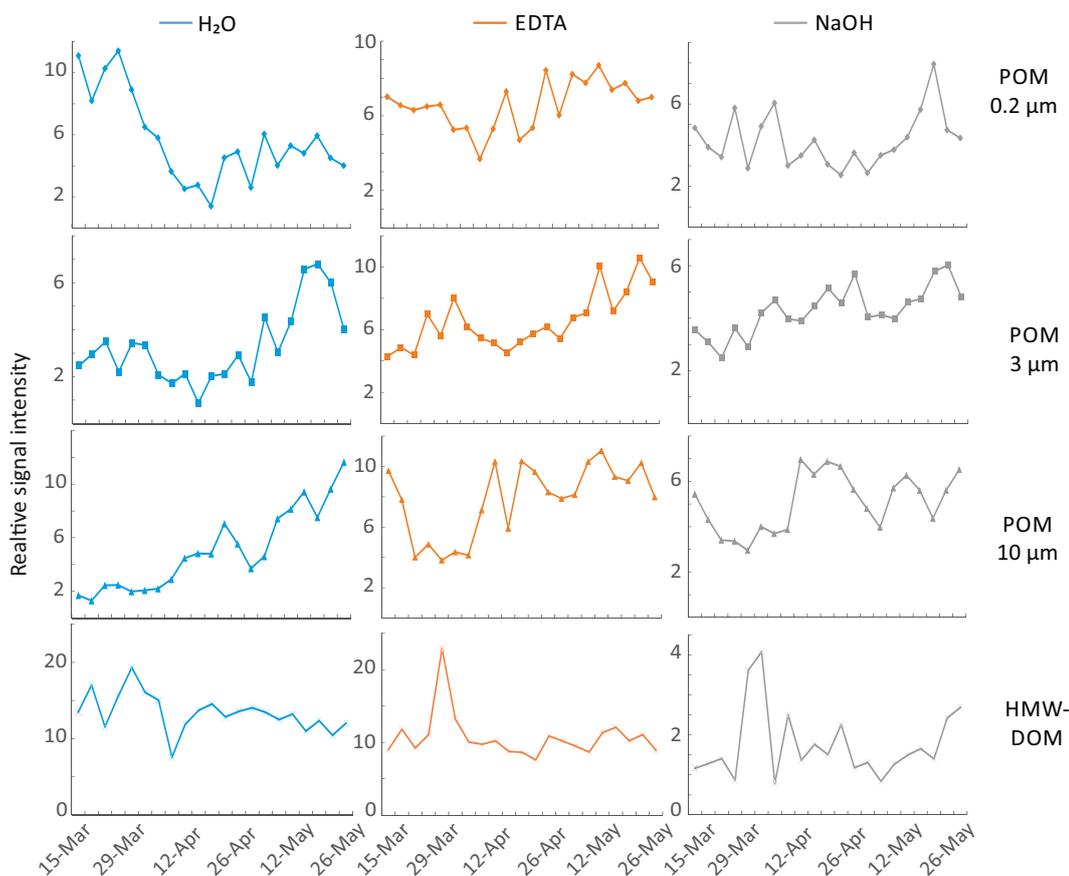


Figure 2.3.4. Intensity of the microarray signal of SusD 9 towards extracts of POM and HMW-DOM from the North Sea, Helgoland. Plots showing the SusD 9 signal detected with microarray analysis for the H₂O, EDTA and NaOH extracts of the samples collected during the Helgoland spring bloom in 2016. POM samples represent material from the sequential filtration through 10, 3 and 0.2 μm filters, whereas the flow-through was concentrated and presented as HMW-DOM samples.

Discussion

In the examination of novel probes, we focused on the marine SusD-like proteins extracted from *Bacteroidetes* present during diatom blooms in the North Sea, Helgoland. We targeted this group of proteins for multiple reasons. The metagenomic and metaproteomic data from the Helgoland sampling campaign showed that SusD-like proteins are among the highest expressed proteins occurring in the response to diatom blooms. This allows us to confidently hypothesize that these proteins specifically target polysaccharide originated from the same ecological niche. Numerous modifications, such as sulfation, may challenge the recognition of marine polysaccharides by most of the current available probes - which are plant-related. Marine *Bacteroidetes* evolved to use algal-derived polysaccharides as a carbon source (Popper *et al.*, 2011) and their main carbohydrate-binding SusD-like proteins most probably recognize those marine substrates. Characterization of the marine SusD-like proteins could help not only to develop novel probes, but furthermore to get information about new polysaccharides. To achieve that, we could combine microarray analyses of extracted polysaccharides with other techniques, for example monosaccharide analysis and epitope detection chromatography (Cornuault *et al.*, 2014). Another reason to examine SusD-like proteins as probes is the way this group of proteins possibly bind their substrate. It is known that the canonical starch specific SusD recognizes the helical structure of starch rather than its monomeric composition. Specificity towards the shape of polymer is a valuable feature from the perspective of developing new probes. Finally, the substrate concentration in the ocean is probably lower than in the gut environment. A recent calculation shows that laminarin, one of the most abundant marine polysaccharides, is present in the ocean at a concentration in range of mg L⁻¹ (Becker, 2018) whereas for example concentration of starch in the guts is probably higher. It brought us to the assumption that marine SusD-like proteins might present higher affinity toward their substrates comparing to the SusD-like proteins from other environments. Therefore, we investigated so far unknown SusD-like proteins in the context of specific polysaccharide-binding.

The selected collection of SusD-like proteins predicted to be specific for different substrates was expressed as recombinant proteins and tested with binding assays. Analysis by affinity gel electrophoresis showed binding of four proteins: SusD 1 to α -mannan, SusD 19 to β -mannan, SusD 46 and CBM 1 towards laminarin (Figure 2.3.2). However, these binding specificities were not consistent when the same proteins were analyzed by defined microarrays (Figure 2.3.3A). As α -mannan was not printed on the array we did not predict binding of SusD 1. Unexpectedly, SusD 19 did not show binding on β -mannans from the microarray even though the gluco- and galactomannan were included in the analysis. We

hypothesize that lack of laminarin binding by SusD 46 in the array is because this substrate was not immobilized onto the printing surface - nitrocellulose membrane. Polymers of short length, like oligosaccharides, have to be coupled to proteins to get immobilized on the membrane (Pedersen *et al.*, 2012). This is most probably the case for laminarin, which usually have a molecular weight 4–5 kDa (Menshova *et al.*, 2014). This assumption is in agreement with positive controls that were included in other analysis, such as BS 4002 antibodies binding β -1,3-glucans (Meikle *et al.*, 1991), which did not show binding towards laminarin. Nonetheless, microarray analysis revealed that nine out of eleven tested SusD-like proteins displayed binding towards different types of fucoidan and seven out of eleven to both fucoidan and carrageenan. SusD 25 additionally recognized pectin and alginate (Figure 2.3.3A). Therefore, a majority of the examined SusD-like proteins were not monospecific. Bioinformatic predictions of GHs encoded in the vicinity of each selected SusD-like protein do not indicate binding of any of these substrates, except for SusD 25 that is encoded in a fucose-containing polysaccharide PUL.

The common feature of all polysaccharides with which binding was detected is the presence of either sulfate or carboxyl groups, which makes them negatively charged. Protocol optimization including various blocking solution and temperature, did not change the result. One of our hypotheses was that SusD-like proteins targeted the casein present in the blocking solution. Casein is a protein commonly found in milk that undergoes multiple post-translational modifications resulting in a highly glycosylated form (O’Riordan *et al.*, 2014). However, an experiment where blocking was performed with BSA, did not show any additional binding results. Non-specific binding excludes usability of selected proteins as glycan probes.

In parallel, binding analysis with arrays populated with polysaccharide extracts from lab cultures and environmental samples were performed. As the initial test with defined arrays revealed that the majority of the evaluated probes were not monospecific, we cannot conclude which exact polysaccharide structure were SusD-like proteins bind to in the samples containing mixtures of polysaccharides. For the arrays containing microalgae and macroalgae polysaccharide extracts (Figure 2.3.3B) the most intensive signal detected was for polysaccharides extracted with water from brown and red algae. It is established that macroalgae cell wall is rich in anionic polysaccharides such as carrageenan and agar (Anderson *et al.*, 1965). We expected the highest concentration of negatively charged polysaccharides not in the H₂O but in the EDTA extracted polysaccharides. However, previous experiments in our lab showed that fucose-rich polysaccharide is highly present in algae H₂O extracts. Signal intensity of SusD 9 towards extracts of POM and HMW-DOM from the North Sea, Helgoland, was plotted for the following days of the microalgae bloom (Figure 2.3.4). Our

study revealed the binding of four novel CBPs, but it also emphasizes that some of the analyzed SusD proteins bound several negatively-charged polysaccharides.

The results showing that the majority of the analyzed SusD-like proteins binds to negatively charged polysaccharides and does not display specificity towards the predicted substrates, may be explained by the fact that exploited SusD-like proteins represent relatively low binding affinity. Another interpretation could be that we do not have a desired substrate for the tested proteins, but it seems less likely. At the time when experiments for this study were planned and performed, the SusCD complex structure was not known. As mentioned before, Glenwright *et al.* showed that SusD works like a lid, which pushes the polysaccharides through the SusC channel (Glenwright *et al.*, 2017). In the light of proposed mechanism for SusCD cooperation, high affinity is probably not needed for the SusD functions. Interestingly, the growth of model *Bacteroides thetaiotaomicron* and its mutants on longer polymers of starch indicates that SusD presence is more important than its binding ability. SusD is also involved in the sensing and activation of sus operon (Cameron *et al.*, 2014). Growing research on this group of proteins shows that apart from the binding, their primary role is to work together with other surface-glycan binding proteins of the outer membrane complex (Tauzin *et al.*, 2016; Foley *et al.*, 2018). For the so far characterised SusD-like proteins it is known that they have relatively weak affinity (Koropatkin *et al.*, 2008) and contrary to our initial assumption, it is also the case for the tested marine SusD-like proteins.

Application of recombinantly expressed CBPs as probes has a great potential to extend the field of marine glycobiology, but probably other proteins than the ones investigated in this study should be targeted, for example CBMs. The selected SusD-like proteins either had too low affinity towards the substrates predicted based on the genomic PUL composition or we did not have the correct ligands. Furthermore, our data showed that most of them may bind unspecifically to negatively charged polysaccharides. Thus, in the next study we decided to characterize more in detail SusD 46, called GMSusD, specific to laminarin (Mystkowska *et al.*, 2018). We targeted this protein because its substrate - laminarin is a neutral polysaccharide (Painter, 1983), which is one of the best characterised in the marine environment. SusD-like proteins from the ocean are an interesting subject to study not only for their application potential, but also to better understand the protein-carbohydrate interactions that are essential for the polysaccharide utilization and therefore, aids our knowledge of the marine carbon cycle.

Material and Methods:

Design and transformation of recombinant plasmids

The recombinant plasmids pET28 (a+) containing forty-seven SusD-like proteins and one CBM (Table 1) were obtained by gene synthesis (GenScript). For each protein, signal peptide was excluded from the final construct. Competent *E. coli* BL21 (DE3) cells (New England Biolabs Ipswich, Massachusetts, United States) were transformed with these plasmids. The recombinant proteins were designed to contain double his-tags at the N- and C-terminal protein ends. Sequences of particular constructs are showed in the table 2.3.1.

Small-scale heterologous protein expression

Initially, protein expression screening was perform using deep well blocks. Two ml of autoinduction medium with 150 µg/ml kanamycin were inoculated with a single bacterial colony and incubated at 20 °C for four days (Studier, 2005). Cells were lysed in Bugbuster detergent (Novagen, Madison, WI) in 50 mM Tris-HCl, pH 8.0. Cell debris were separated by centrifugation at 2500 g for 30 min at 4 °C. The supernatants consisted of the soluble expressed fractions. A pellet containing insoluble fractions was resuspended in 6 M urea and used for the inclusion bodies analysis. All fractions were analyzed on SDS-PAGE.

Expression and purification of selected probes

Based on the small-scale expression, we selected 11 proteins representing the soluble expression. 0.5 ml of the preculture was used to inoculate 2 L of ZYP-5052 expression medium containing 150 µg ml⁻¹ kanamycin (Studier, 2005). Cells were growing in autoinduction medium at 20°C with shaking (150 rpm) for 4 days and then cells were harvested by centrifugation (45 min, 4500 g, 4 °C). Cells were suspended in 25 ml of sucrose solution [(25% w/v), 50 mM Tris-HCl, pH 8.0]. Ten milligrams of lysozyme were added followed by mixing for 15 min with a magnetic stir bar at 20 °C. 50 ml of a deoxycholate solution [1% deoxycholate (w/v), 1% Triton X-100 (v/v), 20 mM Tris-HCl, pH 7.5, 100 mM NaCl] was added followed by stirring for 10 min. One ml of 1 M MgCl₂ and DNase were added and stirred for 10 minutes. Cell debris was removed by centrifugation at 16,000 g for 45 min at 4°C.

Purification by Immobilized Metal Ion Affinity Chromatography (IMAC) and Size Exclusion Chromatography (SEC)

The selected recombinant SusD-like proteins and CBM were purified using a 5 mL Hi-Trap metal affinity column (GE Healthcare, Chicago, Illinois, United States). For each of the proteins, the cell lysate was loaded in buffer (20 mM Tris, 500 mM NaCl, pH 8.0) and was eluted with an imidazole gradient (15-500 mM). Fractions were analyzed by the SDS-PAGE, those containing the protein of interest were collected and concentrated in a stirred ultrafiltration unit (Amicon, Sigma-Aldrich, St. Louis, Missouri, United States). The concentration value was determined by the absorbance at 280 nm using the extinction coefficient for the specific proteins. Size exclusion chromatography – SEC, with a 120 ml HiPrep 16/60 Sephacryl S-200 HR column, elution was performed with 20 mM Tris-HCl, 150 mM NaCl pH 8, flow rate 0.5 ml min⁻¹. The recombinant proteins after SEC purification were concentrated again in a stirred ultrafiltration unit (Amicon, Sigma-Aldrich). The concentrated protein was centrifuged for 20 min at 12000 g in order to remove aggregates.

Polysaccharide substrates

The following polysaccharides were used for binding studies: arabinoxylan, xylan, galactomannan, glucomannan, mannan, lichenan, β -glucan barley, β -glucan yeast, pachyman, hydroxyethyl cellulose, laminarin (*L. digitata*), xyloglucan, arabinan (sugar beet), lemon pectin (DE=16%), pectin galactan, lemon pectin (DE=81%), fucoidan (*Fucus vesiculosus*), fucoidan (*Sargassum*), fucoidan (*Laminaria*), galactofucan, gum arabic, amylopectin, alginate and carrageenan. Polysaccharides were purchased from Megazyme (Bray, Ireland), Sigma-Aldrich, DuPont Nutrition Biosciences (Brabrand, Denmark), Elicityl (Crolles, France), Carbosynth and Glycomix (Compton, United States).

Affinity gel electrophoresis (AGE)

12% native acrylamide gels were prepared, gels without polysaccharide as control and others containing 0.5% concentration of different polysaccharides. Gels with α -mannan, galactomannan, and laminarin were tested. Ten micrograms of particular CBPs and BSA were loaded onto each gel. The motilities of the proteins were quantified between the native gel and gel containing one of the respective polysaccharides. Electrophoresis was performed for 2 h at a constant voltage of 80 V on ice. The proteins were visualized by staining with Coomassie Blue. The graphs were prepared in Prism 7 assuming one specific binding site (GraphPad Software).

Carbohydrate microarray analysis – defined arrays

Defined polysaccharides were dissolved in water to 5 mg ml⁻¹, except for pachyman and mannan, which were dissolved in 4 M NaOH with 0.1% (w/v) NaBH₄. Subsequently, the polysaccharide solutions were diluted in printing buffer (55.2% glycerol, 44% water, 0.8% Triton X-100) by a 2-fold dilution and added to wells of a 384-microwell plate. The plate was centrifuged for 10 min at 4000 rpm and its content was printed at 22°C with 55% humidity onto nitrocellulose membrane with a pore size of 0.45 µm (Whatman, Sigma-Aldrich) using a microarray robot (Sprint, Arrayjet, Roslin, UK). Each polysaccharide was printed in triplicate. The microarrays were analyzed as previously described (Vidal-Melgosa *et al.* 2015). Briefly, microarrays that contained the printed polysaccharides were blocked for 1 h in PBS (140 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.7 mM KH₂PO₄, pH 7.5) with 5% (w/v) low fat milk powder (MPBS) or PBS. Next, the arrays were incubated for 2 h with the putative probes: his tagged proteins produced in the *E. coli* expression system (soluble proteins listed in Table 6) at 50 µg ml⁻¹ diluted in MPBS. Then, microarrays were washed in PBS and incubated for 2 h with an anti-his tag secondary antibody conjugated to alkaline phosphatase (Sigma-Aldrich) diluted 1:1500 in MPBS. In case of optimization, all described steps if developing the microarray were performed at 4 °. After washing in PBS and deionized water, microarrays were developed in a solution containing 5-bromo-4-chloro-3 indolylphosphate and nitro blue tetrazolium in alkaline phosphatase buffer (100 mM NaCl, 5 mM MgCl₂, 100 mM diethanolamine, pH 9.5). These substrates develop an insoluble purple product on the membrane when they are exposed to alkaline phosphatase conjugated antibodies. Therefore, this reaction allows detection of the binding of a probe to its target antigen, if the antigen was bound to the membrane. Once developed, the arrays were scanned at 2400 dots/inch and probe binding signals were quantified using pertinent software (Array-Pro Analyzer 6.3, Media Cybernetics, Rockville, Maryland, United States). The highest mean spot signal intensity detected was set to 100 and the rest of values normalized consequently. Defined array data is shown in a heat map where detection of binding is indicated by yellow color.

Carbohydrate microarray analysis – CoMPP arrays

The CoMPP arrays populated with microalgae and macroalgae extracts and the ones containing extracted material from samples obtained during a microalgae bloom at the North Sea were provided by Silvia Vidal-Melgosa. Biomass from microalgae lab cultures of *T. weissflogii*, *E. huxleyi* and *T. pseudonana*, of red and brown macroalgae collected at the coast of Roscoff (France), or from POM and HMW-DOM samples from the North Sea was sequentially extracted with: autoclaved milliQ water, 50 mM EDTA pH 7,5 and 4 M NaOH with

0.1 % NaBH₄. For each 10 mg of biomass 300 µl of solvent were used for extraction, which was performed by rotation for 2 h at room temperature. Samples were centrifuged at 4000 rpm for 15 min, supernatant (the extract) was collected and the pellet was resuspended in the next extracting solvent following the same extraction protocol. Extracts were diluted in printing buffer by a 2-fold dilution and added to wells of a 384-microwell plate. Microarray printing, probing and quantification was performed as described above. CoMPP data are shown in a heat map where green color intensity is correlated to mean spot signal value. A cut-off of 5 arbitrary units was applied.

POM and HMW-DOM samples from the North Sea

POM and HMW-DOM samples were collected in 2016 during a microalgae bloom from the 15th of March until the 26th of May at the station Kabeltonne at the North Sea (54°11.3'N, 7°54.0'E) near the island of Helgoland, Germany. A total of 100 L of surface seawater (1 m depth) were collected twice a week with the research vessel Aade. The seawater was sequentially filtered through 10, 3 and 0.2 µm pore size polycarbonate filters with air pressure pumps by < 2 bar. Filters were replaced once clogged. Filters were stored at -80 °C until further analysis. The 0.2 µm-filtrate obtained after the sequential filtration was concentrated to 0.5 L by tangential flow filtration (TFF) using a cut-off of 1 kDa with pressure of < 4 bar. After concentration samples were dialyzed (1 kDa) and freeze dried and used for microarray printing (HMW-DOM samples).

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) grant HE 7217/1-1 to Jan-Hendrik Hehemann and by the Max Planck Society and by a Natural Sciences and Engineering. We would like to thank all people involved in the Helgoland campaign 2016.

General discussion

Proteins involved in the selfish polysaccharide uptake

The primary goal of this dissertation was to elucidate the biochemical details of surface-glycan binding proteins (SGBPs) from marine bacteria. Wide genome sequencing shows that the Sus-like paradigm is a conserved system in a variety of environments (Grondin *et al.*, 2017). Although biophysical conditions differ in multiple ways, for instance, nutrient concentrations, oxygen level or salinity, it seems that principles of the Sus-like system are analogous in marine microbes and in bacteria of the human microbiome. However, there is an urgent need to incorporate in-depth functional studies of Sus-like protein originating from marine bacteria since there are only few functional reports describing these relevant proteins. In my thesis, I provide biochemical and structural analyses on the SGBPs, which are of interest because of their ecological relevance for the global carbon cycle as well as potential application in the field of marine glycobiology.

The term PUL has been coined for the first time by Bjursell, Martens, and Gordon in 2006 (Bjursell *et al.*, 2006), and was initially defined as a cluster of CAZymes genes co-localized with at least one pair of SusCD genes. Nowadays, genomics reveals clusters encoding CAZymes in multiple bacterial species. Consequently, the PUL term is no longer restricted to the locus containing SusCD pairs. Interestingly, SusD-like genes are a unique feature of *Bacteroidetes* (Bjursell *et al.*, 2006) and their products play an important role in the substrate uptake (Koropatkin *et al.*, 2008). SusD together with other SGBPs creates an outer membrane complex, which is most probably connected with the selfish uptake of polysaccharides.

Laminarin substrate recognized by SGBPs

Recent analyses and also data from this thesis emphasize that the carbohydrate structure of laminarin appears in varied forms. Laminarin is predicted to be one of the most common substrates in the ocean and genes responsible for its utilization were present in all metagenomic studies conducted in Helgoland. Predicted laminarin-specific PULs represent approximately 12% of all PULs investigated by Kappelmann *et al.* They determined four variants of PUL organization (named A to D) in tested bacteria (Kappelmann *et al.*, 2019), where each of them was proposed to target different laminarin type. Diatoms produce laminarins of different size and with various branches (Gügi *et al.*, 2015), some reports show even β -1,2-glucose linkages (McConville *et al.*, 1986). Despite the assumed relatively simple structure of β -1,3-glucan, laminarins are rather a heterogeneous group of substrates.

In silico predictions of PULs encoded in particular bacteria species provide valuable information about their substrates. The laminarin PUL from *Gramella* sp. MAR_2010_102, which I investigated in this thesis here was classified as variant D (Kappelmann *et al.*, 2019). The presence of a GH30 gene β -1,6-exoglucanase within this PUL indicated a specificity towards substrates containing β -1,6-glucose moieties. An important question is whether branched or debranched forms of laminarin are transported into the cell. It has been reported that laminarin branches can be up to three-glucose residues long for *E. bicyclis* laminarin (Menshova *et al.*, 2014) and it seems likely that non-linear substrates can create a steric barrier during transport through the membrane. However, our analyses of GMSusD showed high specificity towards the branched form of the substrate meaning that this protein is most probably capable of accommodating these decorations. We suggest further studies on the enzymes and SusC-like protein encoded in the laminarin PUL of *Gramella* sp. MAR_2010_102. Information, whether the protein is located at the cell surface or in the periplasm, is of great value. Experimentally, it could be tested with antibodies targeting this particular protein to accurately localize the predicted debranching GH30 enzyme from the same PUL. Ultimately, bioinformatic predictions combined with biochemical studies on SusD-like proteins involved in the crucial step of glycan uptake will provide more insight into substrate acquisition to the cell of *Gramella* sp. MAR_2010_102.

Results of our research show that GMSusD binds selectively on the AGE to one type of laminarin structure with a high degree of β -1,6-linkage, which was confirmed with additional enzymatic experiments as well as initial attempts of ITC measurements. To extend our analyses of laminarin binding by *Gramella* sp. MAR_2010_102, we also investigated the binding of two other SGBPs- GMSusE and GMSusF. Interestingly, they show variations in affinity towards different laminarin structures, but it is not as pronounced as in case of

GMSusD. Both GMSusE and GMSusF bind laminarins that differ in the content of β -1,6-glucose branches, but GMSusE has higher affinity to laminarin with more β -1,6-glucose branches and GMSusF binds more tightly to linear β -1,3-glucan. Fine differences in the affinity of particular SGBPs towards various laminarin structures illustrate the role of SGBPs in accommodating the polysaccharides at the cell surface. Nevertheless, our analyses of SGBPs strongly indicate that *Gramella* sp. MAR_2010_102 captures the substrate in the form of the β -1,6-glucose branched laminarin.

To fully understand the system, we need to know the exact structure and length of laminarin substrate bound to the surface and transported into the cell. In this study, GMSusD was tested with the laminarin polysaccharides showing very low affinity. An interesting aspect would be to investigate the binding towards structurally different oligosaccharides, shorter chains of polysaccharides. The binding affinity of GMSusE and GMSusF was tested with oligosaccharides, but with the linear β -1,3-linked laminariheptaose and laminarihexose. So far, only linear forms of laminarin oligosaccharides are commercially available. One approach to obtain different oligosaccharides structures would be organic synthesis. The collaboration with the group of prof. Peter H Seeberger is established as they have expertise and equipment for oligosaccharides synthesis. Additionally, a set of well-defined laminarinases is characterized (Becker & Hehemann, 2018), which could be applied to generate oligosaccharides for these types of experiments. Laminarin extracted from various microalgal species could be treated with identified enzymes to generate defined oligosaccharides. They could be applied in the binding assays, such as ITC, but also as substrates for protein co-crystallization with x-ray crystallography, which provides a detailed view of the interaction. Investigation of various oligosaccharides binding by SGBPs separately as well as in the potential protein complexes, for instance of the GMSusE with GMSusD would provide insight into the transported substrate.

We hypothesize that there is an association between bacterial species and particular microalgae. To find a potential diatom producing substrate for the *Gramella* sp. MAR_2010_102, we studied available environmental data. Metagenomic and metaproteomic studies show that proteins with high similarity to GMSusD (>40% amino acid identity) were expressed at the early spring diatom bloom in 2010 in the North Sea. One of the microalgal species – *Chaetoceros debilis* produces a pustulan-like laminarin (Størseth *et al.*, 2006) during spring bloom in the North Sea (Kraberg *et al.*, 2010). As GMSusD is specific to branched laminarin we hypothesize that laminarin from *C. debilis* could be the target for *Gramella* sp. MAR_2010_102. To broaden the picture of this interaction, we suggest experiments of the biochemically characterized GMSusDEF proteins with the laminarin extracted from different diatom species. Material from microalgae collected during the spring diatom bloom could be

General discussion

used for the binding experiments. Additionally, growth experiments of *Gramella* sp. MAR_2010_102 on varied laminarins and oligosaccharides would provide valuable information. That opens a question regarding the bacterial-algal interaction. Are the bacterial species reacting to a particular diatom or rather to a structure of produced polysaccharide? Such knowledge combined with the accurate environmental data about the microalgal succession would allow understanding the relation between this particular bacterium and laminarin substrate.

The work described in manuscript 1 and 2 clearly depicts various binding characteristics of all three described proteins: GMSusD, GMSusE, and GMSusF towards the β -glucan substrate. These slightly different roles indicate the importance of the multiprotein complex assembled on the surface of the bacterial cell, where each protein possibly targets another part of the substrate. As the enzymes anchored in the bacterial cell surface process the polysaccharide, SGBPs constantly adjust to a changing structure of the bound substrate. This is in line with studies conducted on SGBPs from gut microbiota, which emphasize the importance of the surface located complex formation (Cameron *et al.*, 2014; Rogowski *et al.*, 2015; Foley *et al.*, 2018). By creating a complex outer membrane structure, bacteria may also increase the overall affinity by an avidity effect, which would be of great importance for marine bacteria since the ocean is a highly diluted environment. In our approach, SGBPs were tested as recombinant proteins produced in *E. coli*, however, it is likely that the proteins may interact and display different behaviors in the native bacterium. Extraction of a SusCD-like protein complex from native *Bacteroides thetaiotaomicron* resulted in the very tight complex containing additionally two lipoproteins with a structure similar to the SusEF proteins. This close connection emphasizes the importance of complex assembly at the bacterial surface, and it is possible that the overall affinity towards the substrate differs, when SGBPs interact with each other. Structures of SusD-like proteins contain a TRPs motif, which is generally associated with protein-protein interaction. Hence, it represents another subject to investigate in further studies. Elucidating the function of each protein separately and also interacting with each other could contribute to a better understanding of how polysaccharides captured by marine *Bacteroidetes* bind efficiently diluted substrates in the seawater.

Glycan probes

Currently, the marine glycobiology is a highly unexplored field, the analysis of complex polysaccharides suffers from limitations in the available techniques. One major problem is the lack of glycan probes specific to marine polysaccharides. To address this, we investigated a library of SusD-like proteins extracted from the North Sea, near the island Helgoland. One of the main applications suitable for applying glycan probes is carbohydrate microarray technology. This method is well-established for plant-derived polysaccharides and allows for high-throughput screening of hundreds of polysaccharides. Probes are applied for localization of the polysaccharide of interest in the tissue of plants, similar studies could be performed to map occurrence of the polysaccharides in the algal material.

A tested library of proteins exposed no binding to predicted substrates. This result reflects the low affinity of this group of proteins towards their predicted polysaccharide ligand. Additionally, nonspecific binding towards negatively charged polysaccharides was observed. Overall, these two features create a considerable impediment and makes the tested group of proteins unsuitable candidates for glycan probes. Our interpretation and comments on that will be discussed in the following paragraphs.

One of our main hypotheses that affinity of SusD-like proteins in the ocean could be higher is because the substrate is highly diluted in the seawater was most likely incorrect. During the time when experiments were planned and conducted, the structure of SusCD complex was not known. In the light of the proposed pedal bin mechanism low affinity would be in accordance with the proposed role of lid pushing polysaccharide through the pore-like SusC. One has to imagine that tight binding would not be preferable in such case, as it could indispose the transport of the substrate. It leaves unanswered question how marine *Bacteroidetes* keep the substrate attached to the cell surface. Considering our discovery of novel SGBPs: GMSusE and GMSusF, the most reasonable explanation would be that binding of the substrate is performed by a complex of proteins at the cell surface of marine *Bacteroidetes*. We do know that all three of them are binding carbohydrates and now we should understand better how they interact with each other.

As we failed to observed the expected binding signal, we do not exclude that we lack the appropriate substrate. There are some commercially available polysaccharides, but most probably it does not cover the entire collection of algae-derived polysaccharides. One of the biggest challenges in glycobiology is purification of polysaccharides, we possibly do not know the exact structure of the substrate we are looking for. Binding of the glycan probe to the epitope may depend not only on the linkages but also on the position and amount of modifications of polysaccharides. Monoclonal antibodies specific towards fucoidan illustrate differences in the binding depending on the sulfation, where BAM1 identified non-sulfated

and BAM4 sulfated fucoidan epitopes (Torode *et al.*, 2015). As mainly plant-derived polysaccharides were printed on the tested microarrays, maybe there was no target epitope for the tested proteins. For example, many probes from the library were extracted from the predicted mannan PUL and they did not recognize mannan extracted from plants. Potentially it could be a result of unknown modification of target mannan derived from algae. Here, further application of bacterial CBPs as glycan probes raises, where we could use them to extract novel linkages. Assuming that a particular probe binds to the algal extract but does not recognize any of the defined polysaccharides, the experiments could be extended. I would propose further investigation of the same algal extract but chromatographically separated, followed by detection with the tested probe. However, the library of twelve CBPs tested in this study (manuscript 3) revealed nonspecific binding towards negatively charged polysaccharides, which excluded this possibility. Also, twelve soluble proteins were tested, other constructs should be also investigated.

In my Ph.D. dissertation I present an approach for exploring novel candidates for glycan probes in the form of recombinantly produced bacterial CBPs. It has multiple advantages as genetic engineering provides a relatively cheap and easy way to produce proteins of interest. This is a suitable system to manipulate the protein, for example by mutagenesis of the binding site leading to higher affinity. Additionally, it allows for the profiling of the final protein, where potentially we could reduce the size of protein or multiple binding sites. Extraction and purification of the polysaccharides require multiple purification steps and different chromatographic methods, for instant IEX and/or SEC, which often does not result in not entirely homogenic polysaccharide. Other CBPs, for instance, lectins are routinely applied for the affinity chromatography in the plant glycobiology (Merkle & Cummings, 1987). Highly specific probes produced as recombinant proteins could be used as a binding agent on the purification column. This concept provides a relatively simple and efficient system for the exploration of protein-carbohydrate interactions.

Exploring marine organic matter

The ultimate goal of our research is to characterize the bacterial interaction with algal-derived polysaccharides. We would like to know what type of polysaccharides are produced by particular algal species and which bacteria thrive in response to this production. By biochemical analysis of CAZymes and CBPs we could enhance the development of the marine glycobiology field. Nowadays, there is a vast amount of “omics” data but further support with empirical assays for protein characterization are clearly needed (Mills *et al.*, 2015; Grondin *et al.*, 2017). We supplemented bioinformatics with functional studies, which provides more reliable predictions. Protein characterization analyses in parallel with improving computational methods would allow for better understanding of the ecologically important carbohydrate-proteins interactions. One of the big unknowns is which polysaccharides are more “attractive” for microbes. In my thesis, I characterized a set of SGBPs specific to laminarin from planktonic bacteria. Laminarin is a glucan with a relatively simple structure mainly digested at the surface of the ocean. Understanding how marine microbes recognize and bind laminarin is a first step in exploring the field of marine glycobiology. Other questions that need to be addressed are which polysaccharides are rapidly catabolized and which present a more recalcitrant behavior, i.e. are not utilized by bacteria and in consequence create a carbon sink. The work presented in this thesis contributes to our understanding of polysaccharide utilization by marine *Bacteroidetes*. Further comprehensive analysis of marine polysaccharide biology is required to aid the investigation of the above stated questions.

Concluding remarks

The main accomplishment of this thesis has been the first in-detail characterization of marine SGBPs: GMSusD, GMSusE and GMSusF. Additionally, we investigated if proteins from marine bacteria thriving during algal spring blooms in the North Sea could be used as glycan probes for polysaccharide detection. To assess the potential of CBPs as glycan probes and employ them in further applications, molecular characterization of CBPs is required. The foremost outcomes of the research performed during my Ph.D. project are summarized in the following list:

- biochemical characterization of GMSusD protein specific to highly branched laminarin
- exploitation of the obtained structural information about GMSusD to localize other SusD-like proteins with similar shape of binding site in the metagenomes of global surface water datasets
- discovery of two novel laminarin binding proteins: GMSusE and GMSusF, which were previously predicted as hypothetical
- in-depth characterization of the binding ability of GMSusE and GMSusF
- development of a protein library of recombinantly produced SusD-like proteins
- exploring the potential of CBPs, expressed during diatom blooms at the North Sea, as novel probes
- reporting that several of the investigated SusD-like proteins were not mono-specific as they bound negatively charged polysaccharides independently of the protein predicted substrate
- identification of four CBPs binding towards α -mannan, β -mannan, and laminarin.

Overall, the presented thesis has provided molecular details aiding a better understanding of how marine bacteria bind polysaccharides at the cell surface and also contributed to explore CBPs potential as novel glycan probes.

Popular summary

How do marine bacteria recognize the sugars from algae?

Everybody needs to eat, even bacteria. Sugar molecules produced by algae serve as a great food source for marine bacteria. However, the sea is highly occupied with various organisms, including microbes with the same food interest. Each of them has to find their own way to survive in that crowd, so they developed molecular tools to overcome competition. Algal sugars are also quite diluted in the ocean. Wherefore, bacteria have specialized mechanism to recognize and bind their substrate from the surroundings.

In my research, I focus on a particular group of bacteria, which are known to be good at the sugar degradation or more precisely, polysaccharides which are long-chain molecules of sugar units. How do bacteria do that? Microbes possess binding proteins and enzymes, which are anchored in the cell surface and allow them to recognize, initially digest and uptake substrate. This approach is called selfish as bacteria do not share the substrate with others. Some of the sugars are more attractive, some less. The same as for people, we rather prefer chocolate over broccoli.

To understand bacterial preferences and get an overall picture, which allows for understanding ecological implications, we try to characterize each step of bacterial sugar acquisition in detail. During my Ph.D., I studied the molecular nature of proteins involved in the recognition and binding of specific polysaccharide. They are located on the surface of the bacterial cell. I focused on the proteins specific to laminarin, which is sometimes called “starch of the ocean” because it is highly abundant and its chemical structure is relatively simple. From my research, we learnt which proteins are responsible for laminarin binding and that they are very selective as their recognition depends on very fine tuning in the polysaccharide structure. With the technique called X-ray crystallography, we were able to determine the three-dimensional structure of investigated protein at the molecular level. Thanks to this very precise analysis, we figured out what the shape of binding site is and even more accurately which amino acids (building blocks of the protein) are directly interacting with the polysaccharide. We applied obtained biochemical information to the big data set of protein sequences from the ocean. Based on that, we were able to predict proteins with the same specificity located in other parts of the ocean. I also analyzed how tightly these proteins can bind polysaccharides. Surprisingly, the single protein does not show very high affinity, but there are few of them present at the bacterial surface. It seems bacteria have to exploit an entire complex of proteins for this highly efficient selfish uptake. These proteins work together as a team to recognize the substrate of bacterial interest.

Bacterial polysaccharide utilization in the ocean is a meaningful process for all of us. Algae are responsible for about half of the global carbon dioxide fixation meaning that every

Popular summary

second breath you take comes from the ocean. A large part of the fixed organic carbon is in form of polysaccharides. By degrading marine polysaccharides, microbes significantly contribute to the global carbon cycle, the fundamental element supporting all life on Earth. Therefore, in-depth understanding of bacterial polysaccharide utilization is of great importance.

I hope you enjoyed reading this dissertation, which shares my interest and sympathy to the bacterial sugar degraders. If you have questions, please do not hesitate to contact me:

agata.mystkowska@gmail.com

Acknowledgments

I would firstly like to thank you Jan-Hendrik for giving me this great opportunity to do my doctoral thesis in the marine glycobiology group. I admire your engagement and intellect for research. Thank you for always being open for scientific discussion, your comments and suggestions improved my work. I am grateful to be a student of yours.

Thank you, Carol, Nicole, and Thorsten for supporting my research during the last years as my thesis committee. I really appreciate that you shared your precious time for the discussions, advised on my project. Our meetings were always great motivation for me. Special thanks to Nicole for reviewing the thesis and traveling to Germany for my defense.

Thanks to all people involved in POMPU project. Especially to Matthias, thank you for being my external supervisor.

Big thank to all colleagues! I would like to thank my direct supervisor and dear friend Silvia. Your optimism is always a great motivation for me. I really appreciate your support throughout the last few years, especially during tough moments when the results were not quite as we expected them to be. Thank you, Craig and Melissa. You helped me immensely both with my research and support, it is really appreciated. Andi, Nadine, Stefan, and Tao for being the best Ph.D. students' team. It was great to share with you my time here. Gon and Rebecca, thanks for being wonderful office mates. Tina and Alek for organizing lab so well! Thank you Guoyin, Hagen, Jaagni, Matja, Nicola, and Vipul. It was nice to have all of you around!

Special thanks to Burak, Mikkel, Silvia, and Stefan for helping me with thesis writing.

Also, thanks to our collaborators Chiara, Antonio, Andrew, and Al. Al, thank you for inviting me to Victoria for my research stay. It was a great experience. Thanks to all people from Al's lab for always being helpful and providing a super nice working atmosphere. Chelsea, special thanks to you for all your great support.

GLOMAR graduate school- thank you for giving me the opportunity of developing myself throughout multiple courses, a conference on Hawaii and research stay in Canada.

Last but not least thanks to all my friends!!!

Gośka dzięki, że zawsze mogę na Ciebie liczyć.

Mamuś, Tatuś – za bycie najlepszymi rodzicami na świecie i Wasze wsparcie dla każdej mojej decyzji.

Artur, dziękuję za ogromne wsparcie oraz za tak cierpliwie znoszenie tego szaleństwa ☺. Dziękuję Ci, że jesteś tutaj ze mną - to dla mnie nieocenione.

Thank you, dziękuję!

Bibliography

- Alderkamp, A. C., van Rijssel, M., & Bolhuis, H. (2007). Characterization of marine bacteria and the activity of their enzyme systems involved in degradation of the algal storage glucan laminarin. *FEMS Microbiology Ecology*, *59*(1), 108–117. DOI:10.1111/j.1574-6941.2006.00219.x
- Ale, M. T., & Meyer, A. S. (2013). Fucoidans from brown seaweeds: An update on structures, extraction techniques and use of enzymes as tools for structural elucidation. *RSC Advances*, *3*(22), 8131–8141. DOI:10.1039/c3ra23373a
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402. DOI:10.1093/nar/25.17.3389
- Aluwihare, L. I., & Repeta, D. J. (1999). A comparison of the chemical characteristics of oceanic DOM and extracellular DOM produced by marine algae. *Marine Ecology Progress Series*, *186*, 105–117. DOI:10.3354/meps186105
- Anderson, K. L., & Salyers, A. A. (1989). Biochemical evidence that starch breakdown by *Bacteroides-Thetaiotaomicron* involves outer-membrane starch-binding sites and seriplasmic starch-degrading Enzymes. *Journal of Bacteriology*, *171*(6), 3192–3198. DOI:10.1128/jb.171.6.3192-3198.1989
- Anderson, N. S., Dolan, T. C. S., & Rees, D. A. (1965). Evidence for a common structural pattern in the polysaccharide sulphates of the Rhodophyceae. *Nature*, *205*(4976), 1060–1062. DOI:10.1038/2051060a0
- Arnosti, C. (1995). Measurement of depth- and site-related differences in polysaccharide hydrolysis rates in marine sediments. *Geochimica et Cosmochimica Acta*, *59*(20), 4247–4257. DOI:10.1016/0016-7037(95)00247-W
- Arnosti, C. (2003). Fluorescent derivatization of polysaccharides and carbohydrate-containing biopolymers for measurement of enzyme activities in complex media. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, *793*(1), 181–191. DOI:10.1016/S1570-0232(03)00375-1
- Arnosti, C. (2004). Speed bumps and barricades in the carbon cycle: Substrate structural effects on carbon cycling. *Marine Chemistry*, *92*(1–4 SPEC. ISS.), 263–273. DOI:10.1016/j.marchem.2004.06.030
- Baines, S. B., & Pace, M. L. (1991). The production of dissolved organic matter by phytoplankton and its importance to bacteria: patterns across marine and freshwater systems. *Limnology and Oceanography*, *36*(6), 1078–1090.

Bibliography

DOI:10.4319/lo.1991.36.6.1078

- Barraas, R. D., & Stone, B. A. (1969). Beta-1,3-glucan hydrolases from *Euglena gracilis* II. Purification and properties of the -1,3-glucan exo-hydrolase. *Biochemica et Biophysica Acta*, *91*(1969), 342–353. DOI:10.1016/0005-2744(69)90253-8
- Becker, S. (2018). *Carbohydrate polymer in the ocean*. Ph.D. thesis. URN: urn.nbn:de:gbv:46-00106611-18
- Becker, S., & Hehemann, J.-H. (2018). Laminarin Quantification in Microalgae with Enzymes from Marine Microbes. *Bio-Protocol*, *8*(8). DOI:10.21769/bioprotoc.2666
- 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), 1–14. DOI:10.1128/AEM.03389-16
- Benner, R., Pakulski, J. D., Matthew, M., Hedges, J. I., & Hatcher, P. G. (1992). Bulk Chemical Characteristics of Dissolved Organic Matter in the Ocean. *Science*, *255*(5051), 1561–1564. DOI:10.1126/science.255.5051.1561
- Bennke, C. M., Neu, T. R., Fuchs, B. M., & Amann, R. (2013). Mapping glycoconjugate-mediated interactions of marine Bacteroidetes with diatoms. *Systematic and Applied Microbiology*, *36*(6), 417–425. DOI:10.1016/j.syapm.2013.05.002
- Bertozzi, C. R., & Rabuka, D. (2009). Structural Basis of Glycan Diversity. In *Essentials of Glycobiology*. Old Spring Harbor Laboratory Press.
- Bettarel, Y., Kan, J., Wang, K., Williamson, K. E., Cooney, S., Ribblett, S., ... Coats, D. W. (2005). Isolation and preliminary characterisation of a small nuclear inclusion virus infecting the diatom *Chaetoceros cf. gracilis*. *Aquatic Microbial Ecology*, *40*(2), 103–114. DOI:10.3354/ame040103
- 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. *Journal of Biological Chemistry*, *281*(47), 36269–36279. DOI:10.1074/jbc.M606509200
- Blake, A. W., McCartney, L., Flint, J. E., Bolam, D. N., Boraston, A. B., Gilbert, H. J., & Knox, J. P. (2006). Understanding the biological rationale for the diversity of cellulose-directed carbohydrate-binding modules in prokaryotic enzymes. *Journal of Biological Chemistry*, *281*(39), 29321–29329. DOI:10.1074/jbc.M605903200
- Bolam, D. N., & Koropatkin, N. M. (2012). Glycan recognition by the Bacteroidetes Sus-like systems. *Current Opinion in Structural Biology*, *22*(5), 563–569. DOI:10.1016/j.sbi.2012.06.006
- Boraston, A. B., Bolam, D. N., Gilbert, H. J., & Davies, G. J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *The Biochemical Journal*, *382*(Pt 3),

- 769–781. DOI:10.1042/BJ20040892
- Borch, N. H., & Kirchman, D. L. (1997). Concentration and composition of dissolved combined neutral sugars (polysaccharides) in seawater determined by HPLC-PAD. *Marine Chemistry*, *57*(1–2), 85–95. DOI:10.1016/S0304-4203(97)00002-9
- Buchan, A., LeCleir, G. R., Gulvik, C. A., & González, J. M. (2014). Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nature Reviews. Microbiology*, *12*(10), 686–698. DOI:10.1038/nrmicro3326
- Cameron, E. A., Kwiatkowski, K. J., Lee, B., Hamaker, B. R., Koropatkin, N. M., & Martens, C. (2014). Multifunctional Nutrient-Binding Proteins Adapt Human Symbiotic Enhanced Sensing and Catalysis. *MBio*, *5*(5), 1–12. DOI:10.1128/mBio.01441-14
- Cameron, E. A., Maynard, M. A., Smith, C. J., Smith, T. J., Koropatkin, N. M., & Martens, E. C. (2012). Multidomain carbohydrate-binding proteins involved in *Bacteroides thetaiotaomicron* starch metabolism. *Journal of Biological Chemistry*, *287*(41), 34614–34625. DOI:10.1074/jbc.M112.397380
- Cartmell, A., Lowe, E. C., Baslé, A., Firbank, S. J., Ndeh, D. A., Murray, H., ... Bolam, D. N. (2017). How members of the human gut microbiota overcome the sulfation problem posed by glycosaminoglycans. *Proceedings of the National Academy of Sciences*, *114*(27), 7037–7042. DOI:10.1073/pnas.1704367114
- Chen, J., Robb, C. S., Unfried, F., Kappelmann, L., Markert, S., Song, T., ... Teeling, H. (2018). Alpha- and beta-mannan utilization by marine *Bacteroidetes*. *Environmental Microbiology*, *20*(11), 4127–4140. DOI:10.1111/1462-2920.14414
- Chiovitti, A., Molino, P., Crawford, S. A., Teng, R., Spurck, T., & Wetherbee, R. (2004). The glucans extracted with warm water from diatoms are mainly derived from intracellular chrysolaminaran and not extracellular polysaccharides. *European Journal of Phycology*, *39*(2), 117–128. DOI:10.1080/0967026042000201885
- Cornuault, V., Manfield, I. W., Ralet, M. C., & Knox, J. P. (2014). Epitope detection chromatography: A method to dissect the structural heterogeneity and interconnections of plant cell-wall matrix glycans. *Plant Journal*, *78*(4), 715–722. DOI:10.1111/tpj.12504
- Cowtan, K. (2006). The “Buccaneer” software for automated protein chain tracing. *Acta Crystallographica Section A Foundations of Crystallography*, *62*(1), s87–s87. DOI:10.1107/S0108767306098266
- Cuskin, F., Lowe, E. C., Temple, M. J., Zhu, Y., Cameron, E. A., Pudlo, N. A., ... Pen, M. J. (2015). Human gut *Bacteroidetes* can utilize yeast mannan through a selfish mechanism. *Nature*, *517*(7533), 165–169. DOI:10.1038/nature13995
- D’Elia, J. N., & Salyers, A. A. (1996). Effect of regulatory protein levels on utilization of starch

Bibliography

- by *Bacteroides thetaiotaomicron*. *Journal of Bacteriology*, 178(24), 7180–7186.
DOI:10.1128/jb.178.24.7180-7186.1996
- Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., ... Richardson, D. C. (2007). MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research*, 35(2), 375–383. DOI:10.1093/nar/gkm216
- Druffel, E. R. M., Williams, P. M., Bauer, J. E., & Ertel, J. R. (1992). Cycling of dissolved and particulate organic matter in the open ocean. *Journal of Geophysical Research*, 97(C10), 15639–15695. DOI:10.1029/92JC01511
- Emerson, S., & Hedges, J. (2008). Life processes in the ocean. In *Chemical Oceanography and the Marine Carbon Cycle*. 173-218. New York: Cambridge University Press.
DOI:https://doi.org/10.1017/CBO9780511793202
- 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.
DOI:10.1107/S0907444910007493
- Engel, A., Thoms, S., Riebesell, U., Rochelle-Newall, E., & Zondervan, I. (2004). Polysaccharide aggregation as a potential sink of marine dissolved organic carbon. *Nature*, 428(6986), 929–932. DOI:10.1038/nature02453
- Evans, P. R. (2011). An introduction to data reduction: Space-group determination, scaling and intensity statistics. *Acta Crystallographica Section D: Biological Crystallography*, 67(4), 282–292. DOI:10.1107/S090744491003982X
- Fangel, J. U., Pedersen, H. L., Vidal-Melgosa, S., Ahl, L. I., Salmean, A. A., Egelund, J., ... Willats, W. G. T. (2012). Carbohydrate Microarrays in Plant Science. *Methods in Molecular Biology*, 918, 351–362. DOI:10.1007/978-1-61779-995-2
- Ficko-Blean, E., Préchoux, A., Thomas, F., Rochat, T., Larocque, R., Zhu, Y., ... Michel, G. (2017). Carrageenan catabolism is encoded by a complex regulon in marine heterotrophic bacteria. *Nature Communications*, 8(1). DOI:10.1038/s41467-017-01832-6
- Field, C. B., Behrenfeld, M., Randerson, J., & Falkowski, P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science*, 281(5374), 237–240. DOI:10.1126/science.281.5374.237
- Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., ... Bateman, A. (2016). The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Research*, 44(D1), D279–D285. DOI:10.1093/nar/gkv1344
- Foley, M. H., Déjean, G., Hemsworth, G. R., Davies, G. J., Brumer, H., & Koropatkin, N. M. (2019). A Cell-Surface GH9 Endo-Glucanase Coordinates with Surface Glycan Binding Proteins to Mediate Xyloglucan Uptake in the Gut Symbiont *Bacteroides ovatus*. *Journal*

- of Molecular Biology*, 431(5), 981-995. DOI:S0022283618311033
- Foley, M. H., Martens, E. C., & Koropatkin, N. M. (2018). SusE facilitates starch uptake independent of starch binding in *B. thetaiotaomicron*. *Molecular Microbiology*, 108(5), 551–566. DOI:10.1111/mmi.13949
- Ford, W., & Percival, E. (1965). The carbohydrates of *Phaeodactylum tricornutum*. Part I. Preliminary examination of the organism, and characterisation of low molecular weight material and of a glucan. *Journal of Chemical Society*, 0, 7035–7041. DOI:10.1039/JR9650007035
- Gao, Q., Nilsson, U., Ilag, L. L., & Leck, C. (2011). Monosaccharide compositional analysis of marine polysaccharides by hydrophilic interaction liquid chromatography-tandem mass spectrometry. *Analytical and Bioanalytical Chemistry*, 399(7), 2517–2529. DOI:10.1007/s00216-010-4638-z
- Glenwright, A. J., Pothula, K. R., Bhamidimarri, S. P., Chorev, D. S., Baslé, A., Firbank, S. J., ... van den Berg, B. (2017). Structural basis for nutrient acquisition by dominant members of the human gut microbiota. *Nature*, 541(7637), 407–411. DOI:10.1038/nature20828
- Gobet, A., Barbeyron, T., Matard-Mann, M., Magdelenat, G., Vallenet, D., Duchaud, E., & Michel, G. (2018). Evolutionary evidence of algal polysaccharide degradation acquisition by *Pseudoalteromonas carrageenovora* 9T to adapt to macroalgal niches. *Frontiers in Microbiology*, 9(NOV), 1–16. DOI:10.3389/fmicb.2018.02740
- Grondin, J. M., Tamura, K., Déjean, G., Abbott, D. W., & Brumer, H. (2017). Polysaccharide Utilization Loci: Fuelling microbial communities. *Journal of Bacteriology*, 199(15), 1–15. DOI:10.1128/JB.00860-16
- Gügi, B., Costaouec, T. Le, Burel, C., Lerouge, P., & Helbert, W. (2015). Diatom-specific oligosaccharide and polysaccharide structures help to unravel biosynthetic capabilities in diatoms. *Marine Drugs*, 13(9), 5993–6018. DOI:10.3390/md13095993
- Hehemann, J. H., Boraston, A. B., & Czjzek, M. (2014). A sweet new wave: Structures and mechanisms of enzymes that digest polysaccharides from marine algae. *Current Opinion in Structural Biology*, 28(1), 77–86. DOI:10.1016/j.sbi.2014.07.009
- Hehemann, J. H., Correc, G., Barbeyron, T., Helbert, W., Czjzek, M., & Michel, G. (2010). Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature*, 464(7290), 908–912. DOI:10.1038/nature08937
- Jam, M., Flament, D., Allouch, J., Potin, P., Thion, L., Kloareg, B., ... Barbeyron, T. (2005). The endo- β -agarases AgaA and AgaB from the marine bacterium *Zobellia galactanivorans*: two paralogous enzymes with different molecular organizations and catalytic behaviours. *The Biochemical Journal* 385, 703–713. DOI:10.1042/BJ20041044

Bibliography

- Kabisch, A., Otto, A., König, S., Becher, D., Albrecht, D., Schüler, M., ... Schweder, T. (2014). Functional characterization of polysaccharide utilization loci in the marine Bacteroidetes "Gramella forsetii" KT0803. *The ISME Journal*, 8(7), 1492–1502. DOI:10.1038/ismej.2014.4
- Kabsch, W. (2010). Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D: Biological Crystallography*, 66(2), 133–144. DOI:10.1107/S0907444909047374
- Kadam, S. U., Tiwari, B. K., & O'Donnell, C. P. (2015). Extraction, structure and biofunctional activities of laminarin from brown algae. *International Journal of Food Science and Technology*, 50(1), 24–31. DOI:10.1111/ijfs.12692
- Kappelmann, L., Krüger, K., Hehemann, J. H., Harder, J., Markert, S., Unfried, F., ... Teeling, H. (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. DOI:10.1038/s41396-018-0242-6
- Karsenti, E., Acinas, S. G., Bork, P., Bowler, C., de Vargas, C., Raes, J., ... Zingone, A. (2011). A holistic approach to marine Eco-systems biology. *PLoS Biology*, 9(10), 7–11. DOI:10.1371/journal.pbio.1001177
- Kirchman, D. L. (2002). The ecology of Cytophaga-Flavobacteria in aquatic environments. *FEMS Microbiology Ecology*, 39(2), 91–100. DOI:10.1111/j.1574-6941.2002.tb00910.x
- Kirchman, D. L., Yu, L., & Cottrell, M. T. (2003). Diversity and Abundance of Uncultured. *Applied and Environmental Microbiology*, 69(11), 6587–6596. DOI:10.1128/AEM.69.11.6587
- Kopf, A., Bicak, M., Kottmann, R., Schnetzer, J., Kostadinov, I., Lehmann, K., ... Glöckner, F. O. (2015). The ocean sampling day consortium. *GigaScience*, 4(1). DOI:10.1186/s13742-015-0066-5
- Kormas, K. A., Tivey, M. K., Von Damm, K., & Teske, A. (2006). Bacterial and archaeal phylotypes associated with distinct mineralogical layers of a white smoker spire from a deep-sea hydrothermal vent site (9°N, East Pacific Rise). *Environmental Microbiology*, 8(5), 909–920. DOI:10.1111/j.1462-2920.2005.00978.x
- Koropatkin, Martens, E. C., Gordon, J. I., & Smith, T. J. (2008). Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure (London, England : 1993)*, 16(7), 1105–1115. DOI:10.1016/j.str.2008.03.017
- Koropatkin, N. M., & Smith, T. J. (2010). SusG: A Unique Cell-Membrane-Associated α -Amylase from a Prominent Human Gut Symbiont Targets Complex Starch Molecules. *Structure*, 18(2), 200–215. DOI:10.1016/j.str.2009.12.010
- Koropatkin, N. M., Martens, E. C., Gordon, J. I., & Smith, T. J. (2009). The Structure Of A SusD

- Homolog, BT1043, Involved In Mucin O- Glycan Utilization In A Prominent Human Gut Symbiont. *Biochemistry*, 48(7), 1532–1542. DOI:10.1021/bi801942a
- Kraan, S. (2012). *Algal Polysaccharides, Novel Applications and Outlook*. In Carbohydrates Comprehensive Studies on Glycobiology and Glycotechnology (Vol. 2). 489-532. IntechOpen. DOI:10.5772/32009
- Kraberg, A., Baumann, M., & Dürselen, C.-D. (2010). Coastal Phytoplankton. 42-4. Verlag Dr. Friedrich Pfeil
- Krieg N.R., Ludwig W., Euzéby J., Whitman W.B. (2010) *Phylum XIV. Bacteroidetes phyl. nov.*. In: Bergey's Manual of Systematic Bacteriology. 25-469 Springer. DOI:10.1007/bergeysoutline200405
- Labourel, A., Jam, M., Jeudy, A., Hehemann, J. H., Czjzek, M., & Michel, G. (2014). The β -glucanase ZgLamA from *Zobellia galactanivorans* evolved a bent active site adapted for efficient degradation of algal laminarin. *Journal of Biological Chemistry*, 289(4), 2027–2042. DOI:10.1074/jbc.M113.538843
- Labourel, A., Jam, M., Legentil, L., Sylla, B., Hehemann, J. H., Ferrières, V., ... Michel, G. (2015). Structural and biochemical characterization of the laminarinase ZgLamCGH16 from *Zobellia galactanivorans* suggests preferred recognition of branched laminarin. *Acta Crystallographica Section D: Biological Crystallography*, 71, 173–184. DOI:10.1107/S139900471402450X
- Lahaye, M., & Robic, A. (2007). Structure and function properties of Ulvan, a polysaccharide from green seaweeds. *Biomacromolecules*, 8(6), 1765–1774. DOI:10.1021/bm061185q
- Laine, R. A. (1994). Invited Commentary: A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05×10^{12} structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. *Glycobiology*, 4(6), 759–767.
- Larsbrink, J., Zhu, Y., Kharade, S. S., Kwiatkowski, K. J., Eijsink, V. G. H., Koropatkin, N. M., ... Pope, P. B. (2016). A polysaccharide utilization locus from *Flavobacterium johnsoniae* enables conversion of recalcitrant chitin. *Biotechnology for Biofuels*, 9(1), 1–16. DOI:10.1186/s13068-016-0674-z
- Le Costaouëc, T., Unamunzaga, C., Mantecon, L., & Helbert, W. (2017). New structural insights into the cell-wall polysaccharide of the diatom *Phaeodactylum tricornutum*. *Algal Research*, 26(July), 172–179. DOI:10.1016/j.algal.2017.07.021
- Lee, C., Wakeham, S., & Arnosti, C. (2004). Particulate organic matter in the sea: the composition conundrum. *Ambio*, 33(8), 565–575. DOI:10.1579/0044-7447-33.8.565
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. DOI:10.1093/bioinformatics/btp324

Bibliography

- Liu, H., & Naismith, J. H. (2008). An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnology*, *8*(1), 91.
DOI:10.1186/1472-6750-8-91
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, *42*(D1), 490–495. DOI:10.1093/nar/gkt1178
- Maciel, J., Chaves, L., Souza, B., Teixeira, D., Freitas, A., Feitosa, J., & Paula, R. C. M. (2007). Structural characterization of cold extracted fraction of soluble sulfated polysaccharide from red seaweed *Gracilaria birdiae*. *Carbohydrate Polymers*, *71*(4), 559–565.
DOI:10.1016/j.carbpol.2007.06.026
- Mann, A. J., Hahnke, R. L., Huang, S., Werner, J., Xing, P., Barbeyron, T., ... Teeling, H. (2013). The genome of the alga-associated marine flavobacterium *Formosa agariphila* KMM 3901T reveals a broad potential for degradation of algal polysaccharides. *Applied and Environmental Microbiology*, *79*(21), 6813–6822. DOI:10.1128/AEM.01937-13
- Martens, E. C., Chiang, H. C., & Gordon, J. I. (2009). Mucosal Glycan Foraging Enhances Fitness and Transmission of a Saccharolytic Human Gut Bacterial Symbiont. *Cell Host Microbe*, *4*(5), 447–457. DOI:10.1016/j.chom.2008.09.007
- Martens, E. C., Koropatkin, N. M., Smith, T. J., & Gordon, J. I. (2009). Complex glycan catabolism by the human gut microbiota: The bacteroidetes sus-like paradigm. *Journal of Biological Chemistry*, *284*(37), 24673–24677. DOI:10.1074/jbc.R109.022848
- Martin, M., Portetelle, D., Michel, G., & Vandenberg, M. (2014). Microorganisms living on macroalgae: Diversity, interactions, and biotechnological applications. *Applied Microbiology and Biotechnology*, *98*(7), 2917–2935. DOI:10.1007/s00253-014-5557-2
- McCartney, L., Marcus, S. E., & Knox, J. P. (2005). Monoclonal antibodies to plant cell wall xylans and arabinoxylans. *Journal of Histochemistry & Cytochemistry*, *53*(4), 543–546.
DOI:10.1369/jhc.4B6578.2005
- McConville, M. J., Bacic, A., & Clarke, A. E. (1986). Structural studies of chrysolaminaran from the ice diatom *Stauroneis amphioxys* (Gregory). *Carbohydrate Research*, *153*(2), 330–333. DOI:10.1016/S0008-6215(00)90276-3
- 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*(4), 658–674. DOI:10.1107/S0021889807021206
- Meikle, P. J., Bonig, I., Hoogenraad, N. J., Clarke, A. E., & Stone, B. A. (1991). The location of (1,3)-beta-glucans in the walls of pollen tubes of *Nicotiana glauca* using (1,3)-beta-glucan-specific monoclonal antibody. *Planta*, *185*(1), 1–8. DOI:10.1007/BF00194507
- Meikle, P. J., Hoogenraad, N. J., Bonig, I., Clarke, A. E., & Stone, B. A. (1994). A (1,3;1,4)-beta-

- glucan-specific monoclonal antibody and its use in the quantitation and immunocytochemical location of (1,3;1,4)-beta-glucans. *The Plant Journal*, 5(1), 1–9. DOI:10.1046/j.1365-313X.1994.5010001.x
- Menshova, R. V., Ermakova, S. P., Anastyuk, S. D., Isakov, V. V., Dubrovskaya, Y. V., Kusaykin, M. I., ... Zvyagintseva, T. N. (2014). Structure, enzymatic transformation and anticancer activity of branched high molecular weight laminaran from brown alga *Eisenia bicyclis*. *Carbohydrate Polymers*, 99, 101–109. DOI:10.1016/j.carbpol.2013.08.037
- Merkle, R. K., & Cummings, R. D. (1987). Lectin Affinity Chromatography of Glycopeptides. *Analytical Methods*, 138(1984), 232–259. DOI:10.1016/0076-6879(87)38020-6
- Michel, G., Tonon, T., Scornet, D., Cock, J. M., & Kloareg, B. (2010). The cell wall polysaccharide metabolism of the brown alga *Ectocarpus siliculosus*. Insights into the evolution of extracellular matrix polysaccharides in Eukaryotes. *New Phytologist*, 188(1), 82–97. DOI:10.1111/j.1469-8137.2010.03374.x
- Mills, C. L., Beuning, P. J., & Ondrechen, M. J. (2015). Biochemical functional predictions for protein structures of unknown or uncertain function. *Computational and Structural Biotechnology Journal*, 13, 182–191. DOI:10.1016/j.csbj.2015.02.003
- Møller, E. F. (2007). Production of dissolved in the copepods *Acartia tonsa*, organic carbon by sloppy feeding and *Temora longicornis* Centropages. *Limnology and Oceanography*, 52(1), 79–84. DOI:10.4319/lo.2007.52.1.0079
- Moller, I., Marcus, S. E., Haeger, A., Verherbruggen, Y., Verhoef, R., Schols, H., ... Willats, W. (2008). High-throughput screening of monoclonal antibodies against plant cell wall glycans by hierarchical clustering of their carbohydrate microarray binding profiles. *Glycoconjugate Journal*, 25(1), 37–48. DOI:10.1007/s10719-007-9059-7
- Moller, I., Sørensen, I., Bernal, A. J., Blaukopf, C., Lee, K., Øbro, J., ... Willats, W. G. T. (2007). High-throughput mapping of cell-wall polymers within and between plants using novel microarrays. *Plant Journal*, 50(6), 1118–1128. DOI:10.1111/j.1365-313X.2007.03114.x
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., ... Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography*, 67(4), 355–367. DOI:10.1107/S0907444911001314
- Myklestad, S., & Haug, A. (1972). *Chaetoceros affinis* vail willei. *Journal of Experimental Marine Biology and Ecology*, 9(2), 125–136. DOI:10.1016/0022-0981(72)90041-X
- Mystkowska, A. A., Robb, C., Vidal-Melgosa, S., Vanni, C., Fernandez-Guerra, A., Höhne, M., & Hehemann, J.-H. (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. DOI:10.1111/febs.14674

Bibliography

- Nelson, C. E., Attia, M. A., Rogowski, A., Morland, C., Brumer, H., & Gardner, J. G. (2017). Comprehensive functional characterization of the Glycoside Hydrolase Family 3 enzymes from *Cellvibrio japonicus* reveals unique metabolic roles in biomass saccharification. *Environmental Microbiology*, *19*(12), 5025–5039. DOI:10.1111/1462-2920.13959
- Notredame, C., Higgins, D. G., & Heringa, J. (2000). T-coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*, *302*(1), 205–217. DOI:10.1006/jmbi.2000.4042
- O’Riordan, N., Kane, M., Joshi, L., & Hickey, R. M. (2014). Structural and functional characteristics of bovine milk protein glycosylation. *Glycobiology*, *24*(3), 220–236. DOI:10.1093/glycob/cwt162
- Painter, T. J. (1983). Algal Polysaccharides. In *The Polysaccharides*. 195-285. Elsevier. DOI:10.1016/B978-0-12-065602-8.50009-1
- Pang, Z., Otaka, K., Maoka, T., Hidaka, K., Ishijima, S., Oda, M., & Ohnishi, M. (2005). Structure of beta-glucan oligomer from laminarin and its effect on human monocytes to inhibit the proliferation of U937 cells. *Bioscience, Biotechnology, and Biochemistry*, *69*(3), 553–558. DOI:10.1271/bbb.69.553
- Parsons, T. R., Stephens, K., & Strickland J. D. (1961). On the chemical composition of 11 species of marine phytoplankters. *Journal Of The Fisheries Research Board Of Canada*, *18*(6), 1001–1016. DOI:10.1139/f61-063
- Passow, U., & Alldredge, A. (1995). A dye-binding assay for the spectrophotometric measurement of transparent exopolymer particles (TEP). *Limnology and Oceanography*, *40*(7), 1326–1335. DOI:10.4319/lo.1995.40.7.1326
- Pedersen, H. L., Fangel, J. U., McCleary, B., Ruzanski, C., Rydahl, M. G., Ralet, M. C., ... Willats, W. G. T. (2012). Versatile high resolution oligosaccharide microarrays for plant glycobiology and cell wall research. *Journal of Biological Chemistry*, *287*(47), 39429–39438. DOI:10.1074/jbc.M112.396598
- Phansopa, C., Roy, S., Rafferty, J. B., Douglas, C. W. I., Pandhal, J., Wright, P. C., ... Stafford, G. P. (2014). Structural and functional characterization of NanU, a novel high-affinity sialic acid-inducible binding protein of oral and gut-dwelling Bacteroidetes species. *The Biochemical Journal*, *458*(3), 499–511. DOI:10.1042/BJ20131415
- Pinhassi, J., Havskum, H., Peters, F., Malits, A., Sala, M. M., Guadayol, O., & Marrase, C. (2004). Changes in Bacterioplankton Composition under Different Phytoplankton Regimens. *Applied and Environmental Microbiology*, *70*(11), 6753–6766. DOI:10.1128/AEM.70.11.6753–6766.2004
- Popper, Z. A., Michel, G., Hervé, C., Domozych, D. S., Willats, W. G. T., Tuohy, M. G., ... Stengel,

- D. B. (2011). Evolution and Diversity of Plant Cell Walls: From Algae to Flowering Plants. *Annual Review of Plant Biology*, 62(1), 567–590. DOI:10.1146/annurev-arplant-042110-103809
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. DOI:10.1093/bioinformatics/btq033
- Read, S. M., Currie, G., & Bacic, A. (1996). Analysis of the structural heterogeneity of laminarin by electrospray-ionisation-mass spectrometry. *Carbohydrate Research*, 281(2), 187–201. DOI:10.1016/0008-6215(95)00350-9
- Rees, D. A. (1963). a Note on the Characterization of Carbohydrate Sulphates By Acid Hydrolysis. *Biochemical Journal*, 88(2), 343–345. DOI:10.1042/bj0880343
- Reichenbach, H., & Weeks, O. B. (1981). *The Flavobacterium-Cytophaga group*. GBF monograph series No. 5. Verlag Chemie, Weinheim
- Reintjes, G., Arnosti, C., Fuchs, B. M., & Amann, R. (2017). An alternative polysaccharide uptake mechanism of marine bacteria. *The ISME Journal*, 11, 1640–1650. DOI:10.1038/ismej.2017.26
- Reintjes, G., Arnosti, C., Fuchs, B., & Amann, R. (2018). Selfish, sharing and scavenging bacteria in the Atlantic Ocean: a biogeographical study of bacterial substrate utilisation. *The ISME Journal*, 13, 1119–1132. DOI:10.1038/s41396-018-0326-3
- Rioux, L. E., & Turgeon, S. L. (2015). Seaweed carbohydrates. In *Seaweed sustainability*. 141–192. Elsevier. DOI:10.1016/B978-0-12-418697-2.00007-6
- Rioux, L. E., Turgeon, S. L., & Beaulieu, M. (2010). Structural characterization of laminaran and galactofucan extracted from the brown seaweed *Saccharina longicruris*. *Phytochemistry*, 71(13), 1586–1595. DOI:10.1016/j.phytochem.2010.05.021
- Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, 42(W1), 320–324. DOI:10.1093/nar/gku316
- Rogowski, A., Briggs, J. A., Mortimer, J. C., Tryfona, T., Terrapon, N., Lowe, E. C., ... Bolam, D. N. (2015). Glycan complexity dictates microbial resource allocation in the large intestine. *Nature Communications*, 6, 7481. DOI:10.1038/ncomms8481
- Rydahl, M. G., Kračun, S. K., Fangel, J. U., Michel, G., Guillouzo, A., Génicot, S., ... Willats, W. G. T. (2017). Development of novel monoclonal antibodies against starch and ulvan - Implications for antibody production against polysaccharides with limited immunogenicity. *Scientific Reports*, 7(1), 9326. DOI:10.1038/s41598-017-04307-2
- Saito, H., Misaki, A., & Harada, T. (1968). A comparison of the structure of curdlan and pachyman. *Agricultural and Biological Chemistry*, 32(10), 1261–1269. DOI:10.1080/00021369.1968.10859213
- Salmeán, A. A., Guillouzo, A., Duffieux, D., Jam, M., Matard-Mann, M., Larocque, R., ... Hervé, C.

Bibliography

- (2018). Double blind microarray-based polysaccharide profiling enables parallel identification of uncharacterized polysaccharides and carbohydrate-binding proteins with unknown specificities. *Scientific Reports*, 8(1), 2500. DOI:10.1038/s41598-018-20605-9
- Schattenhofer, M., Fuchs, B. M., Amann, R., Zubkov, M. V., Tarran, G. A., & Pernthaler, J. (2009). Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. *Environmental Microbiology*, 11(8), 2078–2093. DOI:10.1111/j.1462-2920.2009.01929.x
- Shipman, J. A., Berleman, J. E., & Salyers, A. A. (2000). Characterization of four outer membrane proteins involved in binding starch to the cell surface of *Bacteroides thetaiotaomicron*. *Journal of Bacteriology*, 182(19), 5365–5372. DOI:10.1128/JB.182.19.5365-5372.2000
- Shoseyov, O., Shani, Z., & Levy, I. (2006). Carbohydrate Binding Modules: Biochemical Properties and Novel Applications. *Microbiology and Molecular Biology Reviews*, 70(2), 283–295. DOI:10.1128/membr.00028-05
- Steinegger, M., & Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, 35(11), 1026–1028. DOI:10.1038/nbt.3988
- Sternberg, M. J., & Kelley, L. A. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols*, 4(3), 363–371. DOI:10.1038/nprot.2009.2
- Størseth, T. R., Hansen, K., Reitan, K. I., & Skjermo, J. (2005). Structural characterization of beta-D-(1→3)-glucans from different growth phases of the marine diatoms *Chaetoceros mulleri* and *Thalassiosira weissflogii*. *Carbohydrate Research*, 340(6), 1159–1164. DOI:10.1016/j.carres.2004.12.036
- Størseth, T. R., Kirkvold, S., Skjermo, J., & Reitan, K. I. (2006). A branched beta-D-(1-3,1-6)-glucan from the marine diatom *Chaetoceros debilis* (Bacillariophyceae) characterized by NMR. *Carbohydrate Research*, 341(12), 2108–2114. DOI:10.1016/j.carres.2006.05.005
- Studier, F. W. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification*, 41(1), 207–234. DOI:10.1016/j.pep.2005.01.016
- Tancula, E., Feldhaus, M. J., Bedzyk, L. A., & Salyers, A. A. (1992). Location and characterization of genes involved in binding of starch to the surface of *Bacteroides thetaiotaomicron*. *Journal of Bacteriology*, 174(17), 5609–5616. DOI:10.1128/jb.174.17.5609-5616.1992
- Tauzin, A. S., Kwiatkowski, K. J., Orlovsky, N. I., Smith, C. J., Creagh, A. L., Haynes, C. A., ... Koropatkin, N. M. (2016). Molecular dissection of xyloglucan recognition in a

- prominent human gut symbiont. *MBio*, 7(2), e02134–15. DOI:10.1128/mBio.02134-15
- Teeling, H., Fuchs, B. M., Becher, D., Klockow, C., Gardebrecht, A., Bennke, C. M., ... Amann, R. (2012). Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science*, 336(6081), 608–611. DOI:10.1126/science.1218344
- Teeling, H., Fuchs, B. M., Bennke, C. M., Kruger, K., Chafee, M., Kappelmann, L., ... Amann, R. I. (2016). Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. *ELife*, 5(e11888), 1–31. DOI:10.7554/eLife.11888
- Terrapon, N., Lombard, V., Gilbert, H. J., & Henrissat, B. (2015). Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. *Bioinformatics*, 31(5), 647–655. DOI:10.1093/bioinformatics/btu716
- Thornton, D. C. O. (2014). Dissolved organic matter (DOM) release by phytoplankton in the contemporary and future ocean. *European Journal of Phycology*, 49(1), 20–46. DOI:10.1080/09670262.2013.875596
- Torode, T. A., Marcus, S. E., Jam, M., Tonon, T., Blackburn, R. S., Hervé, C., & Knox, J. P. (2015). Monoclonal antibodies directed to fucoidan preparations from brown algae. *PLoS ONE*, 10(2), 1–19. DOI:10.1371/journal.pone.0118366
- Torode, T. A., Siméon, A., Marcus, S. E., Jam, M., Le Moigne, M. A., Duffieux, D., ... Hervé, C. (2016). Dynamics of cell wall assembly during early embryogenesis in the brown alga *Fucus*. *Journal of Experimental Botany*, 67(21), 6089–6100. DOI:10.1093/jxb/erw369
- Turnbull, J. E., & Field, R. a. (2007). Emerging glycomics technologies. *Nature Chemical Biology*, 3(2), 74–77. DOI:10.1038/nchembio0207-74
- Unfried, F., Becker, S., Robb, C. S., Hehemann, J. H., Markert, S., Heiden, S. E., ... Schweder, T. (2018). Adaptive mechanisms that provide competitive advantages to marine bacteroidetes during microalgal blooms. *ISME Journal*, 12(12), 2894–2906. DOI:10.1038/s41396-018-0243-5
- Unfried, F., Harder, J., Teeling, H., Hahnke, R., Markert, S., Kappelmann, L., ... Schweder, T. (2016). *The role of the Bacteroidetes strains Formosa sp. Hel3_A1_48 and Hel1_33_131 in polysaccharide utilization during North Sea phytoplankton blooms*. Poster session presented at the Annual Conference 2016 of the Association for General and Applied Microbiology (VAAM), Jena.
- Veldhuis, M., Kraay, G., & Timmermans, K. (2001). Cell death in phytoplankton : correlation between changes in membrane permeability , photosynthetic activity , pigmentation and growth. *European Journal of Phycology*, 36, 167–177. DOI:10.1080/09670260110001735318
- Vidal-Melgosa, S., Pedersen, H. L., Schückel, J., Arnal, G., Dumon, C., Amby, D. B., ... Willats, W.

Bibliography

- G. T. (2015). A new versatile microarray-based method for high-throughput screening of carbohydrate-active enzymes. *Journal of Biological Chemistry*, 290(14), jbc.M114.630673. DOI:10.1074/jbc.M114.630673
- Wang, D., Kim, D. H., Seo, N., Yun, E. J., An, H. J., Kim, J.-H., & Kim, K. H. (2016). A novel glycoside hydrolase family 5 beta-1,3-1,6-endoglucanase from *Saccharophagus degradans* 2-40T and its transglycosylase activity. *Applied and Environmental Microbiology*, 82(14), 4340–4349. DOI:10.1128/AEM.00635-16
- Wang, Y., Chen, X., Guo, W., & Zhou, H. (2018). Distinct bacterial and archaeal diversities and spatial distributions in surface sediments of the Arctic Ocean. *FEMS Microbiology Letters*, 365(24), 1–10. DOI:10.1093/femsle/fny273
- Williamson, S. J., Rusch, D. B., Yooseph, S., Halpern, A. L., Heidelberg, K. B., Glass, J. I., ... Venter, J. C. (2008). The sorcerer II global ocean sampling expedition: Metagenomic characterization of viruses within aquatic microbial samples. *PLoS ONE*, 3(1). DOI:10.1371/journal.pone.0001456
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., ... Wilson, K. S. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography*, 67(4), 235–242. DOI:10.1107/S0907444910045749
- Wu, D., Wu, M., Halpern, A., Rusch, D. B., Yooseph, S., Frazier, M., ... Eisen, J. A. (2011). Stalking the fourth domain in metagenomic data: Searching for, discovering, and interpreting novel, deep branches in marker gene phylogenetic trees. *PLoS ONE*, 6(3), 1–12. DOI:10.1371/journal.pone.0018011
- Xie, G., Bruce, D. C., Challacombe, J. F., Chertkov, O., Detter, J. C., Gilna, P., ... McBride, M. J. (2007). Genome sequence of the cellulolytic gliding bacterium *Cytophaga hutchinsonii*. *Applied and Environmental Microbiology*, 73(11), 3536–3546. DOI:10.1128/AEM.00225-07
- Xing, P., Hahnke, R. L., Unfried, F., Markert, S., Huang, S., Fuchs, B. M., ... Teeling, H. (2014). Niche separation of two *Polaribacter* strains isolated from the German Bight of the North Sea during a spring diatom bloom. *The ISME Journal*, 9(6), 1410–1422. DOI:10.1038/ismej.2014.225
- Xu, S. Y., Huang, X., & Cheong, K. L. (2017). Recent advances in marine algae polysaccharides: Isolation, structure, and activities. *Marine Drugs*, 15(12), 388. DOI:10.3390/md15120388

Appendix

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



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*}

¹Center for Marine Environmental Sciences, University of Bremen (MARUM), Bremen 28359, Germany

²Max 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

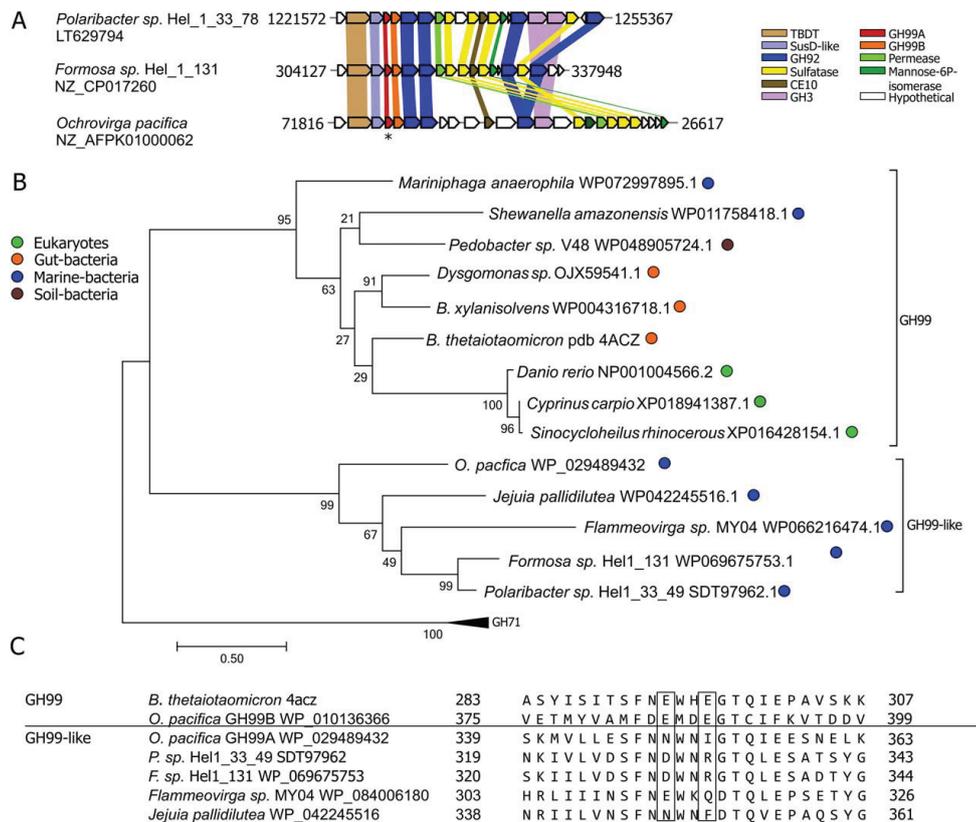


Figure 1. Phylogenetic comparison of OpGH99A with related proteins from marine and terrestrial organisms. (A) The putative sulfo-mannan PULs of the three marine isolates have 15 genes in common. The colors correspond to the function of the protein and proteins with high sequence identity have been linked 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 **ExxE**, 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 ₁ 21
Unit cell <i>a</i> , <i>b</i> , <i>c</i> (Å)	102.05, 102.05, 80.11
Resolution range, (Å)	88.38–2.40 (2.53–2.40)
<i>R</i> _{merge}	0.112 (0.741)
Completeness (%)	100.0 (100.0)
Redundancy	9.6 (9.8)
<i><I>/σ(I)></i>	13.6 (3.0)
No. of reflections	184,087 (27,039)
No. unique	19,223 (2761)
Mosaicity	0.23
Refinement	
<i>R</i> _{work} / <i>R</i> _{free} (%)	0.166/0.215
No. of atoms	3005
Protein	2846
Bromide	5
Water	164
<i>B</i> 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 Pacific¹⁴ 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 P3₁21 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 (α/β)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

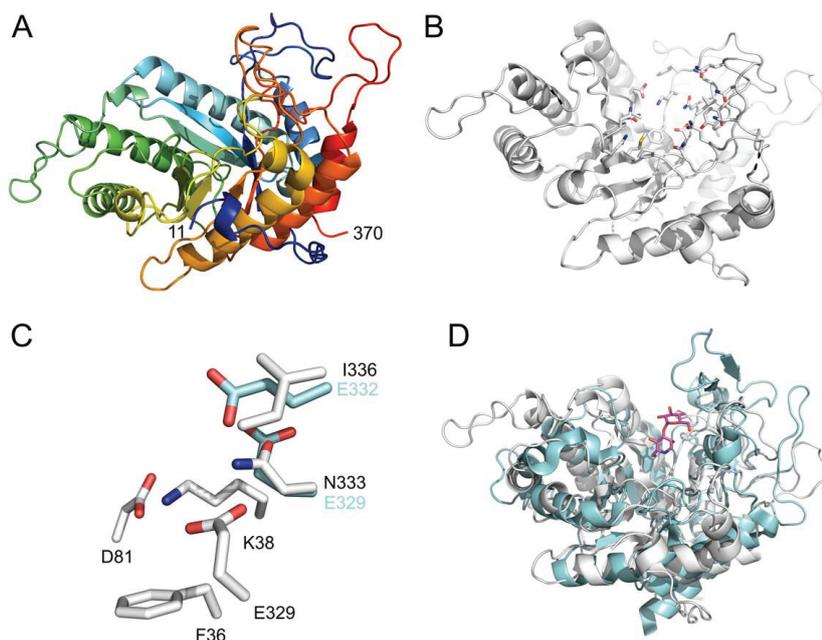


Figure 2. Structure of OpGH99A. (A) The structure of OpGH99A shown as a cartoon representation color ramped from the N-terminus (blue) to the C-terminus (red). OpGH99A adopts a $(\alpha/\beta)_8$ fold and is shown centering on the putative functional site located in the C-terminal end of the barrel. (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* DH5 α (New England Biolabs). Clones were sequenced in house by Sanger sequencing using Big-dye (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 MgCl₂ to a final concentration of 1 mM and DNase to a concentration of 1 mg/mL. The resulting lysate was centrifuged at 16,000g 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 P13²²; 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 PHASER²⁵ in CCP4²⁶ and a Phyre¹⁵ model based on BtGH99¹⁰ (pdb id: 4acy) modified by chainsaw²⁷ 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, Gerdt 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, Gerdt 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, Tausin AS, Spadiut O, Klinger 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, Karpies 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 D* 66:125–132.
 24. Evans PR (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Cryst D* 67:282–292.
 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 D* 50:760–763.
 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 D* 66:486–501.
 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 D* 67:355–367.
 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 D* 66:213–221.
 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 D* 66:12–21.

Versicherung an Eides Statt / Affirmation in lieu of an oath

**gem. § 5 Abs. 5 der Promotionsordnung vom 18.06.2018 /
according to § 5 (5) of the Doctoral Degree Rules and Regulations of 18 June, 2018**

Agata Mystkowska, Borgfelderstr. 20, 28215 Bremen

Ich / I, _____

(Vorname / First Name, Name / Name, Anschrift / Address, ggf. Matr.-Nr. / student ID no., if applicable)

versichere an Eides Statt durch meine Unterschrift, dass ich die vorliegende Dissertation selbständig und ohne fremde Hilfe angefertigt und alle Stellen, die ich wörtlich dem Sinne nach aus Veröffentlichungen entnommen habe, als solche kenntlich gemacht habe, mich auch keiner anderen als der angegebenen Literatur oder sonstiger Hilfsmittel bedient habe und die zu Prüfungszwecken beigelegte elektronische Version (PDF) der Dissertation mit der abgegebenen gedruckten Version identisch ist. / *With my signature I affirm in lieu of an oath that I prepared the submitted dissertation independently and without illicit assistance from third parties, that I appropriately referenced any text or content from other sources, that I used only literature and resources listed in the dissertation, and that the electronic (PDF) and printed versions of the dissertation are identical.*

Ich versichere an Eides Statt, dass ich die vorgenannten Angaben nach bestem Wissen und Gewissen gemacht habe und dass die Angaben der Wahrheit entsprechen und ich nichts verschwiegen habe. / *I affirm in lieu of an oath that the information provided herein to the best of my knowledge is true and complete.*

Die Strafbarkeit einer falschen eidesstattlichen Versicherung ist mir bekannt, namentlich die Strafandrohung gemäß § 156 StGB bis zu drei Jahren Freiheitsstrafe oder Geldstrafe bei vorsätzlicher Begehung der Tat bzw. gemäß § 161 Abs. 1 StGB bis zu einem Jahr Freiheitsstrafe oder Geldstrafe bei fahrlässiger Begehung. / *I am aware that a false affidavit is a criminal offence which is punishable by law in accordance with § 156 of the German Criminal Code (StGB) with up to three years imprisonment or a fine in case of intention, or in accordance with § 161 (1) of the German Criminal Code with up to one year imprisonment or a fine in case of negligence.*

Bremen, 30.04.2019

Ort / Place, Datum / Date

Unterschrift / Signature