Structural and functional organisation of the agarolytic enzyme system of the marine flavobacterium Zobellia galactanivorans

Jan-Hendrik Hehemann

Avertissement

THESE DE DOCTORAT DE
L'UNIVERSITE PIERRE ET MARIE CURIE

Spécialité: Glycobiologie marine

(Ecole doctorale IVIV)

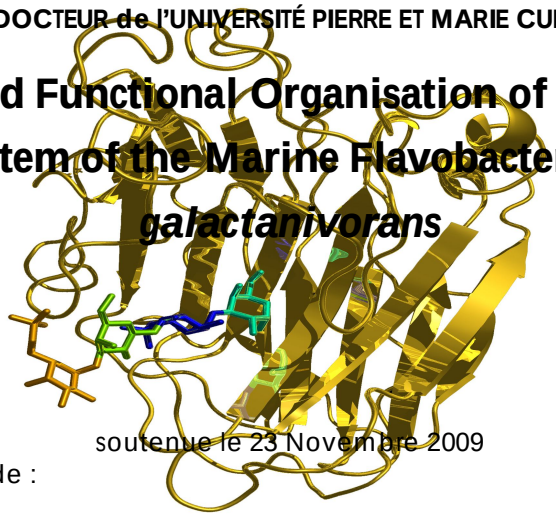
Présentée par

Jan-Hendrik Hehemann

Pour obtenir le grade de

DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE

Structural and Functional Organisation of the Agarolytic Enzyme System of the Marine Flavobacterium *Zobellia galactanivorans*



soutenue le 23 Novembre 2009

devant le jury composé de :

M. Gideon J. Davies	Rapporteur
Mme. Margarete Schüler	Rapporteur
M. Bernard Henrissat	Examinateur
M. Lars Redecke	Examinateur
M. Bernard Kloareg	Examinateur
Mme Mirjam Czjzek	Directrice de thèse

Université Pierre & Marie Curie - Paris 6
Bureau d'accueil, inscription des doctorants et base de données
Esc G, 2ème étage
15 rue de l'école de médecine
75270-PARIS CEDEX 06

Tél. Secrétariat : 01 42 34 68 35
Fax : 01 42 34 68 40
Tél. pour les étudiants de A à EL : 01 42 34 69 54
Tél. pour les étudiants de EM à MON : 01 42 34 68 41
Tél. pour les étudiants de MOO à Z : 01 42 34 68 51
E-mail : scolarite.doctorat@upmc.fr

Für meinen Vater und meine Mutter

Remerciement

Title:

Thesis production despite surfing as continuous source of distraction in Roscoff, France

Authors

Gaelle Correc, Tristan Barbeyron, Etienne Rebuffet, Sylvie Rousvoal, Diane Jouanneau, Francois Thomas, Murielle Jam, Alexandra Jeudy, Justyna Rzonca, Phillipe Potin, William Helbert, Ludovic Delage, Catherine Boyen, Emeline Combot, Simon Dittami, Andres Ritter, Michele Barbier, Agnes Groisillier, Miguel Frada, Valeria Oppliger, Marion Ballenghien, Maryvonne Saout, Dennis Freudenreich, Celine Manceau, Stephanie Salaun, Aude le Bail, Murielle Jam, Sabine Genicot, Jerome Dabin, Stefan Raimund, Peter von Dassow, Gildas Le Corguille, Daniela Mella, Bernard Kloareg, Stephane Hourdez, Lionel Cladiere, Colomban de Vargas, Gurvan Michel, Mirjam Czjzek.

Université Pierre et Marie Curie, Paris 6, CNRS, Station Biologique de Roscoff, F 29682, Roscoff, France

Thank you all!

Thank you Mirjam for importing me to France, and for being a perfect supervisor in life and work. Here I had three fantastic years and could meet amazing people. Gaelle thank you for your incredible help everytime, you are a star. From the lab I especially want to thank Murielle and I am sure, if there would be someone like you in every lab in France, the reverse translatase will certainly be discovered in France. I also thank a lot to Alex, Diane, Francois, Stephanie, Aude, Justyna, Tristan and Sabine for your substantial support in the lab and the office, and for your patience with my "chaotic and cleptomanic" behaviour regarding pencils. You should all receive a decoration from Mr. Sarkozy himself. Etienne you know the *in silico* world, thank you for sharing your knowledge with me. Gurvan thank you for writing, discussing and working with me even though I stressed a lot. I also wish to thank Harry J. Gilbert who was an unoffical

supervisor for me for his advice and support during the three years.

An outstanding thanks is awarded to Gideon Davies and Marga Bauer for their help and for their couragous and brave attemt to crawl through the dark and muddy waters of the first version of this manuscript.

Merci Mirjam de m'avoir importé en France, et d'avoir été une parfaite directrice dans la vie et le travail.
J'ai passé ici trois années fantastiques et j'ai eu la chance de rencontrer des personnes exceptionnelles. Gaëlle, merci pour ton aide incroyable tu es une etoile. Du laboratoire, je veux remercier particulièrement Murielle, ainsi qu' Alex, Diane, François, Stéphanie, Aude, Justyna, Tristan et Sabine : merci pour votre soutien considérable au laboratoire et au bureau. Pour votre patience avec mon comportement chaotique et de cleptomane, vous devriez recevoir une décoration par Mr. Sarkozy lui-même. Etienne, tu connais le monde in silico, alors merci d'avoir partager tes connaissances avec moi. Gurvan, merci d'avoir écrit avec moi et d'être resté cool même quand je stressais beaucoup.

Table of content

“Land comprises 29.2 percent of the surface of the globe; all the rest is occupied by the sea, where the principal mass of green living matter exists and most of the luminous solar energy is transformed into active chemical energy...

These (productive marine ecosystems) are chemical laboratories, with energy more powerful than the most massive forests of solid earth.”

V.L. Vernadsky, The Biosphere (1926), §55-translated by D.B.Langmuir, 1997, Springer Verlag.

Summary

Glycoside hydrolases catalyze one important reaction on earth: the transformation of photosynthetically derived polysaccharides into small building blocks, which constitute the major energy and carbon source at all levels of the food chain. This process is highly efficient in the marine ecosystem, because both turnover of primary production and the subsequent microbial recycling of polysaccharides is accelerated, compared to the terrestrial environment. This makes the microbial glycoside hydrolases a key player in marine and global carbon cycle because polysaccharides do not substantially accumulate in the ocean.

However, and in strong contrast to the enzymes used by terrestrial microbes for plant degradation, the enzymes used by marine microbes for the degradation of algae are largely unknown (the "knowledge gap" of marine glycoside hydrolases). This limits our ability to interpret results from new emerging technologies, such as environmental transcriptomics and metagenomics, to extend our understanding of the function of microbes in the marine ecosystem. We also ignore numerous glycoside hydrolases with potential for interesting new biotechnological applications. Recent efforts to catch up with this "knowledge gap" have in the first instance led to diverse genome sequencing projects, including that of the marine

heterotrophic bacterium *Zobellia galactanivorans*.

The major aim of my thesis was therefore to exploit the genome data of *Z. galactanivorans* and to identify, and subsequently characterize through biochemical and structural methods, the potentially new functions of marine glycoside hydrolase enzymes that were discovered by the expert annotation of the genome (Tristan Barbeyron *et al.* in preparation).

More precisely, during the three years of my thesis in Roscoff, I studied the agarolytic enzyme system selected from the genome data of the marine Bacteroidetes *Z. galactanivorans.* The bacterium contains two glycoside hydrolase family 16 (GH16) enzymes that have previously been identified as β-agarases. The bioinformatic analysis of the genome of *Z. galactanivorans,* carried out at the beginning of my thesis, revealed the presence of 16 family GH16 enzymes in total. Nine of these genes grouped together phylogenetically and contained the two previously identified agarases. This high number of nine related but divergent GH16 enzymes led to the hypothesis that they might be part of an agarolytic system, containing different specificities for the degradation of the chemically diverse red algal agars. Therefore my work concentrated on the analysis of the seven agarase related enzymes to test this hypothesis.

By using a strategy going from the "gene to the structure and function", I applied medium throughput heterologous protein expression, preparative protein purification with HPLC, biochemical and functional characterization, protein crystallization and structural analysis by crystallographic methods, in order to analyze and characterize in detail the selected gene products. Three of the novel family GH16 enzymes could be produced in a recombinant form. One I have identified as being another β-agarase with modified substrate specificity and the other two enzymes were identified as the first β-porphyranases. Their substrate ,porphyran, is a sulfated polysaccharide that belongs to the agar class and is predominantly found in red algae of *Porphyra* spp.. Notably, this polysaccharide has been known for decades (at least since 1961) but porphyranases had never been described before. During my thesis work I have elucidated the crystal structures of these two β-porphyranases as well as the novel β-agarase, and the results will be presented together with their biochemical characterization. The discovery of these new functions together with the new β-agarase shows that the nine GH16 enzymes form indeed a

system for complex agar degradation containing diverse specificities. Moreover, and profiting of the detailed characterization of the determinants leading to the new enzyme class “porphyranases”, we have performed data mining and phylogenetic analysis of all homologous genes present in public databases. As expected, these genes occur in many marine microbes for porphyran and agar degradation from red algae. However, one GH16 porphyranase gene together with another GH16 agarase gene were identified in a human gut bacterium, sampled from Japanese individuals. Our detailed analysis of the gene context of these different gene occurrences allowed the conclusion that a horizontal gene transfer from marine to human gut bacteria occurred. Such a transfer did not occur to bacteria from the Americans or the Europeans because bacterial samples from these populations do not contain porphyranase or agarase genes. In our opinion this transfer was induced by the difference in habit of nutrition, since the diet of Japanese historically contains seaweed such as *Porphyra* also known as nori (the algae in sushi).

I. Introduction

I.1 The marine ecosystem

I.1.1 Heterotrophic bacteria as a key player in marine carbon cycle

Primary production is the photosynthetic conversion of carbon dioxide and water into oxygen and organic compounds such as sugars by plants. Two-thirds of the earth surface is covered by water and approximately one half of the global primary production occurs in the oceans. The produced particulate organic matter is transported into the ocean's interior where it is eventually sequestered in a manner known as the biological pump. This sequestration has generated a carbon sink and has reduced atmospheric carbon dioxide in an ancient process that literally fuels today's industrial world. Nevertheless, the amount of stored material compared to the initially produced biomass is minimal since over 98% is recycled by heterotrophic bacteria (Smith *et al.* 1992; Hedges *et al.* 2001).

The importance of microbial re-mineralization in the oceanic ecosystem is illustrated by the fact that even though one half of the global primary production occurs in the ocean, it is driven by only ~1% of the total marine biomass (Azam et al. 2007). This non intuitive small number is explained by the extreme turnover of marine microalgae, which is the reason for the high level of biological production in the sea. For example, the turnover rate in marine microalgae is 2-6 days compared to 19 years for terrestrial plants (Field et al. 1998). Algal blooms can appear rapidly and they produce high levels of biomass (Figure 1), then die shortly after, for instance through viral attack (Suttle 2005; Frada et al. 2008) and the remains become available for instantaneous consumption by marine microbes.

Figure 1: *Phaeocystis* algal bloom in the North Sea in front of Scotland illustrating the high productivity of the marine ecosystem. Algal blooms quickly produce high amounts of biomass and dissappear giving rise to the high turnover in the oceanic ecosystem (www.visibleearth.nasa.gov). The produced organic matter is to a high extent recycled by marine microbes that transform high molecular weight into low molecular weight compounds with enzymes.

Photosynthetically produced carbon and biogenic material from other levels of the marine food web is either soluble or becomes part of the particulate organic matter (POM) pool. Particulate organic matter also known as marine snow, creates a microhabitat and carbon source for heterotrophic bacteria (Smith *et al.* 1992) (Figure 2).

The other huge fraction, which is the soluble biogenic material in the ocean, is called dissolved organic matter (DOM). Microbes have a key role in cycling of DOM because they are the only organisms that can access this solute carbon pool. These microbes become a nutritional source for planktonic grazers which in turn are consumed by fish. Therefore the microbes funnel carbon and energy to higher trophic levels (plankton/fish) a process which is called the microbial loop (Azam *et al.* 1983; Azam 1998). The microbes are hence ecologically essential by promoting nutrient availability at the base of the food web via the microbial loop. Furthermore

heterotrophic bacteria help directly to release inorganic nutrients otherwise exported from the euphotic surface layer by sequestration to the depth of the oceans. For example, Bidle *et al.*, discovered that bacterial proteolysis of the proteins that protect diatom silica frustules leads to increased dissolution of silicate, and thus hinders the sequestration of this growth limiting nutrient. This indirect release of silicate which is not used as a nutrient by the bacteria thereby possibly impacts the further growth of Diatoms (Bidle et al. 1999). This process in which the algal growth limiting nutrient silicate is liberated by proteolytic activity may equally occur with trace metals liberated by glycoside hydrolases from cation chelating anionic polysaccharides (see below and Figure 2).

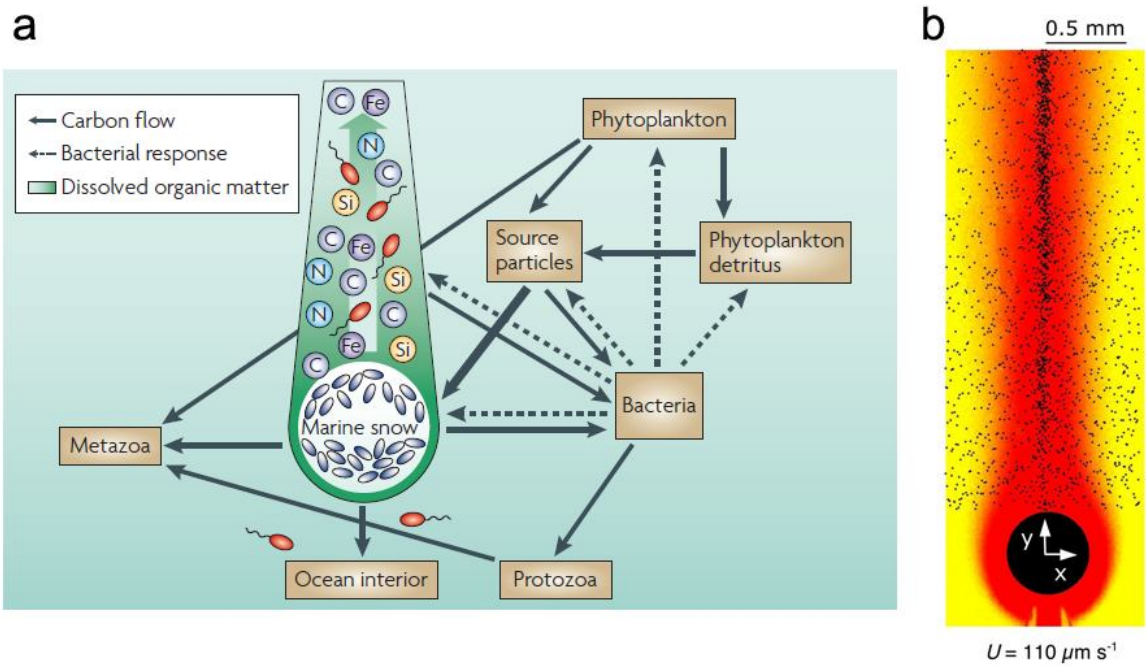


Figure 2: Carbon and nutrient flow from and to particulate organic matter in the sea. (a) The aggregation of particles and their bacterial decomposition are a central part of the carbon cycle in the marine ecosystem. Marine snow is highly colonized by bacteria, and transports carbon, nitrogen, phosphorus, iron and silicon into the ocean's interior. The enzymatic activities of bacteria liberate nutrients which become non sinking dissolved organic matter (DOM) and form plumes in the ocean. Free-living bacteria are attracted to such DOM-rich hot spots to exploit this carbon and energy source. They in turn produce biomass which is fed into the food web (the microbial loop). Copied from (Azam *et al.* 2007). **(b)** Microfluidic experiment showing how nutrient hot spots attract marine heterotrophs created for instance by enzymatic attack from sedimenting particles. Accumulation of *P. haloplanktis* in the nutrient plume of a particle sinking at *U* _ 110 _m s_1. The flow is from bottom to top, the bacteria are represented by the black dots. Red indicates higher nutrient concentration, reproduced from (Stocker *et al.* 2008).

I.1.2 Polysaccharides are an important part of the marine DOM and POM pools

The "Sweet Ocean"

It is estimated that ~40% of the organic matter produced by primary producer's cycles through the dissolved organic matter (DOM) pool (Azam et al. 1983; McCarthy et al. 1996). With ~10^8 g, DOM represents one of the largest dynamic carbon pools on our planet (McCarthy et al. 1996) ranging in the same magnitude as all living vegetation on the earth continents (Hedges 1992). Within DOM, carbohydrates are an abundant component and account for 15-30% of the total mass (Benner et al. 1992). Even though highly diluted (~1mg/L), if one imagines that all DOM carbohydrates (DOM with 30% carbohydrates) present in 1 L of seawater were glucose units in form of pyranose rings and linked in one single polysaccharide chain through β-1,4 linkages (1 unit ~5 Å, measured in cellulose penta-oligosaccharide, pdb; 3I50), the length of this molecule would cover ~3 times the distance from our earth to the moon (1126617 km of 384403 km core to core, Figure 3). This exemplifies the abundance of carbohydrates in the marine ecosystem and shows why their turnover is of such an importance for the global carbon cycling. Furthermore it points to the importance of understanding the mechanisms by which microbial carbohydrate catabolism takes place in the sea and why bacteria are so successful in rapidly degrading these enormous amounts of carbon stored in complex marine polysaccharides.

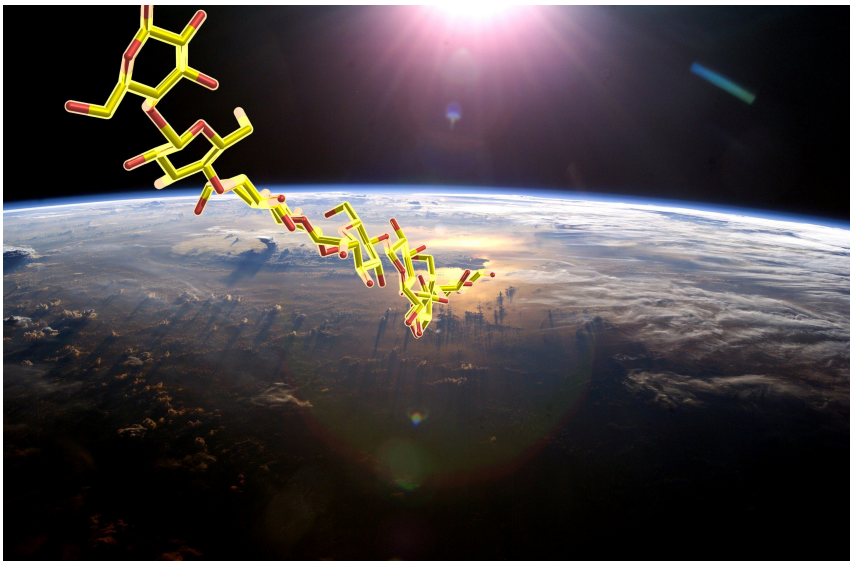


Figure 3: A conceptual image "The Sweet Ocean". The polysaccharide chain illustrates the theoretical carbohydrate chain, of all sugars in DOM of one litre of seawater, covering the distance between the earth and the moon (background: NASA ISS007 Sunrise Pacific Ocean).

I.1.3 Bacterial enzymatic decomposition of marine organic matter

As illustrated and described above, organic carbon matter in marine environments are to a major extent made up of polysaccharides (Benner 1992, Scoog et al. 1997). Unicellular and multicellular algae as well as bacteria have been suggested to be the major contributors to the production of carbohydrates found in DOM and POM (Biersmith et al. 1998). In plants and macroalgae, carbohydrates occur predominantly in the form of cell wall polysaccharides (Gilbert et al. 2008) or storage polysaccharides, whereas microalgae secrete extracellular polysaccharides with diverse functions (antiviral, grazing control, or to lower the sedimentation coefficient and many others) (Myklestad 1995; Wotton 2004). For instance, microalgae produce polysaccharides as carbon shunt, when excess production occurs during photosynthesis (Smetacek et al. 1986) a process not described for terrestrial plants. All these polysaccharides serve as carbon and energy source for bacteria, which degrade high molecular weight compounds including polysaccharides,

DNA and proteins with specific enzymes to reduce molecular mass. This is a critical factor for nutrition uptake in bacteria since only molecules under 6 kDa can pass the cell wall (Weiss et al. 1991).

I.1.4 Why enzymatic degradation of gel forming marine galactans impacts the carbon cycle

The ocean is a "Sweet Jelly"

In the ocean, production and storage of photosynthetically derived organic matter are separated in space. Up to ten thousand meters of water divides surface from seafloor, a distance to be conquered by sedimenting particles. Consequently, processes modulating the sedimentation coefficient of particles, either positive or negative, influence carbon cycle.

Jellifying polysaccharides have a unique role in the marine ecosystem since they are involved in gel particle formation in the sea. These abundant colloidal gel particles (~10% of DOM) are important for the microbial loop, sedimentation processes, element cycling, carbohydrate chemistry and particle dynamics in the ocean (Passow 2002; Verdugo et al. 2004). Marine gels are three-dimensional networks of biopolymers imbedded in seawater. They are abundant in seawater (Alldredge et al. 1993) and their size ranges from single macromolecules over intermolecular assemblies of several hundreds of microns to hectares of sea surface. A macro gel that is to a high extent composed of polysaccharides is the mucilage event, covering huge areas of sea-surface of the Adriatic Sea. It has been proposed that a limitation of microbial degradation by unknown reasons is the factor for this “constipated” carbon flow in the marine ecosystem (Figure 4d) (Umani et al. 2007; Sartoni et al. 2008).

Marine gels can form within minutes to hours from dissolved organic matter released by algae or bacteria. It was the dynamic light scattering (DLS) based laboratory study carried out by Chin et al (Chin et al. 1998), which resolved the previously proposed hypothesis that seawater DOM can spontaneously aggregate to form micro gels of increasing size and therefore create a conduit from DOM to POM (McCarthy et al. 1996). It is assumed that anionic polysaccharides are

essential in this process since they promote micro gel formation due to their stickiness (Chin et al. 1998; Engel et al. 2004). This mechanism can be controlled with EDTA, which chelates divalent cations and abolishes any coagulative activity. In the ocean, the analogous metal chelating ability of anionic residues, most often carboxyl or sulfate groups, on marine polysaccharides are responsible for this coagulation. This implies that anionic polysaccharides influence trace metal cycling as they chelate divalent cations, which are removed from the surface water layers when sedimentation occurs (Hedges 1992; Hedges et al. 2001). Since aggregation leads to increased particle size which results in higher sinking rates, anionic polysaccharides were suggested as a potential carbon sink by increasing the transport of photosynthetically derived organic matter and trace metals to the sea floor (Engel et al. 2004). Bacterial activity may have the diametrical opposite effect, when enzymatically degrading polysaccharides and thereby reducing the downward flux. One can infer that the microbial community (1 million bacteria/ml) contains all the necessary tools (glycoside hydrolases) to degrade the vast complexity of polysaccharides in the oceans. This is because over 98% of the organic matter is recycled and no evidence for selective preservation for instance of polysaccharides was found in sediments (Hedges et al. 2001). However, and despite their crucial role, most of the enzymes that are used by marine bacteria to degrade the oceanic gel phase (dominantly constituted of polysaccharides) are unknown to date.

To study enzymatic degradation of gel forming particulate organic matter is difficult due to the chemical complexity as well as the limited availability of colloidal aggregates. As an alternative to naturally occurring aggregates, agarocolloids from macrophytic red algae and the agarolytic system from the marine Bacteroidetes *Zobellia galactanivorans* are used in this study as a model system for the enzymatic degradation of the oceanic gel phase (Figure 4c).

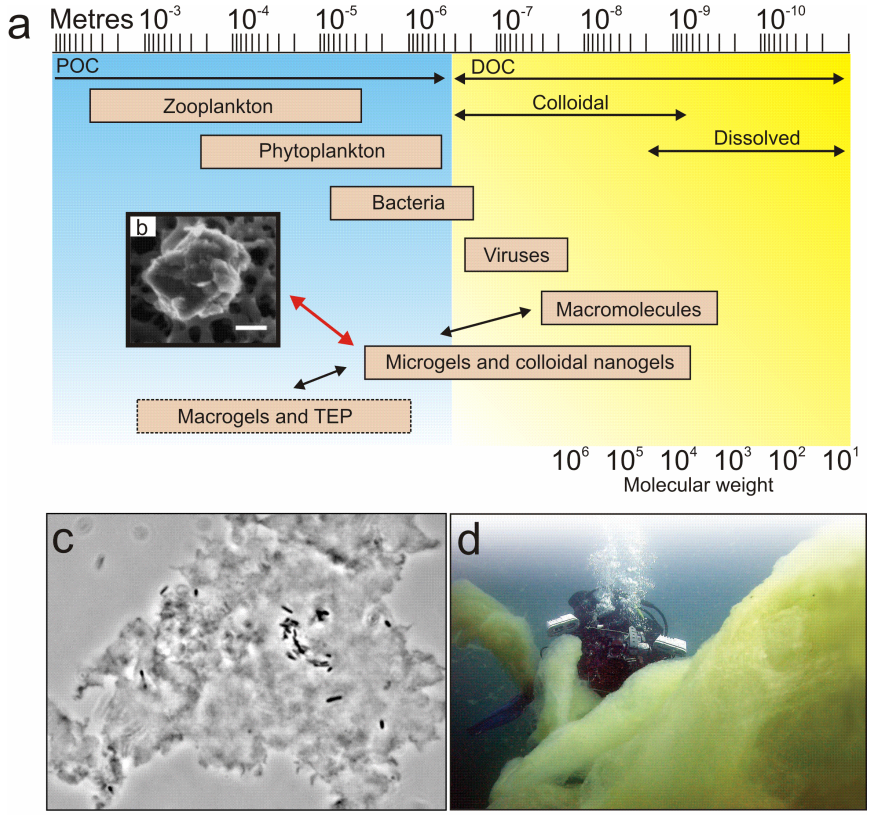


Figure 4: The transition from dissolved to particulate organic matter can be induced by jellifying anionic polysaccharides which are abundant in the marine ecosystem. (a) Here, the full size range of organic matter is shown, from monomers, polymers, colloids and gel particles to traditional particles (divided into particulate organic carbon (POC) and dissolved organic carbon (DOC). TEP, transparent exopolymer particle copied and modified from (Azam et al. 2007). **(b) Inset:** Environmental scanning electron microscopy photograph of a fully hydrated microgel from seawater (scale bar, 2µm) copied from (Chin et al. 1998). **(c)** Microphotograph of an agarose microgel associated with and degraded by the heterotrophic bacterium *Z. galactanivorans*. **(d)** A diver examining the mucilage event in the Adriatic Sea (polysaccharide dominated macrogel).

I.2 Marine polysaccharides

In the first part of the introduction I stressed the importance of microalgae as carbon source in the marine ecosystem. They are significant producers of the polysaccharides which are a major component of the oceanic gel phase, particulate organic and dissolved organic matter. To understand the bacterial tools (enzymes) which allow the high turnover in the marine ecosystem

without a "constipated" carbon flow is therefore of ecological interest. Nevertheless, even though several studies about polysaccharide composition in microalgae have been carried out (Myklestad 1977; Paulsen et al. 1977; Alderkamp et al. 2007), much less is known about their chemical structure and abundance in comparison to the polysaccharides found in marine macro algae. The ecological importance of these macrophytes is not negligible because they potentially function as carbon sinks in contrast to microalgae of which carbon is rapidly recycled (Smith 1981; Gattuso et al. 1998).

Therefore I will focus here on the cell wall polysaccharides of marine macrophytes, many of which are well characterized and commercially available to perform biochemical or microbiological experiments which is not the case for polysaccharides from microalgae.

I.2.1 The cell walls of marine macrophytes

Plant and marine algal cell walls can be seen as a two-phase system in which a crystalline phase, the skeleton, is embedded in an amorphous phase, the matrix. However, the ratio of skeletal over matrix polysaccharides in the cell walls of marine algae is inversed with respect to that in plants; while in plants cellulose generally represents the major fraction, the amount of jellifying matrix polysaccharides of the algal cell walls can reach 80 % of the total mass. Furthermore, these matrix polysaccharides in algae are highly anionic (sulfate or carboxylate groups) whereas neutral polysaccharides are a minor component. In addition, most of these matrix polysaccharides form gels and their abundance in marine macrophytes (Figure 5), or as less characterized transparent exopolysaccharides from unicellular algae or bacteria, indicates the general importance of gel forming polysaccharides in the marine ecosystem. It has been suggested that these anionic gel forming polysaccharides (alginate, carrageenans, agars, fucoidans, ulvans see Figure 6) may be a convergent adaptation to the marine environment (Kloareg et al. 1988).

One common modification of gel forming algal polysaccharides that is completely absent in terrestrial plants is the sulfate-esterification. The diversity of marine algal polysaccharides is

summarized in Table 1.

Figure 5: Algal diversity near Roscoff in the rocky shore intertidal zone. (1) *Porphyra umbilicalis* (red algae), **(2)** *Palmaria palmata* (red algae), **(3)** *Ulva lactuca* (green algae), **(4)** *Fucucs serratus* (brown algae)

Table 1. Polysaccharide diversity in marine algae. Unlike terrestrial plants and most freshwater algae, photosynthetic algae and plants living in the marine environment contain sulfated polysaccharides.

	Chlorophyta (green algae)	Rhodophyta (red algae)	Heterokonts (brown algae & diatoms)
Neutral polysaccharides	Cellulose β-1,3-xylans β-1,4-mannans	Cellulose β-1,3-xylans β-1,4-mannans	Cellulose chitin (diatoms)
	Starch	Floridean starch	Laminarin
Anionic polysaccharides	Macroalgae		
	Ulvans	Agars Carrageenans	Alginates Fucans
	Microalgae		
	Capsulans (*Prasinococcus*)	Sulfated gluco-galactoxylans (*Porphyridium*)	Carboxylated and sulfated polysaccharides (*Phaeodactylum*)

I.2.2 The skeleton component of algal cell walls

Similar to terrestrial plants, most macroalgae contain neutral and linear polysaccharides as cell wall skeleton, the most common polysaccharides being cellulose, xylans and mannans. Pure and highly crystalline cellulose exists in some algal species for instance in *Siphonocladales* (Cronshaw et al. 1958), *Cladophorales* (Frei et al. 1961) and in the *Chlorococcales*. This is generally rare, since alternative building blocks, such as xylose, increase the heterogeneous character of the molecules present in algal cell walls. For example, the cellulose of *Rhodymenia palmata* contains up to 50% of β-1,4-xylose (Cronshaw et al. 1958). The thalli of *Rhodophyceae* and *Phaeophyceae* contain between 1 and 8% cellulose in contrast to higher plants which contain around 30 % (Preston, 1974).

In conclusion, marine algae contain cellulose to a lesser extent than terrestrial plants and most often of a more heterogeneous character (Kloareg et al. 1988). I will further focus on the gel forming, anionic matrix polysaccharides from marine algae, since they are particular to the marine ecosystem, and skeletal cell wall polysaccharides will not further be discussed here.

I.2.3 The matrix component of algal cell walls

In contrast to skeletal polysaccharides the matrix polysaccharides are in most cases extractable with hot water and for the cation chelating matrix polysaccharides additives such as salts, alkaline, acid or chelators are needed. Several of these matrix polysaccharides are used as food additives or for biotechnological applications due to their high jellifying ability (agar, carrageenan, alginate). Sulfated polysaccharides such as fucans, carrageenans and porphyran found recent interest due to their pharmacological activities which include antiviral (Buck et al. 2006), antithrombic (Franz et al. 1996; Alban et al. 1997), antiinflammatory and antioxidative effects (Zhang et al. 2004; Hatada et al. 2006).

A common feature of algal polysaccharides is their high heterogeneity resulting from variable sulfate position (agars, carrageenans, fucans), sometimes not clear evidence of repeating units

(ulvans) and the common presence of branching positions. These modifications which challenge bacterial degradation complicate in the same time the complete structural elucidation of these anionic polysaccharides and limit pharmacological application (Groth et al. 2009). The reason for this highly variable structure is unclear but one possible explanation may be a first defence against microbial attack by limiting enzymatic degradation systems through chemical complexity.

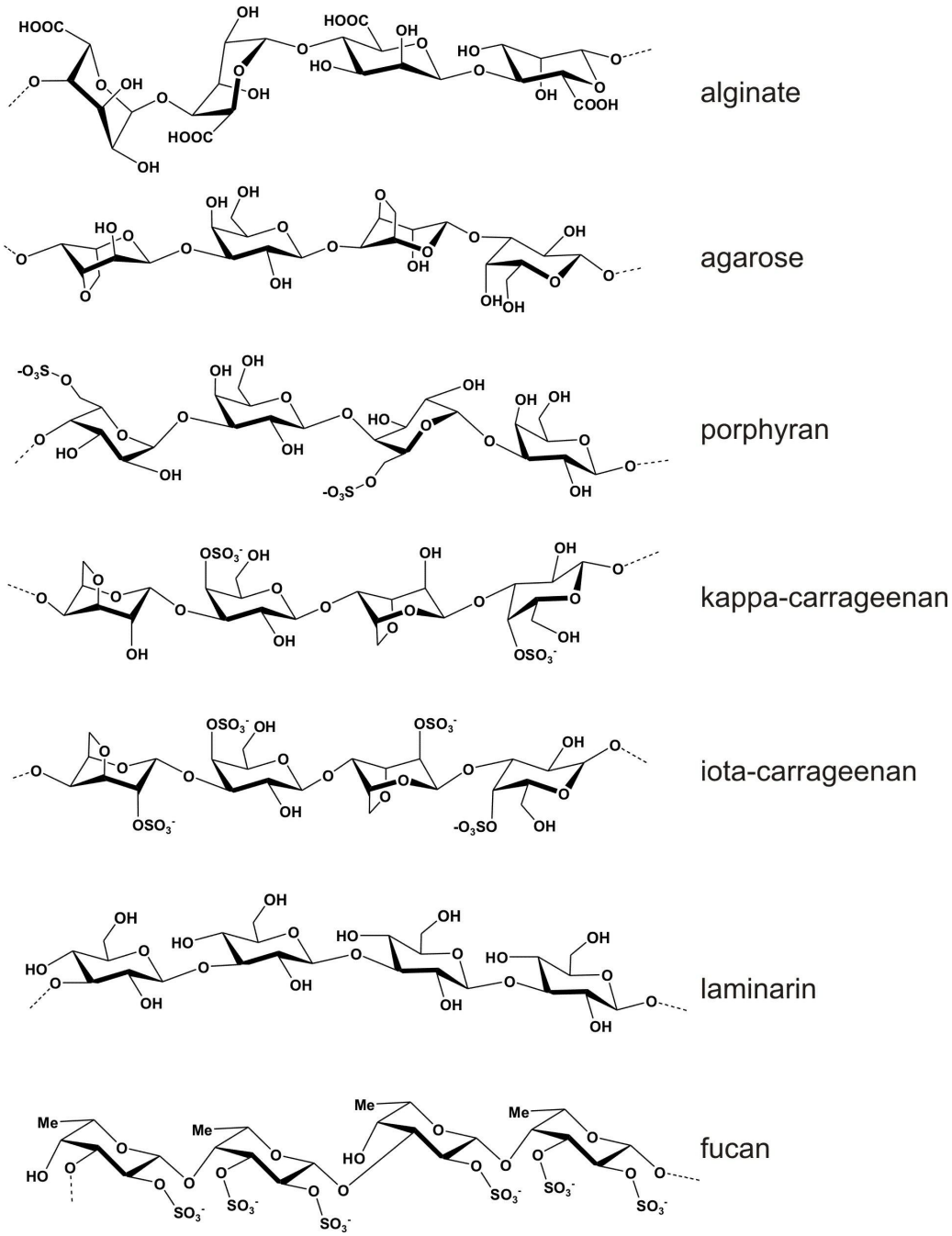


Figure 6: Complex marine algal polysaccharides. Most of these polysaccharides form gels except laminarin which is a storage glucan and is soluble in water.

I.2.4 The green algae

The *Chlorophyceae* contain cell wall matrix polysaccharides reminiscent of the hemicelluloses and pectin found in higher plants though they are sulfated and heavily ramified (Percival 1979). The matrix polysaccharides of *Cladophorales* and *Codiales* are comprised of sulfated xylo-galacto-arabinans of L-arabinose blocks separated by D-galactose residues and with D-xylose branching (Percival et al. 1971). *Ulvales*, *Acrophoniales* and *Urosporales* contain glucurono-xylorhamnans which are also sulfated and additionally carboxylated. They are frequently comprised of repetitive L-rhamnose-2-sulfate and D-glucuronate units, alternatively linked by α-1,3 and β-1,3 glycosidic bonds and further ramified by β-1,4 linked D-xyloses (Percival 1979).

I.2.5 The brown algae

Brown algae are commonly known and exploited for their alginate content, a polysaccharide with calcium dependant gelling ability. Alginate which is widely used as a matrix in medical applications (Reis et al. 2006), as an ingredient in modern food (Mc Hugh 2003) and other applications, such as heavy metal detoxification, due to its high metal ion chelating ability (Davis et al. 2003; Davis et al. 2003). This polysaccharide is one of the recent highlighted ingredients of the "molecular cuisine" used by Ferran Adria (El Bulli four year winner as the world's best restaurant) to create his surprising apple and orange caviar. Alginate is a heterogeneous polymer of the two uronic acids β-D-1,4-mannuronate (M) and its C5 epimer α-L-1,4-guluronate (G). These residues appear as homopolymeric MM and GG blocks or in a heteropolymeric distribution GM (Haug et al. 1967; Haug et al. 1969). The gelling behaviour of this polysaccharide depends on the distribution of G and M units in the polysaccharide chain. This distribution can be modulated by C-5 epimerases which convert the mannuronic into guluronic acid and are therefore of biotechnological interest (Morch et al. 2008).

Brown algae especially the Fucales contain another group of matrix polysaccharides, the fucoidans. These are sulfated polysaccharides, generally of an α-L-fucose backbone. The polysaccharide can be substituted with sulfate-esters, masked with ramifications and also may

contain other monosaccharide residues. In nature, these anionic polymers are of high heterogeneity since they contain a continuous spectrum of highly ramified polysaccharides, ranging from high uronic acid, low-sulfate–containing polymers with significant proportions of D-xylose, D-galactose, and D-mannose to highly sulfated homofucan molecules with approximately three substituents per disaccharide (Kloareg et al. 1988; Colin et al. 2006; Descamps et al. 2006).

I.2.6 The cell wall matrix of marine red algae: agars and carrageenans

Rhodophyta contain mainly linear sulfated galactans either of the carrageenan class (*Carrageenophyta*) or the agar class (*Agarophyta*). The red algal galactans, agars and carrrageenans are widely exploited in various industries due to their special rheological properties (Mc Hugh 2003; Michel et al. 2006). Both polysaccharide classes form gels, the strength of which is modulated by substitutions on the linear molecular backbone. This backbone consists of alternating 1,4-linked β-D-galactose (G unit) and 1,3-linked α-L-galactose (L unit agar) or 1,3-linked α-D-galactose (D unit carrageenan), following the nomenclature by Knutsen et al. (Knutsen et al. 1994) (Figure 7).

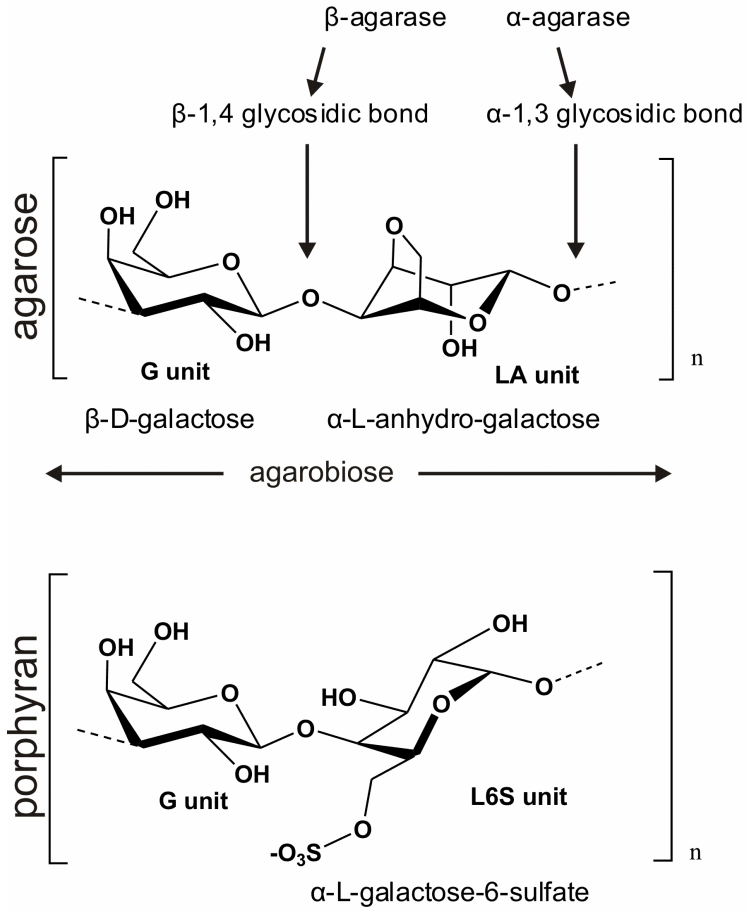


Figure 7: The repetitive motifs in agarose and in porphyran and the linkage specificity of the different agarases (see below). Please note the difference between G-LA or G-L6S motifs which are different from the neoagarobiose or neoporphyrobiose motifs LA-G or L6S-G, respectively.

In agarose, the neutral fraction of agar, the L unit is modified by a 3,6-anhydro bridge (LA) (Craigie 1990) that stabilizes the molecule by allowing helix formation. In natural agars from red algae, the G and L units can be variably substituted by methyl and sulfate groups (Figure 8), (Maciel et al. 2008). Composition and regularity of the primary structure of carrageenans and agars directly affect their gelling behaviour. The presence of disaccharide units, where the 3,6-anhydro bond is replaced by sulfation on C6 leading to L-galactose-6-sulfate units, markedly affect gelling behaviour in agars and in carrageenans. Such modified structures are considered as

biological precursors, found especially in actively growing tissues of red algae, where this sulfated sugar is enzymatically converted into the anhydro sugar by the action of a sulfoeliminase (Rees 1961). The fate of the precursor agarocolloid pool has been studied in *Gracilaria chilensis* fed with ^{13}C enriched CO_2. In this study it was shown by mass spectrometry that the majority of the produced L6S units were subsequently transformed into LA units (Hemmingson et al. 1996). Nevertheless, a substantial amount of residual L6S units that remained in the developed tissue indicates that the potential biological function of this modification is more than solely a precursor, even though still not understood. Moreover, in certain red algal species, such as in developed *Porphyra* spp., the majority of L-galactoses in the agar primary structure occur as L6S units, underlining the latter argument.

For biotechnological applications, agars are chemically (alkaline) treated to obtain the neutral agarose with high gelling quality. Indeed, agars and carrageenans form thermo-reversible gels in solution and their respective melting temperatures and the gel rigidity are strongly influenced by ester sulfates masking the polysaccharide chain. In the natural context, as cell wall polysaccharide, agars are modified by various substitutions such as methyl-ethers or ester-sulfonic groups (Craigie 1990). More rarely one can also find pyruvic acid modifications, which form a ketal cycle with the D-galactose, strongly influencing the physico-chemical properties of the extracted polysaccharides. Agars can be extracted from *Gelidium*, *Gracilaria*, *Ahnfeltia*, *Acanthopelis* and *Pterocladia*. The highest quality of agar is extracted from *Gelidium* which contains a low amount of modifications and high amount of LA units resulting in high gelling quality. Therefore *Gelidium* species are industrially exploited as agar source. But the discovery of an alkaline pre-treatment, which transforms the sulfated agars into neutral agarose allowed to industrially exploit low quality agars from *Gracilaria* spp. (Nisizawa et al. 1987).

A highly sulfated polysaccharide is the earlier mentioned porphyran, the major cell wall matrix component found in *Porphyra* spp. and *Bangia* spp.. Porphyran contains predominantly α-L-galactose-6-sulfate (L6S unit) instead of α-L-3,6-anhydro-galactose and is further highly methylated on the C6 of the G units. Due to the L6S modification, the polysaccharides lack the high gelling capacity of agarose therefore porphyran solutions are viscous. The extracted

polysaccharides from *Laurencia* or *Polysiphonia* are even further from the ideal agarose structure, since their structural units that resemble porphyran α-L-galactose units are further methylated in position 2 and sulfated in position 6 (Batey et al. 1975). The polysaccharides of *Ceramium rubrum* are similar to those in *Laurencia* but the L-galactose units are seldom sulfated (Turvey et al. 1976)(Figure 8).

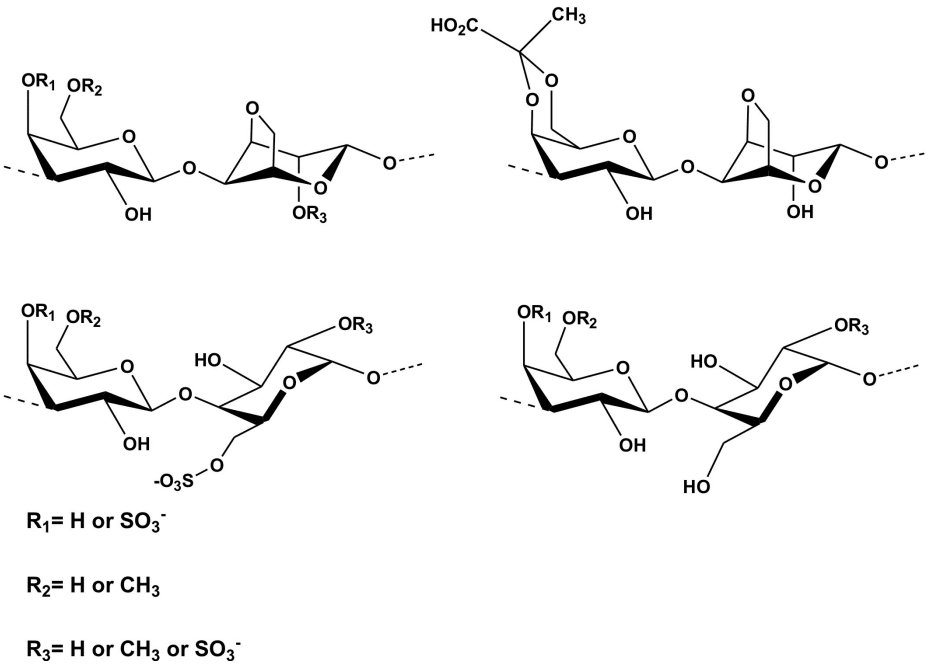


Figure 8: Agarocolloid diversity in marine red algae. Different substitutions such as methylations, sulfatations or pyrovate groups have been reported to decorate the neutral disaccharide agarose unit.

I.2.7 Agarose 3D structure

It is still matter of debate whether agarose molecules form single or double antiparallel helices that anneal and lead to the gel state of agarose. In their early work, Arnott et al. analyzed X-ray diffraction patterns of agarose fibres (Arnott et al. 1974) and proposed a double helical structure as interpretation of their results (Figure 9). In the double helical conformation the antiparallel agarose chains would form left handed helices with a pitch of 1.9 nm (length of repeating unit

h=0.633nm) and threefold symmetry. This double helical structure would have an inner cavity of 0.45 nm in diameter, enough to host water molecules that would be involved in the stabilization of the strands through hydrogen bonds with O-2 of the G and O-5 of the LA unit. In contrast, Ford and Atkins interpreted their experimental data obtained from dried films with a single helical structure of extended helices and a pitch ranging from 0.89 to 0.97 nm (Ford et al. 1989). These experimental results were supported by modelled agarose strands based on the agarobiose units (Jiminez-Barbero et al., 1989) and the interpretation of UV-spectra recorded from dried films (Arndt et al. 1994). On the other hand, the double helical structure initially proposed by Arnott et al., received recent support by Allouch et al., who modelled the double helical structure by extending two agarose oligosaccharides that were bound in the active site and at a secondary binding site in their protein crystal structure of the β-agarase AgaA mutant from *Z. galactanivorans* (Allouch et al. 2004).

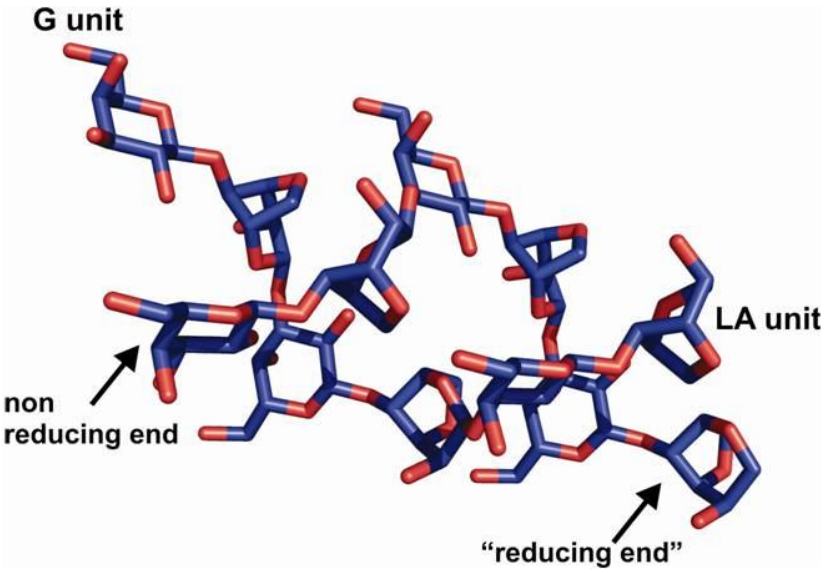


Figure 9: Double helical structure model proposed by Arnott *et al.* (Arnott *et al.* 1974) (pdb code 1AGA).

Both models highlight the importance of the anhydro-bridge that stabilizes the helical structure by inter and/or intramolecular hydrogen bonds and are in agreement with the fact that LA

substitution of C6 (L6S) with sulfate-ester abolishes gel formation which can be explained by the different backbone structures between agarose and porphyran (Rees et al. 1982; Morrice et al. 1984) (Figure 10). In conclusion, the precise quaternary assemblage of agarose chains in the gel state is still to be resolved.

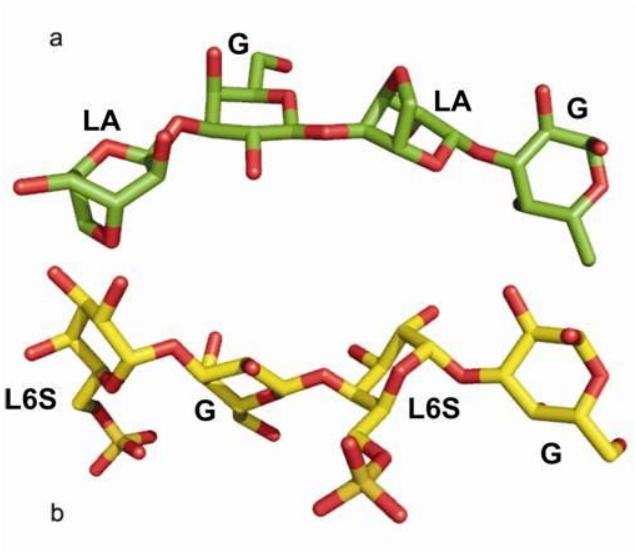


Figure 10: Neoagarose-tetrasaccharide compared with the neoporphyrose tetrasaccharide. (a) In agarose the 3,6-anhydro bond locks the pyranose ring in the $^{1}C_{4}$ conformation leading to the regular agarose backbone structure. **(b)** In porphyran the anhydro bond does not exist due to the sulfation on C6. The 3,6-transannular interaction between C6 and the C3-OH when the L-galactose pyranose ring is in the $^{1}C_{4}$ conformation is energetically unfavourable and the ring switches from $^{1}C_{4}$ to the $^{4}C_{1}$ conformation.

In summary two factors can be defined which render these polysaccharides recalcitrant against microbial degradation. First of all, the various chemical modifications that mask the polysaccharide chains may result in adapted degradation systems with enzymes of different specificities. Secondly, the physical gel forming behaviour (Helbert et al. 2006) of these matrix galactans render these polysaccharides recalcitrant against enzymatic degradation. This heterogeneous phase problem is also found in fibrillar or crystal forming polysaccharides, the most recalcitrant case is cellulose, and results from the fact that a dissolved enzyme has to degrade an insoluble substrate. Terrestrial bacteria have evolved highly complex degradation machineries such as cellulosomes (Bayer et al. 1998), or degradation systems including enzymes

with different mode of action or enzymes with different modules, to efficiently degrade the heterogeneous phase (Ferreira et al. 1997; Fernandes et al. 1999). This heterogeneous phase problem may equally play a role in the aggregation of jellifying polymers, a process that limits enzyme accessibility.

On can assume that microbes adapted to the chemical variability in agars and target this complexity by secretion of numerous hydrolytic enzymes with varying specificies and/or modes of action. This has been described for neutral agarose (Ekborg et al. 2005) but never for chemically diverse galactans in the sea. As mentioned in the first paragraph, the recycling of marine polysaccharides is mainly performed by marine bacteria and their enzymes are largely unknown.

I.3 The marine heterotrophic bacteria

I.3.1 Marine bacteroidetes

Autotrophic bacteria, aside heterotrophic bacteria dominate the bacterial biomass in marine surface waters. Among the heterotrophic bacteria the most abundant groups are the *Proteobacteria* and the *Cytophaga-Flavobacteria* cluster (Kirchman 2002). The importance of these bacteria in the degradation of high molecular weight polymers and therefore in the carbon cycle has been shown (Cottrell et al. 2000). Members of *Bacteroidetes* are frequently associated with marine particles (Delong et al. 1993) and their comparative genome analysis has recently been subject of a "Marine Genomics Europe" project aiming at understanding their environmental impact. One of the sequenced genomes, from the marine bacteroidetes *Gramella forsetii*, already indicated its ability to degrade polymeric organic matter (Glöckner et al. 2003; Bauer et al. 2006).

Within the same comparative genomic project, the genome of the marine bacterium *Z. galactanivorans* was analyzed and the thorough manual annotation was performed by members

of the group (namely Tristan Barbeyron and Gurvan Michel; manuscript in preparation).

I.3.2 Zobellia galactanivorans

This heterotrophic, marine bacteroidetes was isolated from the surface of the marine red algae *Delesseria sanguinea* (Barbeyron et al. 2001) (Figure 11). *Z. galactanivorans* is a Gram-negative ,non-spore-forming, rod shaped bacterium (0.2 by 1.5-8.0 μm). The colonies are yellow or orange due to a flexirubin-type pigment. The cells do not possess flagella but display gliding locomotion, and they are chemo-organotrophic and heterotrophic. The metabolism is respiratory, not fermentative and strictly aerobic, with oxygen as the final electron acceptor. The bacterial cells were found to efficiently hydrolyze the galactans from red seaweeds such as agar and diverse carrageenans, observed by the strong gel hydrolysing ability of colonies on these gels. In recent experiments, where marine algal polysaccharides and others were used as sole carbon source, it was shown that *Z. galactanivorans* can grow on a multitude of algal substrates, many of which are sulfated. An additional interesting point is that the use of cellulose, a major land plant polysaccharide, as substrate, did not result in any growth (Barbeyron T. and Michel G. personal communication).

Due to its high ability to degrade diverse marine algal galactans, it has been included in the comparative genome sequencing project. The genome analysis allowed the identification of several new agarase-like sequences, which were selected to be subject of the present study (see below).

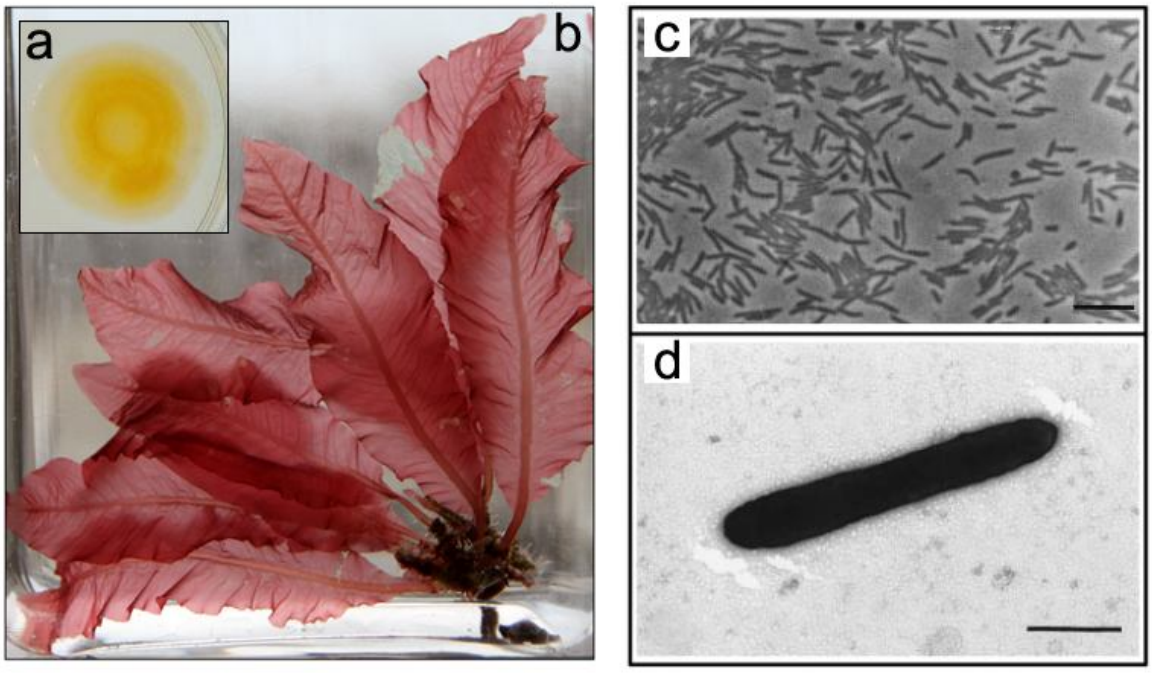


Figure 11: The marine heterotrophic bacterium *Z. galactanivorans* collected from the red algae *Delessaria sanguinea*. (a) A colony of *Z. galactanivorans.* **(b)** The red algae *Delessaria sanguinea*. Microphotograph of *Z. galactanivorans*. **(d)** Electron microphotograph of *Z. galactanivorans* (Barbeyron *et al.* 2001).

An overall genome analysis revealed that *Z. galactanivorans* contains 143 glycoside hydrolases. This high number reflects its ability to decompose marine algal polysaccharides. The genome also contains 72 sulfatases and such a high number of sulfatases are only outnumbered by one other marine bacterium, the marine planctomycet *Rhodopirellula baltica*, which contains 103 sulfatases (Glöckner et al. 2003; Bauer et al. 2006). In contrast to the high number of sulfatases found in these marine bacteria, the number of sulfatases in soil bacteria is marginal. For instance, *Flavobacterium johnsoniae* contains 12 and many soil bacteria entirely lack sulfatases. One can therefore speculate that the high number of sulfatases is a marine ecosystem specific niche-adaption to cope with the high amount of sulfated polysaccharides (Figure 6) in this special ecosystem. Interestingly the proteobacterium *Saccharophagus degradans* (Weiner et al. 2008) does not contain sulfatases, reflecting its life style in the brackish estuarine environment with

carbon input from land. *Z. galactanivorans* is the marine bacterium with the highest number of algal polysaccharide degrading glycoside hydrolases of which the crystal structures are solved to date.

The detailed bioinformatic analysis of the gene context of glycoside hydrolases and sulfatases in the genome of *Z. galactanivorans* showed that these genes often clustered together in operon-like structures that very often either included TonB-dependant receptors, accompanied by SusD-like proteins, and/or two-component systems. This type of putative operon structure has been found connected to pathogenesis (Blanvillain et al. 2007) and has been named carbohydrate utilization locus, or CUT locus. The study shows that the presence of these operons is important for the adaptation to its host and the authors suggest that an overrepresentation and the presence of CUT loci designate the ability to scavenge carbohydrates.

I.4 The glycoside hydrolases

I.4.1 Enzymatic transformation of HMW compounds

Proteins and DNA are considered as easily degradable due to their limited number of building blocks (two pyrimidines and two purines in RNA and DNA) (usually 23 L-amino acids in proteins) and the uniform kind of linkages. For example RNA rapidly degrades on the bench, due to the ubiquity of RNAses. In contrast, in polysaccharides and other glycan structures post translational modifications on glycoproteins challenge bacterial degradation systems due to their high diversity and complex structure of assemblage. For instance, for a theoretical hexasaccharide sequence formed by different sugar units, more than 1.05x1012 possible oligosaccharide structures can be calculated. In contrast, a set of six amino acids generates only 46 656 different structures, which is 7 orders of magnitude lower (Laine 1994). This enormous theoretical and observed carbohydrate diversity resides from seven major structural features.

These are i) epimers of the D and L configuration, ii) linear primary structure or branching, iii) ring size, iv) anomeric configuration, v) linkage position, vi) branching positions and vii) reducing terminal attachment, which combine all into various structures with identical mass. This huge chemical variation found in glycans is also the reason for the large abundance of glycoside hydrolases displaying different substrate specificities, and that are classified into families based on their sequence similarity (httpd://www.cazy.org) (Cantarel et al. 2008).

I.4.2 Glycoside hydrolases and their sequence based classification

Glycoside hydrolases (GH) are the enzymes responsible for degradation of carbohydrates to produce mono- and oligosaccharides from substrates with higher degree of polymerization. Glycoside hydrolases are ubiquitous throughout all kingdoms of life where they catalyse the cleavage and formation of glycosidic bonds, reactions which are important in a myriad of biological processes.

The high diversity of their substrates is reflected by the number and diversity of carbohydrate acting enzymes or CAZymes, which are found with highly divergent sequences, different folds, and catalytic mechanisms. However, since the carbohydrate diversity exceeds by far the number of protein folds, CAZymes have evolved from a limited number of ancestors by acquiring novel specificities at substrate and product level (Cantarel et al. 2008).

Carbohydrate active enzymes have been classified into families based on their amino-acid sequence similarity. The carbohydrate active database (CAZy database) merges the present knowledge on 115 glycoside hydrolase, 91 glycosyltransferase, 19 polysaccharide lyase and 52 carbohydrate-binding module families (http://www.cazy.org/).

I.4.3 The catalytic mechanism of glycoside hydrolases

The glycosidic bond is hydrolysed by two critical residues: the acid/base and the base/nucleophile residue that are generally aspartates or glutamates (Davies et al. 1995). The hydrolytic attack occurs via two possible mechanisms which result either in retention or inversion of the configuration of the anomeric carbon (Koshland 1953; Sinnott 1990; Vocadlo et al. 2008). The distance between the two involved catalytic residues differs between these two mechanisms, although exceptions do exist. If the catalytic mechanism involves retention of the anomeric carbon the distance is around 5.5 Å, whereas this distance is generally bigger in enzymes (6.5Å-9Å) of inverting mechanism, since here the activated water molecule requires further space between the nucleophile residue and the anomeric carbon. Inverting glycoside hydrolases proceed via a direct displacement mechanism, in which an oxocarbenium-ion like transition state is formed. The two carboxylic active site residues are placed at an appropriate distance to provide the base catalytic assistance activating a water molecule (nucleophile), while the second catalytic residue (proton donor) provides acid assistance by protonating the glycosidic bond oxygen at the point of cleavage. The catalysis in retaining glycoside hydrolases proceeds via a double-displacement mechanism, in which a covalent glycosyl-enzyme intermediate is formed and subsequently hydrolysed by an oxocarbenium ion-like transition state (Withers 2001) (Figure 12).

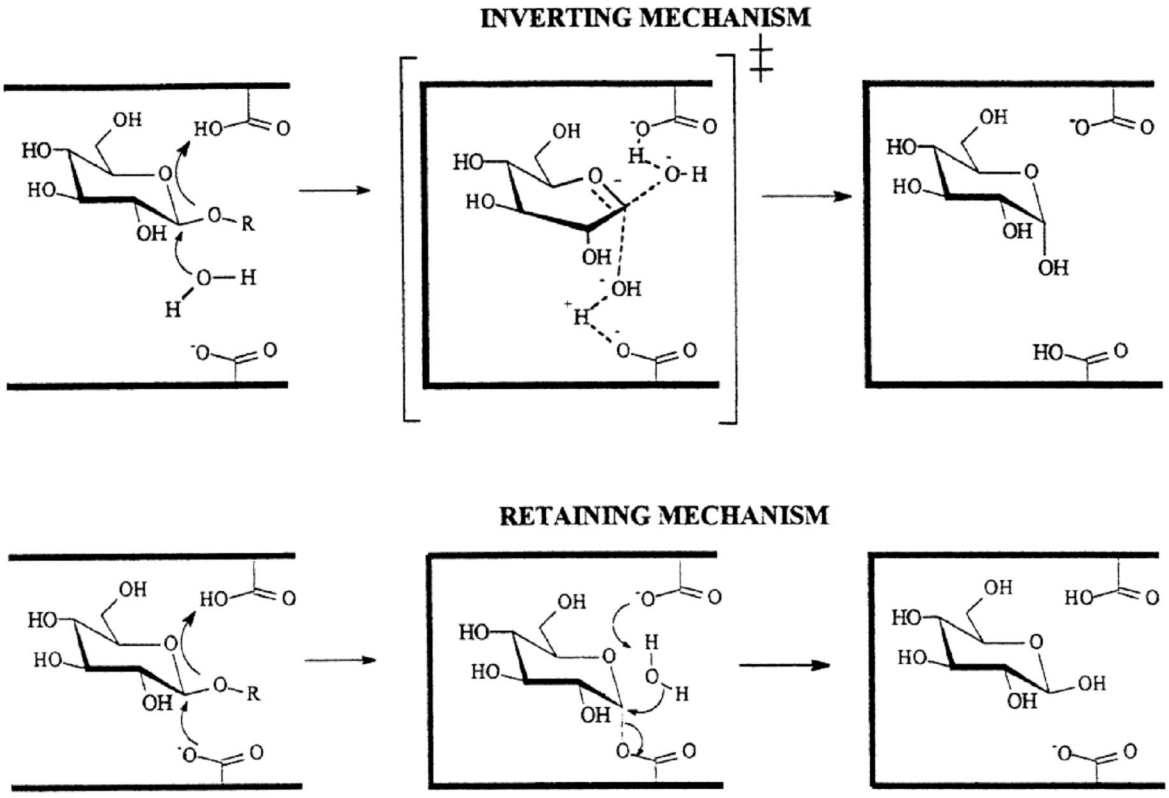


Figure 12: Glycoside hydrolase mechanism of inverting and retaining enzymes (from Withers 2001).

I.4.4 Mode of action in glycoside hydrolases

Besides their catalytic mechanism, which separates retaining from inverting enzymes, glycoside hydrolases can be additionally classified by their mode of action. Three modes of action are generally defined by the outcome of the biochemical analysis. Keeping in mind that these definitions are not completely strict in nature, one can define exo-, endo- or a processive mode of action. Exo enzymes specifically attack chain ends of the substrate and then dissociate from their substrate, while endo enzymes cleave anywhere along the chain and dissociate after the reaction. Processive enzymes usually divided into endo- and exo- processive; catalyze several successive

rounds of hydrolysis before dissociating from the substrate. The precise characterization of the mode of action of an enzyme is not straightforward, but generally the protein structure analysis can complement the biochemical data. An enzyme with a pocket topology points towards an exo-activity, the presence of a catalytic groove can randomly bind on a polymeric substrate and is indicative of enzymes with an endo-mode of action, but is equally found in processive enzymes. Finally, in the tunnel or tyroid topology that has been described for enzymes with a processive mode of action, large loops or domains close down on the polymer chain that is funnelled through the enzyme, which most probably decreases the dissociation rate.

Within family GH16 examples for all these different topologies have been described. The recent crystal structure of a β-galactosidase from *Clostridium perfringens* (Tempel et al. 2005) revealed the pocket topology, while both β-agarases display the open-groove catalytic site and the κ-carrageenase CgkA possesses the tunnel topology, which is reminiscent of processive hydrolases (Divne et al. 1994; Divne et al. 1998) (Figure 13).

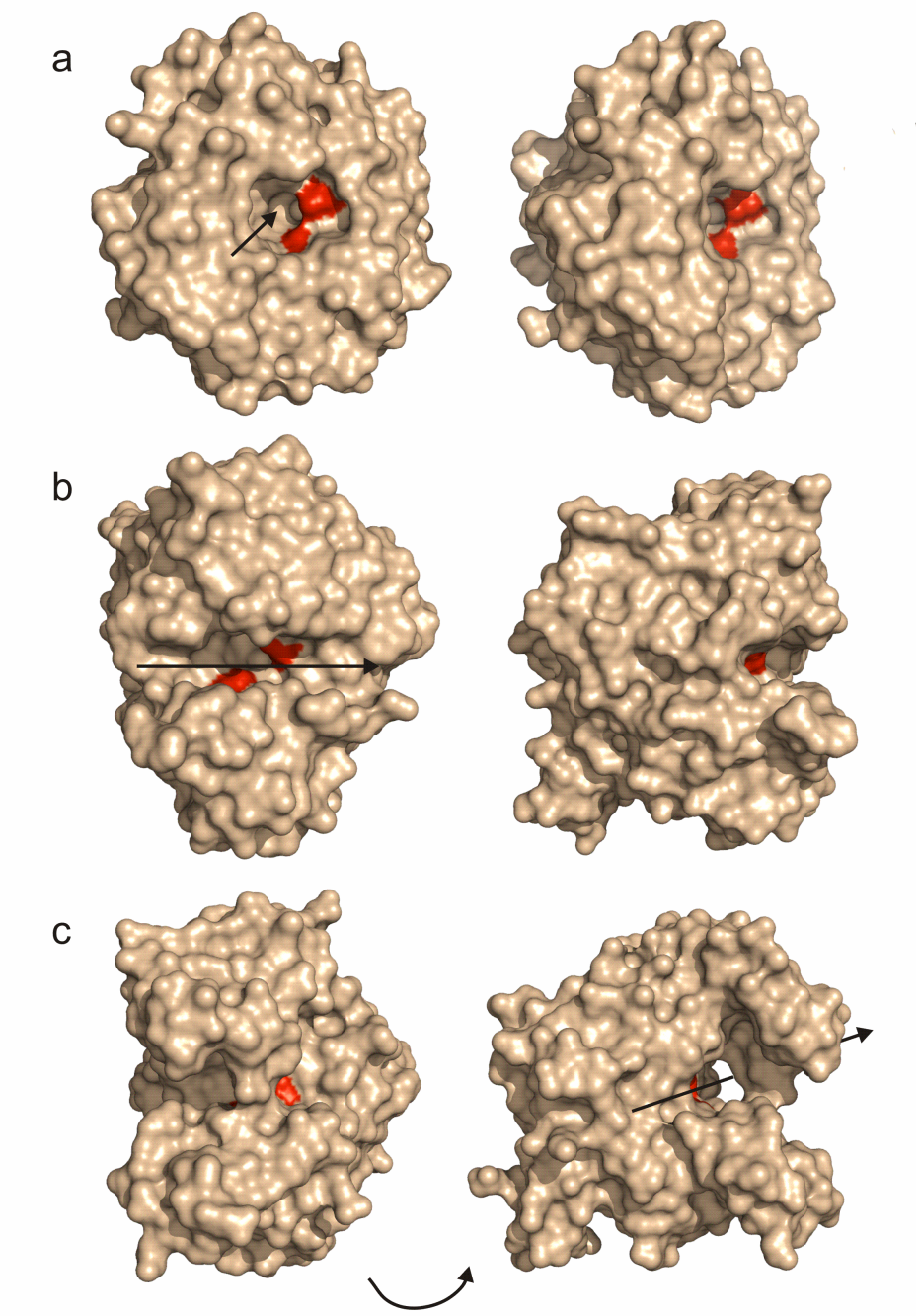


Figure 13: These three GH16 enzymes present the different active cleft topologies pocket, open cleft and the tyroid. (a) The β-galactosidase with the pocket topology. **(b)** The β-agarase A with the open cleft. **(c)** The κ-carrageenase with the tunnel topology. The arrows indicate the substrate binding with the tip pointing towards the reducing end of the substrate.

I.4.5 Subsites in glycoside hydrolases

Within the substrate binding cleft of an enzyme, positive (+) and negative (-) substrate binding subsites can be defined, with respect to the catalytic residues, at the point of cleavage. This nomenclature has been adopted for all glycoside hydrolases, whereby definition, –n subsites represent enzyme/sugar ring interactions with the glycon (new reducing end) and +n subsites represent the interactions with the aglycon (new non-reducing end) (Davies et al. 1997). By definition the point of glycosidic bond cleavage is situated between the -1 and +1 binding sites (Figure 14).

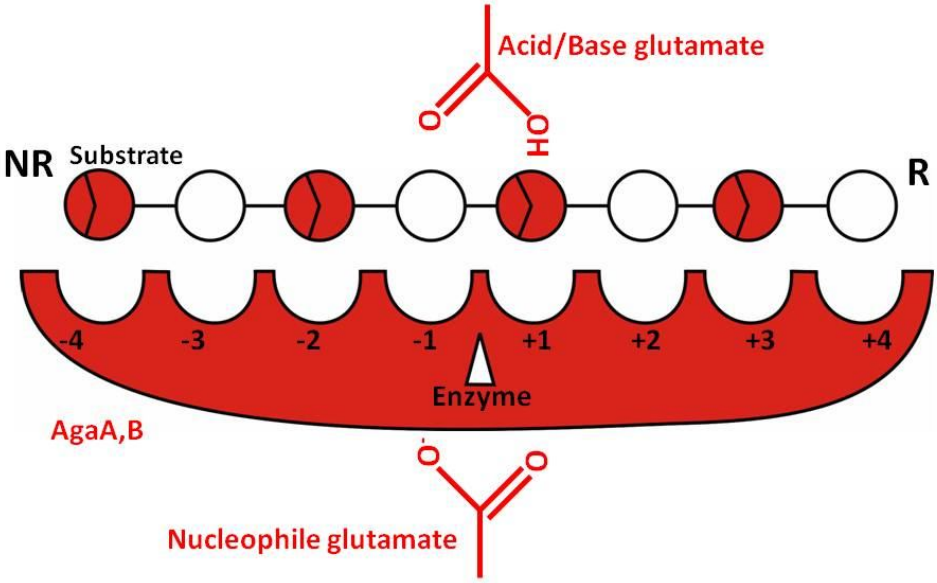


Figure 14: Subsite nomenclature in glycoside hydrolases. Here shown for the β-agarases AgaA and AgaB from *Z. galactanivorans*. Red circles show the L-galactose and white circles the D-galactose rings found in agarose. **NR** stands for the non reducing end and **R** for the reducing end of the substrate oligosaccharide.

I.4.6 Agarolytic bacteria and glycoside hydrolases

The first agar-degrading microorganisms were isolated from a Norwegian fiord by Gran in 1902. Since then at least 30 microorganisms with agarolytic activity have been reported. The vast majority of these bacteria are of marine origin belonging to *Cytophaga* (Duckworth et al. 1968; Duckworth et al. 1969; Duckworth et al. 1969; Van der Meulen et al. 1975), *Microbulbifer* (Ohta et al. 2004), *Pseudomonas* (Ha et al. 1997; Kang et al. 2003), *Pseudoalteromonas* (Belas 1989; Vera et al. 1998; Ivanova et al. 2003; Schroeder et al. 2003), *Microscilla* (Zhong et al. 2001), *Vibrio* (Aoki et al. 1990; Sugano et al. 1993; Sugano et al. 1994; Sugano et al. 1994; Araki et al. 1998), *Alterococcus* (Shieh et al. 1998), *Alteromonas* (Potin et al. 1993), *Thalassomonas* (Ohta et al. 2005) , *Saccharophagus* (Ekborg et al. 2005; Ekborg et al. 2006) and *Zobellia* (Allouch et al. 2003; Jam et al. 2005). Agarase activity has also been observed in terrestrial organisms, such as *Paenibacillus* (Hosoda et al. 2003) and *Streptomyces* (Bibb et al. 1987); genes of agarases have been found in soil by metagenomics (Voget et al. 2003), and interestingly agarase activity was found in an unidentified hospital contaminant (Swartz et al. 1959). Currently there is only one report of an agarase purified from an eukaryote, the mussel *Littorina mandshurica* (Usov et al. 1975). Because agarolytic bacteria have been isolated from various far eastern mussels, it is likely that the agarase activity reported for *L. mandshurica* results from an associated bacterial symbiont.

Agars serve as carbon source for these diverse microorganisms, which degrade this class of polysaccharides with glycoside hydrolases specific for the alternating α-1,3 und β-1,4-glycosidic linkages. Based on their specificity, agarases can be classified into two major categories, β- and α-agarases. β-Agarases cleave the β-1,4 linkages to produce oligosaccharides of the neoagarobiose series with D-galactose at the reducing end. The α-agarases cleave the α-1,3 glycosidic linkages and produce oligosaccharides of the agarobiose series with 3,6-anhydro-L-galactose at the reducing end.

One of the most intensively characterized agarolytic systems originates from *Pseudoalteromonas atlantica* (Day et al. 1975; Groleau et al. 1977). This system consists of an extracellular endo-β-agarase I, which depolymerises agarose to neoagarotetraose. The end-products are subsequently processed by two periplasmic enzymes, the β-agarase II and a neoagarobiose hydrolase, finally yielding 3,6-anhydro-L-galactose and D-galactose (Morrice et al. 1983; Morrice et al. 1983). A similar mode of agar degradation was reported for *Pseudomonas elongata* (Vattuone et al. 1975) and *Flavobacterium flevense* (Van der Meulen et al. 1975). In contrast, *Vibrio* sp. JT0107 secretes two β-agarases that depolymerise agarose to neoagarobiose and neoagarotetraose, respectively which are subsequently degraded by a periplasmic α-L-galactosidase. This enzyme cleaves the α-1,3-linkages of the neoagaro-oligosaccharides from the nonreducing ends (Sugano et al. 1993; Sugano et al. 1994). Recently the agarolytic system from *Saccharophagus degradans* (Ekborg et al. 2005) has been characterized. This heterotrophic proteobacterium was isolated from decomposing cord grass and was the subject of a genome sequencing project, due to its biotechnological potential. Functional studies have shown that this bacterium degrades a wide array of complex polysaccharides including those originating from algal, fungal and plant sources (Weiner et al. 2008). The agarolytic system present in this microbe has been extensively studied on a genomic and proteomic level (Ekborg et al. 2006), and has been found to be composed of five agarases belonging to three different glycoside hydrolase families, namely GH16, GH50 and GH86.

I.4.7 Family classification of agarases

The sequence based glycoside hydrolase classification has the advantage of relating the structural fold to the function of the enzymes. Consequently, divergent evolution leads to polyspecific enzyme families. On the other hand, this classification will separate enzymes that display the same substrate specificity but that have convergently evolved on different protein folds. As an example, cellulases can be found in 12 different families. In the same manner, the actually known β-agarases are classified into three glycoside hydrolase families, GH16, GH50 and GH86

(http://www.cazy.org) and various β-agarases have been biochemically characterized (for review (Michel et al. 2006).

Interestingly, although active on the same substrate polysaccharide, the three α-agarases known to date fall into their own family, GH96, and are much less abundant, (Flament et al. 2007). Indeed, the environmental abundance of ß-agarases appears much higher when inferred from the available number of β-agarase like sequences in the actual databases (NCBI, protein sequence database: 233 β-agarase sequences compared to only three α-agarases). Traditionally, agarases are screened on agar plates and the positive agar lyses from environmental, from expression clone libraries with *E. coli* as host, or from purified enzymes can be easily detected by the depression created in the gel by enzymatic activity (Figure 15). This method equally detects α-agarase or β-agarase activity and is thus unbiased, compared to the reducing sugar assay that gives negative results with α-agarases. Since the number of reported β-agarases outnumbers the α-agarases, for which the screening method is most probably not the reason, one can assume that α-agarases are less abundant in the marine environment.

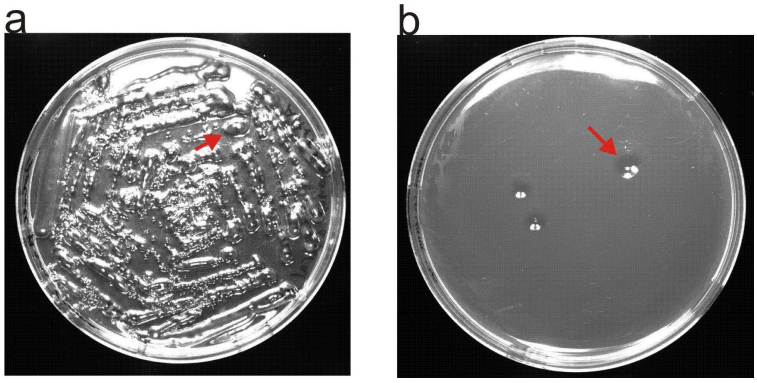


Figure 15: Agarolytic activity demonstrated on an agar plate. (a) *Alteromonas agaralytica* streaked out on an agar plate strongly degrades the agar around the colonies. **(b)** The recombinantly expressed and purified α-agarase from the same bacterium has the same effect and leads to a hole in the substratum. The arrows indicate the degradation of agar by the enzymatic activity.

The β-agarase family GH16 contains the most agarases in the databases (Tristan Barbeyron personal communication). β-Agarases, and the here presented β-porphyranases provide an interesting model system to investigate the molecular bases of the heterogeneous gel forming polysaccharide degradation by marine heterotrophic microbes.

As mentioned above (The cell wall matrix of marine red algae: Agars and carrageenans) certain components of agar, such as porphyran, may consist mostly of sulfated sugar units. Yet it is unclear whether this type of agarocolloid is degraded by β-agarases displaying very broad substrate specificity, or if specific enzymes have evolved to degrade these agar-variant structures. Until the start of my thesis, no porphyran specific activity had been reported, although various β-agarases from marine bacteria have been used to analyze the primary structure of red algal galactans especially of porphyran from *Porphyra* spp. (Morrice at al. 1983).

I.4.8 The glycoside hydrolase family GH16

As mentioned previously, the family classification of glycoside hydrolases is sequence based. A major consequence of this is that the catalytic mechanism is in general the same throughout a GH-family of enzymes (Davies et al. 1995; Henrissat et al. 1997). However, recent exceptions have been reported (Gloster et al. 2008). Another consequence is that while some families contain enzymes displaying one and unique specificity, as family GH6 for cellulases or GH11 for xylanases, others including GH1, GH5 or GH13 covers a wide range of substrate specificities. The glycoside hydrolase family 16 (GH16) is such a polyspecific family with eight described specificities and includes more than 900 sequences to date. Numerous structures have been described, two of which are active on red algal galactans, namely the β-agarases and a κ-carrageenase. All GH16 enzymes contain the same catalytic residues and share the retaining reaction mechanism with retention of the configuration at the anomeric carbon (Keitel et al. 1993). The eight enzyme activities that have been described so far cover lichenases, xyloglucan endotransferases, keratan-sulfate endo-1,4-beta-galactosidases, glucan endo-1,3-beta-D-

glucosidases, endo-1,3(4)-beta-glucanases, xyloglucanases, β-agarases and κ-carrageenases to which we can add the new activity of β-porphyranases described in my thesis work (Table 2).

Table 2: The different specificities described for GH16 enzymes.

Activity	EC number	PDB structures
Endo-1,3(4)-β-D-glucanase	3.2.1.6	1
Laminarinase	3.2.1.39	1
Lichenase	3.2.1.73	3
β-agarase	3.2.1.81	2
κ-carrageenase	3.2.1.83	1
Keratan-sulfate endo-1,4-β-galactosidase	3.2.1.103	-
Xyloglucanase	3.2.1.151	1
Xyloglucan:xyloglucosyltransferase	3.2.1.207	2
β-porphyranase	3.2.1.-	2

Both β-agarases, AgaA and AgaB from *Z. galactanivorans* as well as the κ-carrageenase CgkA from *P. carrageenovora* share the jellyroll fold, formed of two β-sheets with of seven β-strands each. These two β-sheets are stacked in a sandwich like manner and this sandwich is twisted around the potential carbohydrate substrate to create an extended binding cleft. This cleft is surrounded by strand connecting surface loops which influence specificity and the mode of action of the enzyme (see below). The catalytic event in retaining enzymes takes place via a double-displacement mechanism that involves two strictly conserved carboxylic residues (Koshland 1953; Sinnott 1990). These catalytic residues, located in the centre of the substrate binding cleft, are characterized by the conserved sequence pattern ExDx(x)E of active GH16.

The first glutamate of this pattern could be identified as the nucleophile and the last glutamate as the acid/base in the 1,3-1,4-β-glucanase from *B. macerans* (Hahn et al. 1995). The central aspartate residue, which is also strictly conserved, seems to maintain the nucleophile glutamate in the negatively charged state (Kleywegt et al. 1997). Even though the catalytic machinery of different enzyme structures within family GH16 is conserved, local residue substitutions modulate substrate specificity at the active site.

I.4.9 The β-agarases from *Z. galactanivorans*

The first GH16 enzyme purified from *Z. galactanivorans was* a κ-carrageenase which was expressed by the bacterium in presence of carrageenan as substrate. This enzyme could be purified to homogeneity from the culture supernatant and was subsequently cloned and characterized (Potin et al. 1991; Barbeyron et al. 1998). When *Z. galactanivorans* is cultivated in the presence of agar the culture supernatant contains β-agarase activity specific for the β-1,4 glycosidic linkages in the polysaccharide. At least two agarase genes coding for different β-agarases were identified by functional cloning in *Z. galactanivorans* (Jam et al. 2005). These genes were named *agaA* and *agaB*. The product of gene *agaA* codes for a protein with 539 amino acids and contains a N-terminal sequence of 19 residues which probably targets the gene product into the extracellular medium (von Heijne 1983) and is cleaved of in the mature protein (Figure 16).

Further sequence analysis revealed that the gene product of *agaA* is modular, and contains three domains which are the catalytic GH16 module coupled to two C-terminal modules of unknown function. The high affinity of the native AgaA enzyme to sepharose beads (cross linked agarose) suggests that these non characterized modules might consist of agarose specific carbohydrate binding modules (Murielle Jam personal communication).

The gene product of *agaB* is a protein of 40.7 kDa molecular weight and 335 residues and the 20 first residues encode a hydrophobic signal peptide which does not contain the peptidase I specific

cleavage site (Nakai et al. 1991) (PSORT) but a possible lipid anchor, suggesting that this protein remains cell wall associated after its expression (Jam et al. 2005). The sequence analysis showed that AgaB is a monomodular enzyme, consisting only of the GH16 catalytic module.

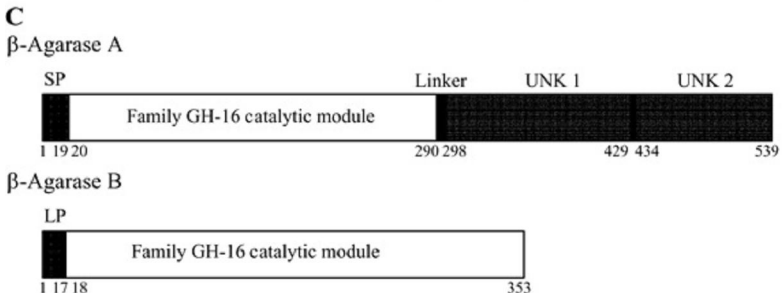


Figure 16: The gene organisation of the two first described β-agarases from *Z. galactanivorans* copied from (Jam et al. 2005).

The biochemical analysis of these β-agarases revealed that the major reaction products of both are neo-agarohexaose and neo-agarotetraose and that AgaB can further degrade the neo-agarohexaose into neo-agarotetraose and neo-agarobiose. The degradation efficiency of both enzymes was highest on neo-agarooctaose indicating that these enzymes possess a substrate binding cleft consisting of eight subsites, four on the non reducing end (-) and four on the reducing end (+). This interpretation was also confirmed by the crystal structures of both enzymes (Allouch et al. 2003).

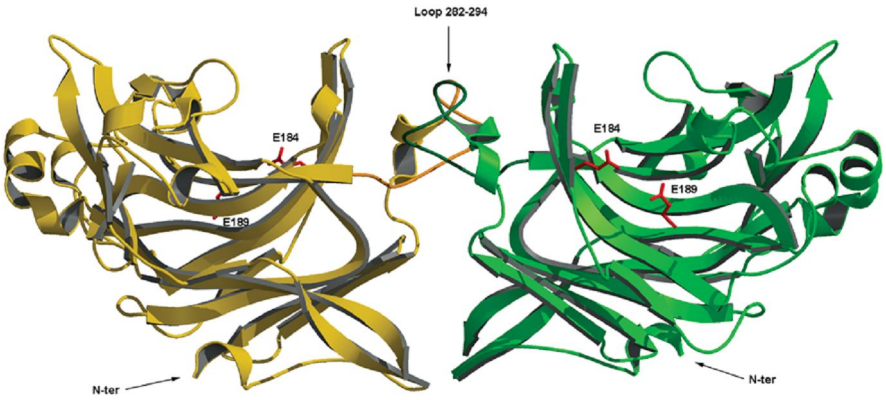


Figure 17: The β-agarase AgaB from *Z. galactanivorans* forms a biological dimer copied from (Jam et al. 2005).

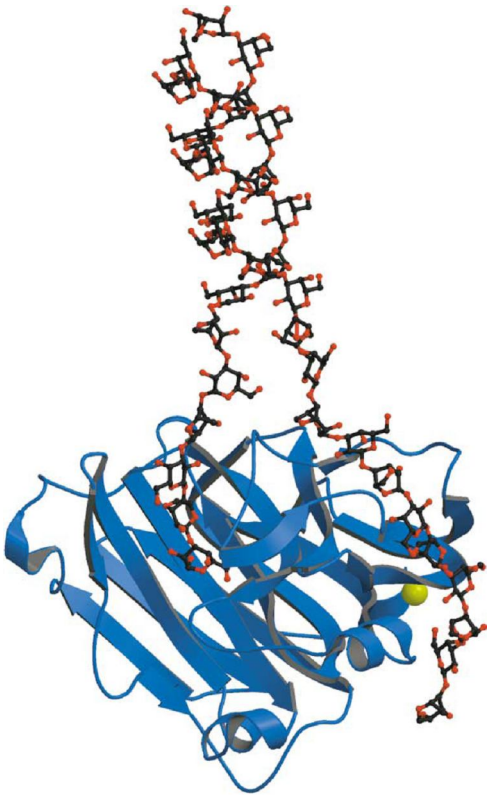


Figure 18: The β-agarase AgaA from *Z. galactanivorans* contains a second binding on the back of the enzyme possibly involved in the unwinding of the double helical agarose chain as here modelled by extending the two bound oligosaccharides copied from (Allouch et al. 2004)**.**

Furthermore the crystal structure analysis revealed some particular features of both enzymes. AgaB forms a biological dimer induced by an unusual loop in GH16 between residues 282-294 (Figure 17). This dimer may be due to the biological role of AgaB as cell wall bound glycoside hydrolase and it was confirmed by gel filtration experiments.

When the activity of the β-agarases were compared on solid state agarose gel it revealed that AgaA had a significantly higher efficiency in agarose gel degradation than AgaB and the protein crystal structure analysis gave an elegant and surprising explanation for this phenomenon. AgaA has a second agarose binding site which is located at the opposite side of the enzyme in respect to the substrate binding cleft (Figure 18). By extending the oligosaccharides bound to both binding sites the double helical agarose model, proposed by Arnott and Fulmer et al. 1974, could be modelled without disturbing the stereochemistry of the glycosidic linkages and the sugar planes. It was postulated that this new mode of binding could allow AgaA to unwind the double helical agarose which would explain why this enzyme is so active on agarose gel in respect to AgaB (Allouch et al. 2004).

I.5 The "knowledge gap" of marine polysaccharides degrading enzymes

In the past, advances in microbiology, including marine microbiology, depended mostly upon culturing. The new age of metagenomics enables the study of the vast majority of microbial species which are as yet unable to be cultivated in the laboratory. These technologies and the analyses they enable (comparative (meta)genomics, (meta)transcriptomics, (meta)proteomics, metabolomics, high throughput gene disruptions, etc) have ushered in a new era of biology with fundamental implications for basic research and biotechnological advances. But, they also pose challenges, especially in marine biology, since this flow of data also highlight the lack of knowledge concerning marine specific metabolisms.

Although projects as the Sorcerer II Global Ocean Sampling Expedition have highlighted the fact that much of the phylogenetic and biochemical diversity of life on Earth is present in its marine microbes, they also highlight that we are way behind in understanding the diversity and functioning in the marine environment. - Up to 60 % of the sequences have no equivalent to previously studied proteins. There are huge numbers of putative genes, the function of which is often unknown and at best only deduced from sequence comparisons. Because more is often known about the genetics and physiology of terrestrial organisms, the number of unknown/putative genes is overwhelming for marine samples because there is so little experimental data on marine model organisms. As an example, despite their ecological importance the number of biochemically and structurally described "marine" glycoside hydrolases is low compared to "terrestrial" glycoside hydrolases (Elifantz, Waidner et al. 2008). Moreover, when identified, these enzymes often constitute completely new protein families: ι- and λ-carrageenases (Michel et al. 2003), α-agarases (Flament et al. 2007) or fucanases (Colin, 2006). Therefore, these enzymes are only accessible through the application of standard biochemical approaches, since in any genomic approach they would have been annotated as “conserved hypothetical proteins” or given incorrect substrate specificities. The keyword search "glycoside hydrolase" in the Protein Data Bank (21.09.09) gave 413 structure hits of which 170 can be attributed to cellulase or related (CBM) functions. In contrast, a search for marine polysaccharide specific glycoside hydrolases such as agarases results in five structural hits and a search for carrageenase resulted in three hits. Crystal structures for unexploited polysaccharides ulvan or fucoidan do not yet exist, not mentioning enzymes active on unknown substrates from ecologically important unicellular algae. This bias towards biotechnologically important terrestrial enzymes follows an economical rationale but does not reflect the ecological importance of bacterial polysaccharide processing in the sea. The economical interest in marine glycoside hydrolases may change since marine algae currently gain in interest as possible producers of biofuels, increasing the possible applications of new enzymes. Moreover, biochemical characterization of new marine enzymes may help to understand the functioning of the ecosystem.

Most enzymes currently used in biotechnology are of microbial origin. The microbial world

contains the greatest fraction of biodiversity in the biosphere. It is thought that the marine environment, covering more than 70 percent of the earth's surface, contains ~4 x 10^{30} microorganisms. Commercial expectations are that microbes will deliver the greater part of enzyme diversity and the majority of new applications. However, the well-known dilemma of microbes, whatever their origin , that the majority cannot be cultivated, limits the application of the traditional means of enzyme discovery. Here it was therefore decided to use a different approach and to exploit the available genomic data of one marine bacterium, sequenced because of its algal polysaccharide degrading capacity. Indeed, with this approach that consisted in the combination of knowledge based genome mining, medium throughput cloning and expression, natural substrate screening as well as classical biochemical and structural characterization, it was possible to identify and fully characterize a new glycoside hydrolase function.

I.6 Aim: finding new glycoside hydrolases by analysing the agarolytic system of *Zobellia galactanivorans*

New glycoside hydrolases and enzymes in general can be screened in a number of ways. Expression clone libraries, in which genetic information is linked with an expressed protein, can be screened for a certain substrate. This was demonstrated for the agarases AgaA and AgaB, and numerous other reported agarolytic enzymes. This strategy is especially convenient with agarases since their activity can be directly monitored. An activity based screening approach has the potential to detect new glycoside hydrolase families but depends on available substrates and activity assays. The latter is rarely available for biotechnologically non-exploited polysaccharides from marine algae. If algal polysaccharides were available, marine bacteria could be screened for their ability to grow on these substrates for instance as sole carbon source. The active protein can be purified and the sequence partially determined by classical biochemical approaches such as edman or mass spectrometry protein sequencing. The obtained sequence information can be used to amplify the target gene with degenerated primers to obtain the gene sequence allowing further molecular biological analysis (Hehemann et al. 2008). This approach

is laborious but it enables the discovery of new glycoside hydrolases belonging to new families.

Another possibility is to choose sequences of known GH families in genome data bases that display significant divergence from enzymes of known function, hinting at new specificities. Combined with the technical advances (such as medium throughput cloning strategies) that have been reached through structural genomics, one can select a high number of promising targets, analyse the soluble produced enzymes ‘from the genome to the biochemical characterization’ and eventually identify new specificities/functions through this method. This is the knowledge based approach that was applied in this study.

II. Results

A rational approach based on genome analysis and phylogenetic analysis of the known GH family 16 was used to find new activities.

The genome analysis, carried out by Tristan Barbeyron and Gurvan Michel, revealed the existence of 16 GH16 enzymes in the genome of *Z. galactanivorans* (Figure 19a)(Table 3). The phylogenetic analysis of the 16 GH16 shows that they form four clades. Clade1 contains two enzymes without predicted function. Clade2 contains four enzymes which are predicted to be laminarinases. Clade3 contains a κ-carrageenase and Clade4 contains the two previously described β-agarases from *Z. galactanivorans* (Allouch et al. 2003; Allouch et al. 2004; Jam et al. 2005) together with seven phylogenetically related enzymes. The five enzymes Zg2600_PorA, Zg1017_PorB, Zg3376_PorC, Zg3628_PorD and Zg3640_PorE do not contain the key residues for agarose recognition, which are conserved in AgaA and in AgaB. Therefore, these five enzymes could not be predicted as agarases but are possibly acting on agarocolloids. In contrast, the new Zg4243 AgaD enzyme, which is closely related to AgaA and AgaB, contains the key residues which were shown to interact with the agarose pyranose-rings in the AgaA oligoagarose complex structure (Allouch et al. 2004). Therefore this enzyme could be confidently predicted to be a β-agarase.

For the following analysis a special focus was put on Clade4, because of the unusually high number of seven agarase related GH16 enzymes. This suggested a new function to allow the

bacterium a broader exploitation of the chemically diverse red algal galactans. Nevertheless the new enzymes from all four clades (except the previously described AgaA, AgaB and CgkA) were selected for a medium throughput strategy (see below) with the aim to analyse them from gene to function. Therefore a total of 12 (AgaC was not compatible with the restriction enzymes used in this study) from the 16 enzymes were selected.

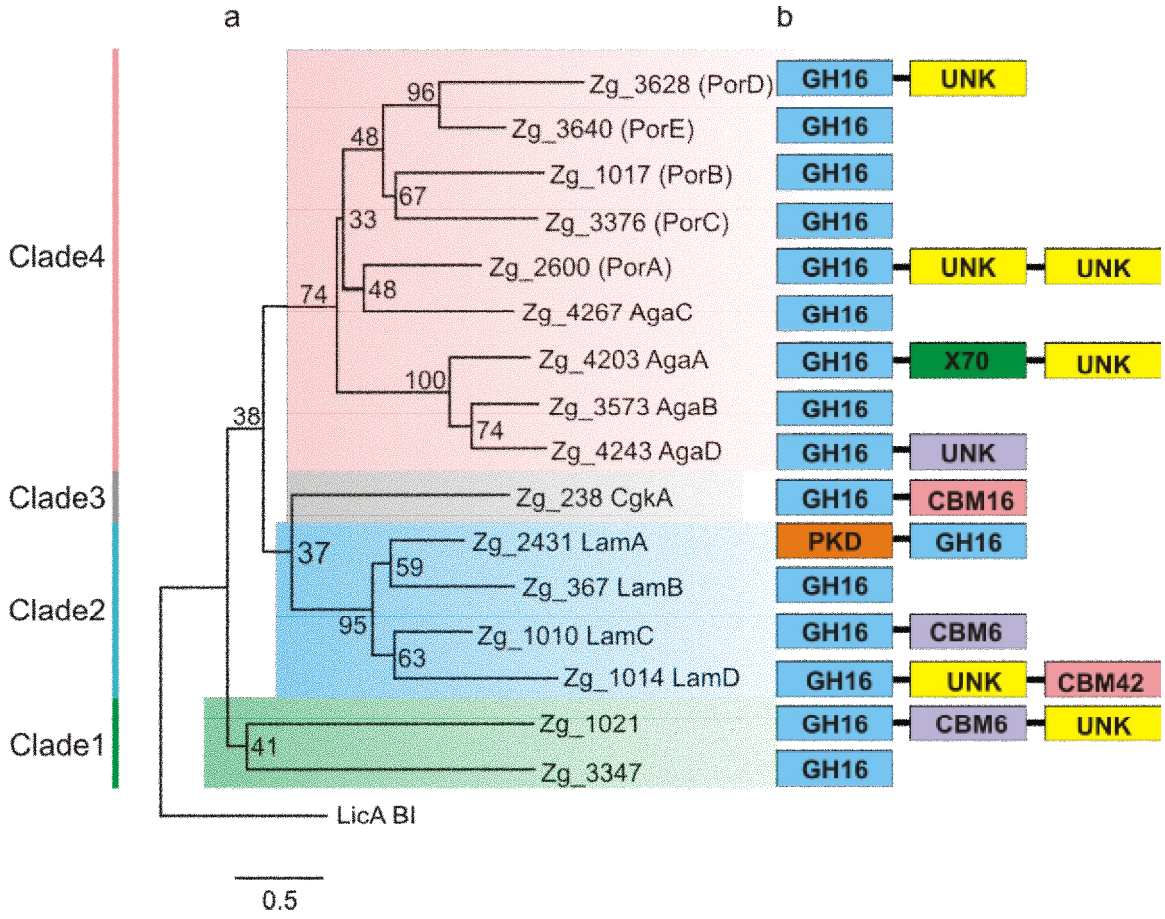


Figure 19: Genomic analysis of *Z. galactanivorans* reveals 16 GH16 enzymes. (a) The 16 GH16 sequences were used to construct a phylogenetic tree. **(b)** The sequence analysis revealed that nine of the 16 enzymes are multimodular and contain the catalytic GH16 module plus one or more modules of unknown function (UNK) or carbohydrate binding modules (CBM) belonging to different families (http://www.cazy.org).

Table 3: Phylogeny matrix showing identity and similarity between all agarolytic GH16 modules (Clade4) of *Z. galactanivorans*. The identity is coloured in red and the similarity in yellow**.**

	1	2	3	4	5	6	7	8	9	10
1. Zg4203_AgaA		35.6	19.6	32.1	25.2	19.8	27.6	22.8	20.4	24.2
2. Zg3573_AgaB	54.2		25.5	35.5	22.0	23.1	21.7	20.0	21.0	18.4
3. Zg4267_AgaC	37.9	46.1		20.6	27.5	22.6	23.9	22.8	21.9	17.9
4. Zg4243_AgaD	47.0	56.3	38.0		21.8	20.2	21.5	17.9	18.0	21.2
5. Zg2600_PorA	43.2	38.7	44.7	36.1		31.8	32.7	26.9	25.1	23.3
6. Zg3628_PorD	37.6	39.6	40.2	35.5	46.2		38.2	33.1	27.1	20.8
7. Zg3640_PorE	43.5	36.9	38.6	34.6	52.1	52.1		36.2	25.9	23.7
8. Zg1017_PorB	42.5	37.2	39.5	34.4	48.5	50.7	50.0		30.0	21.0
9. Zg3376_PorC	34.4	43.9	40.9	36.1	39.8	43.0	39.8	44.5		19.4
10. Zg236_CgkA	40.7	36.9	40.8	34.6	40.0	41.7	39.6	43.2	36.2	

II.1 Medium throughput cloning and expression strategy

In order to produce recombinant enzymes for a subsequent biochemical and structural analysis, the GH16 enzymes were analysed in respect to their modular architecture. Seven genes revealed modular arrangement of a catalytic module and one or more accessory modules (Figure 19b). Some of these additional modules could be identified as carbohydrate binding modules (Boraston

et al. 2004) whereas others are of unknown function.

These unknown modules, CBMs and catalytic modules are connected via flexible linker regions. Since protein flexibility is a known problem hindering protein crystallisation, the proteins with several modules were truncated by PCR. The CBMs and unknown modules were cloned seperately for for further biochemical characterization.

This led to separated catalytic, unknown and carbohydrate binding modules which were cloned in parallel. Among the 13 enzymes, AgaC (Zg_4367) was not compatible with the cloning strategy and for LamB (Zg_367) and Zg_3376 no positive clones could be obtained even after repeating the experiment a second time. Therefore 10 of the 13 enzymes could be cloned and together with their modules or truncated versions this led to a total of 22 constructs which were subsequently analysed by protein expression studies in *E. coli*. The expression analysis revealed that fourteen of the 22 constructions led to soluble expression in Bl21 (DE3) pLysS cells (Figure 20). It should be noted here that no alternative expression strains were tested. Especially in cases where not even insoluble expression occurred codon optimized strains may be a further alternative, while for the cases where the entire produced protein occurred in an insoluble form, folding optimized strains may help to obtain a higher amount of soluble expressed protein. Another aspect is the mode of induction which was accomplished by auto induction (Studier 2005) a method that sometimes does not give the same level of host response leading to lower expression compared to the induction by IPTG. For weakly expressed candidates IPTG may therefore help.

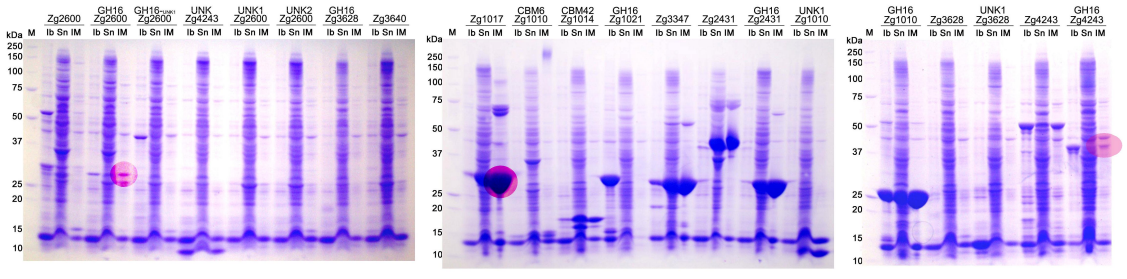


Figure 20: SDS-PAGE expression analysis of all targets cloned within the medium throughput expression

strategy of this study. 5 ml of the Bl21 (DE3) pLysS cells with the constructs were cultivated at 20 ºC in auto induction medium for three days. The cells were lysed by a lysozyme supplemented lysis buffer. They were centrifuged and the cell lysis supernatant separated from the pellet. **(Ib)** This pellet was extracted with 6 M urea and the extracted material corresponded to the inclusion body fraction loaded into the first lane of each expression experiment. **(Sn)** The cell lysis supernatant contains the soluble protein and all co extracted *E. coli* proteins shown in the central lane of each experiment. **(IM)** the cell lysis supernatant was than supplemented with 50 µl of nickel cellulose affinity matrix (HyperCell) to purify the his-tagged proteins in batch. After a short incubation period followed by centrifugation the supernatant was discarded and the protein was eluted with a small amount (200 µl) of 500 mM imidazole from the affinity matrix pellet. This crude purified fraction is shown in the right lane of each experiment. The three catalytic modules further analyzed in this work are highlighted by pink spots.

Table 4: Summary of the medium throughput strategy. ( ) No experiment. (-) Experiment gave a negative result. (+) Experiment gave positive result.

Protein	Expr. Temp	Medium	Inclusion bodies	Soluble	Purified	Concentrated	Crystals
PorA_all	20	Autoinduction	+	+			
PorA_cat	20	Autoinduction	+	+	+	+	+
PorA_cat_UNK1	20	Autoinduction	+	?			
AgaD_CBM	20	Autoinduction	+	+	+	aggregation	-
PorA_UNK1	20	Autoinduction	-	-			
PorA_UNK2	20	Autoinduction	-	-			
PorD_cat	20	Autoinduction	-	-			
PorE	20	Autoinduction	-	-			
PorB	20	Autoinduction	-	+	+	+	+
LamC_CBM6	20	Autoinduction	-	+	+	+	-
LamD_CBM42	20	Autoinduction	+	+			
Z1021_cat	20	Autoinduction	+	-			
Z3347_all	20	Autoinduction	-	+	+	+	-
LamA_all	20	Autoinduction	-	+			
LamA_cat	20	Autoinduction	-	+			
LamC_UNK1	20	Autoinduction	-	+			
LamC_cat	20	Autoinduction	+	+			
PorD_all	20	Autoinduction	-	-			
PorD_UNK1	20	Autoinduction	-	-			
AgaD_all	20	Autoinduction	+	+	+	+	
AgaD_cat	20	Autoinduction	+	+	+	+	+
LamD_cat	20	Autoinduction	+				

Special section: Macromolecular crystallisation

The first obstacle encountered for the procedure to obtain a protein structure is the production and purification of a protein, either soluble expressed in an expression host like *E. coli*, refolded from a denatured protein, or directly purified from the natural source. The second obstacle, and often considered as "the bottleneck", in protein structure determination is the production of diffracting protein crystals. Crystallisation can be seen as two phases which are nucleation and crystal growth. In the nucleation phase the macromolecules, proteins, must overcome an energy barrier to form an ordered three dimensional array of critical size that is thermodynamically stable. In the second phase this initial aggregate will ideally attract further molecules from the supersaturated solution and thereby grow until a crystal of exploitable size is obtained. Several critical factors which influence these phases have been described. Homogeneity is considered as the most important factor since contaminations may disturb formation of the target molecules into a crystal lattice. These contaminations may be co extracted proteins shown as additional bands on a SDS-PAGE gel. Further the different populations (dimers, aggregates) of the protein of interest can be considered as contamination since they may equally interfere. Therefore a two step purification strategy which first purified for affinity (IMAC) and then for size (size exclusion chromatography) was used throughout this study. The proteins were routinely assayed for purity by SDS-PAGE and for monodispersity by dynamic light scattering (DLS) before crystallisation experiments were carried out. When monodispersity was not given due to concentration/aggregation problems a solubility screen was used to find solubility promoting buffer/salt combinations. In contrast to salts which can be crystallized by quick evaporation this does usually not work for macromolecules because the intermolecular interactions that lead to the periodic array are small compared to the size of the molecules. Therefore a subtle way to promote super saturation of the macromolecule must be employed. In order to create a supersaturated protein solution which eventually leads to crystallisation, the vapour diffusion methods was used. In this method a drop (0.5-3 µl) of the protein solution is mixed with an equal volume of the crystallisation solution (variable buffers, salts, polymers, etc.). This mixed drop is than equilibrated against a bigger volume of the crystallisation solution called reservoir (500-1000 µl) in an enclosed space, which is sealed with silicone grease. In the relatively small drop which is usually hanging on a cover slide above the crystallisation solution, the concentration of the crystallisation solution in general is 50 % because it was diluted with the protein solution. This difference in concentration leads to water exchange towards the crystallisation reservoir via the gas phase and to a steady increase of protein concentration in the drop. This slow supersaturation in combination with the right crystallisation solution can result in protein crystals = student satisfaction.

II.2 The targets for further characterization: AgaD, PorA and PorB

I used the strategy of knowledge based parallel expression to rapidly obtain a protein sample of at least one of the secreted targets to perform structural studies in parallel to their biochemical characterization.

Without trying to increase the total number of expressed targets, I hence concentrated on the three enzymes that showed reasonable (Zg_2600 PorA, the catalytic module), medium (Zg_4243 AgaD, the catalytic module) and high expression (Zg_1017 PorB),(Figure 20) and belonged to the Clade4 (Figure 19).

Whereas Zg_2600 PorA and Zg_4243 AgaD are multi modular the enzyme Zg_1017 PorB naturally contains only the catalytic module. For large scale protein production the expression cells carrying the plasmids containing the catalytic domains of Zg_2600 PorA, Zg_4243 AgaD and the entire enzyme of Zg_1017 PorB were used to produce the recombinant enzymes.

It is noteworthy that the expression cells (*E. coli*-Bl21 (DE3) pLysS) and the cloning cells (E. coli-DH5alpha) carrying Zg_4243 AgaD formed depressions clearly identifiable after one week at 4ºC around the bacterial colonies (Figure 21a). This effect was even more pronounced when the purified protein was placed on an agar plate, on which strong agarolytic activity was observed (Figure 21b). This first biochemical result confirmed the phylogenetic analysis and shows that Zg_4243 AgaD is a true agarase. In contrast, the cells carrying the plasmids with Zg_2600 PorA and Zg_1017 PorB did not dig holes in the agar plate, even after several weeks of incubation at 4ºC, indicating that these enzymes do not cleave agarose.

The three enzymes could be produced in larger scale and in sufficient amounts for crystallisation and biochemical characterization. For the three enzymes a two step purification strategy was applied in which the protein was first purified by affinity chromatography (IMAC) and then polished by gel filtration. Since agarases interact with sepharose (cross linked agarose), as shown

previously for AgaA (Murielle Jam, personal communication), sephacryl (S-200) was used here as gel filtration matrix (cross linked allyl dextrose).

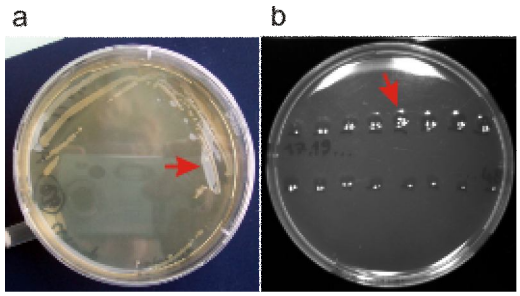


Figure 21: AgaD is an agarase as shown by its agarolytic activity on an agar plate. (a) *E. coli* DH5α bacteria transformed with the AgaD plasmid led to a depression in the agar around the colonies. **(b)** The recombinant enzyme was purified by gel filtration and the fractions were screened for agarolytic activity by adding 2 µl of each on an agar plate. The arrows point towards the holes that are formed by the agarolytic activity of the enzyme.

Interestingly the enzymes Zg_2600 PorA and Zg_4243 AgaD were also bound by this resin and eluted with the injected salt-front in contrast to Zg_1017 PorB that eluted at its expected molecular weight (data not shown). After the final polishing step, the proteins were concentrated for crystallisation purposes. In the following two chapters first the results obtained for the new agarase Zg_4243 AgaD and than for the new enzymes Zg_2600 PorA and Zg_1017 PorB will be presented.

III. The Structural and Biochemical Characterization of the new β-agarase AgaD

The activity assay on an agar plate revealed agarolytic activity of the enzyme Zg_4243 AgaD. Since *Z. galactanivorans* already posesses the two previously characterized β-agarases AgaA and AgaB we questioned the biological role of this additional enzyme. One possible explanation may be that different substrate specificities which render such an agarolytic system more efficient for the heterogeneous agarocolloids found in agarophytes. To analyze the function of this new agarase, I used a parallel strategy of protein crystallography and structure determination, combined with biochemical characterization. In the following section the results of the crystallization and the data collection will be described which led to the first manuscript.

III.1 Introduction for manuscript1: Protein crystallization of AgaD

The enzyme Zg_4243 AgaD started to precipitate in a standard Tris-HCl buffer with a pH of 7.5 and 50 mM NaCl and the protein concentration did not exceed 2 mg/ml. In order to improve the solubility of this enzyme, a solubility screen was used (Collins et al. 2004) (Figure 22). This screen revealed that protein solubility was substantially improved in tri-sodium-citrate, which was subsequently used as storage buffer. Using this buffer the protein could be concentrated beyond 8 mg/ml.

The optimized buffer not only improved solubility, but possibly also crystallisation efficiency, since the protein started to crystallize in its storage buffer at 4ºC (not shown). The protein was screened against three times 96 crystallisation solutions (namely the JCSG+ and PACT from

Qiagen and PEGI screens from Molecular dimensions) and crystals were observed in 18 conditions.

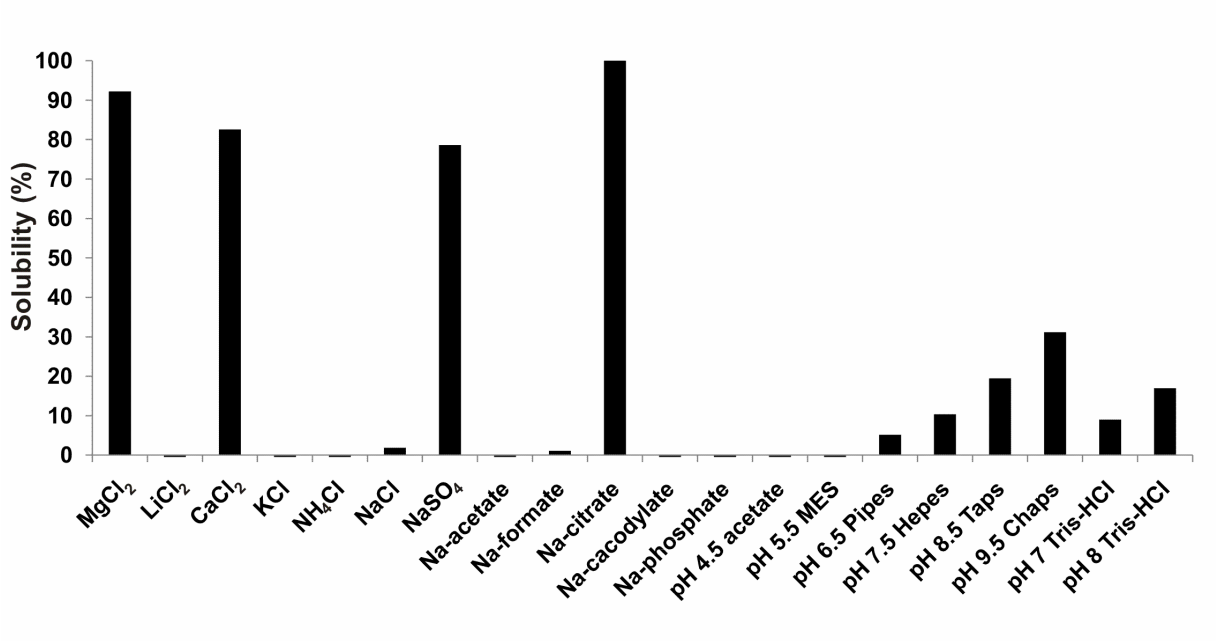


Figure 22: A solubility screen revealed tri-sodium-citrate as best buffer/salt to promote solubility of the enzyme AgaD. The solubility is shown in percent. The protein was precipitated by dialysing against water. The precipitate was separated in 20 tubes and centrifuged. The pellet was redissolved in a 100 mM solution of the above indicated salts and buffers. The amount of redissolved protein was measured by Bradford and the compound resulting in the highest protein concentration (Na-citrate) was set as 100 %.

Within these 18 different crystallization conditions, the crystal forms were predominantly plate-like, hexagonal shape. Single crystals suitable for diffraction analysis could be obtained in various conditions, predominantly with PEG as precipitant and with varying buffers, with or without salts. For the detailed composition of these solutions I refer to the detailed experimental procedure including the exact conditions described in manuscript 1. The tested hexagonal shaped crystals were of space group $p3_121$ and diffracted to 2.2 Å resolution at the ID 14 beamline (ESRF, Grenoble).

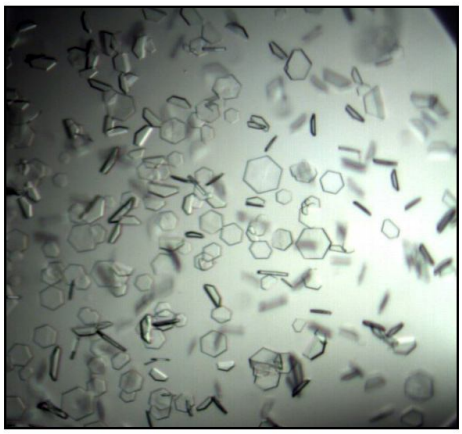


Figure 23: Hexagonal crystals of AgaD obtained in the initial screening and by the subsequent optimization.

Even though molecular replacement with AMORE and AgaB as search model gave a solution, the refinement did not result in an ameliorated map, even after several rounds of model building with Coot and refinement with Refmac.

Data treatment from these crystals indicated merohedral twinning, explaining why the R_{factor} would not decrease during the refinement procedure. Due to the high number of promising crystallisation conditions, I decided to try modifying the crystal form with additives instead of solving the problem by twin refinement that would also have been possible. Volatile additives are suggested, in particular by Therese Bergfors, to resolve twinning problems, since they finely tune the dielectric constant. Indeed, by using volatile additives, such as dioxane or isopropanol, the crystal shape could be changed from hexagonal to baguette-like shape (Figure 24) and the space group from hexagonal to orthorhombic (see below). With this strategy better diffracting protein crystals could be obtained for Zg_4243 AgaD, and the structure was finally solved by molecular replacement, using a complete data set that was collected at 1.5 Å resolution on an orthorhombic crystal indexed in space group $p2_12_12_1$ instead of $p3_121$. Data processing with SCALA did not indicate twinning when taking the cumulative intensity distribution into account.

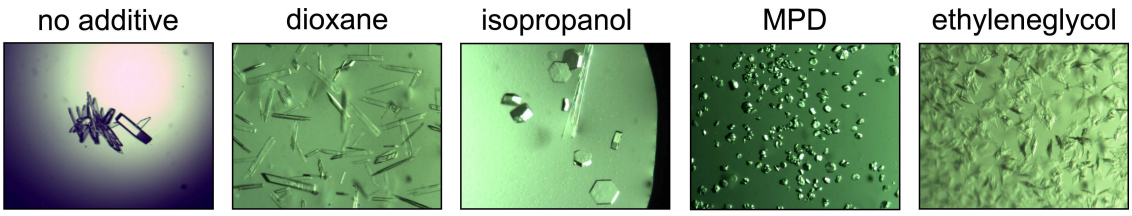


Figure 24: The crystal form of AgaD could be modified with isopropanol and dioxane from hexagonal to baguette-like morphology. Notably the drop with isopropanol contains one elongated baguette-like crystal besides several hexagonal crystals.

III.2 Manuscript 1

Expression, purification and preliminary X-ray diffraction analysis of the catalytic module of a ß-Agarase from the Flavobacterium *Zobellia galactanivorans*

Jan-Hendrik HEHEMANN[1,2*], Gurvan MICHEL[1,2], Tristan BARBEYRON[1,2] and Mirjam CZJZEK[1,2]

[1]Université Pierre et Marie Curie-Paris 6, Unité Mixte de Recherche 7139 "Marine Plants and Biomolecules", Station Biologique, F-29682 Roscoff Cedex, Bretagne, France

[2]Centre National de la Recherche Scientifique, Unité Mixte de Recherche 7139 "Marine Plants and Biomolecules", Station Biologique, F-29682 Roscoff Cedex, Bretagne, France

Keywords: β-agarase, glycoside hydrolase, crystallization, heterologous expression, *Zobellia galactanivorans*

Running title: crystallization of the catalytic module of β-agarase D from *Z. galactanivorans*

[*]Corresponding author:

Tel (33) 298 29 24 75, Fax (33) 298 29 23 24, E-mail: czjzek@sb-roscoff.fr

Published in

Acta Cristallogr. F

III.2.1 Abstract

Marine bacteria secrete specific glycoside hydrolases such as agarases to access polysaccharides from algal cell walls as carbon and energy source. In the attempt to identify agarases with variable degradation patterns, we have performed the expression, purification and crystallisation of a novel family GH16 *ß*-agarase from the marine bacterium *Zobellia galactanivorans*. The purified enzyme crystallized in two distinct forms that were grown by the hanging-drop vapour-diffusion method with polyethylenglycol as precipitant. Hexagonal crystals of the space group $P3_121$ diffracted to 2.2 Á resolution whereas orthorhombic crystals belonging to the space group $P2_12_12_1$ diffracted up to 1.5 Á resolution.

III.2.2 Introduction

Agarose is a marine polysaccharide, present as a cell wall matrix compound in red algae (Craigie et al. 1974), that can chemically be described as a neutral unbranched polysaccharide composed of repeating neoagarobiose units ((3,6)-anhydro-L-galactose-α(1,3)-D-galactose) joined by β(1,4)-bonds (Rees 1969). Industrial processes use alkaline treatment to produce an almost pure agarose with high gelling quality and without chemical modifications. In contrast, the galactose subunits of the chemically variable class of natural agarocolloids (Craigie 1990), alternatively also named agarans (Knutsen et al. 1994), are methylated, pyruvylated, sulfated or glycosylated to form various derivatives with different gelling properties and solubility characteristics (Rees et al. 1962; Morrice et al. 1983; Rochas et al. 1986). These variations occur in algal cell walls as a function of species, geographical origin, stage of life cycle, physiological state or as a function of age (Lahaye 2001). To be able to degrade these chemically different polysaccharide structures to completion, marine bacteria must secrete a set of agar degrading enzymes, the precise interplay and synergy of which has not yet been studied in detail. A recent genomic and proteomic analysis of *Saccharophagus degradans*, a rod-shaped, aerobic, bacterium isolated from the surface of decomposing saltwater cord grass in marshland (Ekborg et al. 2005), has revealed the presence of an agarolytic system with five agarases belonging to three different glycoside

hydrolase (GH) families (Ekborg et al. 2006). Several enzyme copies belonging namely to families GH50, GH86 and one endolytic β-agarase from GH16 (http://www.cazy.org/Cazy; (Henrissat et al. 1997)) were identified, however, only unmodified neutral agarose degradation activity was described (Ekborg et al. 2006). In contrast, *Zobellia galactanivorans* (formerly *Cytophaga drobachiensis*) is a marine heterotrophic Flavobacterium capable of degrading more complex polysaccharides, such as agarocolloids (Michel et al. 2006), with GH16 enzymes only. Indeed, the ongoing genome project of this marine bacterium (Michel G, personal communication) has revealed a high number of family GH16 enzymes, some of which are clearly related to the previously characterized β-agarases A and B (AgaA, AgaB) (Allouch et al. 2003; Jam et al. 2005). No family GH50 or GH86 β-agarase has been detected. Nevertheless, the high number of family GH16 agarase-like sequences suggests a finely tuned system for complex agarocolloid degradation. Sequence analysis of one of these GH16 β-agarases (AgaD) revealed that, although closely related (32 % identity with AgaA and 39 % identity AgaB), the sequence of AgaD represents a third family GH16 β -agarase, containing several large insertions with respect to AgaA and AgaB (Figure 25). The sequence variations raised the question whether AgaD would perform catalysis with a different mode of action (i.e. exolytic or processive) or even display variation of substrate specificity with respect to the chemical modifications present in the naturally occurring substrate.

Moreover, agar is not only of great ecological importance, but due to its gelling properties, it is also widely used in food, cosmetic, pharmaceutical and biotechnological industries. The access to multiple agarose active enzymes with varying substrate specificities is therefore of biotechnological advantage. In contrast to the impact of its substrate, structural details describing the substrate recognition and specificity have so far only been determined for two true endo β-agarases AgaA and AgaB (Allouch et al. 2003; Allouch et al. 2004; Jam et al. 2005).

In order to analyse and understand the variability of substrate specificities in family GH16 β-agarase, we have performed the production and crystallization of this third β-agarase AgaD from *Z. galactanivorans*. Moreover, structural and biochemical comparison with the existing β-agarase structures will lead to an extended understanding of the agarolytic system of the

heterotrophic marine bacterium *Z. galactanivorans*, a model organism for marine organic matter transformation.

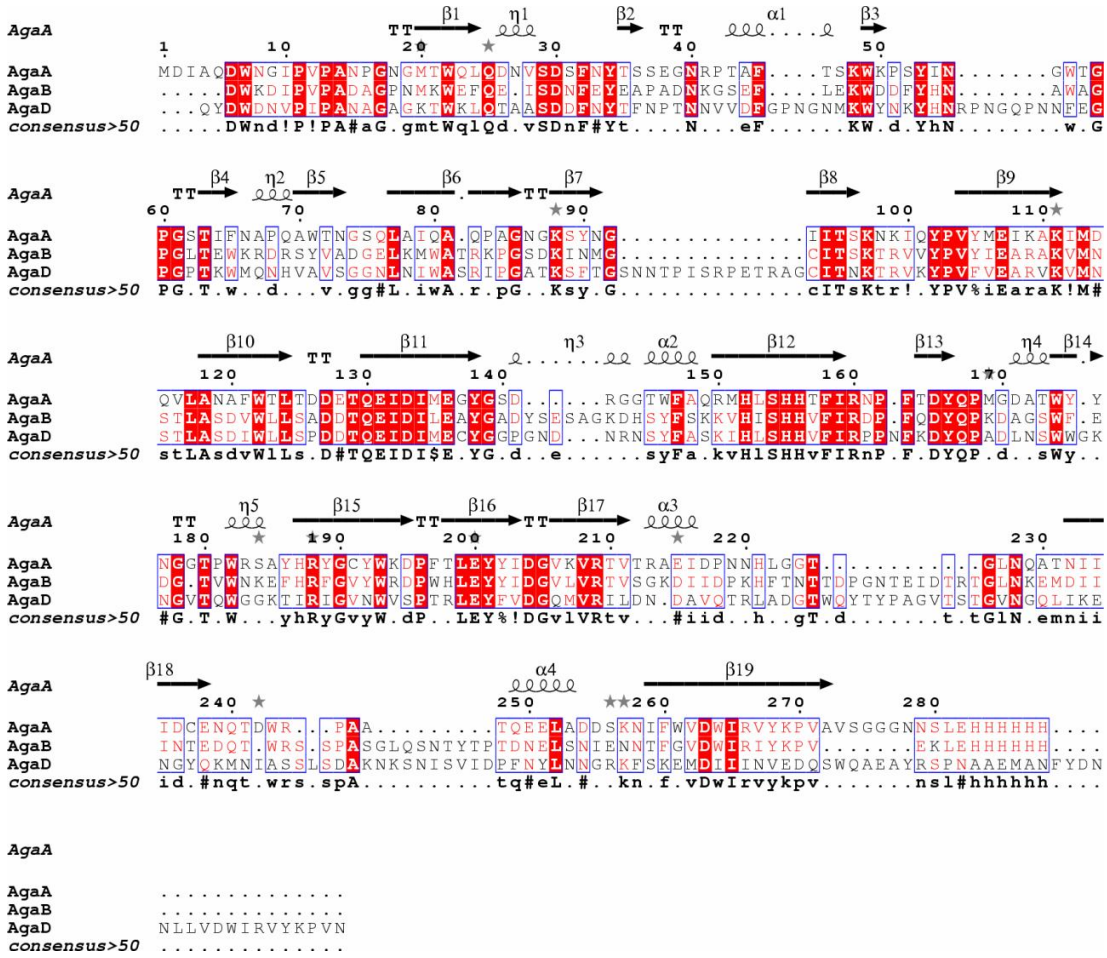


Figure 25: Sequence alignment of three β-agarases (A, B and D) from *Z. galactanivorans*. The secondary structure elements of β-agarase A are indicated above the alignment. The alignment was produced using Multalign (Corpet 1988) and the figure was drawn using Espript (Gouet et al. 2003).

III.2.3 Material and Methods

Expression and purification

Bioinformatics analysis (data not shown) of the open reading frame coding for β-agarase D revealed that it is a bimodular protein composed of a C-terminal domain of unknown function and at the N-terminus the catalytic domain which belongs to the GH16 family of glycoside hydrolases. In the past, bimodular glycoside hydrolases have shown to be recalcitrant to crystallization, and we therefore have cloned a construct containing the catalytic domain only (hereafter called AgaD_cat).

The nucleotide sequence corresponding to the catalytic domain was amplified by PCR from *Z. galactanivorans* genomic DNA using a set of primers (Forward 5'-gggggAgATCTCAATACGATTGGGACAACGTGCC-3' and reverse 5'-CAATACGATTGGGACAACGTGCC-3'). The obtained PCR product was purified with a Qiagen QIAquick purification Kit and digested with *Bgl*II/*Eco*RI (50/30 ends) in NEB2 buffer (BioLabs) at 310 K for 3 h. Digested and purified PCR product was ligated with an equally digested and purified pFO4 vector (derivative of pET15) 4°C by T4 DNA Ligase (overnight). The ligation mixture was transformed into *E. coli* DH5α strains by chemical transformation. The obtained colonies were screened by colony PCR and a positive colony with the correct fragment length was used to prepare plasmid for expression studies. The plasmid was extracted with a Wizard Plus SV Minipreps kit (Promega) and used to transform chemically competent *E. coli* BL21 (DE3) cells. One fresh colony was picked to innoculate a 5 ml LB-medium preculture with 100 µg ml^{-1} ampicillin, which has been incubated overnight. One ml of the preculture was used to innoculate 1 L of a ZYP-5052 expression culture (Studier 2005) (100 µg ml^{-1} amp) until saturation of the culture (final OD 600 nm: ~ 16) was obtained. The cells were harvested by centrifugation (4000g, 277 K, 20 min) and the cell pellet was resuspended in buffer A (20 mM Tris, 200 mM NaCl, 20 mM imidazole, pH 7.5). The collected cells were lysed for 30 minutes on ice, followed by sonication and the lysate was cleared by centrifugation (50 000g, 277 K, 30

min) and a subsequent filtration with a 0.2 µm filter (Millipore). The filtrated solution was loaded onto a 10 ml IMAC HyperCell resin column (Pall Corporation), which was charged with $NiSO_4$. The column was equilibrated with buffer A without Lysozyme and DNAse. After a step of washing with buffer A (ten column volumes), the protein was eluted with a 60 ml linear gradient from buffer A to 60 % of buffer B (20 mM Tris, 200 mM NaCl and 500 mM imidazole) at a flow rate of 1 ml min^{-1}. The fractions were analysed by SDS-PAGE and fractions which contained AgaD_cat were pooled. The volume was reduced to 4 ml by ultra filtration on an Amicon membrane (polyethersulfone, 30 kDa cutoff). Preliminary trials to concentrate the enzyme beyond a nominal concentration of 2 mg/ml failed because the protein precipitated from the solution. Since efficient crystal growth depends on protein concentration and solubility in the first place, we used a solubility screen according to Collins et al. to find an optimal buffer combination for crystallisation and for protein storage (Collins et al. 2004). A buffer composed of 100 mM sodium citrate and also containing 1 mM $CaCl_2$ and 1 mM $MgCl_2$ at pH 8 (buffer C) was used in the final gel filtration polishing step as well as protein storage buffer.

For the gel filtration we used a Sephacryl S-200 column (GE Healthcare) pre-equilibated with buffer C at a flow rate of 1 ml min^{-1} (Figure 2). The purified enzyme was concentrated to ~8 mg ml^{-1} by ultra filtration on an Amicon membrane (10 kDA cutoff). All chromatography were carried out on an ÄKTA Explorer Chromatography system (GE-Healthcare).

Crystallisation

The first crystallization-screening experiments were carried out at 292 K with the sitting-drop vapour diffusion method in 96 well Corning plates. The crystallisation screening was performed with JCSG+, PACT and the PEG I screens from Qiagen, which makes a total of 288 conditions in three 96-well plates (Corning no. 3551, one well sitting drop). A crystallization robot (Proteomics Solutions, Honeybee961) was used for pipetting and the drops contained 300 nl protein solution (~8 mg ml-1 in buffer C) that was mixed with 150 nl of reservoir solution. After visual identification of initial crystallisation conditions, these were further optimised in 24 well Linbro plates by the hanging-drop vapour diffusion method. The drops were prepared on

siliconized cover slips by mixing 2 µl of protein solution with 1 µl of well solution and the equilibration was performed against 500 µl reservoir solution at room temperature.

Data collection and X-ray diffraction analysis

X-ray diffraction data were collected from a crystal of recombinant AgaD_cat at 100 K on beamline ESRF_ID14-2 (Grenoble, France) with an ADSC Quantum 4 detector and a detector distance of 222.9 mm and on ESRF_ID14-1 with a ADSC Quantum Q210 Detector with a detector distance of 142.88 mm. For cryo protection, crystals were soaked for one minute in crystallisation solution complemented with 6 % glycerol before they were flash frozen within the cryo stream. The datasets were collected for the hexagonal crystal form at a wavelength of 0.933 Å at the ESRF beamline ID14-EH2 and for the orthorhombic crystal form at a wavelength of 0.934 Á at beamline ID14-1. All raw data were processed using the program MOSFLM (Leslie 1992). The data were then merged and scaled using the program SCALA (Collaborative Computational Project Number 4 1994). Detailed data statistics for both crystal forms are reported in Table 1. Molecular replacement was carried out using the program AmoRe (Navaza 2001) with the β-Agarase B protein structure from *Z. galactanivorans* (PDB code 1o4z) as search model.

III.2.4 Results and discussion

After initial purification and concentration the tendency for low solubility (<2mg/ml) associated with high aggregation was observed for AgaD_cat. Therefore, we performed a solubility screen following the strategy of Collins et al. (Collins et al. 2004) and a buffer combination that strongly improved solubility could be identified. Thereby we obtained a monodispersive solution at a concentration of 8 mg/ml of AgaD_cat (Figure 26). Subsequently, three crystallisation screens were used for screening of crystallisation conditions (PEGI, PACT and JCSG+).

Hexagonal and orthorhombic single crystals grew in a total of 18 different conditions (Table 6), in general within one or two weeks at 292 K. Subsequently, the conditions which produced the largest single crystals (JCSG+, Nr. 60) were scaled up in 24-well Linbro plates with the hanging-drop vapour-diffusion method. Hexagonal crystals suitable for diffraction studies could be grown after further optimisation trials with a crystallisation solution of 28 to 34 % PEG 8000 (w/v) and 0,1 M imidazole pH 8 at 292 K.

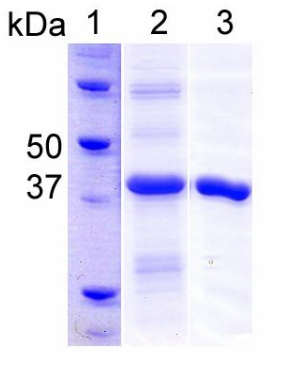


Figure 26: SDS-PAGE analysis of purified recombinant β-Agarase D (AgaD_cat). The recombinant β-agarase was produced in *E. coli* bacteria and purified by IMAC (Immobilised metal affinity chromatography) and gel filtration from the bacterial cell lysate. Lane 1: Marker. Lane 2: Agarase after nickel affinity chromatography (IMAC-cellulose). Lane 3: Agarase after gel filtration chromatography on Sephacryl (S-200) resin.

The crystals used for diffraction studies grew within one week, were of hexagonal shape with dimensions of about 0.3 x 0.3 x 0.05 mm and diffracted to 2.2 Á. A complete dataset was collected with these crystals and data statistics are reported in Table 1. We succeeded to obtain phases by molecular replacement with β-Agarase B from *Z. galactanivorans* (PDB code 1o4z) as search model and two molecules per asymmetrical unit. However, the cumulative intensity distribution given by SCALA (Collaborative Computational Project Number 4 1994) indicated twinning and the R_{free} factor would not decrease even after several rounds of refinement. To avoid complex refinement of twinned data, and having identified numerous crystallization

conditions, we first tempted to obtain a variant crystal form of AgaD_cat. Volatile additives such as dioxane are known to reduce twinning (Bergfors 2003). Therefore we used 4 % dioxane or isopropanol together with the initial optimised crystallisation condition (28-34 % PEG 8000 and 0,1 M imidazole pH 8). Both additives changed the crystal morphology from hexagonal plates to needle or rod shaped crystals (Figure 27). Several crystals of the new crystal form were tested and one crystal (grown with isopropanol as additive) diffracted up to 1.5 Á resolution. The dataset was treated as described above and all crystallographic data are described in Table 5. The addition of dioxane or isopropanol changed the space group from hexagonal $P3_121$ to orthorhombic $P2_12_12_1$, with unit cell parameters of a=53.25, b=77.27, c=83.7 Á, respectively. Moreover, the data quality could be strongly improved. Assuming a molecular weight of 40.24 kDa, the calculation of the Matthews coefficient resulted in a V_M value of 2.14 and a solvent content of 42.2% for the presence of one molecule in the asymmetrical unit.

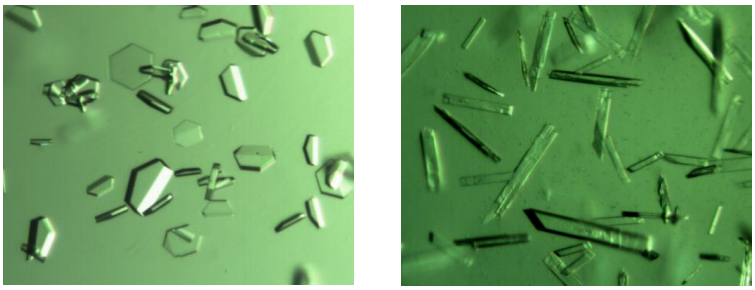


Figure 27: Crystals of β-AgaD_cat. Left panel: Crystals of β-AgaD_cat displaying a trigonal crystal form with space group $P3_121$. The average size of the crystals is 0.2 mm × 0.2 mm × 0.08 mm. Right panel: Crystals of β-AgaD_cat obtained by addition of dioxane to the crystallization condition leading to orthorhombic crystals with space group $P2_12_12_1$. The average size of the crystals is 0.6 mm × 0.1 mm × 0.08 mm.

III.2.5 Conclusion

A new ß-agarase from the heterotrophic marine bacterium *Z. galactanivorans* was cloned,

expressed, purified and crystallised. This agarase contains three sequence insertions which do not occur within the previously described ß-Agarases A and B from the same organism. The enzyme showed activity on agarose and the sequence divergence may be explained by a different mode of action or a different specificity for natural agarocolloids. The systemic characterisation of agarase-like enzymes from *Z. galactanivorans* will extend our knowledge of marine organic matter degradation by glycoside hydrolases.

III.2.6 Acknowledgments

JHH was supported by a FP-6 Marie Curie Fellowship (MEST-CT-2005-020737). This work was also supported by the "Region Bretagne" through the program Marine3D. We are indebted to the European Synchrotron Radiation Facilities (ESRF, Grenoble, France) for accession to beam-time through a beam allocation group and we thank all beamline scientists and staff for help on the beamline.

III.2.7 References

Allouch, J., Helbert, W., Henrissat, B. & Czjzek, M. (2004). Parallel substrate binding sites in a beta-agarase suggest a novel mode of action on double-helical agarose. *Structure* 12, 623-632.

Allouch, J., Jam, M., Helbert, W., Barbeyron, T., Kloareg, B., Henrissat, B. & Czjzek, M. (2003). The three-dimensional structures of two beta-agarases. *Journal of Biological Chemistry* 278, 47171-47180.

Bergfors, T. (2003). Seeds to crystals. *Journal of Structural Biology* 142, 66-76.

Collaborative Computational Project Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallographica. Section D, Biological Crystallography* 50, 760-763.

Collins, B. K., Tomanicek, S. J., Lyamicheva, N., Kaiser, M. W. & Mueser, T. C. (2004). A preliminary solubility screen used to improve crystallization trials: crystallization and preliminary X-ray structure determination of Aeropyrum pernix flap endonuclease-1. *Acta*

Crystallographica Section D-Biological Crystallography 60, 1674-1678.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* 16, 10881-90.

Craigie, J. (1990). Cell Walls. In *Biology of the red algae* (Cole, K. & Sheath, R., eds.), pp. 221-257. Cambridge university Press, Cambridge.

Craigie, J. & Leigh, C. (1974). carrageenans and agars. 109-133.

Ekborg, N. A., Gonzalez, J. M., Howard, M. B., Taylor, L. E., Hutcheson, S. W. & Weiner, R. M. (2005). *Saccharophagus degradans* gen. nov., sp. nov., a versatile marine degrader of complex polysaccharides. *International Journal of Systematic Evolutionary Microbiology* 55, 1545-1549.

Ekborg, N. A., Taylor, L. E., Longmire, A. G., Henrissat, B., Weiner, R. M. & Hutcheson, S. W. (2006). Genomic and proteomic analyses of the agarolytic system expressed by *Saccharophagus degradans* 2-40. *Applied and Environmental Microbiology* 72, 3396-3405.

Gouet, P., Robert, X. & Courcelle, E. (2003). ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* 31, 3320-3.

Henrissat, B. & Davies, G. (1997). Structural and sequence-based classification of glycoside hydrolases. *Curr Opin Struct Biol* 7, 637-44.

Jam, M., Flament, D., Allouch, J., Potin, P., Thion, L., Kloareg, B., Czjzek, M., Helbert, W., Michel, G. & Barbeyron, T. (2005). The endo-beta-agarases AgaA and AgaB from the marine bacterium *Zobellia galactanivorans*: two paralogue enzymes with different molecular organizations and catalytic behaviours. *Biochemical Journal* 385, 703-713.

Knutsen, S., Myslabodski, D., Larsen, B. & Usov, A. (1994). A modified system of nomenclature for red algal galactans. *Botanica Marina* 37, 163-169.

Lahaye, M. (2001). Developments on gelling algal galactans, their structure and physico-

chemistry. *Journal of Applied Phycology* 13, 173-184.

Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* 26.

Michel, G., Nyval-Collen, P., Barbeyron, T., Czjzek, M. & Helbert, W. (2006). Bioconversion of red seaweed galactans: a focus on bacterial agarases and carrageenases. *Applied Microbiology and Biotechnology* 71, 23-33.

Morrice, L. M., McLean, M. W., Long, W. F. & Williamson, F. B. (1983). Porphyran primary structure. An investigation using beta-agarase I from Pseudomonas atlantica and 13C-NMR spectroscopy. *Eur J Biochem* 133, 673-84.

Navaza, J. (2001). Implementation of molecular replacement in AMoRe. *Acta Crystallographica. Section D, Biological Crystallography* 57, 1367-1372.

Rees, D. (1969). Structure, conformation, and mechanism in the formation of polysaccharide gels and networks. *Adv. Carbohydr. Chem. Biochem.* 24, 267-332.

Rees, D. A. & Conway, E. (1962). Structure and Biosynthesis of Porphyran - a Comparison of Some Samples. *Biochemical Journal* 84, 411-&.

Rochas, C., Lahaye, M. & Yaphe, W. (1986). Sulfate Content of Carrageenan and Agar Determined by Infrared-Spectroscopy. *Botanica Marina* 29, 335-340.

Studier, F. W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification* 41, 207-234.

Table 5: Data collection statistics of the two crystal forms of AgaD_cat

Wavelength / Beamline	0.933 / ID14-EH2	0.934 / ID14-EH1
Space group	$P3_121$	$P2_12_12_1$
Unit cell parameters (Å)	a=98.49, b=98.49, c=181.21	a=53.25, b=77.27, c=83.7
Resolution range (Å)	62.137-2.2 (2.32-2.2)	43.85-1.5 (1.58-1.50)
No. of observations	302205 (30196)	385296 (55310)
No of Unique reflections	49464 (6817)	55977 (8035)
Completeness (%)	96.3 (91.8)	99.9 (99.9)
Mean I/sigma(I)	15.4 (3.5)	21.3 (9.1)
Rmerge (%)	10.5 (27.1)	5.3 (24.9)
Redundancy	6.1 (4.4)	6.9 (6.9)

Table 6: Crystallization conditions for AgaD_cat identified from the initial screening

Condition	Salt	Buffer	pH	Precipitant
JCSG+: 9	0.2 M Ammonium chloride	None	6.3	20 % PEG 3350
JCSG+: 25	0.2 M Sodium chloride	0,1M Phosphate/citrate	4.2	20 % PEG 8000
JCSG+: 60	None	0.1 M Imidazole	8.0	10% PEG 8000
JCSG+: 95	0.2 M Magnesium chloride	0.1 M BIS TRIS	7.5	25 % PEG 3350
PACT: 1	None	0.1 M SPG	4.0	25 % PEG 1500
PACT: 22	0.2 M Magnesium chloride	0.1 M MES	6.0	20 % PEG 6000
PACT: 48	0.002 M Zinc chloride	0.1 M TRIS	8	20 % PEG 6000
PEG I: 6	None	0.1 M Sodium acetate	4.6	25 % PEG 2000 MME
PEG I: 25	None	0.1 M Sodium acetate	4.6	25 % PEG 3000
PEG I: 26	None	0.1 M Sodium acetate	4.6	25 % PEG 4000
PEG I: 27	None	0.1 M Sodium acetate	4.6	25 % PEG 6000
PEG I: 28	None	0.1 M Sodium acetate	4.6	25 % PEG 8000
PEG I: 49	0.2 M Sodium fluoride	None		20 % PEG 3350
PEG I: 53	0.2 M Magnesium chloride	None		20 % PEG 3350
PEG I: 80	0.2 M Magnesium sulfate	None		20 % PEG 3350
PEG I: 87	0.2 M Sodium phosphate	None		20 % PEG 3350
PEG I: 91	0.2 M Ammonium phosphate	None		20 % PEG 3350

PEG I: 96	0.18 M tri-Ammonium citrate	None		20 % PEG 3350

III.3 The crystal structure of AgaD

AgaD is a multimodular enzyme composed of a N-terminal GH16 catalytic domain and a C-terminal domain of ~10 kDa with unknown function. The N-terminus of the gene product contains 20 amino acids, which serve as a signal peptide to promote secretion of the protein, and which are absent in the mature protein. For biochemical and structural analysis the catalytic domain was cloned separately and further analyzed in this study. This domain, consisting of 357 residues with a molecular weight of 40.2 kDa and a theoretical isoelectric point of 8.35 (www.expasy.org), was expressed in *E. coli*. The protein was crystallized and the structure was solved by molecular replacement with AMORE using AgaB as a search model and in the resolution range from 10-4 Å. The result gave one contrasted solution with a correlation factor of 25.4 and a R_{factor} of 50.9% (for refinement statistics see Table 7).

The AgaD protein structure possesses the GH16 jelly roll fold, similar to AgaA and AgaB from *Z. galactanivorans* and described for various GH16 enzymes with divergent substrate specificities. The jelly roll fold is composed of two sandwich-like, stacked β-sheets which are twisted to form the substrate binding cleft that extends across the surface of the enzyme to harbour a polysaccharide chain or oligosaccharide substrate molecules. The core backbone chain of the β-agarase AgaD is almost perfectly superimposable with AgaA (rmsd 1.264 Å) and AgaB (rmsd 1.203 Å), (Figure 28). The sequence identity between AgaD and AgaB is 35.5 % while it is 32.1 % between AgaD and AgaA, values that are in the normal range for enzymes of one sub-family (http://www.cazy.org). The higher identity between AgaD and AgaB explains why the molecular replacement was successful with AgaB as search model and not with AgaA. Whereas the core structures of the three agarases are almost perfectly super imposable the surface loops are rather different in conformation. Moreover AgaD contains three loop variations with respect to AgaA and AgaB (Figure 29). One loop elongates the distance of the binding cleft in AgaD which measures 42 Å in contrast to 34 Å in AgaA and 33.5 Å in AgaB. Surface loops

modulating enzyme-substrate interaction for a GH16 enzyme have been firstly described for the lichenase (Keitel et al. 1993) and possibly the loops described in AgaD play a role in enzyme-substrate interaction, differently to AgaA and AgaB.

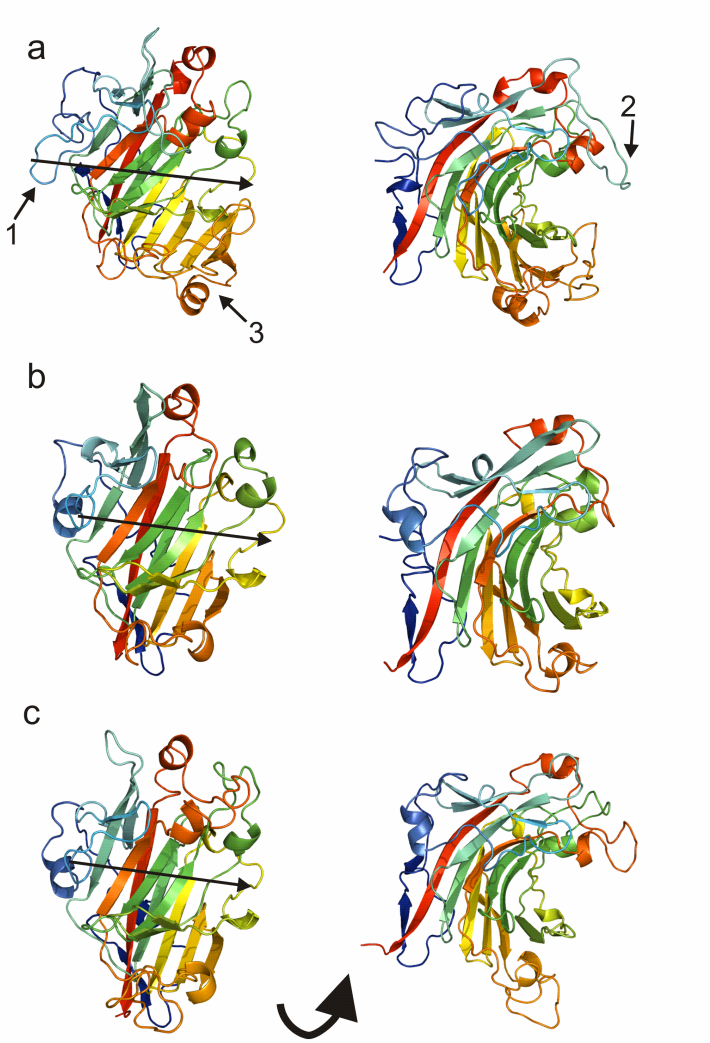


Figure 28: Crystal structure of the β-agarase AgaD in comparison to AgaA and AgaB from *Zobellia galactanivorans*. The cartoon plots are colour ramped from the N-terminus in blue to the C-terminus in red. **(a)** Front view of the substrate binding cleft of AgaD and from the side. **(b)** Front view of the substrate binding cleft of AgaA and from the side. **(c)** Front view of the substrate binding cleft of AgaB and from the side. The long arrow indicates binding of substrate in the binding cleft with its tip directing towards the reducing end of the substrate chain. The small arrows shown in **(a)** indicate the major differences found in AgaD.

The homology alignment, illustrated in Figure 29b shows that the three loops are created by insertions that do not exist in AgaA or AgaB. These insertions are clearly located around the substrate binding cleft and include residues N76–P82 (loop1), S118-T132 (loop2) and D260-K311 (domainX) (Figure 29a).

The loop1 insertion extends the substrate cleft on the -n binding site of the enzyme, when compared to the AgaA substrate complex structure. Loop2 extends above the substrate binding cleft, almost on top of the catalytic residues E174 and E179, and leads to a more, tunnel-like topology when compared to the crystal structures of AgaA and AgaB. The loop insertion 3, ranging from residues D260 to K311, can be described as a supplemental domain with three sequential β-strands, one alpha-helix and the interconnecting loops and is called domainX. These three β-strands extend the lower of the two cleft forming β-sheet at the +n binding sites of the enzyme. The loop which connects the two outermost β-strands forms a new wall of the substrate binding site and thereby creates a more pronounced cleft at this side of the enzyme, compared to AgaA and AgaB. From the various GH16 agarase-like sequences found in public databases, only one other enzyme (MS116) has a long insertion similar to AgaD at the same position. This enzyme is from the marine Bacteroidetes *Microscilla*. Interestingly, a similar domain, but at the C-terminus is found in the family GH16 xyloglucan endo-transferase and xyloglucanase enzymes the structure of which have been recently described (Johansson et al. 2004; Baumann et al. 2008).

The superimposition of the three agarase structures shows that the residues determinant for agarose recognition in AgaA and AgaB are not only conserved in sequence but most of them also in position and conformation (Figure 30a). Nevertheless, R77 located on loop1 and Y296 in domainX are in correct positions to interact with the agarose substrate, as shown in (Figure 30b). I decided to analyse enzyme substrate interaction of AgaD by co crystallisation of an active site mutant with oligosaccharides. To suppress catalytic activity and thereby trap the substrate for non productive binding, I mutated E174 to a serine and used this inactive variant for co crystallisation with octa and deca-oligoagarose. Diffracting crystals could be obtained but ligand

electron density could not be observed in the substrate binding cleft. However a crystal structure of AgaA in complex with a tetra-oligosaccharide bound to the -n subsites was previously published by Allouch et al. (2004).

Table 7: X-ray data collection and refinement statistics of AgaD

Data collection	
Space group	$P2_12_12_1$
Resolution range (Å)[a]	56.7 – 1.5 (1.58-1.50)
No of observations	385296 (55310)
No of unique reflections	55977 (8035)
Completeness (%)	99.9 (99.9)
$<I/\sigma(I)>$	21.3 (9.1)
Redundancy	6.9 (6.9)
R_{sym} (%)	5.3 (24.9)
R_{pim} (%)	2.2 (10.2)
Refinement	
Resolution (Å)	1.8 (43.9)
R_{work}/R_{free}	14.01 (18.51)
Number of atoms	
Protein	2874
Water	600
B-factors	
Protein	9.36
Water	11.35
R.m.s deviations	
Bond lengths (Å)	0.022
Bond angles (°)	1.92

[a]Values in paranthesis are in the highest resolution shell.

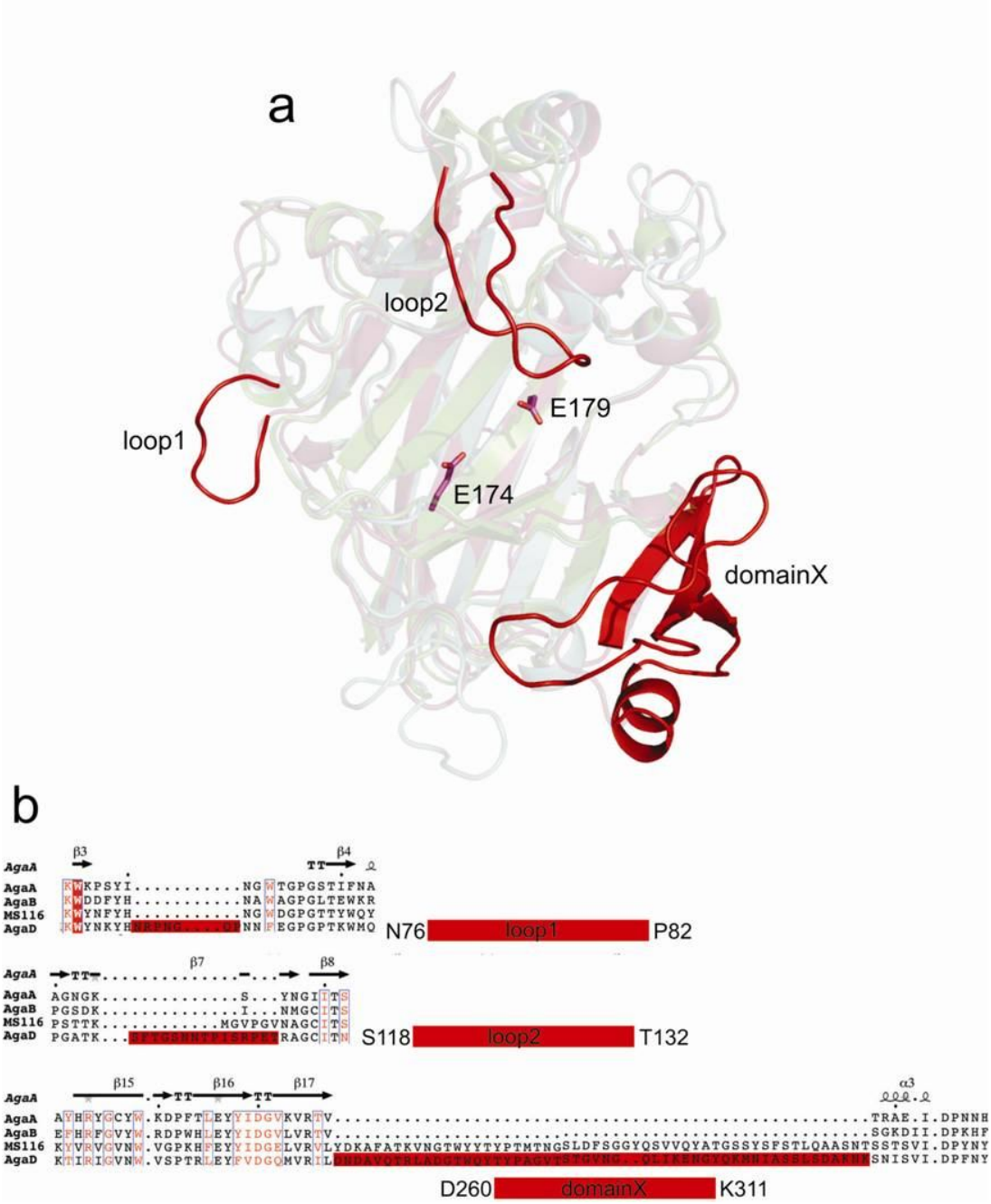


Figure 29: AgaD contains three insertions around the substrate binding cleft that do not occur in AgaA and in AgaB. Cartoon representation of the crystal structure of AgaD superimposed with the crystal structures of AgaA and AgaB. **(a)** AgaD in magenta and insertions in red, AgaB in blue and AgaA in green. The two catalytic glutamates are shown as stick presentations. **(b)** Homology alignment between the three agarases from *Z. galactanivorans* and one agarase from *Microscilla*.

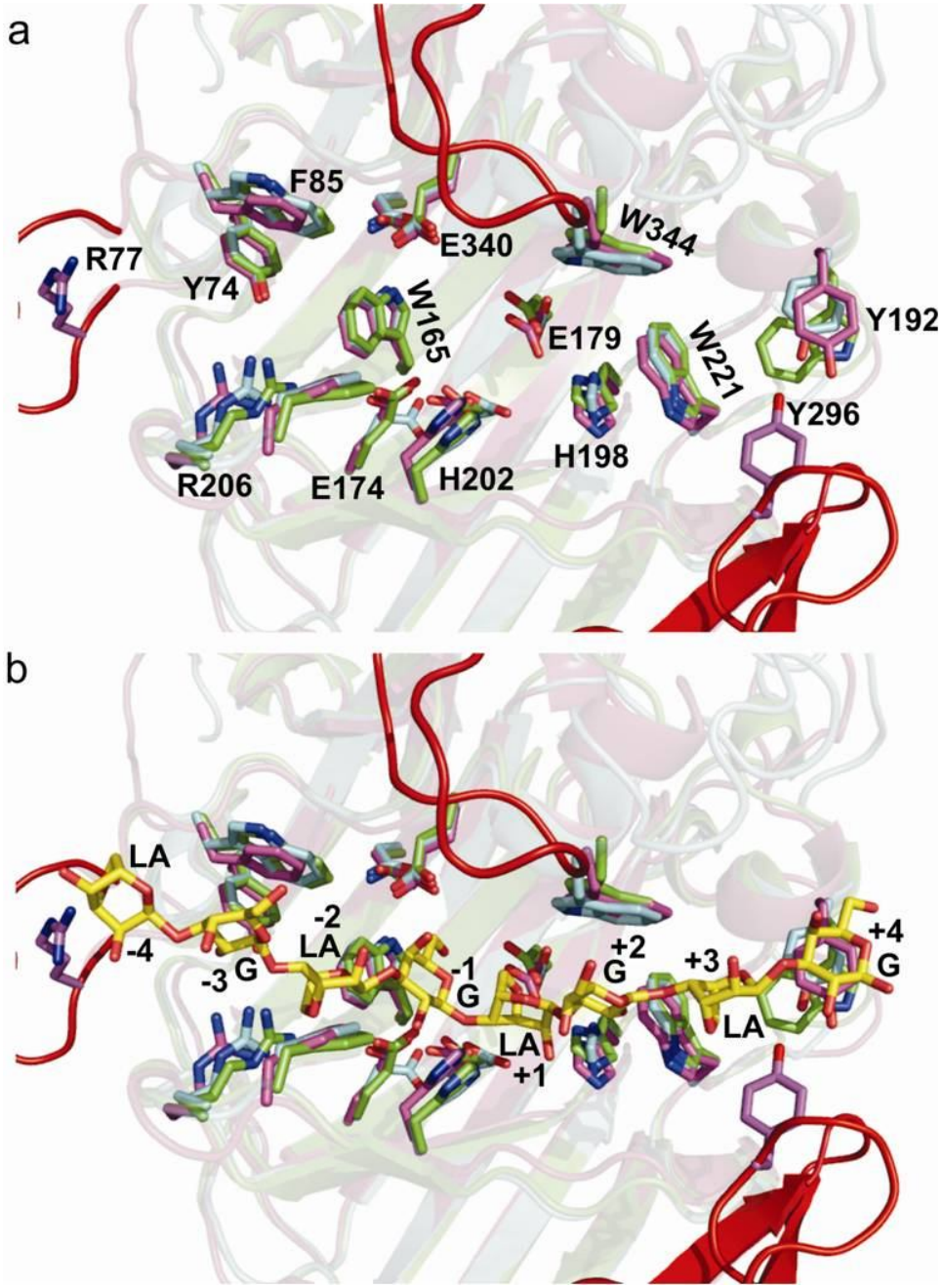


Figure 30: The active side residues of AgaD are conserved with AgaA and AgaB from *Z. galactanivorans* but the loops may constitute new substrate contacts. (a) The three structures were superimposed and the residues that were identified for agarose recognition in AgaA and AgaB are highlighted and shown as stick representation. **(b)** A complex structure of AgaB with a bound neoagarooctasaccharide (unpublished data) was superimposed and shows that AgaD has a longer substrate binding cleft with additional -n and +n subsite contacts.

Since the involved residues are conserved between the enzymes AgaA, B and D, one can assume that the orientation of the substrate will not dramatically differ. Therefore, I modelled the oligo tetrasaccharide into the four -n subsites of the active cleft of the three superposed enzymes. For the four +n subsites, I used an AgaB octa-oligosaccharide complex structure in which the four +n subsite sugar units could be modelled with low occupation (Czjzek M. data not published). The result in Figure 30b shows that loop1 and especially residue R77 creates further contacts with the LA unit bound in subsite -4. With the here modelled oligosaccharide the contact between R77 and the LA unit is too close, implying that this sugar ring and probably also the adjacent G unit bound to the -3 subsite, or the arginine must be oriented slightly different in this enzyme. Furthermore, this loop clearly creates a longer contact interface and probably one or two more negative binding subsites with a possibly higher activity on longer oligosaccharides (Figure 30).

III.4 Biochemical characterization of AgaD

In view of the observed differences though subtle in the structural comparison, I performed a detailed biochemical characterization to find out in which way AgaD may be complementary to AgaA and AgaB.

III.4.1 Catalytic behaviour of AgaD

When the three agarases were used to degrade agarose in superfusion (44 ºC) the initial velocity of degradation was almost identical for AgaD, AgaA and AgaB. Nevertheless, a plateau was reached rapidly with AgaD, and therefore only ~50% of oligosaccharides with respect to AgaA and AgaB were released at the final stage of degradation (Figure 31). Further addition of AgaD to the agarose resulted in additional amounts of released oligosaccharides (data not shown). This indicated that it is not an intrinsic property of the substrate but of the enzyme that accounts for the lower amount of released oligosaccharides. Several reasons may explain this behaviour, such as lower thermo stability or higher proteolytic sensibility. I determined the melting point of AgaD experimentally by dynamic light scattering. Indeed a melting temperature of 32 ºC could

be observed, explaining the lower amount of total released sugars at 44 ºC (Figure 32).

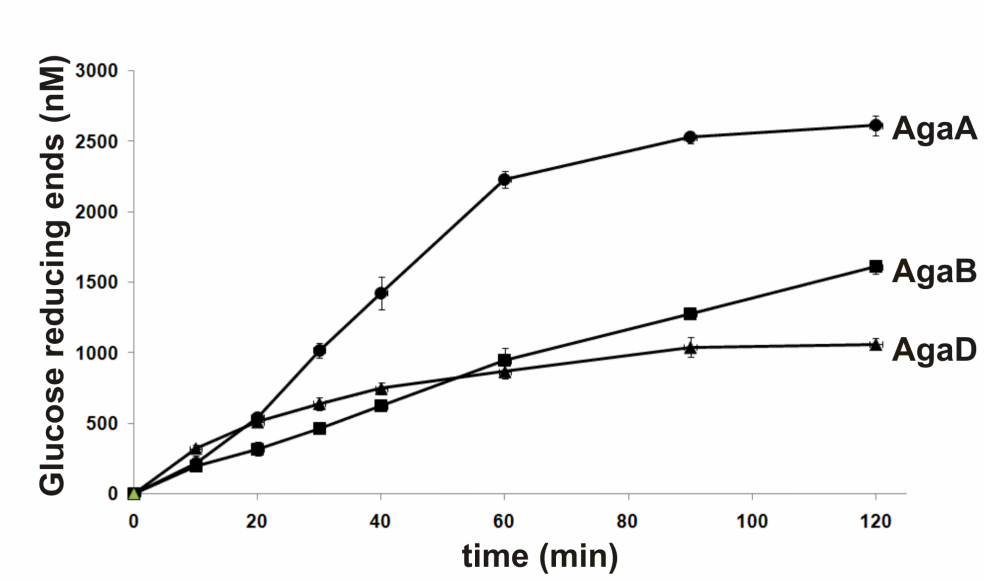


Figure 31: AgaD is less efficient on agarose than AgaA and AgaB. A solution with 0.125% agarose in superfusion was incubated with the three enzymes and aliquots were analysed by the reducing sugar assay under identical conditions.

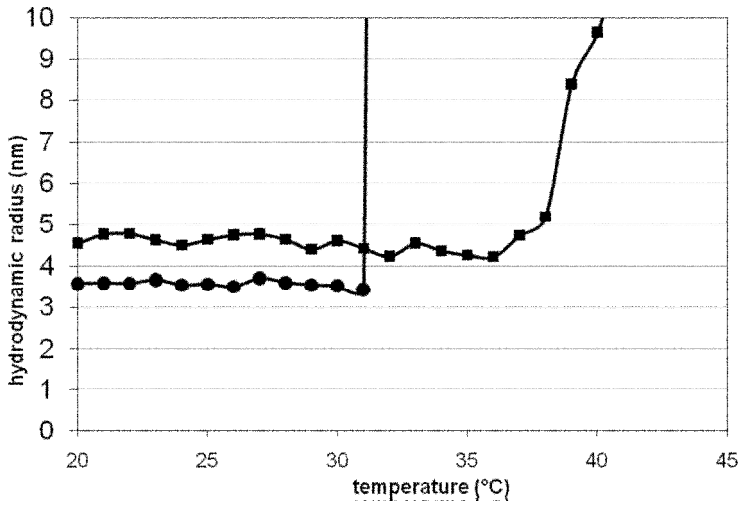


Figure 32: The enzyme AgaD has a lower thermo-stability than AgaB, as measured by dynamic light scattering. The melting temperature was experimentally determined for AgaD (circles) to be 32 ºC, compared to the melting temperature of AgaB 39 ºC (squares). The low temperature stability is shown by the rapid increase of particle size above 32 ºC for AgaD, which is due to the induced denaturation and subsequent aggregation.

This certainly is one reason for the lower efficiency of AgaD, since experiments with non jellyfied agarose have to be carried out above 40 ºC to avoid gelation. However, even when AgaD was used to degrade agarose gel at lower temperature, by adding enzyme until no further change in detectable reducing ends could be measured AgaD did degrade only 70% of the gel, compared to AgaA (not shown). Therefore, I assume that the low thermo stability of AgaD may not be the only factor influencing the difference in amount of released sugars produced on neutral agarose. The difference could also be due to globally production of longer chains, resulting in a lower net amount of reducing ends.

The determination of the enzymatic parameters of AgaD certainly was complicated through the lower thermo stability of this enzyme. With defined oligosaccharides of different degree of polymerization (DP) the kinetic constants and the contribution of subsites to catalysis, can be measured at lower temperatures, since oligosaccharides are soluble. However, at this stage of my thesis I did not have any defined agar-oligosaccharides since these compounds are not commercially available and we had to produce all substrate molecules by ourselves.

Therefore I used agarose in superfusion as described for AgaA and AgaB in Jam et al. (2005) to measure the initial velocity of AgaD. By this method I determined the kinetic parameters to be K_m=6.5 mM, k_{cat}=248 s^{-1} and k_{cat}/Km=48 $mM^{-1}s^{-1}$. These values compare well to those previously described for AgaA (K_m=2 mM, k_{cat}=150 s^{-1}, k_{cat}/Km=75 $mM^{-1}s^{-1}$) and AgaB (K_m=1 mM, k_{cat}=100 s^{-1}, k_{cat}/Km=105 $mM^{-1}s^{-1}$). However due to the low thermo stability, even though the values for AgaD were obtained from the initial rate of degradation, they may be underestimated. Under the above defined conditions, the enzymatic parameters of AgaD are in the same range of magnitude as those of AgaA and AgaB.

III.4.2 AgaD is an endo β-agarase cleaving the β-1,4 linkages in agarose

AgaD has an endo type of degradation behaviour, as shown by the analysis of the oligosaccharide reaction products. The enzyme was incubated with agarose gel at 30 ºC and aliquots were taken for analysis by fluorophore labelled polyacrylamid gel-electrophoresis

(PACE). The produced oligosaccharides were labelled with ANTS and appeared as discrete fluorescent bands. The gel shows that at the beginning of the degradation kinetic, random sized oligosaccharides are produced by AgaD (Figure 33). This type of pattern, with random sized oligosaccharides produced at the very initial step of the reaction, has been reported to be typical for an endo mode of action. These intermediate sized oligosaccharides were progressively degraded to obtain agarose oligosaccharides of the neoagarobiose series. This enzyme produced hexa-oligosaccharides (DP6) and tetra-oligosaccharides (DP4) as final reaction products. Prolonged incubation with high concentrations of AgaD did not liberate smaller oligosaccharides from DP6 or DP4. This shows that the smallest final reaction product of AgaD is a tetrasaccharide and that the hexasaccharide, produced from decasaccharide or longer oligos, is not further degraded by this enzyme.

A similar degradation behaviour has been reported for AgaA, but it contrasts that of AgaB, which is able to degrade DP6 into DP4 and DP2 (Allouch et al. 2003).

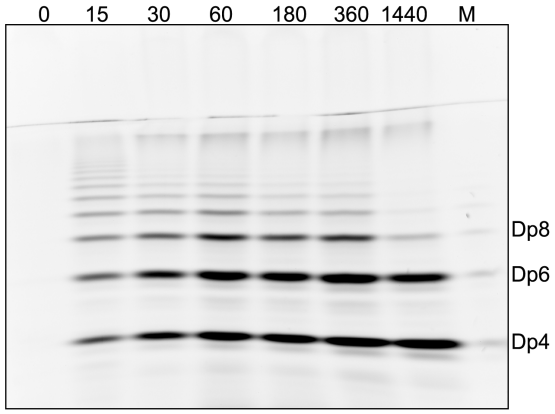


Figure 33: The β-agarase AgaD degrades agarose with an endo mode of action as shown by polyacrylamid-gel electrophoresis analysis of fluorophore labelled oligosaccharide products. The agarose was degraded with AgaD and at different time points (shown in minutes) aliquots were taken. The reaction products obtained during the degradation were labelled with the fluorophore ANTS and separated on a 30% polyacrylamidgel. M= oligosaccharides of the neoagarobiose series produced by AgaB: Dp4= LA-G-LA-G, Dp6=LA-G-LA-G-LA-G and labelled with ANTS. The time is shown in minutes.

III.4.3 Agarase specificities are different on natural substrates extracted from the agarophytes *Gelidium*, and *Porphyra*

The enzymatic properties on neutral agarose did not differ significantly between AgaD, AgaA and AgaB, despite the observed structural differences. Therefore we tested natural substrates extracted from two marine red algal species as substrate and compared the activity of the three enzymes. We chose *Gelidium spinosum* because they are economically exploited. These agarophytes contain high quality agars due to the low amounts of substitutions and produce gels with elevated gel strength, since the natural structure is close to the ideal agarose primary structure. This lack of modifications, reduces the need for further alkaline processing which is an economical factor in agar and agarose production (Mc Hugh 2003). As a second species we chose *Porphyra umbilicalis*, which in contrast to *Gelidium spp*. contains a highly sulfated agarocolloid (soluble agar-like) named porphyran. In porphyran the LA units are sulfated on C6, leading to L6S units, while the G unit can be methylated on C6. These modifications are highly variable in amount and distribution (Rees et al. 1962). The L6S modification has a drastic effect on the rheological behaviour of the polysaccharide since porphyran is soluble and forms viscous solution while agarose forms gels. Nevertheless besides the characteristic L6S-G motifs, natural porphyran also contains variable amounts of ideal agarose LA-G motifs, which result in the fact that agarolytic activity can be observed on porphyran, since these motifs occur in block structures. The degradation of porphyran with agarases has extensively been studied with enzymes from *Cytophaga* species (Turvey et al. 1967; Turvey et al. 1967; Duckworth et al. 1969) and with the β-agarases I and II from *Pseudomonas atlantica* (Morrice et al. 1983; Morrice et al. 1983). These various enzymes were shown to produce agarose oligosaccharides and porphyra/agarose hybrid oligosaccharides, in which the porphyran motifs are located at the non-reducing end. To date however, an agarolytic enzyme that produces porphyran oligosaccharides, consisting of L6S-G motifs, has not been described.

To perform the experiments on natural substrates the cell wall polysaccharides were extracted from the different algal species by hot water (see below). The gel-forming polysaccharides

extracted from *Gelidium spinosum* will further on be named agars, while the viscous solution-forming polysaccharides from *Porphyra umbilicalis* will be named porphyran.

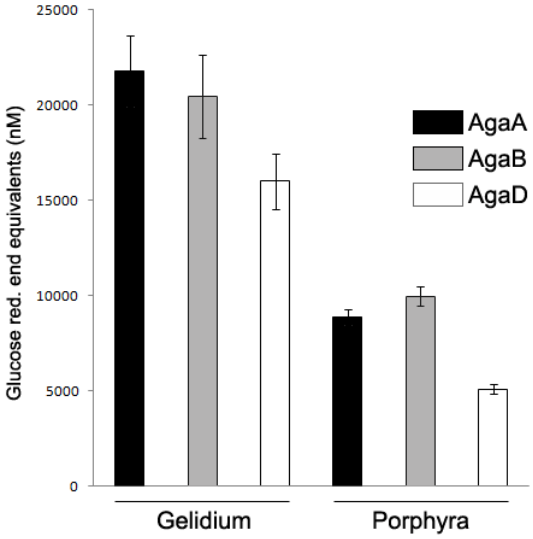


Figure 34. Activity of AgaA, AgaB and AgaD on red algal polysaccharides from *Gelidium* and *Porphyra*. The polysaccharides were extracted with hot water. These extracts were used as substrate for activity measurements with the enzymes.

All three enzymes had highest relative activity on extracts from *Gelidium spinosum* (Figure 34). The high amount of ideal agarose motifs (LA-G) within the agar from *Gelidium spp*. explains the high activity of agarases, which is lower on porphyran where all three agarases produced less oligosaccharide. The difference in released quantity of oligosaccharides for AgaD is only slightly lower on *Gelidium.* This is reminiscent of the AgaD activity on pure agarose gel, in which ~70% sugars are released when compared to AgaA. This difference of AgaD is even more pronounced on porphyran, where only 50% of oligosaccharides are released compared to AgaA and AgaB. This result indicated that AgaA and AgaB may accept modifications such as sulfations within the substrate binding cleft to allow productive binding. In contrast AgaD may be less promiscuous and more specific for longer agarose blocks in the polysaccharide chain.

III.4.4 FACE and HPLC analysis of porphyran degradation by AgaA, AgaB and AgaD

Since AgaD was less active than AgaA and AgaB on porphyran, indicating different specificities on this substrate, I decided to analyze the produced oligosaccharides of the three enzymes by FACE. For FACE experiments one can choose between different fluorophores. With ANTS three negative charges are added per reducing end and therefore any anionic and neutral oligosaccharides that are produced migrate to the anode and become visible on the gel. Therefore this label does not allow deciphering between charged and uncharged oligosaccharides. Thus the neutral 2-amnoacridone (AMAC) was used as a label so that only the anionic oligosaccharides migrated on the gel. In this way neutral oligosaccharides were "subtracted" since they remained in the gel stacking. With this technique subtle differences in enzyme specificity on heterogeneous polysaccharides can be identified (Goubet et al. 2002). Here the reaction products of the three agarases were compared with the reaction products produced by a porphyranase (PorA) from *Z. galactanivorans* which will be described below in manuscript 2. AgaA and AgaB produce a number of discrete bands of which one double band had the same velocity as a porphyran/agarose hybrid oligosaccharide (L6S-G-LA-G-L6S-G manuscript 3) produced by PorA (Figure 35a). H-NMR analysis (data not shown) revealed that these bands produced by AgaA and AgaB correspond to the hybrid oligosaccharide L6S-G-L6S-G-LA-G and that variable methylation on C6 of the G-units accounts for the double band character. This double band is also produced by AgaB and with both enzymes, AgaA and AgaB, additional bands of lower velocity can be identified which were not further characterized. The degradation pattern produced by AgaD is clearly different than that of AgaA and AgaB. None of the major bands produced by AgaA and AgaB appeared at this intensity with AgaD even after adding extensive amounts of enzyme. Compared to AgaA and AgaB, AgaD produced a smear of bands which migrate with lower velocity and may be interpreted as higher hybrid oligosaccharides. This degradation pattern was verified by HPLC and again none of the lower hybrid oligosaccharides

that are produced by AgaA or AgaB appeared as peaks in the chromatogram of AgaD (Figure 35b). The production of L6S-G-L6S-G-LA-G with AgaA and AgaB shows that both enzymes accept an L6S unit on the +1 and on the -4 binding subsite, a fact that has been previously described by (Duckworth et al. 1969) and (Morrice et al. 1983). We further identified the reaction product L6S-G-LA-G-LA-G produced by AgaA and AgaB but neither this oligosaccharide nor L6S-G-L6S-G-LA-G was produced when porphyran was degraded by AgaD. AgaD therefore has a much higher specificity for non substituted blocks of agarose as indicated by the degradation pattern and the protein structures. Whereas in AgaA and AgaB, the L6S unit in -4 is solvent exposed and this unit may interact with loop1 in AgaD, which therefore creates higher substrate specificity for none substituted neutral agarose. Additionally, the domainX may restrict specificity for agarose since it closes off the substrate binding cleft on the side of the positive binding subsites, in particular a tyrosine pointing at the 3,6-anhydrobridge of the LA unit located at the +3 binding site (Figure 30). This has to be confirmed by modelling or co crystallisation studies. However this is the first description of a strict agarase. The more closed and elongated substrate binding cleft may be an adaption to degrade longer neutral oligosaccharides generated from agarose gels by AgaA. Further experiments such as expression analysis or kinetic studies on defined oligosaccharides are needed to have a better understanding of the specific functions in this system of agarolytic proteins.

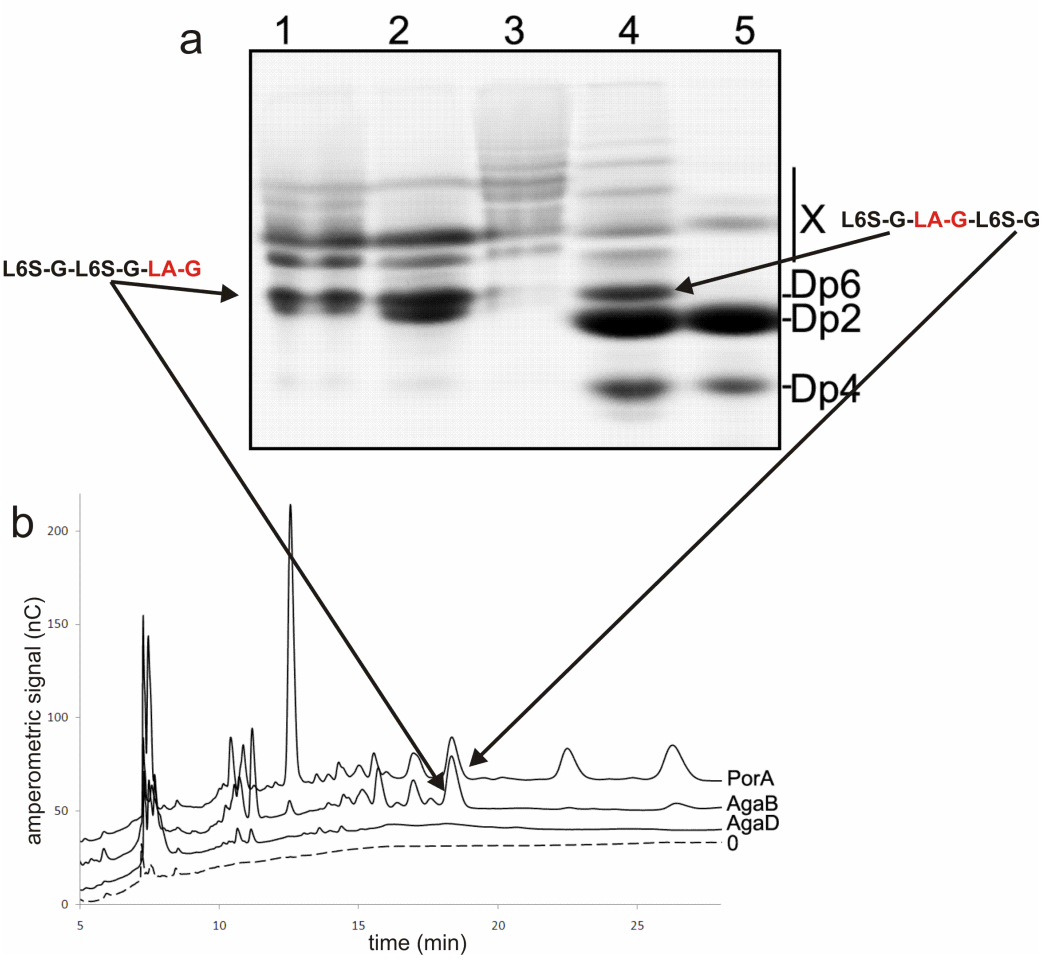


Figure 35: AgaA and AgaB produce sulfated oligosaccharides which are not produced by AgaD. (a) AMAC-PACE of the oligosaccharides produced by porphyran degradation with the different agarases and a porphyranase. Lane1: AgaA, lane2: AgaB, lane3: AgaD, lane4: PorA, lane5: PorA on AgaB pre-digested porphyran. **(b)** HPLC-analysis of the oligosaccharides produced by porphyran degradation.

III.5 Conclusion: The new β-agarase AgaD together with AgaA,B as part of the agarolytic enzyme system of Z. galactanivorans

The elongated substrate binding cleft and the additional domainX render AgaD more specific for non substituted agarose blocks in heterogeneous agarocolloids. This is the reason for the lower activity of AgaA on porphyran compared to AgaA and AgaB which release smaller hybrid oligosaccharides from this type of substrate. The higher activity on longer agarose oligosaccharides may be accounted by synergistic action with AgaA which has been previously shown to be efficient in agarose gel degradation. AgaA may release blocks of agarose which may subsequently be efficiently processed by AgaD. Agarose is the preferred substrate for AgaD which was further verified by mRNA expression analysis in which porphyran repressed the expression of AgaD which was elevated with agarose as substrate (data not shown). The absence of L6S-G motifs on the reducing end of oligosaccharides that are released by AgaA and AgaB from porphyran shows that the -2 and -1 binding subsites are selective for the LA-G agarose motive in agarases and therefore critical for productive binding. In contrast a L6S unit bound in subsite +1 is possible because L6S-G units are encountered on the non reducing end of produced hybrid oligosaccharides. The longer substrate binding cleft of AgaD together with the additional domainX render this enzyme more specific for longer non substituted agarose stretches whereas the shorter and more open substrate binding clefts of AgaA,B allow a broader substrate specificity (Figure 36).

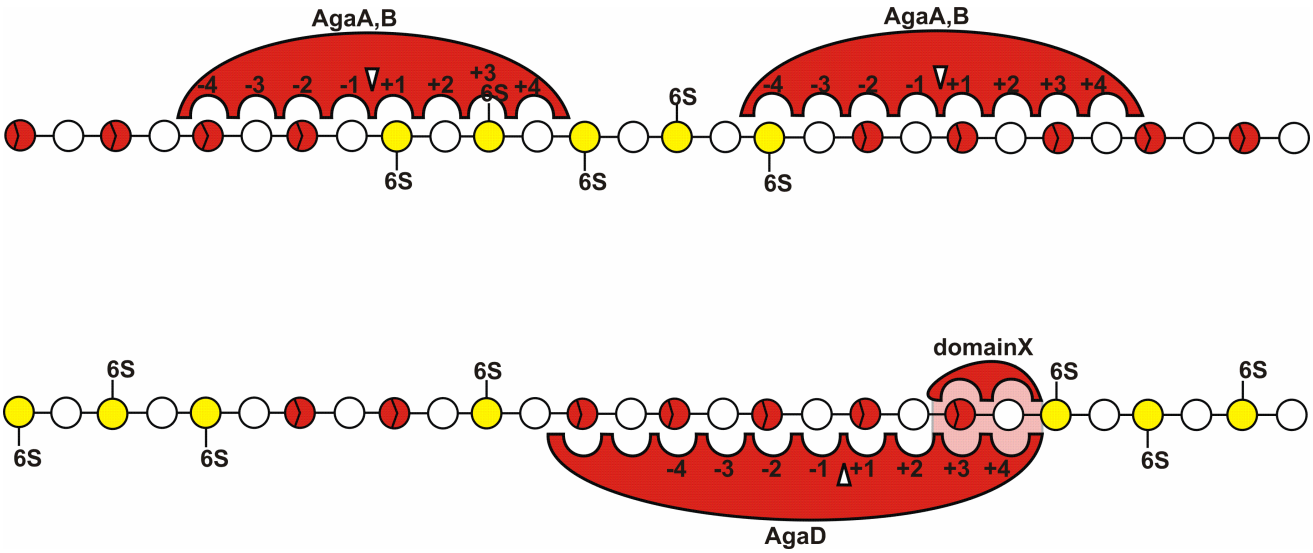


Figure 36: Substrate specificity of agarases from *Z. galactanivorans*. The red circles illustrate 3,6-anhydro-L-galactopyranose (LA), the white circles D-galactopyranose and the yellow circles L-galactopyranose-6-sulfate.

IV. The first β-porphyranases PorA and PorB

The enzymes Zg_2600 PorA and Zg_1017 PorB which both fall into the agarase-like clade did not posess any activity on an agarose plate and on purified agarose oligosaccharides (data not shown). In order to reveal their function a biochemical screening was carried out which will be described below and in parallel I aimed for the crystallization and structural analysis of both enzymes which will be described in the following section.

IV.1 Crystallisation of PorA

As mentioned before, a scaled up production of heterologous expressed protein samples were performed in order to have enough protein for a thorough biochemical and structural

characterization.

The enzyme Zg_2600 PorA could be purified and showed minor problems when trying to concentrate it which was resolved by adding 5 % of glycerol. This protein started to form visible single crystals of perfect shape immediately after setting up the drops. These crystals grew overnight to exploitable single crystals which could be directly used for X-ray experiments without further optimization (Figure 37).

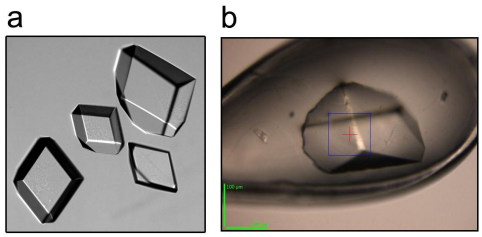


Figure 37: The protein crystals of PorA. Drops of this enzyme were produced with almost the same crystallisation solution as obtained in the screen. Crystals appeared instantaneously and reached final size within one to two days. They were of high diffracting power and a dataset with 1.1 Å resolution could be obtained (data not shown).

PorA was crystallized by the hanging drop method in 18-20% PEG 4000, 0.2 M ammonium sulfate and 0.1 M sodium acetate pH 4.6 at 20 °C and for 1-3 days (Figure 37). The crystals belonged to space group $P3_121$ with unit cell parameters of a=b=70.17 Å, c=92.78 Å, $\alpha=\beta=90°$, $\gamma=120°$.

Drops were created with 2 µl of 2.6 mg/ml protein solution plus 1 µl of crystallization solution and equilibrated against 500 µl crystallization solution. For cryoconservation crystals were soaked in crystallization solution with 10% glycerol prior to flash freezing at 100 K. Native crystals diffracted to 1.1 Å resolution and a complete data set was collected on beamline ID23-EH1 equipped with an ADSC Quantum Q315r detector (ESRF, Grenoble, France) at a wavelength of 0.82655 Å.

Since all trials failed to solve the structure by molecular replacement, using various family GH16 structures as templates, and given the fact that the diffraction power of the crystals was excellent, I decided to generate heavy atom derivatives by crystal soaking experiments prior to producing a

selenomethionine labeled protein. Moreover, I selected to test compounds for which the heavy atom possesses an absorption edge easily accessible by the tuneable synchrotron beamlines, to eventually use a SAD/MAD or SIRAS to determine the 3D structure. In order to generate a heavy atom derivative, various native crystals were soaked for 24 hours in the crystallization solution supplemented with 5 mM potassium-tetrachloro-aureate (L-III edge around 1.04 Å), 5 mM $Yb(NO_3)_3$ (L-III edge around 1.63 Å) and 5 mM $Er(NO_3)_3$, (L-III edge around 1.4834 Å), respectively. After the rather long incubation period the crystals soaked with potassium-tetrachloro-aureate were of yellow colour without being effected in their morphology (Figure 38).

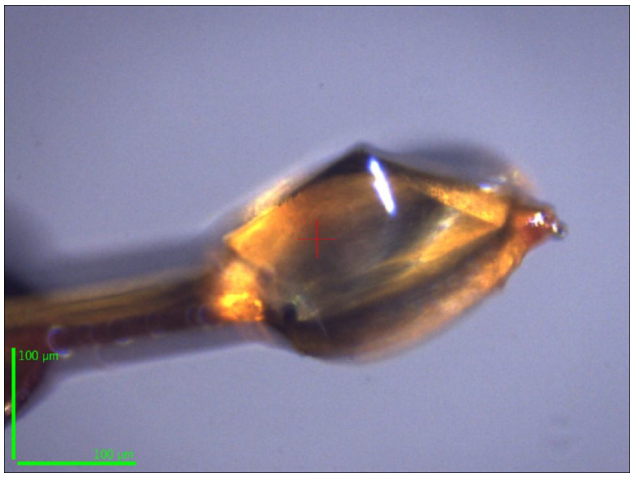


Figure 38: Single crystal of PorA Au-derivative on the beamline prior to the MAD data collection.

IV.1.1 3D structure solution of PorA using a gold derivative

The datasets necessary to perform the MAD experiment were collected at the ESRF (European Synchrotron Radiation Facilities, Grenoble, France) beamline ID23-EH1, equipped with an ADSC Quantum Q315r detector.

On the beamline, the presence of heavy atoms was first verified using the automatic fluorescence

measurement, a setup that is today a standard implementation on all MX beamlines. This preliminary measurement unambiguously showed the presence of gold in a first derivative crystal. Since this gold derivative crystal of PorA, displayed an excellent diffraction power (in the same manner as the native crystals), I decided to collect a complete MAD data set at four different wavelengths. An energy-scan was then measured on the PorA-Au crystal, in order to determine the exact wavelengths around the Au L-III absorption edge (Figure 39) at which the successive data sets were to be recorded. Accordingly, the dataset corresponding to the absorption peak energy 11925.1 eV was collected at wavelength 1.03965 Å, the inflection point at 11921.2 eV (1.04005 Å), the low remote at 11975.1 eV (1.03530 Å) and the high remote at 12000 eV (1.03315 Å). The data collection strategy was chosen to be able to extract the maximum anomalous information from the data sets, in case the crystal 'died' due to irradiation damage. The data sets were therefore collected in the order peak, inflection, low energy remote and high energy remote.

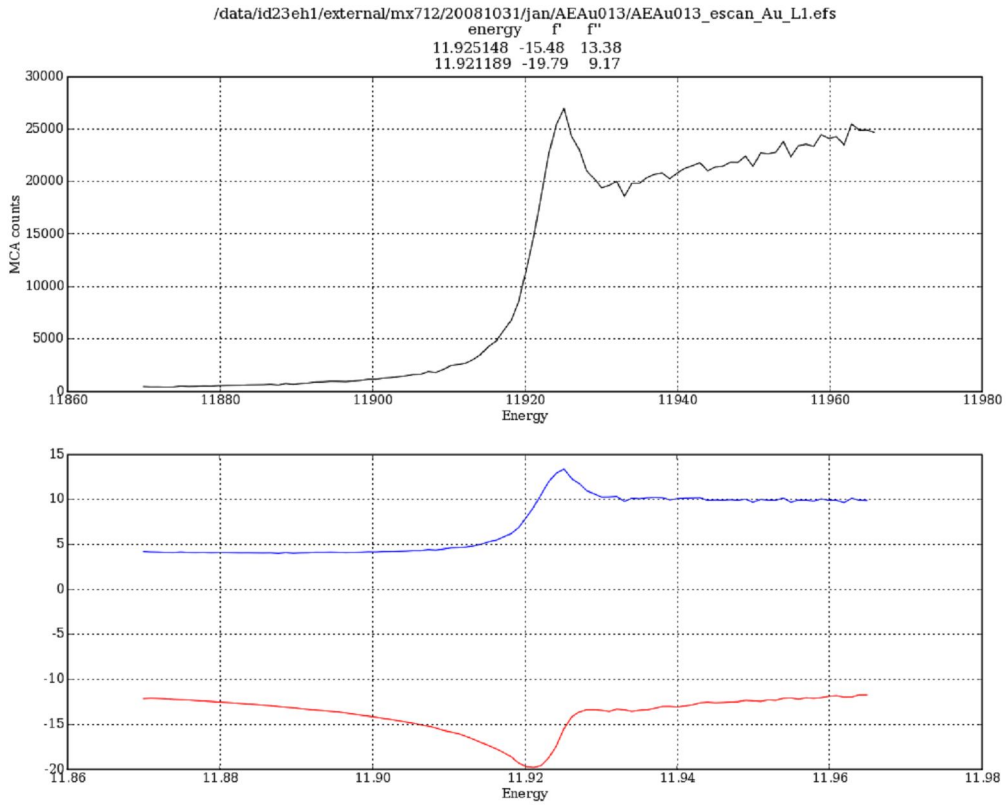


Figure 39: Energy scan around the Au-absorption edge of the PorA-Au derivative crystal.

Table 8. XPREP analysis of "peak" dataset (normalized intensities)

```
Anomalous signal/noise ratios (1.0 is random).  The first line is based on
  input sigmas, the second on variances of F+ and F- (if not already averaged):
 Inf - 8.0 - 6.0 - 5.0 - 4.0 - 3.5 - 3.0 - 2.8 - 2.6 - 2.4 - 2.2 - 2.0 - 1.8 A
    4.50  3.64  2.73  2.22  2.07  2.24  2.46  2.46  2.32  2.02  1.76  1.51
    3.37  2.92  2.14  1.67  1.80  1.91  1.74  1.71  1.73  1.79  1.73  1.62

    81.2  Neighbors used on average for F+/F- local scaling
 Rint(anom) =   0.1020  before and   0.1011  after local scaling
```

Anomalous data were individually processed with XDS and scaled with XSCALE (Kabsch 1988). To estimate the phasing power of the collected MAD data, a first analysis on the unmerged "peak" dataset was performed using XPREP (G. Sheldrick, part of the Bruker

structure determination suite). The normalized intensity measurements unambiguously displayed an important anomalous signal, covering the complete resolution range (Table 8). The anomalous Patterson map that was subsequently calculated also using XPREP, showed the presence of two to three atoms per asymmetric unit (Figure 40).

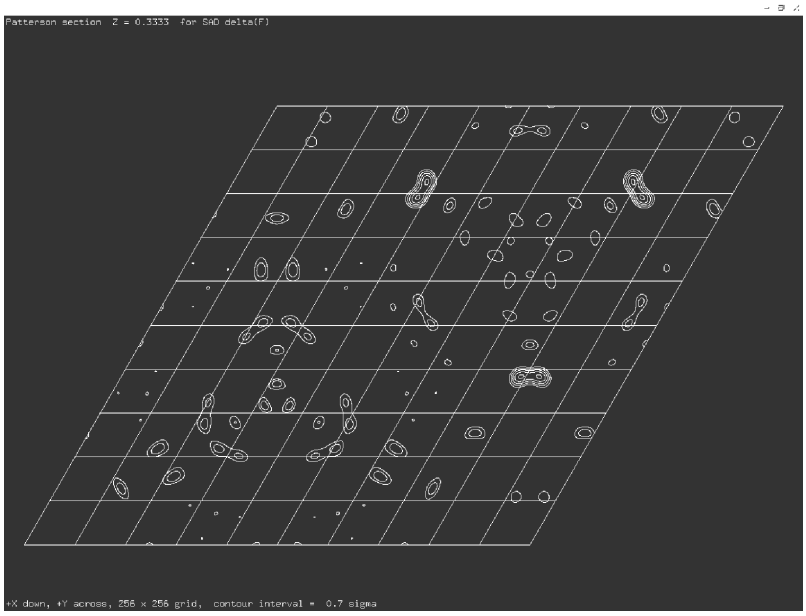


Figure 40: Harker-plot at section Z=0.3333 of the anomalous Patterson-map, calculated with the normalized intensities of the "peak" data and contoured at 0.7s levels. The six symmetry related Harker vectors of one heavy atom position per asymmetric unit can unambiguously be identified. The two following peaks are identified with hkl2map. The respective coordinates in the Patterson map are:

	u	v	w	relative height/sigma
First peak	0.2101	0.3797	0.3333	4.78
Two following peaks	0.7815	0.1558	0.3333	2.73
	0.7999	0.3173	0.3333	2.56

The crystallographic structure solution was calculated using the SHELX-suite through the graphical interface HKL2MAP (Pape et al. 2004). After scaling the data sets together with SHELXC, the heavy atom sub-structure was determined with SHELXD, followed by the phasing step and density modification with SHELXE. The program-suite automatically tests both 'hands'

for the determined heavy atom positions but does not test enantiomorphous space groups (Figure 41).

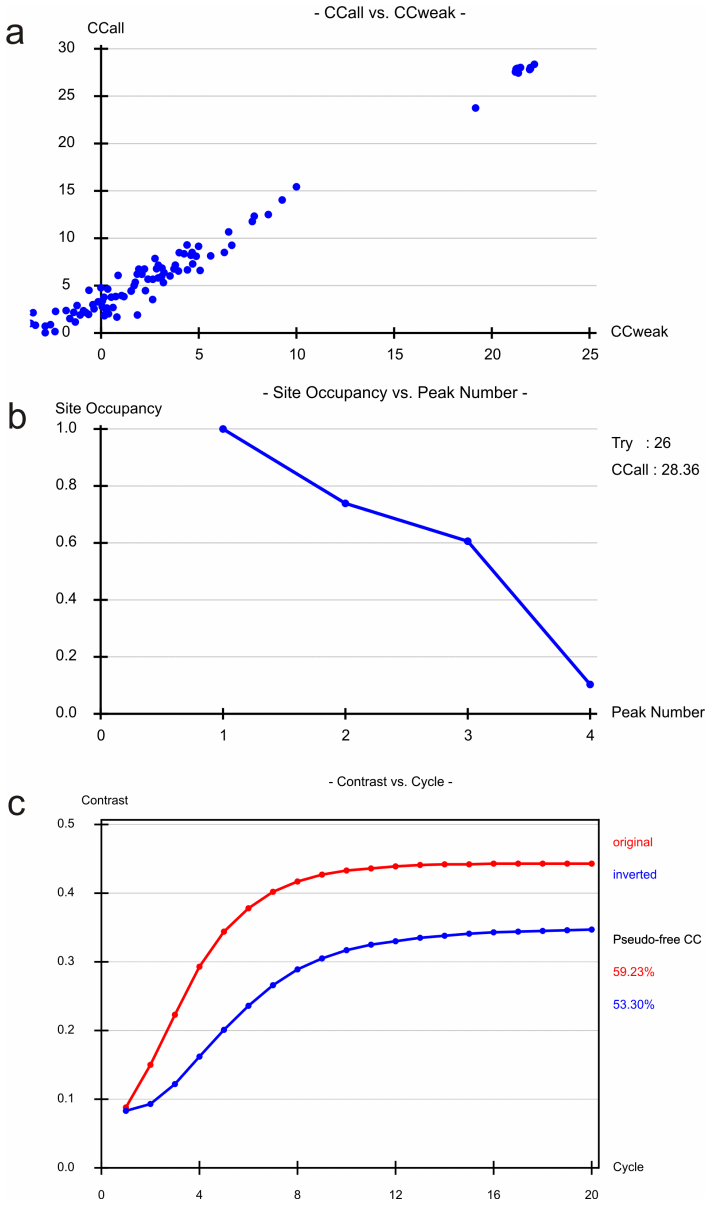


Figure 41: Graphical display of the phasing statistics as determined by the ShelX suite. (a) The plot display shows the contrast of the Patterson map deconvolution, trial by trial. The overall correlation coefficient is plotted as a function of the correlation function of the weakest trial. Each spot represents one trial. (**b)** The site occupancy plotted as a function of site number. The structure contains one major Au binding site and two secondary binding sites. (**c)** The pseudo-free correlation coefficient is plotted as a function of phasing trial for the two coordinate

'hands' that are tested. The evident contrast between the two solutions is a sign that the phasing procedure was successful.

These steps were therefore repeated for both enantiomorphous space groups $P3_121$ and $P3_221$ for this crystal system. Best phases and electron density maps were obtained with space group $P3_121$. The resulting electron density map was further improved by solvent flattening with the program DM (CCP4) Figure 42). Starting phases were used for the initial model which was build with ARP/*w*ARP and REFMAC (Murshudov et al. 1997; Perrakis et al. 1999) as part of the CCP4 suite (Potterton et al. 2003). The refinement was carried out with REFMAC and model building with Coot (Potterton et al. 2004). Water molecules were added automatically with REFMAC-ARP/*w*ARP and visually verified. All further data collection and refinement parameters are given in Table 1 of the Supplemental Material of manuscript 2.

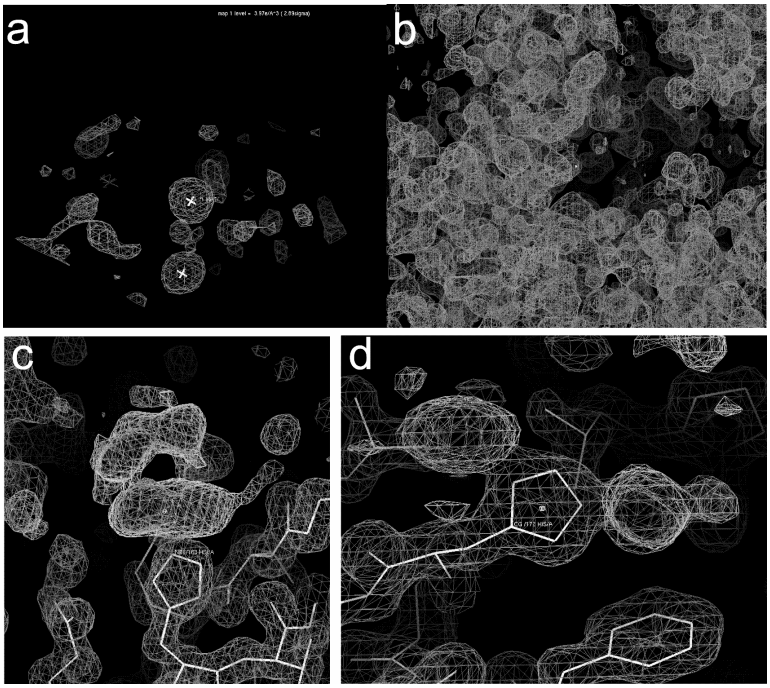


Figure 42: Experimental and refined electron density maps, illustrating the phasing procedure. (a) The first electron density map calculated from the phases obtained for the 'original hand' in Shelxe, contoured at 3 σ level. **(b)** Overall view of the first electron density map contoured at a 1 σ level, showing distinct protein and solvent regions. **(c)** Fourier difference density map (Fo-Fc) calculated after structural refinement of the Au-data set showing the major gold-chlorate position at a histidine of the active site cleft of the enzyme. **(d)** Fourier difference density

map (Fo-Fc) calculated after structural refinement of the Au-data set showing the two minor gold-chlorate position, both attached to the same histidine residue.

IV.1.2 Crystallization of PorB

The enzyme Zg_1017 PorB was produced at high levels and could be purified and concentrated without any problems to 8 mg/ml in 20 mM Tris-HCl buffer pH 8. The first crystallization-screening experiments were carried out at 292 K with the sitting-drop vapour diffusion method in 96 well Corning plates. The crystallisation screening was performed with JCSG+, PACT and the PEG I screens from Qiagen, which makes a total of 288 conditions in three 96-well plates (Corning no. 3551, one well sitting drop). A crystallization robot (Proteomics Solutions, Honeybee961) was used for pipetting and the drops contained 300 nl protein solution that were mixed with 150 nl of reservoir solution. After visual identification of initial crystallisation conditions, these were further optimised in 24 well Linbro plates by the hanging-drop vapour diffusion method. The drops were prepared on siliconized cover slips by mixing 2 µl of protein solution with 1 µl of well solution and the equilibration was performed against 500 µl reservoir solution at room temperature.

One promising screening condition was composed of 200 mM lithium-sulfate, 100 mM sodium acetate pH 4.5 and 50 % PEG 400 and the second condition of 200 mM lithium sulfate, 100 mM sodium acetate pH 4.5 and 30 % PEG 8000. These conditions were varied in concentration of salt or PEG, as well as the molecular weight of PEG in grid screens to reproduce the crystals obtained in the initial screen. Small needle-like crystals or crystal balls (Figure 43) could be reproduced when 300-400 mM lithium sulfate and 34-36 % of PEG 1000 with 100 mM sodium-acetate pH 4.5 was used as crystallisation solution. Some of the crystals were bigger than others and when these were tested on the in house X-ray source they appeared to be salt. This results most probably from the high salt-PEG concentration that was necessary to induce crystal growth. At this point, it was not sure whether all the crystals in the drop were salt, or if they were

constituted of protein. The small size of these needles did not allow to use coomassie based staining, that colour protein crystals blue. Therefore it was not obvious if any optimization effort was worthwhile. To clarify the situation and inspired by the use of UV lamps for protein crystal identification (Carsten Dirks and Arne Meyer, pers. communication) I used a microscope with a 280 nm light source.

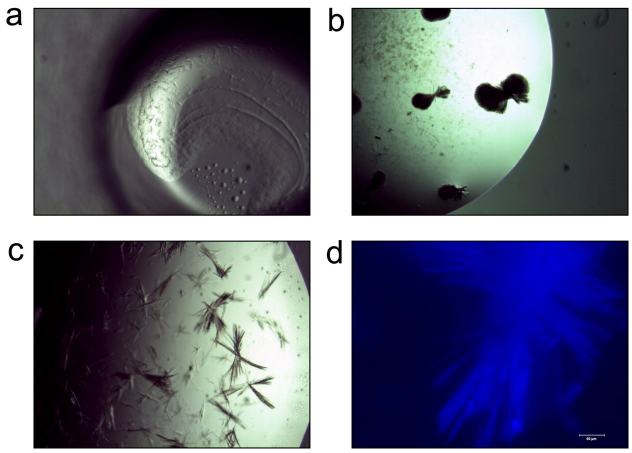


Figure 43: Crystallization screening led to small needles. **(a)** One of the two wells of screening conditions in which crystals appeared. **(b,c)** The crystals could be reproduced in larger scale with a urchin or needle shape. **(d)** To verify that these crystals were constituted of protein before further optimization attempts were made, a microphotograph at 280 nm was taken and the blue glowing crystals could thereby be identified as protein crystals.

With this approach the small needle clusters could be identified as protein crystals because they appeared blue due to the tryptophan fluorescence (Figure 43d). Therefore further attempts were made to optimize their size and shape. After further optimization and by changing the buffer to morpholino-ethane-sulfonic acid and varying the pH from 5.5 to 6.5, triangular single crystals grew from phase separation (Figure 44). These were then used for diffraction studies (see below).

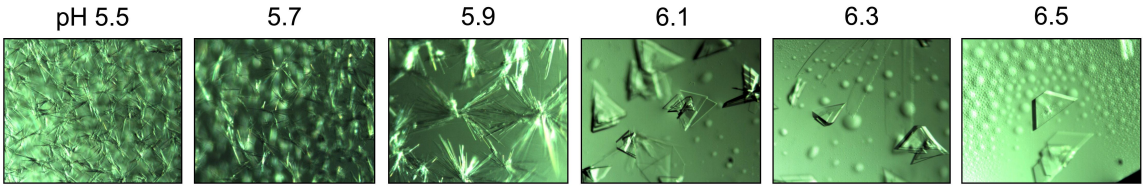


Figure 44: Crystal optimization of PorB. Changing the buffer to MES and adjusting the pH resulted in larger single crystals**.** These crystals belong to space group $P3_121$ and diffracted to 1.8 Å resolution.

Total sitting drop size was 3 µl which was composed of 2µl protein solution (5.6 mg/ml in 20 mM Tris pH 7.5) and 1 µl of crystallization solution. Drops were equilibrated against 500 µl crystallization solution at RT. For cryo-conservation crystals were brought to 10 % glycerol in 5 % increments in the crystallization solution. The structure of PorB was solved by molecular replacement Phaser (CCP4) with the 3D-structure of PorA as search model. The solution was unique and the initial R factor was 55.4% in the resolution range 45-1.8 Å. After 10 initial rounds of manual refinement with REFMAC and model building with Coot, phases and model were used for automatic model construction with ARP/*w*ARP. The final model was manually finished with REFMAC and Coot. For all datasets, five percent of the observations were flagged as free (22) and used to monitor refinement procedures and a summary for data collection and final refinement statistics are given in Table 1 (Supplementary information manuscript 2). All crystallography figures were created with the PYMOL Molecular Graphics System (http://www.pymol.org).

IV.1.3 The crystal structures of PorA and PorB

In comparison to the agarases AgaA and AgaB the substrate binding cleft of PorA and PorB contains several basic residues (arginine, lysine and histidine) (Figure 45) rendering the substrate binding cleft basic around the catalytic residues. Both enzymes PorA and PorB share the jelly roll fold but are nevertheless remarkably different since the cleft is generally more open in PorB and closed in PorA. This derives from a loop that is positioned above the +1 and +2 subsites in the substrate binding cleft of PorA. The purpose of this loop is not yet understood but its

structural arrangement indicates that motion must occur for productive binding because any modelled substrate would be sterically hindered by the loop in this position (data not shown). This loop has low B-factors and a structure in complex with substrate will hopefully give more insights. All further details are given in manuscript 2. Both structures were deposited at the Protein Data Bank with accession codes: 3ILF and 3JUU for PorA and PorB respectively.

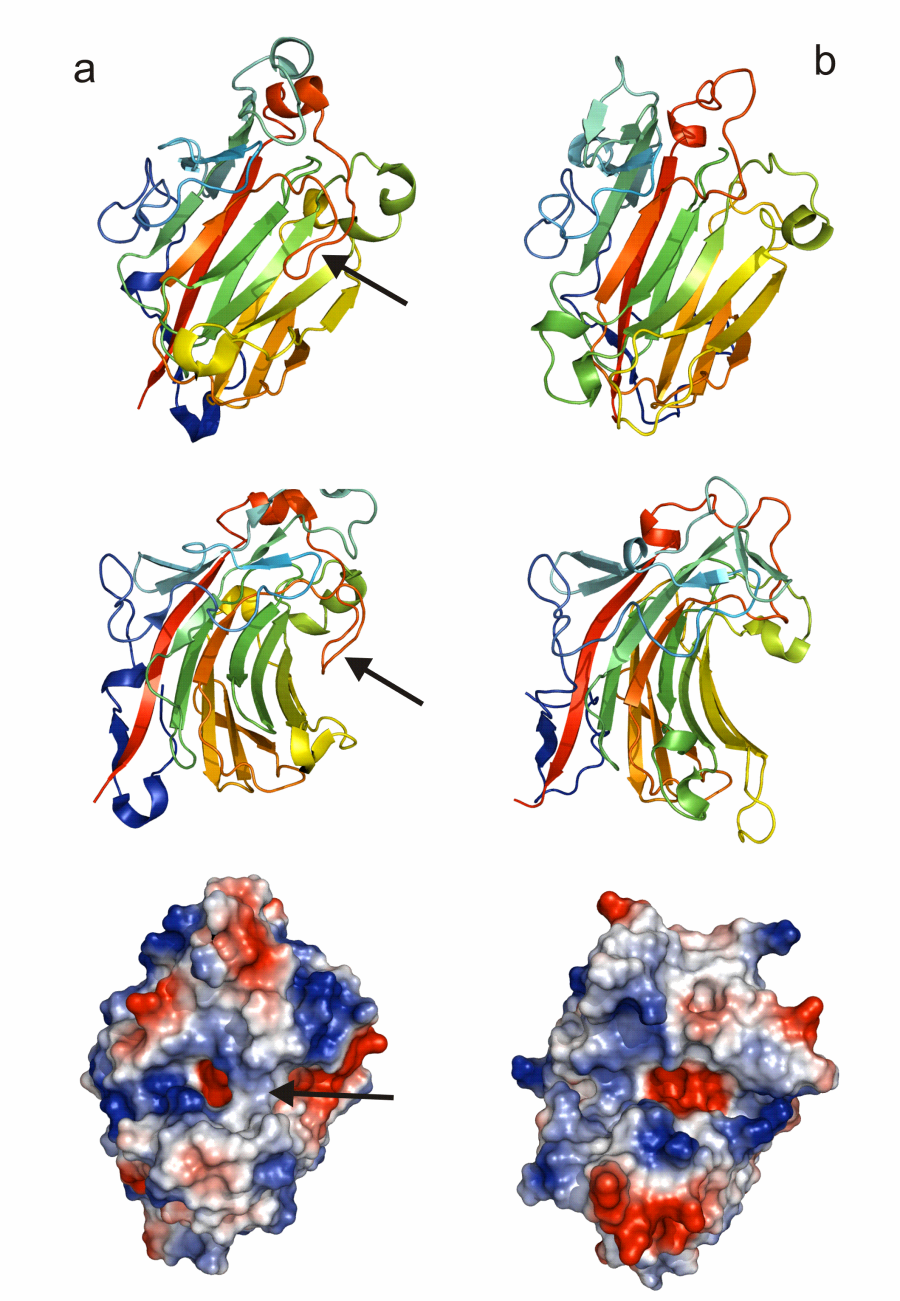


Figure 45: The crystal structures of PorA and PorB. Cartoon plots and electrostatics surface representation calculated for PorA **(a)** and for PorB **(b).** The arrows point towards the loop that extends into the substrate binding cleft of PorA.

IV.2 The discovery of the β-porphyranase activity

During my thesis I obtained crystals of PorA and PorB before knowing the function of these proteins. When PorA and PorB were tested on agarose or on agarose oligosaccharides no activity could be measured. Therefore I searched for algal polysaccharides as substrates to screen for enzymatic activity. The amount of commercially available algal polysaccharides is limited and often they are chemically modified and different from their original chemical composition. For example this is the case for agar and agarose which are usually alkali treated to remove the sulfate modifications and to convert L6S into LA. Therefore I decided to use natural polysaccharides, directly extracted from the marine algae, as samples for activity screening.

Because the phylogenetic analysis shows that both enzymes are related to agarases I decided to screen their activity on natural polysaccharides, including agarocolloid, from red algae that I collected close to Roscoff in the intertidal zone. I screened for activity on hot water extracts from the different red algae *Gelidium spinosum*, *Gracilaria gracilis*, *Laurencia pinnatifida*, *Porphyra umbilicalis*, *Delesseria sanguinea* (from which *Z. galactanivorans* was isolated) and *Zostera marina* a marine angiosperm which has been recently shown to contain mixed linked sulfated galactans (Aquino et al. 2005). The result of such a screen is shown in Figure 46 and shows that both PorA and PorB have highest activity on the polysaccharide extracted from *Porphyra umbilicalis*. These enzymes are indeed the first described β-porphyranases. Their crystal structures and biochemical characterization are described in manuscript 2. The biochemical and structural description of these two new enzymes allowed us to define the new sub family of β-porphyranases within GH16. By data mining these enzymes were found in many marine bacteria. But surprisingly porphyranase genes were also found in human gut bacteria, in particular in *Bacteroidetes plebeius* collected from feaces of Japanese individuals.

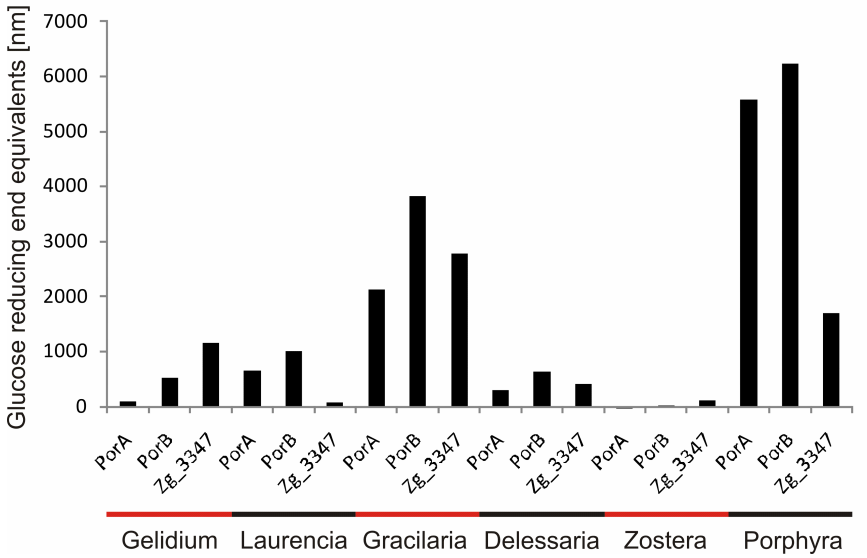


Figure 46: Activity of the enzymes PorA, PorB and Zg_3347 on hot water extracts from marine algae. The same quantity of dry weight algal powder was used for each algae and extensively extracted with 95ºC water. The extract was centrifuged and the different enzymes were added. After incubation the amount of released sugars was measured by reducing sugar essay and compared with a glucose calibration curve. Notably, the enzyme Zg_3347 released a comparatively high amount of reducing ends from *Gracilaria gracilis*, an activity which was not further analysed. Presented activities are not normalized against the amount of extracted dry weight.

IV.3 Introduction for manuscript2: Porphyranases and agarases constitute the first example of a nutrition derived CAZyme update into human gut bacteria

The human gut harbours the vast number of tens of trillions of microbes. This intestinal microbial community, also referred to as the gut microbiome, is constituted mainly of the bacterial domain but also archae and even eukaryotes are encountered. Most bacteria colonize the human colon where among their other functions they digest dietary polysaccharides to extract energy and carbon. Microbial polysaccharide fermentation and the bacterial production of short chain fatty acids in the intestine contributes to 10% of our daily calorie harvest. Without the microbes many polysaccharides would be non digestible for the host because the CAZymes used to break down the polysaccharides are not encoded in the human genome.

How during the course of human evolution, was the diversity of CAZymes found in gut bacteria obtained and did differences in diet lead to differen CAZymes in indiviual humans?

The phylogenetic detection of a marine porphyranase in a Japanese gut bacterium pointed towards the novel route of gene transfer from a seaweed associated marine bacteria to the human gut. The description of this CAZyme update, the porphyranases and the consequences of their occurence in gut bacteria are presented in manuscript2.

IV.4 Manuscript 2

The "Sushi factor" enabled porphyranase gene transfer from the oceans to Japanese gut bacteria

Jan-Hendrik Hehemann[1,2], Gaëlle Correc[1,2], Tristan Barbeyron[1,2], William Helbert[1,2], Mirjam Czjzek[1,2*] & Gurvan Michel[1,2]

[1]*Université Pierre et Marie Curie, Paris 6, Végétaux marins et Biomolécules UMR 7139, Station Biologique de Roscoff, F 29682, Roscoff, France*

[2]*Centre National de la Recherche Scientifique, Végétaux marins et Biomolécules UMR 7139, Station Biologique de Roscoff, F 29682, Roscoff, France*

**Corresponding author*

Published in

Nature

IV.4.1 Abstract

Gut microbes supply us with energy from dietary polysaccharides by providing carbohydrate active enzymes, or CAZymes, which are not encoded in our genomes. These enzymes target polysaccharides from terrestrial plants, which dominated diet in mammalian evolution. Studies show that glycosidases are exchanged between gut bacteria by horizontal gene transfer. Yet whether CAZyme updates are transferred from bacteria living outside the gut has never been shown. Here we present the first porphyranases from a marine bacterium, *Zobellia galactanivorans*. Their structural and biochemical characterization describes the unambiguous substrate specificity for porphyran, the cell wall polysaccharide from the red seaweed *Porphyra spp*. also known as nori in Japan. This sulfated polysaccharide has been known since decades but porphyranases have never before been described. Their identification provides phylogenetic evidence that genes coding a porphyranase, and associated proteins used for *Porphyra* degradation, have been transferred to the gut bacterium *Bacteroides plebeius* which was isolated from Japanese individuals. The presence of a porphyran degradation system in human gut bacteria can be explained by the Japanese life style, since *Porphyra* is a significant ingredient in Japanese cuisine and consumed raw in Sushi. The marine origin of these novel enzymes, shows that contact with environmental microbes attached to non sterile food sources is a factor of CAZyme diversity in human gut microbes.

The human gut is dominated by microbes belonging to two bacterial phyla, *Bacteroidetes* and *Firmicutes*. An exemplary well studied gut microbe is *Bacteroides thethaiotaomicron*, which breaks down complex polysaccharides in vitro (Xu et al. 2003), in vivo (Sonnenburg et al. 2005) and which has one of the largest repertoires of CAZymes among sequenced microbes. Its genome contains 237 glycoside hydrolases, and 15 polysaccharide lyases, as well as 163 paralogs of two outer membrane proteins involved in binding and import of starch (SusC/SusD) (Koropatkin et al. 2008). A fundamental question is how was such gene diversity acquired in gut

bacteria? Evolutionary factors that influence gene content and diversity include gene duplication, gene loss and horizontal gene transfer (HGT) (Lozupone et al. 2008). To resolve a HGT event is a widely acknowledged problem (Nakamura et al. 2004). In the gut ecosystem lack of sufficient sequencing depth within the *Bacteroidetes* complicates distinguishing transfer from biased sampling (Xu et al. 2007) and it is difficult to resolve origin.

The repertoire of CAZymes in herbivorous and omnivorous gut microbes shows functional convergence with plant polysaccharide-degrading bacteria found in the terrestrial ecosystem (http://www.cazy.org). This resemblance, which complicates the identification of HGT of new genes, has been shaped by millions of years of gut microbiome evolution (Ley et al. 2008) with plants from the green lineage as major source of food. The situation may change when an unusual polysaccharide becomes a sugar source in human nutrition; red macroalgae, which contain unusual sulfated polysaccharides, are traditionally consumed in Japan where seaweeds contribute significantly to the daily diet (15 g/person/day (Fukuda et al. 2007)). The red alga *Porphyra tenera* is the most consumed seaweed, eaten in various dishes, especially wrapped around rice in sushi. About 400 000 wet tonnes of *Porphyra* are cultivated per year in Japan and processed into ~10 billion nori sheets (Mc Hugh 2003). Thus porphyran, the main cell wall polysaccharide of *Porphyra*, is a daily natural dietary fibre for the Japanese. Agars, carrageenans and other red algal cell wall polysaccharides are economically exploited food additives and part of the modern diet. Sulfate substitutions are a common chemical modification of these seaweed polysaccharides (Craigie 1990), yet terrestrial plants do not produce any sulfated polysaccharides (Kloareg et al. 1988). Therefore, glycoside hydrolases acting on sulfated algal polysaccharides are specific for marine ecosystems and used by marine bacteria to gain energy and carbon (Michel et al. 2006). Several agarases and carrageenases from marine bacteria have been recently characterized due to their biotechnological interest (Michel et al. 2001; Michel et al. 2001; Allouch et al. 2003; Flament et al. 2007), whereas little is known about glycoside hydrolases specific for unexploited algal polysaccharides, such as porphyran. The environmental genomics revolution produces enormous amounts of sequence data. For example, at a single sampling site in the North Atlantic Ocean over 1.2 million new genes were identified by metagenomics (Venter et al. 2004). Without new biochemical screening techniques to fill the gap of functional

knowledge, such virtual information remains un-exploitable. This limits our understanding of ecosystem function, for instance by environmental transcriptomic approaches (Shi et al. 2009).

Our research focuses on the identification of new enzymes specific for sulfated algal polysaccharides by a multifaceted post-genomic approach. These efforts led to the characterization of a new class of glycoside hydrolases specific for porphyran and allowed to resolve for the first time a HGT event between microbes from the ocean and the human gut.

IV.4.2 Discovering a new enzyme activity

The marine bacterium Z*obellia galactanivorans* was isolated from the red seaweed *Delesseria sanguinea* for its capacity to degrade agars and carrageenans (Barbeyron et al. 2001). The first structurally and biochemically described β-agarases (AgaA and AgaB) and κ-carrageenase (CgkA) were from this marine *Bacteroidetes* (Barbeyron et al. 1998; Allouch et al. 2003; Jam et al. 2005). However, the recent genome analysis of *Z. galactanivorans* (T.B. et al., in preparation) reveals the existence of five proteins (Zg1017, Zg2600, Zg3376, Zg3628 and Zg3640) which are distant relatives of β-agarases and κ-carrageenases (in average 23% sequence identity with AgaA and CgkA). All of these sequences feature the catalytic signature EXDXXE typical of glycoside hydrolase family 16 (GH16) (Figure 47), but do not possess the critical residues needed for recognition of agarose (Allouch et al. 2003) or κ-carrageenan (Michel et al. 2001). To identify their substrate specificity these novel GH16 enzymes were included in a medium throughput expression strategy. Two recombinant proteins, Zg1017 (272 residues, M_w= 31.3 kDa) and the catalytic module of Zg2600 (256 residues, M_w= 29 kDa), were expressed solubly in *Escherichia coli* (Figure 20).

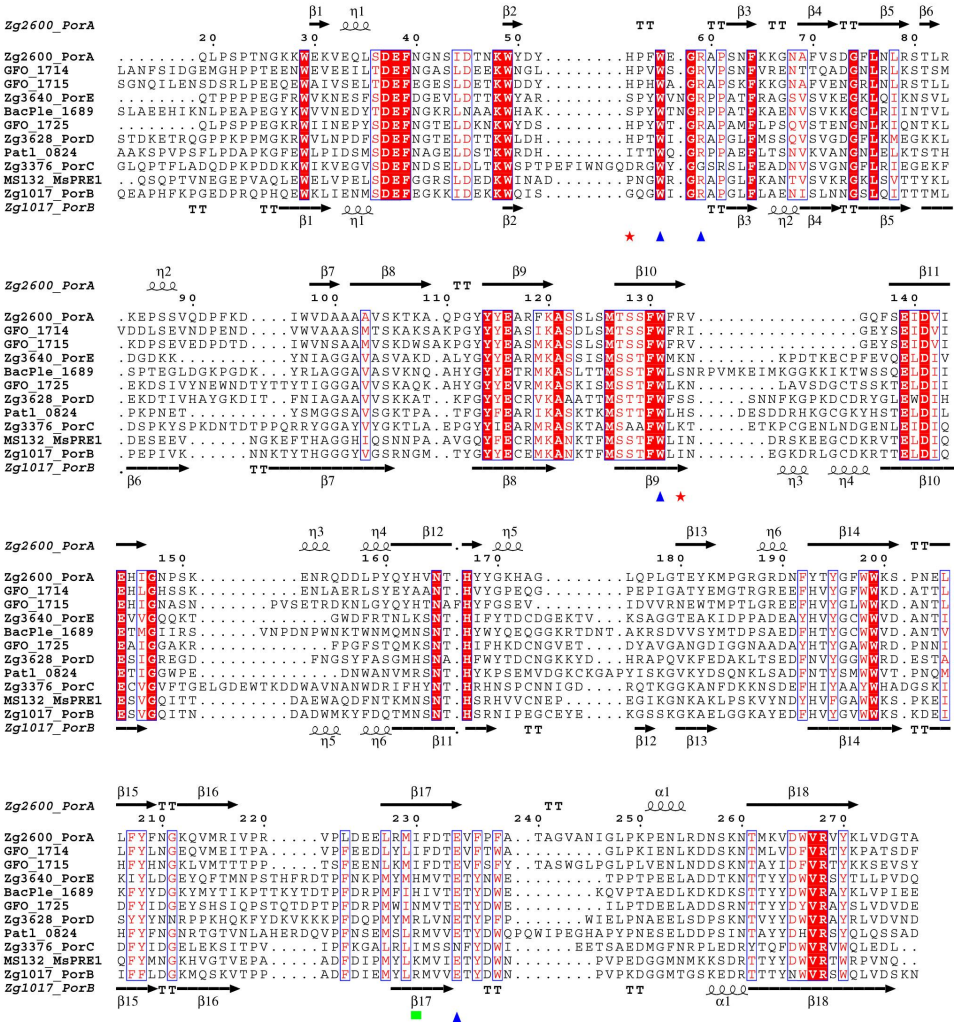


Figure 47: Structural alignment of GH16 proteins identified as β-porphyranases that are found in marine bacteria and in the Japanese gut bacterium *Bacteroides plebeius* (BacPle_1689). The sequences are compared with the secondary structures of PorA (above) and PorB (below). Alpha helices and beta strands are represented as helices and arrows, respectively, and beta turns are marked with TT. The blue triangles marked the residues involved in the recognition of the D-galactose in the subsite -1 of PorA and PorB. The residues involved in the recognition of the L-galactose 6-sulfate in subsite -2 of PorA and PorB are marked by red stars and green squares, respectively. This sequence alignment was prepared using ESPript (Gouet et al. 2003) with the sequence listed in the Supplementary Table 1.

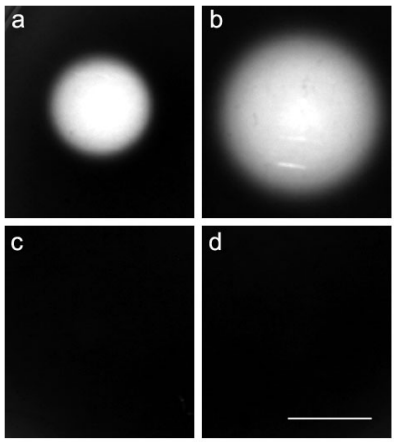


Figure 48: β-Porphyranases are not active on agarose gel compared to β-agarases. (a) 1 µg of AgaA, **(b)** 1 µg of AgaB, **(c)** 1 µg of PorA, (d) 1 µg of PorB were placed on an agarose plate and incubated overnight at 30 ºC. Activity was detected with Lugol and is shown by a bright halo. Inset scale length = 1cm.

As predicted, these enzymes had no activity on commercial agarose (Figure 48) or κ-carrageenan. Consequently, we screened their activity against polysaccharides extracted from various marine macrophytes. Zg2600 and Zg1017 were found to be inactive on extracts from marine angiosperms, brown algae or carrageenophytic red algae. Both enzymes were active on extracts from the agarophytic red algae *Gelidium*, *Gracilaria* and *Porphyra* (Figure 49a). These enzymatic activities correlate with increasing amounts of 4-linked α-L-galactose-6-sulfate (L6S) in the extracted polysaccharides, as shown by ^{1}H-NMR (Figure 49b) (Maciel et al. 2008). This sugar is typical of porphyran, which ideally consists of alternating L6S and 3-linked β-D-galactopyranose units (G) (Turvey et al. 1961) (Figure 50). Occasionally, L6S units are replaced by 4-linked 3,6-anhydro-α-L-galactose units (LA) as found in agarose (Anderson et al. 1965) and O-methylations have been reported (Morrice et al. 1984). *Porphyra* extracts were purified from LA-G motifs by β-agarase pre-treatment since natural porphyran is a heterogeneous polysaccharide with ~ 2/3 L6S-G and ~1/3 LA-G motifs (Zhang et al. 2005) (Figure 49). Zg2600 and Zg1017 showed highest activity on this pure porphyran which contains > 95 % L6S-G motifs, while the agarases were inactive (Figure 50). Therefore Zg2600 and Zg1017 are the first porphyranases described to date, referred to as PorA and PorB, respectively and they represent a new class of glycoside hydrolases.

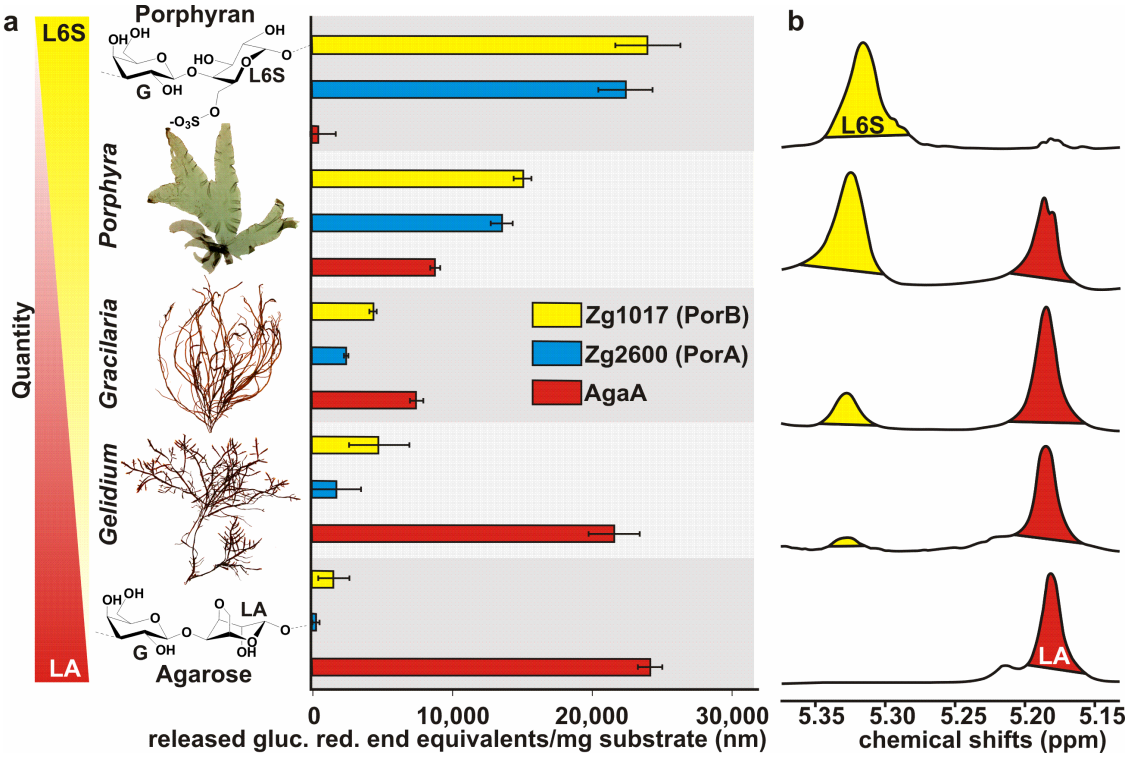


Figure 49: Activity screening on natural algal polysaccharides reveals porphyranase activity. (a) Recombinant enzymes were screened on crude algal extracts and the species on which the enzymes were active are displayed. The enzyme activity is shown in comparison to pure porphyran and agarose. **(b)** ^{1}H-NMR spectra from the substrates corresponding to the left panel show that enzymatic activity correlates with quantity of L6S. By peak integration of the L6S and LA anomeric protons the ratio between L6S-G and LA-G units in the alternating polysaccharide can be calculated. This shows that enzyme activity is highest on pure porphyran with >95% L6S-G units.

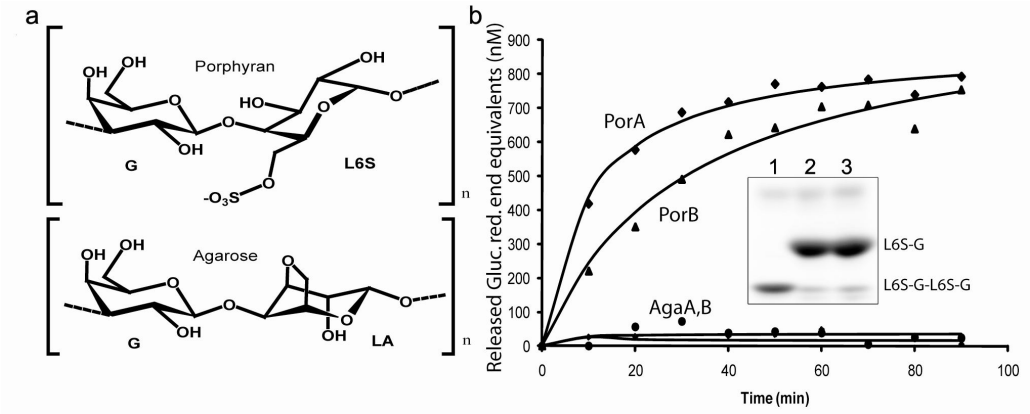


Figure 50: Porphyran is degraded by porphyranases. (a) Porphyran polysaccharide primary structure in comparison with agarose. **(b)** Activity of the β-porphyranases PorA and PorB (~5 nM) on pure porphyran (0,125% w/v) compared with the β-agarases AgaA and AgaB. Inset: The neo-porphyrotetraose is degraded to disaccharide by both porphyranases as shown with fluorophore labelled reaction products analyzed by polyacrylamide-gel electrophoresis. Lane 1 shows the control L6S-G-L6S-G. Lane2 shows that L6S-G-L6S-G is degraded to L6S-G by PorA. Lane 3 shows that L6S-G-L6S-G is also degraded to L6S-G by PorB.

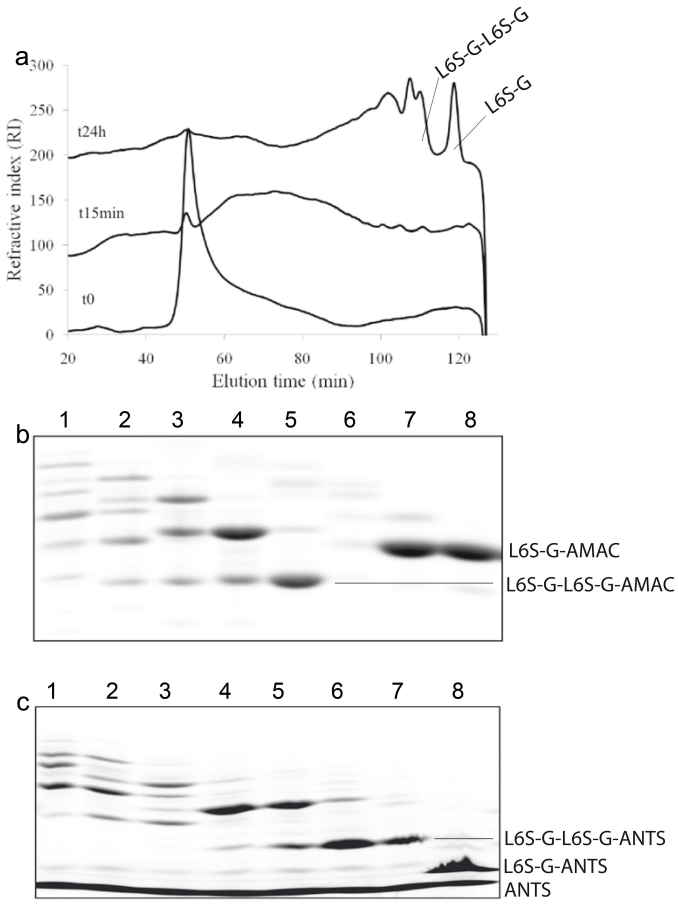


Figure 51: Phorphyran hydrolysis by PorA. (a) Natural porphyran was degraded with PorA. During the time course of degradation aliquots were analyzed by analytical size exclusion chromatography (SEC). PorA produces predominantly L6S-G corresponding to the peak at the end of the elution profile (~120 min). PorB similarly produces L6S-G as major final product (data not shown). After complete degradation of the polysaccharide the released oligosaccharides were purified by preparative SEC that resulted in a chromatogram reminiscent of the one shown here. **(b)** The purified oligosaccharides (preparative SEC, chromatogram not shown) were analyzed by fluorophore assisted polyacrylamide-gel electrophoresis with 2-amnoacridone (AMAC) as a label. Lanes 1-4 correspond to uncharacterized oligosaccharides with degree of polymerization over 4 (>DP4). Lane 5 shows the DP4 of structure L6S-G-L6S-G as revealed by ^{1}H-NMR (see below) and by co crystallisation. Lanes 7-8 show the major product of both porphyranases L6S-G which corresponds to the final peak in the SEC chromatogram. Note that the lower velocity of L6S-G compared to the DP4 L6S-G-L6S-G oligosaccharide is due to a turning point. **(c)** This effect is abolished when ANTS is used as label. In this case the L6S-G disaccharide shown in Lane 8 runs faster as the L6S-G-L6S-G tetrasaccharide on lane 6-7. This is expected due to its smaller mass.

The ^{1}H-NMR analysis of the purified reaction products shows that the major final product of both porphyranases is the disaccharide L6S-G. Additionally, minor amounts of the tetrasaccharide L6S-G-L6S-G and unidentified oligosaccharides with higher mass are produced (Figure 51, Figure 52). The structure of the tetrasaccharide was further verified by co-crystallization with an active site mutant of PorA (see below). Both PorA and PorB hydrolyzed tetrasaccharide L6S-G-L6S-G to the disaccharide L6S-G (Figure 50b inset). Therefore, these enzymes cleave the β-1,4-glycosidic bond between G and L6S units and are defined as β-porphyranases by analogy with the distinction between α- and β-agarases (Flament et al. 2007). Altogether, *Zobellia galactanivorans* contains a complete system for porphyran degradation comprising both β-porphyranases and β-agarases due to the heterogeneous character of the natural polysaccharide.

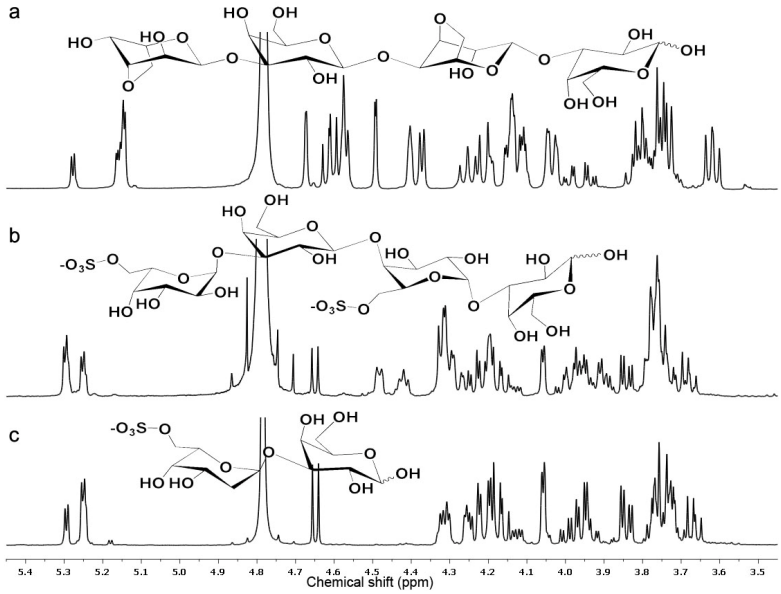


Figure 52: ^{1}H-NMR analysis of oligosaccharides produced by porphyranases in comparison to agarose tetrasaccharide. (a) neo-agarotetraose (LA-G-LA-G); **(b)** neo-porphyrotetraose (L6S-G-L6S-G); **(c)** neo-porphyrodiose (L6S-G). For all spectra the α-anomeric configuration of D-galactose (reducing end) was the most downfield shifted signal at 5.3 ppm (J_{H1H2}= 3.8 Hz), whereas the signal for the β-anomeric configuration was at 4.64

ppm (J_{H1H2}= 7.7 Hz). The anomeric proton signals of L-galactose units allowed unambiguous assignment of the shifts from the 3,6-anhydro-α-L-galactose (LA) of the agarose motive at 5.17 ppm and the α-L-galactopyranose-6-sulfate (L6S) of the porphyran motive at 5.25 ppm. **(b)** The neo-porphyrotetraose spectrum showed 2 additional signals at 4.49 and 4.41 ppm, compared to the neo-porphyrodiose **(c)**, that were respectively ascribed to the internal H1 and H3 of the G residue. **(a)** Logically these signals of an internal D-galactose unit were present in the neo-agarotetraose spectrum as well. For both porphyran oligosaccharides the coupling constant of L6S-H1 was 3.9 Hz which is typical of the α-anomeric configuration. Moreover the HMBC experiments highlighted a correlation between L6S-C1 and Gr-H3 suggesting that the two residues were linked by α (1→3) linkage. All together, these results were consistent with a L6S-G disaccharide structure obtained by the cleavage of the β (1→4) glycosidic bond of porphyran.

IV.4.3 Structural determinants of porphyran active enzymes

The crystal structures of PorA and PorB explain the unambiguous substrate specificity and allow identifying key residues for porphyran recognition in β-porphyranases (For data statistics see Supplementary Table 1). Although quite divergent (27% of sequence identity), both β-porphyranases adopt the family GH16 typical jelly-roll fold and possess the catalytic residues E-139 (mutated to S-139 to make tthe inactive mutant) and E-144 in PorA and E-156 and E-161 in PorB. The structure of the inactive mutant PorA_E139S in complex with L6S-G-L6S-G shows the reducing end bound to the subsites (Davies et al. 1997) -1 and -2 and two sugar units extending beyond the substrate binding cleft (Figure 53). At the cleavage subsite -1, the G unit forms a stacking interaction with W-131 (W-139, PorB) and hydrogen bonds to residues R-59 (R-70, PorB), D-141 (D-158, PorB) and E-234 (E-256, PorB). The presence of G units is conserved in the primary structure of agarose and porphyran and these sugars are indeed almost superimposable at subsite -1 in PorA and the agarase AgaA (Allouch et al. 2004). Thus structural differences for substrate recognition predominantly concern the subsite -2, which binds a L6S unit in PorA and a LA unit in AgaA (Allouch et al. 2004). The L6S unit is surrounded by W-56 (W-67, PorB) on one side, R-133 on the opposite side and by R-59 (R-70, PorB) situated above. While W-56 stacks to one face of the L6S unit, R-59 forms a bidentate hydrogen bond with the L6S-O2 and the α-(1,3)-glycosidic bond oxygen. The sulfate group of the L6S unit is oriented

towards the interior of the cleft and protrudes into a positively charged and hydrophobic pocket (Figure 53b). The negative charge of the sulfate group is neutralized with salt bridges from the side chains of H-53 and R-133. Critical for porphyran recognition, this pocket leaves space for the bulky sulfate group at C6. In contrast, non-adapted glycoside hydrolases, such as β-agarases, impose high sterical constraints at this position illustrated by L6S-G superposed on LA-G in the AgaA structure (Figure 53c). Therefore, productive binding is impossible as reflected by the biochemical activity of AgaA, which cannot degrade pure porphyran. The superposition illustrates how the 3,6-anhydro-bond locks the LA pyranose ring in the energetically unfavoured $^{1}C_{4}$ conformation that switches to $^{4}C_{1}$ by sulfation on C6 (L6S), explaining hydrolysis of the different backbone structures of agarose and porphyran.

In PorB the sulfate recognition of L6S at subsite -2 differs with respect to PorA, since H-53 and R-133 are not conserved. The H-53 of PorA is substituted by G-64 both sequentially and spatially by the side chain of Tyr-111, however, this tyrosine residue does not interact with the sulfate group of L6S. The role of H-53 is carried out in PorB by R-252 that is at the correct distance to form a salt bridge with the sulfate of L6S. This interpretation is supported by a bound morpholino-ethane-sulfonic-acid molecule (MES-buffer) in the subsite -2 of PorB, mimicking the L6S residue. The sulfonic group of the MES molecule is oriented towards the interior of the cleft and interacts with R-252 (Figure 53d). This shows that also in PorB the basically charged pocket is present to harbour the L6S sulfate group as described for PorA. In conclusion, the crystal structures of PorA and PorB explain the unambiguous specificity of porphyranases for the sulfated porphyran from *Porphyra* spp.. Furthermore, the crystal structures of two divergent porphyranases allowed pinpointing regions of variation and definition of the key residues involved in porphyran recognition. This knowledge was exploited in the phylogeny analysis of environmental porphyranase-like genes.

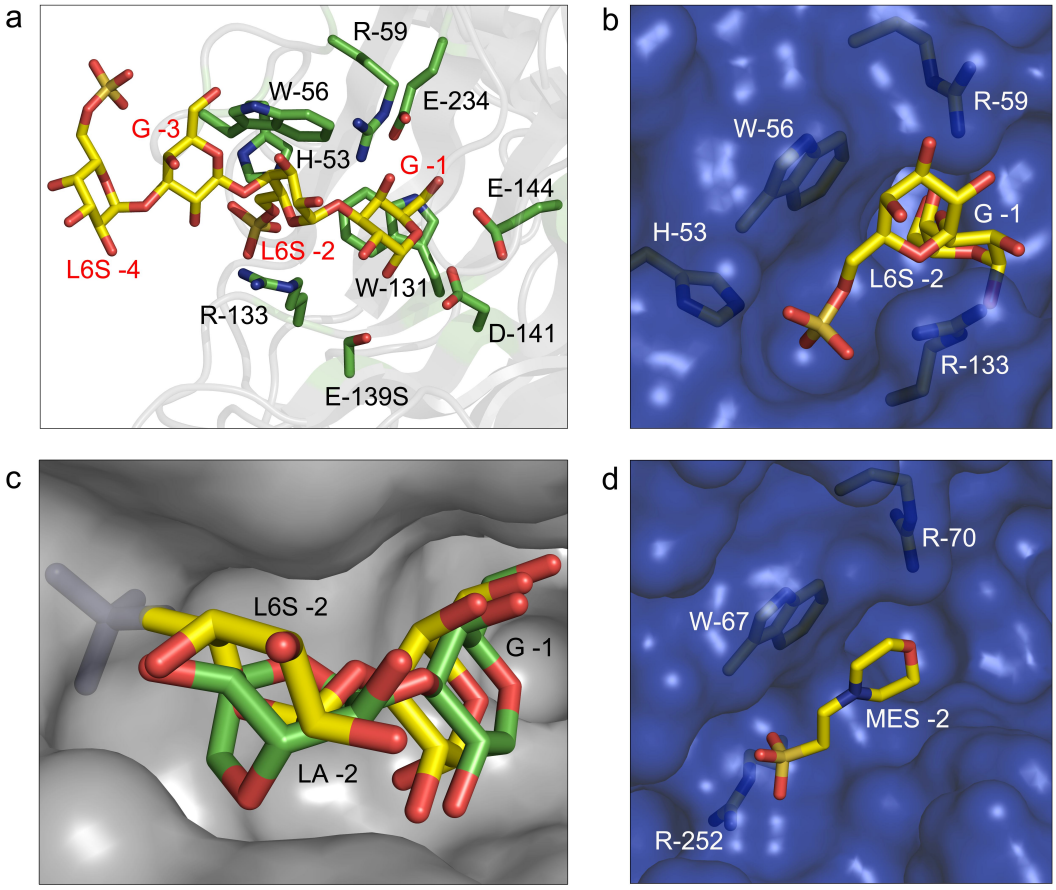


Figure 53: Crystal structures of native PorB and the PorA active site mutant in complex with neo-porphyrotetraose shows substrate specificity of porphyranases. (a) An active site mutant of PorA was generated by PCR mutagenesis and cocrystallized with the purified porphyran tetrasaccharide showing the key residues for porphyran recognition. **(b)** L6S unit bound to subsite -2 of PorA. This hydrophobic and basic pocket explains the strict specificity of PorA for porphyran. **(c)** The subsite -2 pocket of PorA is absent in agarase (AgaA) where the sulfate group of the modelled porphyran repeating unit leads to a sterical clash with the enzyme structure. **(d)** A bound MES molecule mimics the natural L6S at subsite -2 in PorB, indicating an orientation of porphyran similar to that in PorA and shows that the sulfonic group is stabilized by residue R-252.

IV.4.4 β-Porphyranases are abundant in marine bacteria

Sequence homology search against Genbank identified several potential β-porphyranases, mostly annotated as β-agarases or family GH16 glycoside hydrolases. Phylogenetic analysis reveals a coherent monophyletic clade confined by PorA and by PorB together with nine proteins (including Zg3376, Zg3628 and Zg3640) (Figure 54, Clade 2) (for details see Supplementary Table 2). The residues constituting subsite -1 (R-59, W-131, D-141 and E-234, PorA numbering) are well conserved in all these proteins. Regarding the subsite -2, the sequences divide into three groups. The proteins GFO1714 and GFO1715 from the marine *Bacteroidetes Gramella forsetii* (Bauer et al. 2006) are proteins similar to PorA with conserved H-53 and R-133 (~55% sequence identity with PorA). The protein MS132 from *Microscilla sp*. PRE1, another agar-degrading marine *Bacteroidetes* (Zhong et al. 2001), shares 54% sequence identity with PorB and features a conserved basic residue at positions 252 (residue 231 in PorA). The remaining sequences have in average 35% sequence identity with either PorA or PorB and display hybrid subsite -2 organization with mostly basic or polar residues at position 53, 133 and 231. All these additional proteins are thus predicted as β-porphyranases and clade 2 can be defined as a new subfamily of glycoside hydrolases within family GH16. Therefore the *Z. galactanivorans* genes Zg3376, Zg3628, Zg3640 are named *porC*, *porD* and *porE* respectively.

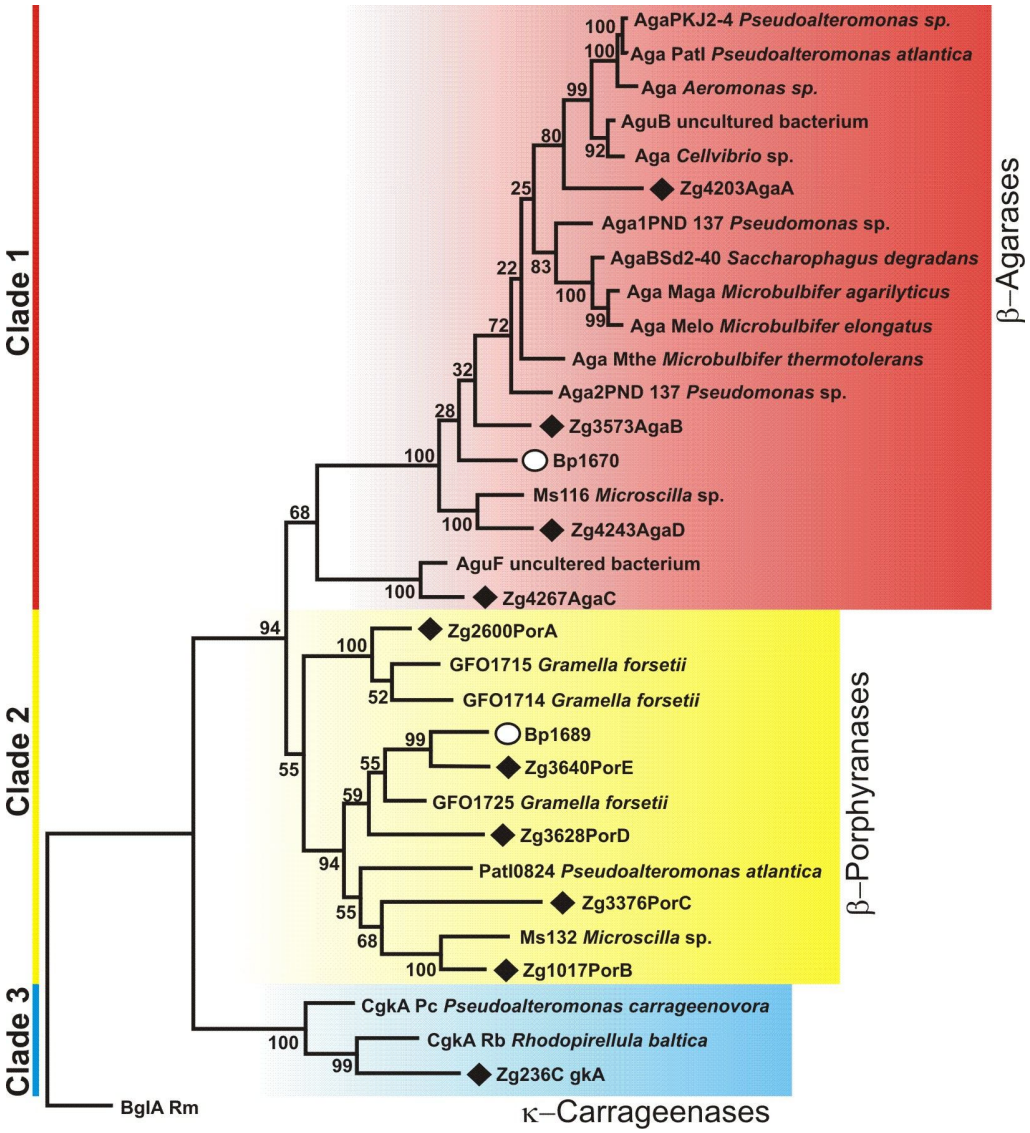


Figure 54: Phylogenetic analysis of family GH16 galactanases reveals porphyranases in many marine bacteria and in the Japanese gut bacterium *Bacteroides plebeius*. Family GH16 enzymes specific for red algal sulfated galactans cluster in three clades: **(Clade 1)** β-agarases, **(Clade2)** β-porphyranases and **(Clade3)** β-carrageenases. Marine bacteria use these enzymes for red algal galactan degradation. The non-marine exception is the Japanese gut bacterium *B. plebeius* which contains a β-porphyranase and a β-agarase (white circles) reminiscent of the porphyranolytic system of *Z. galactanivorans* (black diamond's). This phylogenetic tree was prepared with the sequences listed in the Supplementary Table 2.

IV.4.5 Horizontal gene transfer from marine to human gut bacteria

All β-porphyranases are found in marine bacteria, except Bp1689 which surprisingly originates from the Japanese gut bacterium *Bacteroides plebeius* (Kitahara et al. 2005). This species, sequenced as part of the Human Gut Microbiome Initiative (http://genome.wustl.edu/genomes), also contains a family GH16 β-agarase, Bp1670 (Figure 54, Clade1). The presence of both a β-porphyranase and a β-agarase gene in *B. plebeius* is reminiscent of the porphyran degradation system of *Z. galactanivorans*. In contrast, none of the 22 other *Bacteroides* genome sequences available at NCBI contain any β-porphyranase or β-agarase genes. It is noteworthy that these *Bacteroides* species were isolated essentially from individuals with a western diet. Analysis of the *B. plebeius* genome shows that Bp1670 and Bp1689 are surrounded by other CAZyme genes (Figure 55): two additional β-agarases from family GH86 (Bp1693 and Bp1694), two family GH2 β-galactosidases, which could release galactose from agar/porphyran degradation products (Bp1672 and Bp1673), and a sulfatase potentially involved in the desulfation of porphyran oligosaccharides (Bp1701). This genomic region also contains a gene coding a family 22 carbohydrate binding module (CBM22, Bp1696), preceded by a SusD-like gene (Bp1697) and its associated TonB-dependent receptor gene (TBDR, SusC-like, Bp1698). The closest homologues of all these carbohydrate-related genes occur in *Z. galactanivorans* and in *Microscilla sp.* PRE1 and not in other gut *Bacteroides*. Six of these genes (Bp1670, Bp1671, Bp1689, Bp1693, Bp1694 and Bp1696) are conserved with marine bacteria only. Moreover, Bp1696, Bp1697, and Bp1698 share synteny with two polysaccharide utilization loci (PUL) from *Z. galactanivorans*, which include either *agaB* (Zg3573) or *porE* (Zg3640). Interspersed between all these genes shared with marine bacteria, are genes well conserved with other gut *Bacteroides* (Supplementary Table 3) showing that this bacterium is a regular gut symbiont that received an unusual set of genes. The possible mechanism for HGT could be identified by analyzing the region downstream of the β-agarase Bp1670, which contains genes coding for conserved relaxase/mobilization proteins (Bp1662 and Bp1663/Bp1665), required for initiation and termination of conjugative DNA transfer (Garcillan-Barcia et al. 2009). While, Bp1663 and

Bp1665 are pseudogenes, corresponding to a relaxase gene interrupted by a putative reverse transcriptase (Bp1664), Bp1662 is an intact gene. This suggest that this HGT occurred by a conjugative mechanism. Interestingly, the β-agarase and β-porphyranase genes (Figure 54: Clade1-MS116, Clade2-MS132 and Figure 55: MS110) of *Microscilla sp*. PRE1 are carried by a plasmid, that contains a relaxase gene (MS155) homologous to Bp1662 (Zhong et al. 2001). In conclusion, all these results support that *B. plebeius* acquired a porphyran degradation system that originates from a porphyranolytic marine *Bacteroidetes*.

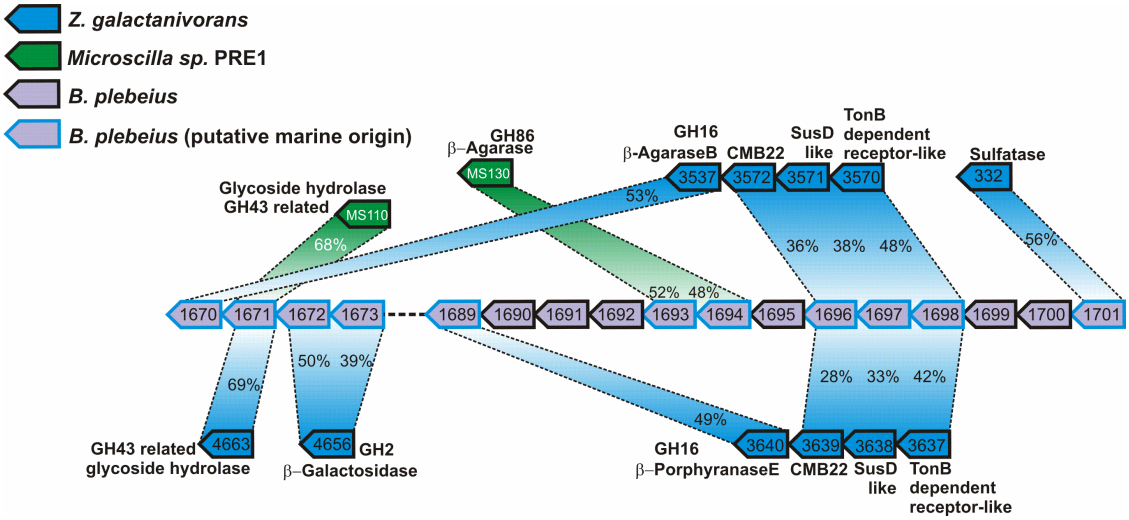


Figure 55: Upstream and downstream of the porphyranase gene (Bp1689) the genome of *Bacteroides plebeius* contains carbohydrate-related genes which share highest identity with proteins used for red algal galactan degradation in two marine *Bacteroides*. Shown is the sequence identity between *B. plebeius*, *Microscilla sp*. PRE1 and *Zobellia galactanivorans* proteins. Six of these genes (Bp1670, Bp1671, Bp1689, Bp1693, Bp1694 and Bp1696) are conserved with marine bacteria but are absent in genomes of other gut *Bacteroides*.

IV.4.6 Discussion

To reveal unknown marine CAZyme functions by screening against defined substrates is technically limited by the complexity of carbohydrates (Laine 1994) and by the lack of commercial algal substrates. Here this problem was circumvented by the use of natural substrates leading to the discovery of a new porphyranase activity. Such a “reverse biochemistry” approach may be widely applicable to new CAZymes, becoming available through structural genomic methods. A significant number of marine microbes are *Bacteroidetes* and their glycoside hydrolases are mechanistic key players in polysaccharide degradation (Elifantz et al. 2008) influencing the carbon cycle (Kirchman 2002). The extreme microbial biomass turnover in the oceanic ecosystem led to the analogy of the "ocean interior as a global heterotrophic digester" (Hedges et al. 2001). Following this concept, our discovery of the first porphyranases allowed detecting genetic exchange between two digestion systems, illustrating the impact of environmental diversity of microbes on humans. Recent gut metagenome data showed high redundancy of metabolic functions in the gut microbiome, which is shared between human individuals of one population (Turnbaugh et al. 2009). Therefore the occurrence of a porphyran degradation system in one *Bacteroides* species from Japanese and the absence in 22 *Bacteroides* species with western origin indicates a lifestyle specific event. The detection of this HGT was possible due to two favourable factors: (i) β-porphyranase genes are rare, if not absent, in terrestrial organisms; (ii) the transfer was recent compared to millions of years of mammalian gut microbiome evolution. This is supported by the high sequence identity of the porphyranase and associated genes in *B. plebeius* with genes found in marine *Bacteroides*. In contrast, acquisition of terrestrial plant-specific CAZyme genes likely occurred early in evolution of herbivorous/omnivorous mammalian, explaining our difficulties to describe HGT today. Thus the question arises when the here described genes were acquired by Japanese gut bacteria. One cannot exclude their occurrence in continental ancestors from East Asia before peopling of the Japanese archipelago occurred. Nevertheless seaweeds play an important role in the Japanese culture for over 1000 years, since tax records from the eight century listed seaweeds as payments to the Japanese government (Nisizawa et al. 1987). Therefore nutrition is the most probable vector for contact with marine microbes that led to the HGT and indeed the only porphyran

source in human nutrition is nori. Traditionally nori is not roasted and thus contact with associated marine microbes is promoted in Japanese sushi (Nisizawa et al. 1987). Consequently the consumption of food with associated environmental bacteria, called here the "Sushi Factor", is the most likely mechanism which promoted this CAZyme update into human gut microbes.

IV.4.7 Methods summary

Porphyra umbilicalis algae were collected in spring 2008 near Roscoff in the intertidal zone. One hundred g of dry weight material was used for the polysaccharide extraction (Ishihara et al. 2005). This polysaccharide is referred to as porphyran. To remove a maximum of LA-G motifs, porphyran was pre-digested with excess of β-agarase B (AgaB) (Allouch et al. 2003). The resistant fraction was separated from oligosaccharides by gel-filtration. This fraction is referred to as pure porphyran.

Enzyme activity assays were carried out as previously described for β-agarases (Jam et al. 2005) but at 30ºC with a porphyranase concentration of ~5 nM and a porphyran solution of 0.125 % (w/v). The oligosaccharides produced during enzymatic digestion of porphyran, pure porphyran, agarose and polysaccharides from various algal species were measured by reducing sugar assay, calibrated against a glucose concentration curve and presented as glucose reducing end equivalents (Kidby et al. 1973).

Recombinant PorA and PorB were obtained following a medium throughput cloning and expression strategy (see Methods). All enzyme constructs are appended by a His_6-tag that is exploited for protein purification following standard affinity- and size-exclusion HPLC methods.

The crystal structure of PorA was solved by MAD on a gold derivative using the programs SHELX (Schneider et al. 2002) and HKL2MAP (Pape et al. 2004). A standard mutagenesis kit from Qiagen was used to introduce the E139S mutation into PorA (for mutagenesis primers Supplementary Table 4). This inactive mutant was used to obtain the complex structure by co-crystallization with purified neo-porphyrotetraose. Refinement was performed with REFMAC

(Murshudov et al. 1997) and model building with Coot (Potterton et al. 2004). The crystal structure of PorB was solved by Phaser (CCP4) (Potterton et al. 2003) and using PorA as search model. All data sets were collected at the European Synchrotron Radiation Facilities, Grenoble, France. For further details see Methods.

The purification of porphyran oligosaccharides was carried out as described in the French patent No. 0955070 (Hehemann *et al*. 2009). ^{1}H-NMR spectra were recorded and the signals were fully assigned using a complete set of correlation spectra (COSY, HMBC and HMQC). Details of the phylogenetic analysis are described in the Methods section.

IV.4.8 Methods

Activity screening on algal extracts

Algae of species (*Ectocarpus silliculosus*, *Delesseria sanguinea*, *Zostera marina*, *Porphyra umbilicalis*, *Gracilaria gracilis*, *Gelidium spinosum*, *Chondrus crispus* and *Laurencia pinnatifida*) were all collected at low tide near Roscoff, extensively washed with freshwater, shock frozen with liquid nitrogen and ground in a mortar until a fine powder was obtained. The algal powder was treated with 80 % ethanol, centrifuged and the supernatant was discarded. Ethanol extraction was repeated until the supernatant remained colourless. The obtained algal solid fraction was pelleted, frozen and subsequently dried by lyophyllisation. For the enzymatic activity tests, 50 mg of algal powder was resuspended in 1 ml distilled water and extracted for 1 hour at 98°C. The suspension was centrifuged for 5 min at 3000 g and the supernatant was transferred into a new tube. This extraction was repeated twice and the three extraction supernatants were finally combined. The dried mass was recovered for each extract to calculate the normalized enzyme activity and to record ^{1}H-NMR-spectra.

One hundred µl of algal extract was incubated with ~1 µg of enzyme at 30°C in 300 mM NaCl and 10 mM sodium acetate, 16 mM tri-sodium citrate (pH 7.2) until complete digestion could be obtained. Complete digestion was monitored by further adding of enzyme and re-measuring the

amount of released oligosaccharides. The samples were centrifuged for 30 min at 14000 rpm and liberated oligosaccharides were quantified by reducing sugar assay (Kidby et al. 1973). A calibration curve with 50 – 300 µg/ml of glucose was used to calculate the amount of reducing ends as glucose reducing end equivalents.

Medium throughput cloning and expression strategy

A total number of 11 family GH16 enzymes from *Z. galactanivorans* were selected for the medium cloning and expression strategy. Several of the 11 genes are multi-modular and consist of catalytic modules appended by carbohydrate binding modules (CBMs) or modules of unknown function (UNK). The multimodular enzymes were cloned as several constructs, full-length and catalytic modules or CBMs only. This led to a total number of 24 different constructs that were cloned and expressed in a parallel strategy. Cloning strategy and primers are given in Supplementary Table 4.

All gene-targets were amplified in parallel by PCR with a set of forward and reverse primers that were calculated to have a melting temperature of 70°C. Purified PCR products were digested with a mix of the corresponding restriction enzyme/buffer system (NEB) for three hours at 37°C. After digestion, the PCR products were again purified with a standard DNA purification Kit and finally eluted with 25 µl of H_2O. The PCR products were ligated with an equally digested pet derived plasmid which was additionally dephosphorylated. For the transformation chemically competent DH5α *E. coli* cells and standard procedures were followed.

The recombinant soluble expression of the different constructs was first screened in a medium throughput manner. To this end, the 22 obtained clones were transformed into recombinant Bl21 (DE3) cells and then used to inoculate a 3 ml LB-medium preculture, supplemented with 100 µg/ml of ampicillin in deep well format. The preculture was incubated overnight at 37°C, 150 rpm and 10 µl were transferred for inoculation of 2 ml ZYP-5052 medium (ampicillin, 100 µg/ml) (Studier 2005). The expression test cultures were incubated at 20°C for three days and harvested by centrifugation. The cells were lysed with 500 µl of lysis buffer composed of 50 mM

Tris pH 8, 300 mM NaCl, 1 mg/ml lysozyme and DNAse. The cell lysate was centrifuged and the supernatant separated for batch affinity purification with HyperCell nickel resin. The remaining cell pellet which contained the insoluble expressed fraction was extracted with 6 M urea for 30 minutes and centrifuged. The supernatant was stored for SDS-PAGE analysis. The cell lysate supernatant was used for batch affinity purification and 50 µl of nickel charged HyperCell resin was added to the supernatant. After incubation for 20 min, the supernatant was discarded and the pellet washed with 500 µl buffer (lysis buffer without lysozyme and DNAse) followed by further centrifugation. The bound proteins were eluted adding 500 mM Imidazole. Fractions of inclusion bodies, soluble cell lysate and batch purified proteins were analysed on 12% SDS-PAGE. The expression procedures were then up scaled to 1 L cultures for protein production prior to structural and biochemical analysis.

Crystallization and Structure Solution

PorA was crystallized by the hanging drop method in 18-20% PEG 4000, 0.2 M ammonium sulfate and 0.1 M sodium acetate pH 4.6 at 20 °C. Drops contained 2 µl of 2.6 mg/ml protein solution plus 1 µl of crystallization solution and were equilibrated against 500 µl crystallization solution. A heavy atom derivative was obtained by soaking a native crystal for 24 hours in mother liquor supplemented with 5 mM potassium tetra-chloro-aureate (III) and 5% glycerol. Before flash freezing at 100 K the crystal was back-soaked in mother liquor with 10 % Glycerol for ~1 minute. MAD data were collected on beamline ID23I. Four datasets were collected from one heavy atom derivative crystal, around the L-III gold absorption edge. Anomalous data were processed with XDS and scaled with XSCALE (Kabsch 1988). The heavy atom substructure solution was determined with SHELXD (Schneider et al. 2002) followed by density modification with SHELXE as part of HKL2MAP (Pape et al. 2004) and the resulting electron density map was improved by solvent flattening with the program DM (CCP4) (Potterton et al. 2003). The crystallization conditions for the PorA_E139S complex structure were adapted from the native protein and 5-10 mM of neo-porphyrotetraose was added. For these crystals the data collection was carried out on beamline BM30A. Data were processed with Mosflm (Leslie et al. 2007) and

SCALA (CCP4) (Potterton et al. 2003).

PorB crystallized within one to two weeks by the sitting drop method with a crystallization solution composed of 33-35 % PEG 1000, 0.4 M lithium sulfate and 0.1 M MES buffer at pH 6.0. The sitting drops were formed by mixing 2µl of protein solution (5.6 mg/ml) and 1 µl of crystallization solution, and equilibrated against 500µl crystallization solution at RT. For cryo-conservation crystals were brought to 10 % glycerol in 5 % increments in the crystallization solution. After solving the phases by molecular replacement, the model building and refinement were carried out as described for PorA. For all datasets, five percent of the observations were flagged as free and used to monitor refinement procedures and a summary for data collection and final refinement statistics are given in Supplementary Table 1.

Purification of the oligosaccharides by preparative size exclusion chromatography

The material for oligosaccharide purification was obtained by incubation at 30ºC of porphyran with 2.6 µg/ml of PorA in water.

The purification of porphyran oligosaccharides was carried out by preparative size exclusion chromatography using three Superdex 30 (26/60) columns in series integrated on a HPLC system liquid injector/collector (Gilson). The freeze-dried hydrolysis product was dissolved in de-ionized water at a concentration of 4 % (w/v). After filtration 4 mL of sample was injected then eluted with 50mM ammonium carbonate at a flow rate of 1.5 mL/min. Fractions of oligosaccharides which could be assigned to distinct peaks were collected, pooled and analyzed by HPLC (data not shown) and fluorophore assisted polyacrylamide gel electrophoresis and the fractions were freeze-dried by lyophilization for further analyses by nuclear magnetic resonance.

Fluorophore-assisted carbohydrate electrophoresis

Fractions obtained after separation of oligosaccharides were analysed by Fluorophore-carbohydrate-PAGE (Starr et al. 1996). Aliquots of 100 µl were labelled with 2 µl AMAC (2-

amnoacridone) solution or with 2 µl ANTS (8-aminonaphthalene-1,3,6-trisulfonate) solution. After dissolution of the oligosaccharide pellets, 5 µl of a 1 M sodium cyanoborohydride in dimethylsulfoxid (DMSO) was added and the mixture was incubated for 16 hours at 37° C. The samples were analysed on a 30 % polyacrylamide running gel with a 4 % stacking gel.

Nuclear magnetic resonance spectroscopy

All samples were solubilised in D_2O and exchanged twice, before being analysed on a 500 MHz Bruker NMR spectrometer. For the polysaccharides extracted from algae that were used as substrates for enzymatic assays ^{1}H-NMR spectra were recorded at 70°C using 64 scans. For oligosaccharides ^{1}H-NMR spectra were recorded at 25°C using 16 scans. Chemical shifts are expressed in ppm by reference to an external standard TSP (trimethylsilylpropionic acid). The NMR signals of the purified oligosaccharides were fully assigned using a complete set of correlation spectra: COSY (doublequantum-filtered correlation spectroscopy), HMBC (heteronuclear multiple bond correlation) and HMQC (heteronuclear single quantum correlation).

Sequence and phylogenetic analyses

Multiple sequence alignments were generated using MAFFT with the iterative refinement method and the scoring matrix Blosum62 (Katoh et al. 2002). The structure-based multiple alignment of the β-porphyranases was displayed using the program ESPript (Gouet et al. 2003). For the phylogenetic analyses of the family GH16 enzymes, the sequences were selected from the CAZY database. The MAFFT alignment of these sequences was manually refined using Bioedit (© Tom Hall), on the basis of the superposition of the structure of the κ-carrageenase of *Pseudoalteromonas carrageenovora* (Michel et al. 2001), the β-agarase AgaA (Allouch et al. 2003) and the β-porphyranases PorA and PorB from *Z. galactanivorans*. Phylogenetic trees were derived from this refined alignment using the Maximum Likelihood method with the program PhyML (Guindon et al. 2003). The reliability of the trees was always tested by bootstrap analysis

using 100 resamplings of the dataset. The trees were displayed with MEGA 3.1 (Kumar 2004).

IV.4.9 References

J. Xu, M. K. Bjursell, J. Himrod et al., A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis, *Science* **299**, 2074 (2003).

J. L. Sonnenburg, J. Xu, D. D. Leip et al., Glycan foraging in vivo by an intestine-adapted bacterial symbiont, *Science* **307**, 1955 (2005).

B. L. Cantarel, P. M. Coutinho, C. Rancurel et al., The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics, *Nucleic Acids Res.* **37**, D233 (2009).

N. M. Koropatkin, E. C. Martens, J. I. Gordon et al., Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices, *Structure* **16**, 1105 (2008).

C. A. Lozupone, M. Hamady, B. L. Cantarel et al., The convergence of carbohydrate active gene repertoires in human gut microbes, *Proceedings of the National Academy of Sciences of the United States of America* **105**, 15076 (2008).

Y. Nakamura, T. Itoh, H. Matsuda et al., Biased biological functions of horizontally transferred genes in prokaryotic genomes, *Nature Genetics* **36**, 760 (2004).

J. Xu, M. A. Mahowald, R. E. Ley et al., Evolution of symbiotic bacteria in the distal human intestine, *Plos Biology* **5**, 1574 (2007).

R. E. Ley, C. A. Lozupone, M. Hamady et al., Worlds within worlds: evolution of the vertebrate gut microbiota, *Nat. Rev. Microbiol.* **6**, 776 (2008).

S. Fukuda, H. Saito, S. Nakaji et al., Pattern of dietary fiber intake among the Japanese general population, *European Journal of Clinical Nutrition* **61**, 99 (2007).

D. J. Mc Hugh, in *FAO Fisheries Technical Paper No 441*, edited by FAO (FAO, Rome, Italy, 2003).

J. Craigie, in *Biology of the red algae*, edited by K. Cole and R. Sheath (Cambridge university

Press, Cambridge, 1990), pp. 221.

B. Kloareg and R. Quatrano, Structure of the cell walls of marine algae and ecophysiological functions of the matrix polysaccharides., *Oceanogr. Mar. Biol. Ann. Rev.* **26**, 259 (1988).

G. Michel, P. Nyval-Collen, T. Barbeyron et al., Bioconversion of red seaweed galactans: a focus on bacterial agarases and carrageenases, *Applied Microbiology and Biotechnology* **71**, 23 (2006).

D. Flament, T. Barbeyron, M. Jam et al., Alpha-agarases define a new family of glycoside hydrolases, distinct from beta-agarase families, *Appl. Environ. Microbiol.* **73**, 4691 (2007).

J. Allouch, M. Jam, W. Helbert et al., The three-dimensional structures of two beta-agarases, *J. Biol. Chem.* **278**, 47171 (2003).

G. Michel, L. Chantalat, E. Duee et al., The kappa-carrageenase of *P. carrageenovora* features a tunnel-shaped active site: a novel insight in the evolution of Clan-B glycoside hydrolases, *Structure* **9**, 513 (2001).

G. Michel, L. Chantalat, E. Fanchon et al., The iota-carrageenase of *Alteromonas fortis*. A beta-helix fold-containing enzyme for the degradation of a highly polyanionic polysaccharide, *J. Biol. Chem.* **276**, 40202 (2001).

J. C. Venter, K. Remington, J. F. Heidelberg et al., Environmental genome shotgun sequencing of the Sargasso Sea, *Science* **304**, 66 (2004).

Y. M. Shi, G. W. Tyson, and E. F. DeLong, Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column, *Nature* **459**, 266 (2009).

T. Barbeyron, S. L'Haridon, E. Corre et al., *Zobellia galactanovorans* gen. nov., sp. nov., a marine species of *Flavobacteriaceae* isolated from a red alga, and classification of, *Int. J. Syst. Evol. Microbiol.* **51**, 985 (2001).

T. Barbeyron, A. Gerard, P. Potin et al., The kappa-carrageenase of the marine bacterium *Cytophaga drobachiensis*. Structural and phylogenetic relationships within family-16 glycoside

hydrolases, *Mol. Biol. Evol.* **15**, 528 (1998).

M. Jam, D. Flament, J. Allouch et al., The endo-beta-agarases AgaA and AgaB from the marine bacterium *Zobellia galactanivorans*: two paralogue enzymes with different molecular organizations and catalytic behaviours, *Biochem. J.* **385**, 703 (2005).

J. S. Maciel, L. S. Chaves, B. W. S. Souza et al., Structural characterization of cold extracted fraction of soluble sulfated polysaccharide from red seaweed *Gracilaria birdiae*, *Carbohydrate Polymers* **71**, 559 (2008).

J. R. Turvey and D. A. Rees, Isolation of L-Galactose-6-Sulphate from a Seaweed Polysaccharide, *Nature* **189**, 831 (1961).

N. S. Anderson and D. A. Rees, Porphyran - a Polysaccharide with a Masked Repeating Structure, *J. Chem. Soc.*, 5880 (1965).

L. M. Morrice, M. W. Mclean, W. F. Long et al., Porphyran Primary Structure, *Hydrobiologia* **116**, 572 (1984).

Q. B. Zhang, H. M. Qi, T. T. Zhao et al., Chemical characteristics of a polysaccharide from *Porphyra capensis* (Rhodophyta), *Carbohydrate Research* **340**, 2447 (2005).

G. J. Davies, K. S. Wilson, and B. Henrissat, Nomenclature for sugar-binding subsites in glycosyl hydrolases, *Biochem. J.* **321**, 557 (1997).

J. Allouch, W. Helbert, B. Henrissat et al., Parallel substrate binding sites in a beta-agarase suggest a novel mode of action on double-helical agarose, *Structure* **12**, 623 (2004).

M. Bauer, M. Kube, H. Teeling et al., Whole genome analysis of the marine *Bacteroidetes* '*Gramella forsetii*' reveals adaptations to degradation of polymeric organic matter, *Environ. Microbiol.* **8**, 2201 (2006).

Z. Zhong, A. Toukdarian, D. Helinski et al., Sequence analysis of a 101-kilobase plasmid required for agar degradation by a *Microscilla* isolate, *Appl. Environ. Microbiol.* **67**, 5771

(2001).

M. Kitahara, M. Sakamoto, M. Ike et al., *Bacteroides plebeius* sp nov and *Bacteroides coprocola* sp nov., isolated from human faeces, *Int. J. Syst. Evol. Microbiol.* **55**, 2143 (2005).

M. P. Garcillan-Barcia, M. V. Francia, and F. de la Cruz, The diversity of conjugative relaxases and its application in plasmid classification, *Fems Microbiology Reviews* **33**, 657 (2009).

R. A. Laine, A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05 x 10(12) structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems, *Glycobiology* **4**, 759 (1994).

H. Elifantz, L. A. Waidner, V. K. Michelou et al., Diversity and abundance of glycosyl hydrolase family 5 in the North Atlantic Ocean, *FEMS Microbiol. Ecol.* **63**, 316 (2008).

D. L. Kirchman, The ecology of *Cytophaga-Flavobacteria* in aquatic environments, *FEMS Microbiol. Ecol.* **39**, 91 (2002).

J. I. Hedges, J. A. Baldock, Y. Gelinas et al., Evidence for non-selective preservation of organic matter in sinking marine particles, *Nature* **409**, 801 (2001).

P. J. Turnbaugh, M. Hamady, T. Yatsunenko et al., A core gut microbiome in obese and lean twins, *Nature* **457**, 480 (2009).

K. Nisizawa, H. Noda, R. Kikuchi et al., The Main Seaweed Foods in Japan, *Hydrobiologia* **151**, 5 (1987).

K. Ishihara, C. Oyamada, R. Matsushima et al., Inhibitory effect of porphyran, prepared from dried "Nori", on contact hypersensitivity in mice, *Bioscience Biotechnology and Biochemistry* **69**, 1824 (2005).

D. K. Kidby and D. J. Davidson, A convenient ferricyanide estimation of reducing sugars in the nanomole range, *Anal. Biochem.* **55**, 321 (1973).

T. R. Schneider and G. M. Sheldrick, Substructure solution with SHELXD, *Acta*

Crystallographica Section D-Biological Crystallography **58**, 1772 (2002).

T. Pape and T. R. Schneider, HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs, *Journal of Applied Crystallography* **37**, 843 (2004).

G. N. Murshudov, A. A. Vagin, and E. J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallographica Section D-Biological Crystallography* **53**, 240 (1997).

L. Potterton, S. McNicholas, E. Krissinel et al., Developments in the CCP4 molecular-graphics project, *Acta Crystallographica Section D-Biological Crystallography* **60**, 2288 (2004).

E. Potterton, P. Briggs, M. Turkenburg et al., A graphical user interface to the CCP4 program suite, *Acta Crystallographica Section D-Biological Crystallography* **59**, 1131 (2003).

J.H. Hehemann, G. Correc, G. Michel et al., Porphyranases et leur utilisation pour hydrolyser des polysaccharides, France (2009).

F. W. Studier, Protein production by auto-induction in high density shaking cultures, *Protein. Expr. Purif.* **41**, 207 (2005).

W. Kabsch, Evaluation of Single-Crystal X-Ray-Diffraction Data from a Position-Sensitive Detector, *Journal of Applied Crystallography* **21**, 916 (1988).

A. G. W. Leslie and H. R. Powell, Processing diffraction data with MOSFLM, *Evolving Methods for Macromolecular Crystallography* **245**, 41 (2007).

C. M. Starr, R. I. Masada, C. Hague et al., Fluorophore-assisted carbohydrate electrophoresis in the separation, analysis, and sequencing of carbohydrates, *Journal of Chromatography A* **720**, 295 (1996).

K. Katoh, K. Misawa, K. Kuma et al., MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform, *Nucleic Acids Res.* **30**, 3059 (2002).

P. Gouet, X. Robert, and E. Courcelle, ESPript/ENDscript: Extracting and rendering sequence

and 3D information from atomic structures of proteins, *Nucleic Acids Res.* **31**, 3320 (2003).

S. Guindon and O. Gascuel, A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood, *Syst. Biol.* **52**, 696 (2003).

S. Kumar, K. Tamura, and M. Nei, MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment, *Brief. Bioinform.* **5**, 150 (2004).

End notes

Supplementary Information will be linked to the online version of the paper. Figures illustrating analytical results described throughout the paper are available in Supplementary Information.

Acknowledgements We thank B. Kloareg, C. de Vargas for critical discussions and reading of the manuscript; The laboratory “Marine Plants and Biomolecules” is funded by the French national research centre (Centre National de la Recherche Scientifique) and the University Marie Curie, Paris VI, France; J.H. Hehemann was supported by a European Marie Curie PhD grant; this work was also funded by the “Region Bretagne” through the program Marine 3D; we thank the beamline scientists and staff at the European Synchrotron Radiation Facilities, Grenoble, France for technical support during the data collections and the NMR Service, University Bretagne Occidentale, Brest, France for access to the Bruker NMR spectrometer.

Author Contributions J.H.H cloned, purified and crystallized the enzymes and extracted polysaccharides; J.H.H. and M.C. collected data and solved the crystal structures; G.C. and J.H.H. purified and characterized oligosaccharides; G.C. and W.H. performed the NMR analysis; M.G., T.B. and J.H.H did bioinformatics analysis; M.C., T.B., G.M and J.H.H designed the study; J.H.H., M.C. and G.M. analysed the data and wrote the paper. All authors discussed the results and commented on the manuscript.

Author information Atomic coordinates and structure factors reported in the paper have been deposited with the Protein Data Bank under accession codes 3ILF (PorA_E139S) and 3JUU (PorB). Correspondence and requests for materials should be addressed to M.C. (czjzek@sb-

roscoff.fr) or G.M. (gurvan@sb-roscoff.fr).

IV.4.10 Supplementary material

Supplementary Table 1. Data reduction, phasing and refinement statistics

Data statistics MAD	**PorA (ID23-1)**	**Peak**	**Edge**	**Remote**
Wavelength (Å)		1.03965	1.04005	1.03530
f' f"		-15.5 13.4	-19.8 9.17	-12.7 10.1
Resolution (Å)		49.9-1.45	49.9-1.45	49.9-1.45
Total Data		167078	160093	148827
Unique Data		82329	78319	72117
Redundancy[a]		2.03 (1.76)	2.05 (1.77)	2.06 (1.85)
Completeness (%)[a]		87.3 (73.4)	85.5 (66.4)	71.2 (36.7)
I/sigI[a]		12.7 (3.08)	12.5 (2.69)	10.9 (2.05)
Rsym (%)[ab]		3.8 (23.8)	4.1 (31.4)	5.1 (42.7)
Phasing (ShelXE)				
CCano (%)		68.2	53.5	58.4
CCiso (%)		76.2	63.7	

FOM[c] before/after DM (to 1.5 Å) 0.22/0.62

Native data statistics	**PorA complex**	**PorB**
Resolution (Å)	49.3-1.8	44.8-1.8
Total Data	180872	235397
Unique Data	27062	62907
Redundancy[a]	6.7 (6.2)	3.7 (3.7)
Completeness (%)[a]	94.6 (91.4)	100.0 (100.0)
I/sigI[a]	20.8 (8.4)	11.0 (3.7)
Rsym (%)[ab]	5.7 (19.5)	8.9 (20.3)

Refinement	**PorA complex**	**PorB**
Resolution range (Å)	49.3-1.8	44.8-1.8
R_{work} (R_{free})[d]	15.9 (20.5)	16.1 (20.1)
RMSD bonds [e](Å)	0.022	0.028
RMSD angles (°)	1.97	2.0

Overall B-factors	**PorA complex**	**PorB**
All atoms	13.5	16.2

Protein	11.5	15.1
Ligand/Ions	19.2	16.5
Water	21.7	21.9

N of atoms	**PorA complex**	**PorB**
Protein	2135	4343
Ligand/Ions	76	68
Water	324	446

[a] Values in parenthesis correspond to the highest resolution shell.

[b] $R_{sym} = \Sigma |I-I_{av}| / \Sigma I$, where the summation is over all symmetry equivalent reflections.

[c] FOM, figure of merit.

[d] R-factor calculated on 5% of data excluded from refinement.

[e] RMSD, Root Mean Square Deviation.

Supplementary Table 2: Genes extracted from Genbank used for phylogenetic tree construction.

label	Definition (GenBank)	Taxonomy	GenBank ID
Aga_Aeromo	β-agarase	*Aeromonas sp.*	AAF03246
Bp1670	hypothetical protein BACPLE_01670	*Bacteroides plebeius* DSM 17135	ZP_03208036
Bp1689	Hypothetical protein BACPLE_01689	*Bacteroides plebeius* DSM 17135	ZP_03208055
Aga_Cellvi	putative agarase	*Cellvibrio sp.* OA-2007	BAH16616
GFO_1714	Secreted glycosyl hydrolase, family 16	*Gramella forsetii* KT0803	CAL66684
GFO_1715	Membrane or secreted glycosyl hydrolase, family 16	*Gramella forsetii* KT0803	CAL66685

label	Definition (GenBank)	Taxonomy	GenBank ID
GFO_1725	Glycosyl hydrolase, family 16	*Gramella forsetii* KT0803	CAL66695
Aga_Maga	Agarase	*Microbulbifer agarilyticus*	BAE06228
Aga_Melo	Agarase	*Microbulbifer elongatus*	BAC99022
Aga_Mthe	Agarase	*Microbulbifer thermotolerans*	BAD29947
MS116	Putative beta-agarase precursor	*Microscilla sp.* PRE1	NP_116804
MS132	Putative beta-agarase	*Microscilla sp.* PRE1	NP_116820
Aga_Patl	β-agarase I	*Pseudoalteromonas atlantica* ATCC 19262	AAA91888
Patl0824	Glycosyl hydrolase, family 16	*Pseudoalteromonas atlantica* T6c	YP_660406
CgkA_Pc	Kappa-carrageenase	*Pseudoalteromonas carrageenovora* ATCC 43555	CAA50624
AgaPKJ2-4	Agarase	*Pseudoalteromonas sp.* KJ 2-4	AAR87712
Aga1PND137	Agarase	*Pseudomonas sp.* ND137	BAD88713
Aga2PND137	Agarase	*Pseudomonas sp.* ND137	BAB79291
CgkA_Rb	Kappa-carrageenase	*Rhodopirellula baltica* SH 1	NP_865103
AgaBSd2-40	Agarase	*Saccharophagus degradans* 2-40	YP_526649
AguB	AguB	Uncultured bacterium	AAP49346
AguF	AguF	Uncultured bacterium	AAP49324
Zg4203AgaA	β-agarase A precursor	*Zobellia galactanivorans* DSIJ	AAF21820
Zg3573AgaB	β-agarase B precursor	*Zobellia galactanivorans* DSIJ	AAF21821
Zg236CgkA	κ-carrageenase precursor	*Zobellia galactanivorans* DSIJ	AC27890

Supplementary Table 3. *Bacteroides plebeius* genes upstream and downstream of the porphyranase are listed against their best BLAST Hits found by Genbank and in the genome of *Zobellia galactanivorans*. Further homologue proteins in *B. thethaiotaomicron* and *B. vulgatus* are presented when existent in their genomes. Blue boxes: the genes with highest identity to genes in *Microscilla* or *Zobellia*. Yellow boxes: Synteny between genes in *B. plebeius* and other gut Bacteroides.

Bacteroides plebeius DSM 17135	Strand	Annotation	Best Blast Hit versus (GenBank + *Z. galactanivorans*)	*B. thetaiotaomicron* VPI-548)	*Bacteroides vulgatus* ATCC 8482
BACPLE_01702	-	L-fucosidase, family GH29	BACCOPRO_00400 (61,3%, *B. coprophilus*)	BT_3665 (34%)	BVU_0217 (35%)
BACPLE_01701	-	sulfatase	Zg332 (56,1%, *Z. galactanivorans*)	BT_3051 (40%)	BVU_0538 (33,8%)
BACPLE_01700	-	Alcohol dehydrogenase	BACCOP_02718 (75,9%, *B. coprocola*)	BT_0535 (28%)	BVU_0760 (73,6%)
BACPLE_01699	-	Two-component system sensor histidine kinase/response regulator	BACCAC_03608 (33,9%, *B. caccae*)	BT_2923 (28,9%)	BVU_0512 (28,5%)
BACPLE_01698	-	TonB-depedent receptor	Zg3570 (48,5%), Zg3637 (42%, *Z. galactanivorans*)	BT_3332 (32,5%)	BVU_1369 (32,3%)

Bacteroides plebeius DSM 17135	Strand	Annotation	Best Blast Hit versus (GenBank + *Z. galactanivorans*)	*B. thetaiotaomicron* VPI-548)	*Bacteroides vulgatus* ATCC 8482
BACPLE_01697	-	SusD-like protein	Zg3571 (37,9%), Zg3638 (33%, *Z. g.*	BT_3025 (27,8%)	BVU_1712 (30,1%)
BACPLE_01696	-	Conserved-lipoprotein containing CBM22	Zg3572 (36,4%), Zg3639 (28%, *Z. galactanivorans*)	No homologue	No homologue
BACPLE_01695	-	Hypothetical protein	No homologue	No homologue	No homologue
BACPLE_01694	-	β-Agarase, family GH86	MS130 (47,9%, *Microscilla sp*. PRE1)	No homologue	No homologue
BACPLE_01693	-	β-Agarase, family GH86	MS130 (51,5%, *Microscilla sp*. PRE1)	No homologue	No homologue
BACPLE_01692	-	Conserved hypothetical protein	BF2161 (30,6%, *B. fragilis* NCTC 9343)	No homologue	No homologue
BACPLE_01691	-	Hypothetical protein	No homologue	No homologue	No homologue
BACPLE_01690	-	Hypothetical protein	No homologue	No homologue	No homologue
BACPLE_01689	-	β-Porphyranase, family GH16	Zg3640 (49,1%, *Z. galactanivorans*)	No homologue	No homologue
BACPLE_01688	-	Hypothetical protein	No homologue	No homologue	No homologue
BACPLE_01687	+	Hypothetical protein	No homologue	No homologue	No homologue

Bacteroides plebeius DSM 17135	Strand	Annotation	Best Blast Hit versus (GenBank + *Z. galactanivorans*)	*B. thetaiotaomicron* VPI-548)	*Bacteroides vulgatus* ATCC 8482
BACPLE_01686	-	Hypothetical protein	No homologue	No homologue	No homologue
BACPLE_01685	-	Conserved hypothetical protein	BACCOP_00861 (28,3%, *B. coprocola*)	No homologue	No homologue
BACPLE_01684	-	Conserved hypothetical protein	BVU_0514 (73,4%, *B. vulgatus*)	BT_3687 (69,9%)	BVU_0514 (73,4%)
BACPLE_01683	-	Conserved hypothetical protein	BACINT_01451 (45,4%, *B. intestinalis*)	No homologue	No homologue
BACPLE_01682	-	Conserved hypothetical protein	EEI93196 (48,3%, *S. spiritivorum*)	BT_2422 (30%)	No homologue
BACPLE_01681	-	Hypothetical protein	No homologue	No homologue	No homologue
BACPLE_01680	-	Conserved hypothetical protein	BT_3666 (73,4%, *B. thetaiotaomicron*)	BT_3666 (73,4%) / BT_1276 (73%)	BVU_0214 (70,1%) / BVU_0190 (69%)
BACPLE_01679	-	Oxidoreductase	BACDOR_00330 (76,5%, *B. dorei*)	BT_3614 (75,2%)	BVU_0219 (75,8%)
BACPLE_01678	-	Conserved hypothetical protein	BACFIN_03900 (76,9%, *B. finegoldii*)	BT_3615 (77,9%)	BVU_0220 (75,2%)

Bacteroides plebeius DSM 17135	Strand	Annotation	Best Blast Hit versus (GenBank + *Z. galactanivorans*)	*B. thetaiotaomicron* VPI-548)	*Bacteroides vulgatus* ATCC 8482
BACPLE_01677	-	Fucose permease	BACFIN_03899 (76,8%, *B. finegoldii*)	BT_3616 (77,5%)	BVU_0221 (75,9%)
BACPLE_01676	-	Sorbitol dehydrogenase	BT_3617 (75,7%, *B. thetaiotaomicron*)	BT_3617 (75,7%)	BVU_0222 (73,1%)
BACPLE_01675	-	Altronate oxidoreductase	BT_0825 (81,2%, *B. thetaiotaomicron*)	BT_0825 (81,2%)	BVU_3075 (80,6%)
BACPLE_01674	-	Altronate hydrolase	PARMER_00240 (70,7%, *Parabacteroides merdae*)	BT_0486 (68,2%)	BVU_3053 (68,6%)
BACPLE_01673	-	β-Galactosidase, family GH2	Zg4655 (39,1%, *Z. galactanivorans*)	BT_4667 (29,9%)	BVU_2162 (30,1%)
BACPLE_01672	-	β-Galactosidase, family GH2	Zg4655 (50,1%, *Z. galactanivorans*)	BT_0461 (29,9%)	BVU_2162 (33,5%)
BACPLE_01671	-	MS110-like protein	Zg4663 (68,8%, *Z. galactanivorans*), MS110 (67,% *Microscilla sp*. PRE1)	No homologue	No homologue
BACPLE_01670	-	β-Agarase, family GH16	Zg3573 (agaB, 52,8%, *Z. galactanivorans*)	No homologue	No homologue

Bacteroides plebeius DSM 17135	Strand	Annotation	Best Blast Hit versus (GenBank + *Z. galactanivorans*)	*B. thetaiotaomicron* VPI-548)	*Bacteroides vulgatus* ATCC 8482
BACPLE_01669	-	Two-component system sensor histidine kinase Nterminal domain	BVU_0682 (51,5%, B. vulgatus)	BT_2619 (27,4%)	BVU_0682 (51,5%)
BACPLE_01668	-	Two-component system sensor histidine kinase Cterminal domain	BVU_0682 (59,6%, *B. vulgatus*)	BT_2619 (38%)	BVU_0682 (59,6%)
BACPLE_01667	-	Two-component system response regulator	BVU_0681 (58,6%, *B. vulgatus*)	BT_2618 (44,6%)	BVU_0681 (58,6%)
BACPLE_01666	+	Conserved hypothetical protein	PARMER_01229 (52,5%, *P. merdae*)	No homologue	BVU_0680 (51,1%)
BACPLE_01665	+	Mobilization protein Cterminal domain	BACSTE_02978 (92%, *B. stercoris*)	BT_2614 (91,4%)	BVU_0679 (91,2%)
BACPLE_01664	+	Reverse transcriptase	BACUNI_01785 (99,6%, Bacteroides uniformis)	BT_2297 (61,6%)	BVU_2134 (62,6%)

Bacteroides plebeius DSM 17135	Strand	Annotation	Best Blast Hit versus (GenBank + *Z. galactanivorans*)	*B. thetaiotaomicron* VPI-548)	*Bacteroides vulgatus* ATCC 8482
BACPLE_01663	+	Mobilization protein Nterminal domain	BT_2614 (84,4%, *B. thetaiotaomicron*)	BT_2614 (84,4%)	BVU_0679 (81,1%)
BACPLE_01662	+	Mobilization protein	BACOVA_00482 (74,1%, *B. ovatus*)	BT_2613 (73,6%)	BVU_0678 (72,6%)

Supplementary Table 4. Primers used for amplification of GH16 glycoside hydrolases and their CBMs from *Z. galactanivorans* and for mutagenesis of PorA

Zg1010 GH16	BamHI/MfeI	ggggggggATCCCAAGATTACAACTTGGTCTGGCAAG	CCCCCCCAATTgTTACTTTTGGTAGACCCTTACGTAATCT
		Primer Forward	**Primer Reverse**
Zg1010 UNK1	BamHI/MfeI	ggggggggATCCGCCTTGAATTCCTCTCTGATAGCT	CCCCCCCAATTgTTACTCTTTGATCAACTGCCGAACAGT
		ggggggggATCCAAATTACAGGTATACCCCGTTCCC	CCCCCCgAATTCTTACTCGGAGATAACGATTTTTTGTTTTAC
Zg3628	BglII/EcoRI	gggggAgATCTCAAGAGCCCCCTAAAACCTATAGT	CCCCCCgAATTCTTATTTGGTCAAATGGATATAGACCGTAC
Zg2600	BamH1/EcoRI	ggggggggATCCCAATTACCATCTCCTACAAACGGG	CCCCCCgAATTCTTAGTTTTTCTGAAACCAGGTTTGTTCAC
Zg3628 UNK1	BglII/EcoRI	gggggAgATCTACCCTAAAACTGTACCACGAAAACG	CCCCCCgAATTCTTATTTGGTCAAATGGATATAGACCGTAC
Zg2600 GH16 (PorA)	BamH1/EcoRI	ggggggggATCCCAATTACCATCTCCTACAAACGGG	CCCCCCgAATTCTTAGTCAACCAATTTATACACCCGTACC
Zg4243	BglII/MfeI	gggggAgATCTCAATACGATTGGGACAACGTGCC	CCCCCCCAATTgTTACTCGGAGATAACGATTTTTTGTTTTAC
		ggggggggATCCCAATTACCATCTCCTACAAACGGG	CCCCCCgAATTCTTATTTCCAAGCAAAAGTTTCCCAACCA
Zg4243 GH16	BglII/MfeI	gggggAgATCTCAATACGATTGGGACAACGTGCC	CCCCCCCAATTgTTAGTTCACAGGTTTGTAAACCCGGAT
		ggggggggATCCGCTCCGATCGGCAGTTACATTTC	CCCCCCgAATTCTTATTTCCAAGCAAAAGTTTCCCAACCA
Zg1014 GH16	BglII/MfeI	gggggAgATCTCAGGTAGGCCAGGTTTTATGGGA	CCCCCCCAATTgTTAGTTGTATTCGGAAACACGCACATAG
		ggggggggATCCGCGCTTTCCGCCCAGTTGGAG	CCCCCCgAATTCTTAGTTTTTCTGAAACCAGGTTTGTTCAC
PorA_E139S	mutagenesis	ccgcgttggccaattttcttcaatcgatgttatcgagcatattgg	ccaatatgctcgataacatcgattgaagaaaattggccaacgcgg
		ggggggggATCCCAAGAGCCCCCTAAAACCTATAGT	CCCCCCgAATTCTTAGTCGTTTACATCGACCAATCGGTA
Zg3640	BamH1/EcoRI	ggggggggATCCCAGACGCCGCCGCCGCCGG	CCCCCCgAATTCTTATTGATCTACAGGAAGCAAGGTATAC
Zg1017 (PorB)	BamH1/EcoRI	ggggggggATCCCAAGAAGCTCCACATTTTAAGCCTG	CCCCCCgAATTCTTAATTCTTTGAATCAACCAATTGCCATG
Zg3376	BamH1/EcoRI	ggggggggATCCTGTAGCAATTCGGGGGATAATGGT	CCCCCCgAATTCTTACAAATCTTCAAGTTGCCATACCCTA
Zg1010 CBM6	BamH1/EcoRI	ggggggggATCCGCTAACACCCTTAAAATCGAGGCG	CCCCCCgAATTCTTACTTTGTAATCCTGATCCAGTTGATAT
Zg1014 CBM42	BamH1/EcoRI	ggggggggATCCACTATTACCTATCGTATTACCATCGA	CCCCCCgAATTCTTAAAAGGTCGCATCATTTCTTGCCGA
Z1021 GH16	BamH1/EcoRI	ggggggggATCCCAAGATGTTATCATGTTCGACGATTT	CCCCCCgAATTCTTAGCGAAATTCCGATACCAAGTTTTCG
Z3347	BamH1/EcoRI	ggggggggATCCTGTAGTGGCCCGGAAGGCCCC	CCCCCCgAATTCTTATAAAATTTCGATTCTGTCGATAATGATT
Zg2431	BamHI/MfeI	ggggggggATCCTGCAGTTCTACGGATTCAGGACAG	CCCCCCCAATTgTTATTGATAGATCCTTACATAGTCTATTTC
Zg2431 GH16	BamHI/MfeI	ggggggggATCCGCCTTTAATACCTTAGTGTTTTCAGA	CCCCCCCAATTgTTATTGATAGATCCTTACATAGTCTATTTC
Zg1010	BamHI/MfeI	ggggggggATCCCAAGATTACAACTTGGTCTGGCAAG	CCCCCCCAATTgTTACTCTTTGATCAACTGCCGAACAGT

IV.5 Introduction for manuscript3: Production of porphyran oligosaccharides with porphyranase

Porphyran is the major cell wall polysaccharide of marine red algae belonging to the genus *Porphyra sp.*. Even though this polysaccharide was identified decades ago, no porphyranases had been reported until now. Thus, only β-agarases from marine bacteria have been used to analyse the primary structure of red algal galactans especially of porphyran from *Porphyra* spp. (Morrice at al. 1983). Indeed, the use of β-agarases was based on the characterization of low abundant degradation products, which were not representative of the overall polysaccharide structure. In contrast to agarases, porphyranases produce oligosaccharides consisting of L6S-G motifs with L6S-G motifs, at the reducing ends of the oligosaccharides. Thus, they degrade the agarase resistant fraction and can be used to analyse the primary structure of porphyran, sulfated agarocolloids and sulfated agars extracted from algae. In addition, sulfated polysaccharides such as fucans, carrageenans and porphyran found recent interest due to their pharmacological activities which include antiviral (Buck et al. 2006), antithrombic (Franz et al. 1996; Alban et al. 1997), antiinflammatory and antioxidative effects (Zhang et al. 2004; Hatada et al. 2006). Specific CAZymes may therefore of interest to analyse such polysaccharides with pharmacological function. Moreover, the precise chemical characterization and description of porphyran oligosaccharides have never been reported so far. The following manuscript (draft version of manuscript3; to be submitted in Carbohydrate Research) therefore reports on the use of β-porphyranases and β-agarases for the production of original porphyran-oligosaccharides. The description of the use of these enzymes for porphyran-oligosaccharide production was also filed as a patent.

IV.6 Manuscript 3

(First draft)

Structure analysis of the degradation product of porphyran by *Zobellia galactanivorans* β-porphyranase A

Gaëlle Correc[†], Jan-Hendrik Hehemann[†], Mirjam Czjzek*, William Helbert*

§ Université Pierre et Marie Curie, Paris VI, CNRS, Marine plants and biomolecules, UMR 7139, Station Biologique, BP 74, F29680 Roscoff Cedex, France

[†]These authors contributed equally to this work

*Corresponding authors: e-mail: helbert@sb-roscoff.fr, czjzek@sb-roscoff.fr

IV.6.1 Abstract

Porphyran from the red seaweed *Porphyra umbilicalis* was degraded with β-porphyranase A from the marine flavobacterium *Zobellia galactanivorans*. β-Porphyranase A produced a new series of sulfated oligosaccharides that are here characterized by NMR, HPAEC and size exclusion chromatography. In contrast to previously used β-agarases, which produce predominantly non-sulfated oligosaccharides of the neoagarobiose series and hybrid oligosaccharides, β-porphyranase A produces oligosaccharides of the α-L-gal*p*-6-sulfate (1→3) ß-D-gal*p* (L6S-G) series the L6S-G disaccharide as major final product. We further used the new β-porphyranase A in combination with the β-agarase B and to examine porphyran structure.

Keywords

Porphyran, porphyranase, agarase, seaweed galactans, algal polysaccharide, glycoside hydrolase

IV.6.2 Introduction

Marine red seaweeds belonging to the genera agarophytes or carrageenophytes contain sulfated cell wall polysaccharides named agars and carrageenans. These linear polysaccharides consist of alternating α(1→3) and β(1→4) linked galactopyranose residues and they are widely used for biotechnological applications and in food industry due to their jellifying properties.

The difference between agars and carrageenans concerns the α(1→3) linked galactose units which contain an unusual 3,6-anhydro bond and are of L-configuration in agars and of D-configuration in carrageenans. For example, the repetitive segment of agarose consists of →3-β-D-gal*p* (G) units (1→4)3,6-anhydro-α-L-gal*p* (LA) →1 (Craigie et al. 1974).

Carrageenans are classified based on the position of the sulfate ester on the sugar units into κ- ι- and λ-carrageenans. This system was not adopted for agars which are more generally classified based on their rheological properties into the agarans which form gels and the agarocolloids which produce viscous solutions (Knutsen et al. 1994).

One sulfated agarocolloid is extracted from the cell wall of red seaweeds belonging to the genus Porphyra (Turvey et al. 1961) and was therefore called porphyran. Whereas agars and carrageenans are modern processed food ingredients and not consumed in form of seaweeds, porphyran is a naturally consumed algal polysaccharide because it is the main component of the edible species *Porphyra yezoensis* and *Porphyra tenera* generally known as nori (Nisizawa et al. 1987; Fukuda et al. 2007). As other sulfated galactans, porphyran may also present some potential pharmacological applications (Pomin et al. 2008) and, for example, it possesses hypolipidemic (Inoue et al. 2009) and anti-allergenic properties (Ishihara et al. 2005).

Extensive structural analysis of the oligosaccharides produced after incubation of porphyran with β-agarases has demonstrated its hybrid–copolymer-structure (Duckworth et al. 1969; Morrice et al. 1983; Morrice et al. 1984). The polysaccharide is composed of α-L-gal*p*-6-sulfate (1→3) ß-D-gal*p* moieties (L6S-G) and to a lesser extend of α-L-3,6-anhydro-gal*p*

(1→3) ß-D-gal*p* (LA-G), as found in agarose. In addition, the polysaccharide backbone may be decorated by methyl ether group (Me) at the C6 position of the G units.

Structural analysis of carrageenans profited from the discovery of κ-, ι- and λ-carrageenases which allowed to dissect the hybrid nature of these complex polysaccharides (Michel et al. 2001; Michel et al. 2001; Guibet et al. 2007). On the contrary, the currently available enzymes for the structural analysis of agarans and agarocolloids are limited in specificity. Only agarose degrading enzymes, called α- or β-agarases as a function of the configuration of the glycosidic bond they cleave, have been reported (Michel et al. 2006). β-agarases, which are specific for LA-G blocks in the polysaccharide, were extensively used as tools to determine the structural diversity of agarocolloids such as porphyran. Nevertheless these investigations were based on the characterization of low abundant degradation products which were not representative of the overall polysaccharide structure. In this context, in depth analysis of the composition as well as the distribution of the various substituted building blocks along the hybrid polysaccharide chain would greatly benefit from new enzymes with alternative specificities.

The marine bacterium *Z. galactanivorans* has been isolated from the surface of the red algae *Delesseria sanguinea* for its ability to growth in minimal medium containing agar, κ- or ι-carrageenans as sole carbon source. The genome analysis indeed revealed several genes showing strong similarities with the known *Z. galactanivorans* β-agarases belonging to the family 16 of glycoside hydrolases (GH16) but also other genes closely related to agarases emerging as new cluster of genes. The corresponding recombinant proteins were inactive on standard agarose (LA-G) but showed high activity on porphyran showing that these enzymes cleave linkages between stretches of L6S-G moieties. Here we took advantage of this recent discovery and used the new recombinant β-porphyranase A and the β-agarase B (Jam et al. 2005) from *Z. galactanivorans* to revisit the porphyran structure from *Porphyra umbilicalis* and to characterize a novel series of porphyran oligosaccharides.

IV.6.3 Material and Methods

Purification of porphyran from *P. umbilicalis*

P. umbilicalis was collected in spring 2008 in the intertidal zone of the Perharidy Bay (Roscoff, France). The algae were extensively washed with freshwater and cleaned from sand and epiphytes. They were dried in an oven at 60 °C. 100 g of dried material was ground with a blender until a fine powder was obtained. The resulting powder was kept for 15 hours in 1 L 7.5 % (v/v) formalin/water solution. Then, an equal volume of water was added and the suspension was boiled under reflux for eight hours. The suspension was filtered through a sieve in order to remove the most of algal fragments. A clear solution was obtained after a centrifugation at 10000 g for 30 min and at 20° C and filtration through diatom earth followed by a second filtration through activated carbon. The pH of the solution was adjusted to pH=7.5 with NaOH (1M). The volume of the sample was reduced to about 1/6 of the starting volume by rotary evaporation (65° C). The polysaccharide was precipitated by adding of four parts of pure methanol, and maintained overnight at 4°C. The precipitate was recovered by filtration and extensively washed with pure methanol and finally with acetone prior to be air dried. This porphyran preparation will be further referred to as native porphyran. To remove agarose moieties from the native porphyran it was extensively digested with β-agarase B in water. The freeze-dried hydrolysis product was resolubilized at 2.5% and the enzyme resistant fraction (ERF) has been recovered by size exclusion chromatography with a Superdex 30 column. This resistant fraction will be further referred to as β-agarase B pre-treated porphyran.

Purification of the enzymes

Recombinant β-porphyranase A from *Z. galactanivorans* was prepared as previously described (manuscript 2). Briefly, *E. coli* BL21(DE3) cells which carried the PFO4 plasmid with ß-porphyranase A gene were grown to high density (Studier 2005). The His-tagged

porphyranase was purified from the cell supernatant by immobilized metal affinity chromatography on 10 ml of a nickel charged IMAC HyperCell resin column (Pall Corporation). The protein was eluted with a linear gradient of 0-500 mM nickel sulfate and fractions which corresponded to the major eluted peak were analyzed by SDS-PAGE. The protein was concentrated to a volume of ~ 5 ml by ultra filtration on an Amicon membrane (polyethersulfone, 30 kDa cutoff). For final protein polishing a Sephacryl S-200 column (GE Healthcare) pre-equilibrated with buffer C at a flow rate of 1 ml min-1 was used. The purified enzyme was concentrated to ~3 mg ml-1 by ultra filtration on an Amicon membrane (10 kDA cutoff). All chromatographic steps were carried out on an ÄKTA Explorer Chromatography system (GE-Healthcare) at 20° C. β-agarase B was produced as previously described (Jam et al. 2005).

Enzymatic degradation

Porphyran (1% (w/v) in deionised water) was incubated for 12 hours at 30°C with β-porphyranase A at an enzyme concentration of 0.15 µg/ml and with the β-agarase B at the same enzyme concentration. Degradation kinetics were followed by reducing end sugar assay and size exclusion chromatography. The completion of the enzyme reaction has been tested by re-adding of enzyme to verify that no further degradation could be induced.

For the agarose hydrolysis a solution of 0.5% (w/v) in deionised water was prepared with commercial agarose powder (Eurogentec). Then β-agarase B was added at 1µg/ml for an incubation of 7 hours at 40°C to prevent jellification.

Analytical Size Exclusion Chromatography

Analytical size exclusion chromatography was conducted with an Ultimate 3000 (Dionex) chromatography system using a Optilab Tex refractive index detector (Wyatt). After filtration through 0.45µm filter membrane, 250µL of sample was injected on two analytical columns Superdex 200 (10/300) and Superdex Peptide HR (10/300), mounted in series (GE Healthcare). Elution was achieved 100mM ammonium carbonate [$(NH_4)_2CO_3$] solution at a

flow rate of 0.3mL/min. All data were acquired with the Chromeleon software (Dionex).

Purification of standard and hybrid oligosaccharides

The freeze-dried hydrolysis product was dissolved in deionized water at a concentration of 4% (w/v). After filtration trough 0.45µm, 4 mL of sample was injected. The purification of porphyran-oligosaccharides was carried out by preparative Size exclusion Chromatography with three Superdex 30 (26/60) columns (GE Healthcare) in series and integrated on a HPLC system liquid injector/collector (Gilson). Detection was achieved by a refractive index detector (Spectra System RI-50). The Gilson system and the detector data were monitored by Unipoint software (Gilson). 50 mM ammonium carbonate [$(NH_4)_2CO_3$] was used a running buffer, at a flow rate of 1.5mL/min for 650 minutes. Fractions of oligosaccharides were collected and frozen until further analysis. The fractions were then freeze-dried before being analyzed by electrophoresis techniques and nuclear magnetic resonance.

High performance anionic exchange chromatography - pulsed amperometric detection (HPAEC-PAD)

For high resolution oligosaccharide analysis the oligosaccharide fractions (filtered on 0.22 µm) were analyzed with a Dionex system (ICS2500) equipped with a column Carbopac PA100 (4/250mm) associated to a PA100 guard-column. The system was equilibrated in 150mM NaOH and the elution of samples (20µL) was performed at 0.5 ml/min flow with 150mM NaOH and a step gradient from 0 to 1M Sodium Acetate (0-5min=0-60% ; 5-10min=60-100%; 10-30=100%). Oligosaccharide detection was carried out with an electrochemical detector (gold electrode) and data acquisition by the Chromeleon Peak Net software.

Nuclear Magnetic Resonance Spectroscopy

All samples were solubilized in D_2O and exchanged twice before being analyzed on the 500 MHz Bruker NMR spectrometer (Service de RMN, University Bretagne Occidentale, Brest,

France). For the porphyran polysaccharide, ^{1}H NMR spectra were recorded at 70°C using 64 scans. For the oligosaccharides, spectra were recorded at 25°C using 16 scans. Chemical shifts are expressed in ppm by reference to an external standard TSP (trimethylsilylpropionic acid). The NMR signals of the porphyran disaccharide were fully assigned using a complete set of correlation spectrums: COSY (doublequantum-filtered correlation spectroscopy), HMQC (heteronuclear multiple quantum correlation) and HMBC (heteronuclear multiple bond correlation). Other oligosaccharides were characterized with the same techniques and in comparison with the structure obtained for the porphyran disaccharide.

IV.6.4 Results and discussion

NMR analysis of porphyran

The porphyran extracted from the cell wall of *P. umbilicalis* was a water soluble polysaccharide with a yield of about 12-15 % per dry weight of algae. The ^{1}H NMR spectra of this polysaccharide (shown in Figure 56) present the characteristic downfield signals at 5.19 and 5.53 ppm from the α-anomeric proton of the 4-linked 3,6-α-L-anhydro-galactose (LA-H1) and α-L-galactopyranose-6-sulfate (L6S-H1) (Murano 1995; Zhang et al. 2005; Maciel et al. 2008). The chemical shift of the corresponding carbons were deduced from heteronuclear ^{1}H/^{13}C chemical shift correlation (HMQC) and the values are similar to those already reported (Usov et al. 1980; Morrice et al. 1983). Relative amounts of the two repetition moieties L6S-G or LA-G were deduced by integrating the signal of the α-anomeric protons. The measured L6S-H1/LA-H1 ratio of 1/0.4 accounts for about one third of LA-G and two thirds of L6S-G repetition moieties (Figure 56A). After incubation with β-agarase B, almost all LA-G moieties were eliminated and solely the L6S-H1 signal remained Figure 56B). Agars and porphyran have been previously shown to carry methyl ether groups (Morrice et al. 1983) (Lahaye et al. 1989) which were detected in the 3.5-3.40 ppm region of ^{1}H NMR spectra. In the case of the here described porphyran, two signals resonating at 3.46 and 3.44 ppm were clearly visible but after the β-agarase B treatment, the polysaccharide was predominantly constituted of L6S-G moieties and contained almost exclusively the methyl

ether at 3.44 ppm.

Chromatographic analysis of enzymatic degradation products

The enzymatic porphyran degradation by β-porphyranase A was monitored by assaying the production of reducing ends and by chromatography. The amount of reducing sugars produced after incubation with β-porphyranase A was about twice more important than after degradation with β-agarase B. In addition, the rate of cleavage was faster with the porphyranase which reached the complete degradation of the polysaccharide β-agarase B (data not shown). When porphyran was completely digested with β-porphyranase A several peaks, could be separated by size exclusion chromatography (SEC) and appeared between 450 and 550 min. Further one peak with a long shoulder appeared below 450 min (Figure 57A). A low amount of undigested high molecular fraction eluting at 220-250 min, which was not further characterized, shows that almost all the polysaccharides were fragmented into oligosaccharides. In order to remove LA-G moieties to reduce the hybrid character of the porphyran, we used β-agarase B and predigested the polysaccharide. We than used the resistant fraction enriched in L6S-G moieties as substrate for β-porphyranase A.

The chromatogram of the digestion of porphyran with β-agarase B (Figure 57C) clearly shows a different peak profile, lower peak size and a pronounced peak corresponding to the L6S-G enriched resistant fraction.

The enzymatic digestion of the enriched L6S-G porphyran with β-porphyranase A lead to a simplified chromatogram (Figure 57B) indicating that the complexity of the chromatogram of the unmodified porphyran (Figure 57A) was correlated to the co-occurrence of LA-G and L6S-G moieties in the polysaccharide. Therefore the additional peaks in Figure 57A can be assigned to oligosaccharides of hybrid structure.

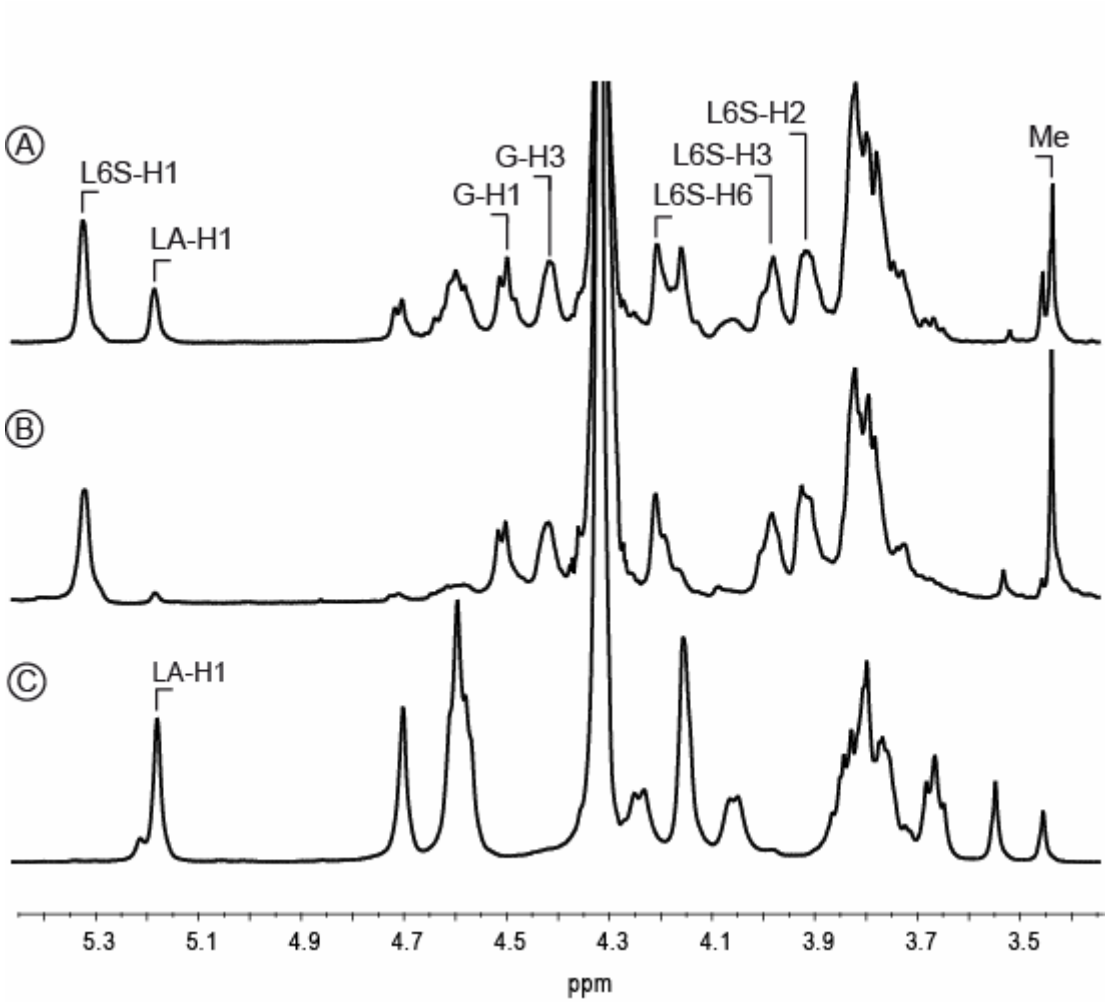


Figure 56: ^{1}H NMR (500MHz) spectra of porphyran from *Porphyra umbilicalis*. (A) Native porphyran without enzymatic pre-treatment. **(B)** High molecular weight fraction that was obtained by digestion of native porphyran with β−agarase B. **(C)** NMR spectrum of agarose. The characteristic anomeric protons of the G-L6S and G-LA moieties are indicated as well as the signal characteristic of the L6S-G moiety.

The most abundant oligosaccharides were purified by preparative SEC at mg scale and their purity was estimated at 90-95% by HPAEC (data not shown). The oligosaccharides which corresponded to the different peaks were analyzed by ^{1}H NMR (Figure 59, Figure 61 for analysis see below). We identified the two major peaks shown in (Figure 57B) as DP2 and DP4 composed solely of the L6S-G moiety (Figure 59) and the peak eluting at 450 min (Figure 57A) corresponds to a hybrid oligosaccharide of L6S-G and LA-G moieties (DP6,

Figure 61).

We also analyzed the sample corresponding to Figure 57A by HPAEC and could separate more distinct peaks. By injecting the purified and characterized oligosaccharides all major peaks could be assigned (data not shown).

The oligosaccharides were separated according to their molecular mass in the SEC experiments however, the correlation between SEC and HPAEC chromatograms was not obvious. We observed that the DP6, containing the neutral LA-G moiety, eluted between DP2 and DP4 in the HPAEC experiment. This effect may derive from the non proportional number of charges/molecular mass. We further identified two additional peaks in the HPAEC profile which were not resolved by SEC. The ^{1}H NMR analysis (see below) showed that they correspond to C6 methylated species of the DP6 and DP4 oligosaccharides whereas a methylated DP2 was never observed.

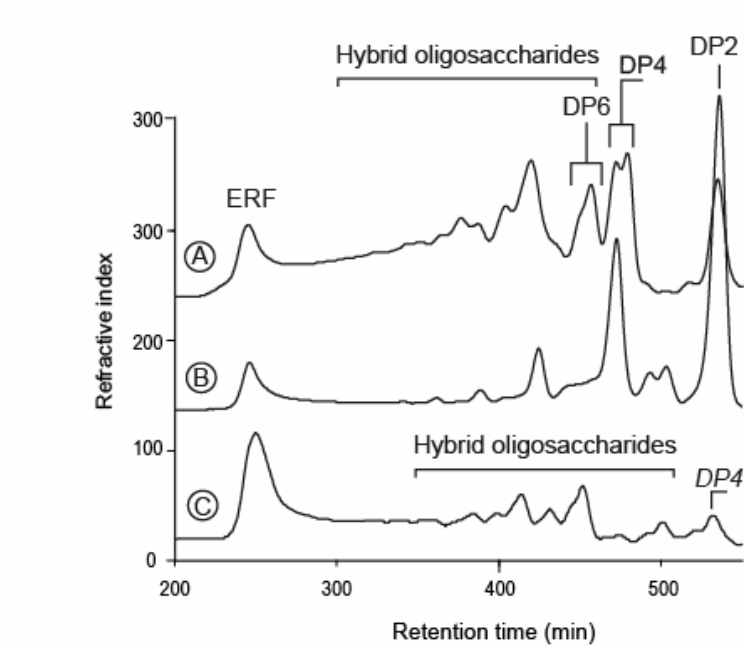


Figure 57: SEC-analysis of the reaction products obtained by digestion of native, and β-agarase B pre-treated porphyran with β-porphyranase A. (A) Chromatogram of native porphyran incubated with the β-porphyranase A. **(B)** Chromatogram of β-agarase B pre-treated porphyran digested with β-porphyranase A. **(C)** Chromatogram of the reaction products of native porphyran digested with the β-agarase B. DP2, DP4 and DP6 designate fractions of di-, tetra- and hexasaccharide. *DP4* refers to neo-agarotetraose. ERF: enzyme-resistant fraction.

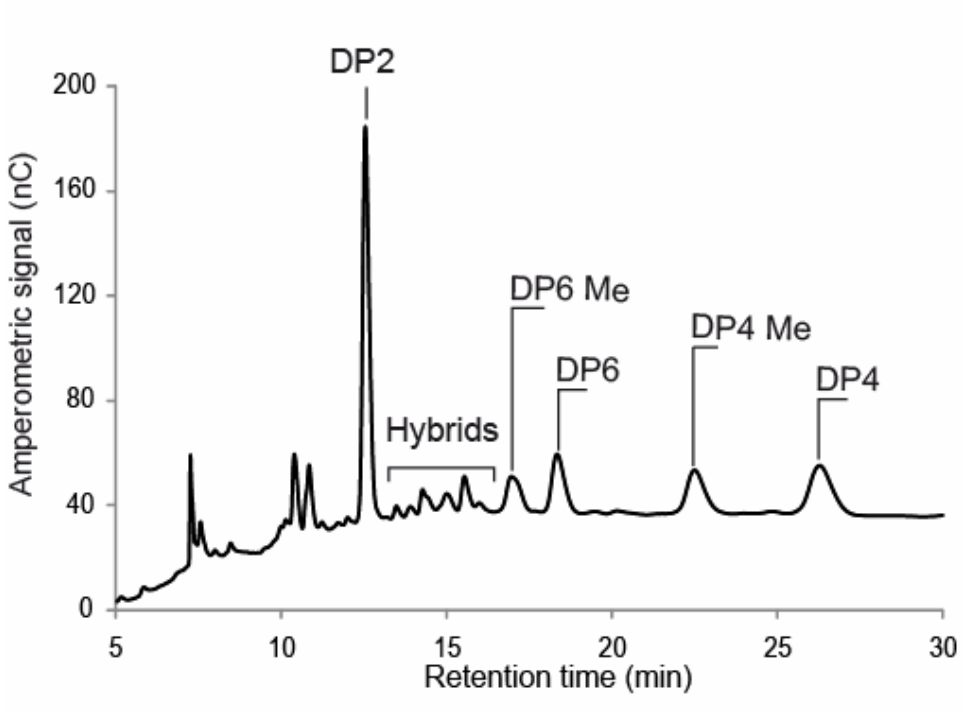


Figure 58: Anion exchange chromatography of the reaction products obtained by degradation of native porphyran incubated with β-porphyranase A. DP2, DP4 and DP6 designate (L6S-G), $(L6S\text{-}G)_2$ and (L6S-G-LA-G-L6S-G) oligosaccharides, respectively. Me: methylated oligosaccharides.

Complete assignment ^{13}C and ^{1}H NMR of standard L6S-G oligosaccharides

The major reaction product of β-porphyranase A was the L6S-G disaccharide. In accordance with the $^{1}H/^{13}C$ correlations and with the integration ratio, the anomeric protons were assigned. The most downfield signals corresponds to the Grα-H1 (5.3 ppm) and L6S-H1 (5.25 ppm) peaks. The coupling constant of these doublets which result from the typical axial-gauche conformation H1/H2 of D-galactose was of about 3.9 Hz. The doublet at 4.64 ppm ($^{3}J_{H1,H2}$ = 7.8 Hz) was assigned to the reducing residue in the β-anomeric configuration (Grβ-H1). The anomeric equilibrium (α/β 0.3/0.7) was determined from the integration of anomeric signals and shown a displacement toward the β-anomer.

The protons of the galactose rings were assigned with the COSY spectra using the anomeric protons as starting point. The H1-H4 correlation systems of the Grα/β and L6S could be identified (Figure 60). The corresponding carbons were deduced from the heteronuclear

$^{1}H/^{13}C$ chemical shift correlation (HMQC, not shown). The connection between H4 and H5-H6 systems were obtained with long range heteronuclear $^{1}H/^{13}C$ chemical shift correlation (HMBC). Evidence that the neutral residue was at the reducing end was provided by the difference of chemical shift encountered at the H6 and C6 positions. The most downfield H6 and C6 were observed for the residue at the non reducing end, which likely carried the sulfate ester group. The L6S-H1 had the α-anomeric configuration suggesting that the two residues were linked by α(1→3) linkage. This was supported by HMBC experiments which allowed connecting, for example, H3 (not H4) of the reducing galactose (Grα-H3 and Grβ-H3) with the anomeric carbon of the L6S residue. All together, these results were consistent with a L6S-G disaccharide structure obtained by the cleavage of the β(1-4) glycosidic bond of porphyran.

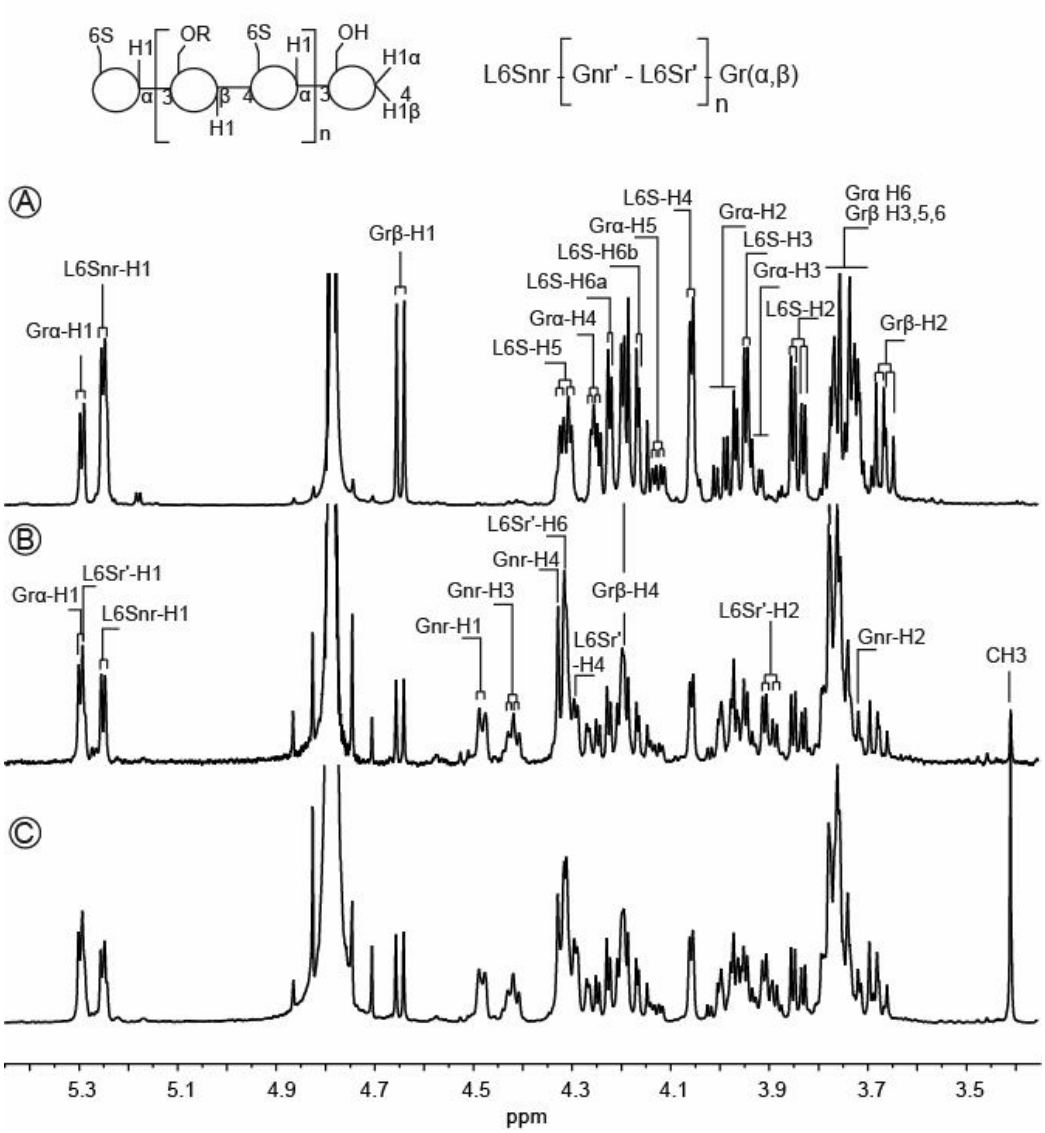


Figure 59: ^{1}H NMR (500MHz) spectra of disaccharide and tetrasaccharide of the L6S-G series. (A): L6S-G disaccharide. **(B):** Non methylated $(L6S\text{-}G)_2$ tetrasaccharide. **(C):** Methylated $(L6S\text{-}G)_2$ tetrasaccharide. Assignments of the disaccharide are indicated in **(A)** and the assignments of the internal G-L6S moiety of the tetrasaccharide in **(B)**. **Inset:** schematic structure of the standard $(L6S\text{-}G)_n$ oligosaccharide series showing H-1 protons and the methyl group position R.

The ^{1}H NMR spectra of the L6S-G di- and tetrasaccharides share several similarities such as the chemical shift of the anomeric protons localized at the reducing and non-reducing end. Based on the DP2 spectra, the three α-anomeric protons were easily ascribed to the internal and non reducing L6S (L6Sr'= 5.29 ppm; L6Snr=5.25 ppm) and to the reducing end

Grα (5.30 ppm). The chemical shift of the β-anomeric proton at the reducing end was found at 4.65 ppm for DP4 and DP2. The additional signal measured at 4.49 ppm was ascribed to the anomeric proton of the internal G residue in the DP4. Like for the DP2, the COSY spectra had allowed connecting the H1-H4 rings protons and HMQC had allowed assignment to the corresponding carbons. We further used HMBC experiments in order to connect the H5-H6 correlation system with the H1-H4 and finally determined the chemical shift of the corresponding carbons.

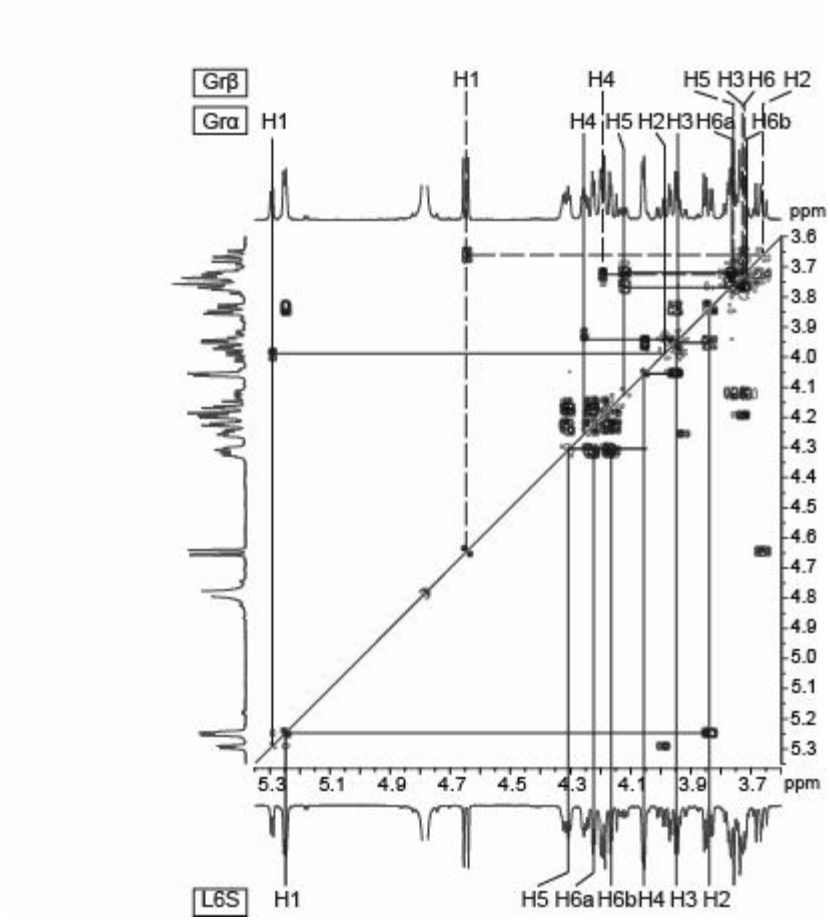


Figure 60: ^{1}H COSY (500MHz) spectrum of the disaccharides (L6S-G)$_1$ at 25°C. Correlation system of the protons, belonging to the L6S and G residues are drawn.

Structure of hybrids L6S-G/LA-G oligosaccharides

The ^{1}H NMR spectra in Figure 61 share similar characteristics with the L6S-G oligosaccharides (Figure 59). We identified the signals corresponding to the α-anomeric protons of the reducing and non-reducing ends, Grα-H1, L6S-H1r' and L6S-H1nr respectively, in the most downfield region of the spectra. Similarly, the β-anomeric protons (Gnr'-H1= 4.62 ppm Grβ-H1= 4.65 ppm) were also identified. The integration of the signal of the anomeric protons which revealed similar ratio than observed in the DP4, suggested that one L6S-G moiety was located at each end of the oligosaccharide. Based on the recorded data on pure porphyran oligosaccharides proton signals as well as carbons were assigned using well-resolved COSY, HMBC and HMQC spectra (data not shown). In addition to the L6S-G signals, the signal resonating at 5.16 ppm which was very close to the LA-H1 signal measured in agarose polymer, was ascribed to the α-anomeric proton of an internal anhydro-galactose residue (LA-H1). The integration of LA-H1 and L6Snr-H1 signals were similar indicating that this hybrid oligosaccharide was an hexasaccharide composed of one LA-G moiety positioned between two L6S-G moieties (L6S-G-LA-G-L6S-G). The protons and carbons of the internal LA-G moiety were completely assigned and the values were comparable to values determined for agarose polymer and for agarose-oligosaccharides (Kazlowski et al. 2008).

Longer size oligosaccharides which eluted between 350-420 min in the SEC experiment (Figure 57A) were not purified to homogeneity. However, fractions were collected and analyzed by NMR as shown in Figure 61C. Interestingly, the NMR spectra revealed that the oligosaccharides were systematically terminated by L6S-G moieties at the reducing and the non reducing end. The integration of the anomeric protons suggested that the internal moieties were composed both LA-G and L6S-G moieties in a ratio of about two LA-G for one L6S-G moiety.

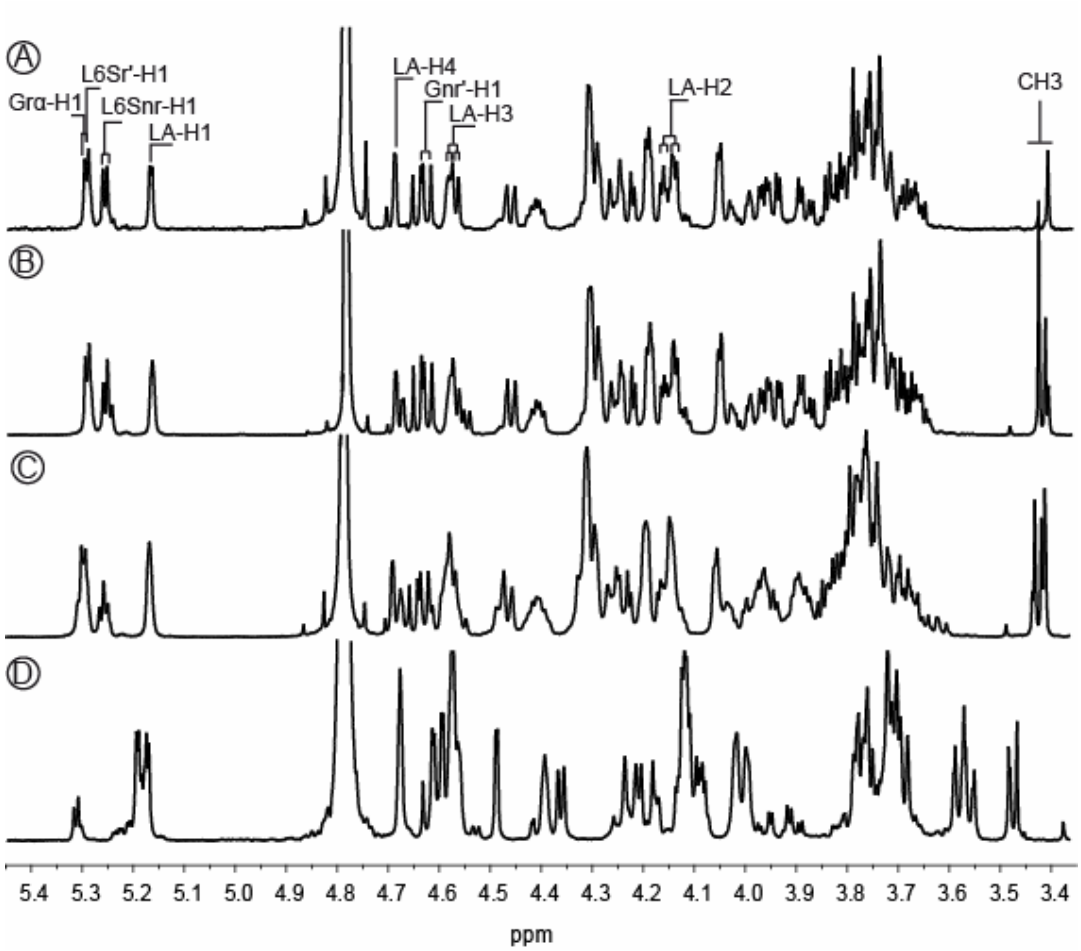


Figure 61: ^{1}H NMR (500MHz) spectra of the non-methylated (A) and methylated (B) L6S-G-LA-G-L6S-G oligosaccharides, high molecular fraction of hybrid oligosaccharides (C), and neo-agarotetraose (D).

Methylation of L6S-G/LA-G oligosaccharides

The extracted porphyran was decorated by methyl group (Figure 56) which were quantified by integration of the ^{1}H NMR signals to substitute about 18 % of D-galactose residues. Similarly, some oligosaccharides obtained after enzymatic digestion were methylated. Notably, a methylated form of the L6S-G-L6S-G tetrasaccharide and a hybrid hexasaccharide eluted as shoulder of their corresponding non-methylated form in the size exclusion chromatography. These two oligosaccharides were purified in order to determine the position of the methyl group along the oligosaccharides.

The protons of the methyl group of the tetrasaccharide were resonating as a single peak at 3.4

ppm in the ^{1}H NMR spectrum and the 61.25 ppm chemical shift of the corresponding carbon was deduced from the HMQC spectrum. The carbon carrying the methyl group was identified as a the C6 carbon of a G unit by using long range $^{1}H/^{13}C$ correlation of the HMBC spectra. The chemical shift of 74.5 ppm of this carbon indicated that the non-reducing end Gnr'-C6 unit was methylated whereas the reducing end Grα and Grβ - C6 signals at 64.3 and 64.1 ppm respectively indicated the absence of methylation.

In the case of the purified fraction of methylated hybrid DP6, three methyl signals were observed with different chemical shift and signal intensities suggesting a mixture of oligosaccharides. The protons of the abundant methyl group had exactly the same chemical shift (3.41 ppm) of that encountered in the tetrasaccharide suggesting than the most abundant hybrid hexasaccharide was likely carrying a methyl group on the non-reducing Gnr' unit. This signal was also present in the spectra of longer size hybrid oligosaccharides but its intensity was then lower than the other signals at 3.42-3.43 ppm. The chemical shifts of this less intense methyl signals were also present in the undigested porphyran.

IV.6.5 Conclusion

Protons and carbon NMR spectra of oligosaccharides of the L6S-G series were assigned for the first time. Taken care of the internal standard, our carbon NMR allowed to assign ^{13}C NMR spectra of L6S-G oligosaccharides. We completed the set of ^{13}C NMR data with the chemical shifts of L6S-G moieties located at the end of the oligosaccharides. ^{1}H NMR spectra of porphyran oligosaccharides were completely elucidated and allowed to further assign the chemical shifts of the proton signal from the polymer. We found that the internal G-L6S moiety of the standard DP4 (L6Snr-G-L6S-Gr) may be seen as a model of repetition moiety encountered in the polymer. Indeed, the chemical shifts of the L6Sr'-H1 (5.33 ppm), L6Sr'-H2 (3.92 ppm) and L6Sr'-H3 (3.98 ppm) were very similar compared to those reported by Maciel et al. (2008). The undefined L6Sr'-H4, -H5 and -H6 in the polymer were straightforwardly measured with the oligosaccharides. Moreover, chemical shifts of the neutral galactose residues were also unambiguously assigned such as the G-H1 (4.5 ppm) and G-H3 (4.42 ppm) which resonated like the internal G unit encountered in the polymer (Table

2).

In contrast to the degradation with β-agarase B, β-porphyranase A almost completely degraded the porphyran. The purification of oligosaccharides and the characterization of the most abundant degradation products by NMR revealed that β-porphyranase A produces predominantly products of the L6S-G series and hybrid L6S-G/LA-G oligosaccharides of Dp6. The smallest oligosaccharide indentified was the standard non-methylated L6S-G disaccharide which can be obtained by degradation of the pure tetrasaccharide of structure L6S-G-L6S-G. In addition, the hybrid oligosaccharides were systematically terminated at both ends by at least one L6S-G moiety. Altogether these results clearly demonstrate that β-porphyranase A selectively cleaves the β(1→4) glycosidic bond between L6S-G moieties and does not productively bind LA-G moieties in binding subsites -1,-2 and +1,+2.

The absence of methylated disaccharides (L6S-G) indicates that C6 methylated D-galactose units cannot bind in subsite -1. However, β-porphyranase A can productively bind methylated oligosaccharides since some DP4 and hybrid DP6 appeared to be methylated on the C-6 of the D-galactose units but never on the reducing end sugar. This suggests that C6 methylated D-galactose can be bound in the subsite -4 and +2 of β-porphyranase A.

The complexity of porphyran is explained by the co-occurrence of two repetition moieties; agarobiose (LA-G) and porphyrobiose (L6S-G), as well as to the decoration of the hybrid – copolymer- chain with methyl groups. No alternative chemical modifications could be identified on the here described oligosaccharides. Previous enzymatic digestion of porphyran using β-agarases highlighted this complexity but on a limited fraction of the polysaccharide since the agarose fraction represents about one third of the porphyran. By using β-agarases Duckworth and Turvey (1969) and later Morrice *et al.* (1983) isolated a low molecular weight fraction which accounted to about 20-30% of the starting material. Similarly, we have observed the low exten of porphyran was degradation by β-agarase B and have shown that an abundant enzyme resistant fraction remains which is of high molecular weight and contains almost exclusively L6S-G moieties. This implies that most of the LA-G moieties are present in the oligosaccharides. We have purified these oligosaccharides and analyzed them by NMR and found neo-agarotetraose and hybrid LA-G/L6S-G oligosaccharides as previously reported (Morrice et al. 1983). The hybrid oligosaccharides

may be terminated by L6S-G moieties at the non-reducing end showing the specificity of β-agarases and β-porphyranases is complementary. Because the amount of pure agarose-oligosaccharide is low compared to the hybrid LA-G/L6S-G oligosaccharides, one can suspect that blocks of agarose if present are very short.

Table 1. ^{1}H and ^{13}C NMR data (δ) ppm of the L6S and G residues at 25°C in standard disaccharide (L6S-G) tetrasaccharide $(L6S\text{-}G)_2$ and polymer $(L6S\text{-}G)_n$.

		(L6S-G)	$(L6S\text{-}G)_2$	$(L6S\text{-}G)_n$			(L6S-G)	$(L6S\text{-}G)_2$	$(L6S\text{-}G)_n$
Grα	H-1	5.29	5.30	-	L6Sr'α/β	H-1	-	5.30	5.33
Grβ	H-1	4.65	4.65	-	L6Snr	H-1	5.25	5.25	-
Gnr'	H-1	-	4.49	4.50	LA	H-1	-	-	5.19
Grα	H-2	3.99	4.00	-	L6Sr'α/β	H-2	-	3.90	3.92
Grβ	H-2	3.66	3.68	-	L6Snr	H-2	3.84	3.84	-
Gnr'	H-2	-	3.75	3.78	LA	H-2	-	-	4.16
Grα	H-3	3.94	3.95	-	L6Sr'α/β	H-3	-	3.98	3.98
Grβ	H-3	3.73	3.76	-	L6Snr	H-3	3.95	3.95	-
Gnr'	H-3	-	4.42	4.42	LA	H-3	-	-	4.60
Grα	H-4	4.25	4.26	-	L6Sr'α/β	H-4	-	4.30	4.27-4.36
Grβ	H-4	4.19	4.20	-	L6Snr	H-4	4.05	4.06	-
Gnr'	H-4	-	4.32	4.27-4.36	LA	H-4	-	-	4.70
Grα	H-5	4.12	4.13	-	L6Sr'α/β	H-5	-	nd	nd
Grβ	H-5	3.72	3.72	-	L6Snr	H-5	4.31	4.30	-
Gnr'	H-5	-	3.76	3.82	LA	H-5	-	-	4.60
Grα	H-6a	3.73	3.72-3.77	-	L6Sr'α/β	H-6a	-	4.31	4.27-4.36
Grβ	H-6a	3.73	3.72-3.77	-	L6Snr	H-6a	4.17	4.17	-
Gnr'	H-6a	-	3.71	3.80	LA	H-6a	-	-	4.06
Grα	H-6b	3.76	3.72-3.77	-	L6Sr'α/β	H-6b	-	4.33	4.27-4.36
Grβ	H-6b	3.76	3.72-3.77	-	L6Snr	H-6b	4.22	4.23	-
Gnr'	H-6b	-	3.71	3.80	LA	H-6b	-	-	4.06

IV.6.6 References

Craigie, J. and C. Leigh (1974). "carrageenans and agars." 109-133.

Duckworth, M. and J. R. Turvey (1969). "The action of a bacterial agarase on agarose, porphyran and alkali-treated porphyran." Biochemical Journal **113**(4): 687-692.

Fukuda, S., H. Saito, S. Nakaji, M. Yamada, N. Ebine, E. Tsushima, E. Oka, K. Kumeta, T. Tsukamoto and S. Tokunaga (2007). "Pattern of dietary fiber intake among the Japanese general

population." European Journal of Clinical Nutrition **61**(1): 99-103.

Guibet, M., S. Colin, T. Barbeyron, S. Genicot, B. Kloareg, G. Michel and W. Helbert (2007). "Degradation of lambda-carrageenan by *Pseudoalteromonas carrageenovora* lambda-carrageenase: a new family of glycoside hydrolases unrelated to kappa- and iota-carrageenases." Biochemical Journal **404**(1): 105-114.

Inoue, N., N. Yamano, K. Sakata, K. Nagao, Y. Hama and T. Yanagita (2009). "The Sulfated Polysaccharide Porphyran Reduces Apolipoprotein B100 Secretion and Lipid Synthesis in HepG2 Cells." Bioscience Biotechnology and Biochemistry **73**(2): 447-449.

Ishihara, K., C. Oyamada, R. Matsushima, M. Murata and T. Muraoka (2005). "Inhibitory effect of porphyran, prepared from dried "Nori", on contact hypersensitivity in mice." Bioscience Biotechnology and Biochemistry **69**(10): 1824-1830.

Jam, M., D. Flament, J. Allouch, P. Potin, L. Thion, B. Kloareg, M. Czjzek, W. Helbert, G. Michel and T. Barbeyron (2005). "The endo-beta-agarases AgaA and AgaB from the marine bacterium *Zobellia galactanivorans*: two paralogue enzymes with different molecular organizations and catalytic behaviours." Biochemical Journal **385**: 703-713.

Kazlowski, B., C. L. Pan and Y. T. Ko (2008). "Separation and quantification of neoagaro- and agaro-oligosaccharide products generated from agarose digestion by beta-agarase and HCl in liquid chromatography systems." Carbohydrate Research **343**(14): 2443-2450.

Knutsen, S., D. Myslabodski, B. Larsen and A. Usov (1994). "A modified system of nomenclature for red algal galactans." Botanica Marina **37**: 163-169.

Lahaye, M., W. Yaphe, M. T. P. Viet and C. Rochas (1989). "C-13-N-M-R Spectroscopic Investigation of Methylated and Charged Agarose Oligosaccharides and Polysaccharides." Carbohydrate Research **190**(2): 249-265.

Maciel, J. S., L. S. Chaves, B. W. S. Souza, D. I. A. Teixeira, A. L. P. Freitas, J. P. A. Feitosa and R. C. M. de Paula (2008). "Structural characterization of cold extracted fraction of soluble sulfated polysaccharide from red seaweed *Gracilaria birdiae*." Carbohydrate Polymers **71**(4): 559-565.

Michel, G., L. Chantalat, E. Duee, T. Barbeyron, B. Henrissat, B. Kloareg and O. Dideberg (2001). "The kappa-carrageenase of *P. carrageenovora* features a tunnel-shaped active site: a novel insight in the evolution of Clan-B glycoside hydrolases." Structure **9**(6): 513-525.

Michel, G., L. Chantalat, E. Fanchon, B. Henrissat, B. Kloareg and O. Dideberg (2001). "The iota-

carrageenase of *Alteromonas fortis*. A beta-helix fold-containing enzyme for the degradation of a highly polyanionic polysaccharide." Journal of Biological Chemistry **276**(43): 40202-40209.

Michel, G., P. Nyval-Collen, T. Barbeyron, M. Czjzek and W. Helbert (2006). "Bioconversion of red seaweed galactans: a focus on bacterial agarases and carrageenases." Applied Microbiology and Biotechnology **71**(1): 23-33.

Morrice, L. M., M. W. McLean, W. F. Long and F. B. Williamson (1983). "Beta-agarases I and II from *Pseudomonas atlantica*. Substrate specificities." European Journal of Biochemistry **137**(1-2): 149-154.

Morrice, L. M., M. W. Mclean, W. F. Long and F. B. Williamson (1983). "Porphyran Primary Structure - an Investigation Using Beta-Agarase-I from Pseudomonas-Atlantica and C-13-Nmr Spectroscopy." European Journal of Biochemistry **133**(3): 673-684.

Morrice, L. M., M. W. Mclean, W. F. Long and F. B. Williamson (1984). "Porphyran Primary Structure." Hydrobiologia **116**(Sep): 572-575.

Murano, E. (1995). "Chemical-Structure and Quality of Agars from Gracilaria." Journal of Applied Phycology **7**(3): 245-254.

Nisizawa, K., H. Noda, R. Kikuchi and T. Watanabe (1987). "The Main Seaweed Foods in Japan." Hydrobiologia **151**: 5-29.

Pomin, V. H. and P. A. S. Mourao (2008). "Structure, biology, evolution, and medical importance of sulfated fucans and galactans." Glycobiology **18**(12): 1016-1027.

Studier, F. W. (2005). "Protein production by auto-induction in high density shaking cultures." Protein Expression and Purification **41**(1): 207-234.

Turvey, J. R. and D. A. Rees (1961). "Isolation of L-Galactose-6-Sulphate from a Seaweed Polysaccharide." Nature **189**(476): 831-&.

Usov, A., S. Yarotsky and A. Shashkov (1980). "^{13}C-NMR spectroscopy of red algal galactans." Biopolymers **19**: 977-990.

Zhang, Q. B., H. M. Qi, T. T. Zhao, E. Deslandes, N. M. Ismaeli, F. Molloy and A. T. Critchley (2005). "Chemical characteristics of a polysaccharide from *Porphyra capensis* (Rhodophyta)." Carbohydrate Research **340**(15): 2447-2450.

V. Material and Methods

Here a short overview is given and most Material and Methods can be found in manuscript 1, 2 and 3

V.1 Expression and purification of PorA and PorB

For biochemical and crystallisation studies of the ß-porphyranase A (PorA) we used the truncated recombinant protein, which consists of the catalytic domain only (residues 19-275), hereafter called PorA. The ß-porphyranase B consists of the family GH16 catalytic domain only which was cloned entirely, hereafter called PorB (residues 22-294). For the generation of these constructs the nucleotide sequence corresponding to the catalytic domain (PorA) and full length mature protein (PorB) was amplified by PCR from *Z. galactanivorans* genomic DNA using a set of primers (for the cloning procedure see manuscript2).

PorA and PorB were expressed in *E. coli* BL21 (DE3) cells. Expression was carried out in 1 L of a ZYP-5052 expression culture with 100 µg ml^{-1} amp at 20 °C. Cells were harvested after 3.5 days with a final OD 600 nm ~16, by centrifugation (4000g; 20 min; 4°C). The cell pellets were resuspended in buffer A (20 mM Tris pH 8, 200 mM NaCl, 20 mM imidazole pH 7.5, lysozyme, DNAse). The resuspended cells were lysed with lysozyme for 30 minutes on ice followed by sonication. The lysate was cleared by centrifugation (50 000g, 30 min, 4°C) and a subsequent filtration with a 0.2 µm filter (Millipore). The filtered solution was loaded onto a 10 ml IMAC HyperCell resin column (Pall Corporation) which was charged with $NiSO_4$. The column was equilibrated with buffer A without lysozyme and DNAse. After washing with buffer A (ten column volumes) the protein was eluted using a 60 ml linear gradient of buffer A up to 60 % of buffer B (20 mM Tris pH 8, 200 mM NaCl and 500 mM imidazole) at a flow rate of 1 ml min^{-1}. The proteins were concentrated by ultra filtration on

an Amicon membrane (polyethersulfone, 30 kDa cutoff) to a final volume of ~5ml. For final polishing a sephacryl S-200 column (GE Healthcare) pre-equilibated with buffer C (20 mM Tris pH 8, 4 % (v/v) glycerol for PorA_CM and 20 mM Tris pH 8 for PorB) at a flow rate of 1 ml min^{-1} was used. Fractions which contained pure protein (as assessed by SDS-PAGE) were pooled and concentrated to 2.6 mg/ml for PorA_CM and ~8 mg ml^{-1} for PorB by centrifugal filtration with an Amicon of 10 kDA cutoff. All chromatography procedures were carried out on an ÄKTA-explorer chromatography system (GE-Healthcare) at RT.

V.2 DNA techniques and plasmid construction

Plasmid DNA minipreparations and agarose gel purification of DNA fragments were performed using Qiagen's QIAprep spin kit and QIAquick gel extraction kit, respectively all other DNA techniques were performed according to Sambrook *et al*.

V.3 The medium throughput cloning strategy

In short: parallel PCR reactions with primers carrying compatible restriction sites for the *E. coli* expression vector pFO4 (pET15b derivative; Ap^r, oriColE1, T7 promoter, used for N-terminal His_6-tagged protein expression) were performed. PCR products were purified using Qiagen's QIAprep spin kit After digestion of vector and PCR products and a further purification step both DNA molecules were ligated and used for the transformation of *E. coli* bacteria (DH5α). Clones were screened for plasmids with correct insert length by colony PCR. Proteins were expressed in Bl21 (DE3) pLysS cells harbouring the different vector constructions.

Due to the high number of different recombinant bacteria, we used auto induction medium to facilitate parallel induction. For the up scaling of the protein expression Erlenmeyer flasks and fermentor cultivation was used. Soluble expressed proteins were purified by immobilised metal affinity chromatography (IMAC). As final purification step a size exclusion

chromatography with sephacryl resin was performed to obtain monodispersive protein solutions. Dynamic light scattering (DLS) was applied to monitor the size distribution. Activity assays were based on the reducing sugar end method and enzyme products, namely agarose oligomers, were analyzed by fluorophore-assisted-carbohydrate electrophoresis (FACE).

V.4 Screening for crystallization conditions

The crystallization screenings were carried out with the honeybee crystallization robot (Genomic solutions) in 96 well sitting drop plates (Greiner) and a drop volume of 300 protein solution and 150 nl of the screening solutions. When crystals appeared, further crystallization experiments in Linbro plates (24 well) in a hanging and sitting drop setup were performed to optimize morphology and size of the crystals. Drops were prepared by mixing 2µl of protein solution with 1 µl of mother solution and the plates were stored at 20 °C and 4 °C. For diffraction studies crystals soaked in 5 % increment steps of Glycerol to a final concentration of 10 % within the crystallization solution. The soaked crystals were flash frozen under a nitrogen cryo stream and measured in-house or at the ESRF (ESRF, Grenoble, France).

V.5 Kinetic studies

Michaelis constant (K_m) and turnover number (k_{cat}) were experimentally determined with agarose in superfusion at a temperature of 44°C. Substrate concentration was varied from 0.0125% to 0.75% agarose in 300 nm NaCl and 50 mM MOPS buffer. All values were determined in triplicate and the amount of released sugars was estimated by the ferric cyanide reducing sugar assay. For each concentration seven time points were analysed during a total reaction time of 24 minutes. The calibration and estimation of cleavage events were made using glucose as standard. The Michaelis constant and the turnover number were determined by a non-linear regression program in excel.

V.6 Sequences and phylogeny

Searches for protein sequence similarities were performed using BLASTp (Altschul et al. 1997) on the GenBank database. Multiple sequence alignments were generated using MAFFT with the iterative refinement method and the scoring matrix Blosum62 (Katoh et al. 2002). The structure-based multiple alignment of the β-porphyranases was displayed using the program ESPript (Gouet et al. 2003). For the phylogenetic analyses of the family GH16 enzymes, the sequences were selected using the CAZY database . The MAFFT alignment of these sequences was manually refined using Bioedit (© Tom Hall), on the basis of the superposition of the structure of the κ-carrageenase of *Pseudoalteromonas carrageenovora* (Michel et al. 2001), the β-agarase AgaA (Allouch et al. 2003) and the β-porphyranases PorA and PorB from *Z. galactanivorans* (This work). Phylogenetic trees were derived from this refined alignment using the Maximum Likelihood method with the program PhyML (Guindon et al. 2003). The reliability of the trees was always tested by bootstrap analysis using 100 resamplings of the dataset. The trees were displayed with MEGA 3.1 (Kumar et al. 2004).

V.7 Fluorophore-assisted carbohydrate electrophoresis analysis (FACE)

Fractions obtained after gel chromatographic separation of oligosaccharides were analyzed by FACE (Starr et al. 1996). Aliquots of 100 µl were dried in a speed vac and the oligosaccharides were labelled with AMAC (2-amnoacridone) or with ANTS (8-aminonaphthalene-1,3,6-trisulfonate). For derivatisation the pellet was redissolved in 2 µl AMAC solution (0.1 M AMAC in acetic acid-DMSO (3:17, v/v) and 5 µl of 1 M sodium cyanoborohydride in water also followed by incubation of 16 hours at 37° C. For derivatisation with ANTS the pellet was redissolved in 2 µl ANTS solution (0.15 M ANTS in

acetic acid-water (3:17, v/v). When the pellet was redissolved 5 µl of a 1 M sodium cyanoborohydride in dimethylsulfoxid (DMSO) was added and the mixture was incubated for 16 hours at 37°C. The samples were analyzed on a 30 % polyacrylamide gel with a 4 % stacking gel and a BioRad gel system. Gels were run for half an hour at 15 mA followed by 40 mA for three hours at 4 °C.

V.8 Enzyme activity essays

Enzyme activity assays were carried out following Jam *et al.* 2005 with enzyme concentrations ~5 nM substrate concentrations of 0.125 % (w/v) and an incubation temperature of 30ºC. Oligosaccharides which were produced during enzymatic digestion of crude porphyran, pure porphyran, agarose (Eurogentec) and by direct activity screening of crude algal extract were determined as described by Kidby and Davidson (Kidby et al. 1973). Aliquots of the reaction medium diluted five to 10 times in 100µl was mixed with 1 mL ferric cyanide solution (300 mg potassium hexocyanoferrate III, 29 g Na2CO3, 1 mL 5 M NaOH, completed to 1 L with water). The mixture was boiled for 15 minutes and its absorbance was read at 420 nm. A calibration curve with 50 – 300 µg/ml of glucose was used to calculate the amount of reducing ends as glucose reducing end equivalents.

VI. Final discussion and outlook

VI.1 The agarolytic system of Z. galactanivorans

During my thesis I studied the agarolytic system of *Z. galactanivorans*. From this system, containing nine GH16 enzymes, I solved the structures and performed the biochemical characterization of three new enzymes. The characterization of the reaction products allowed analysis of the substrate specificities that the new AgaD, the previously described AgaA,B and the two new porphyranases PorA,B possess on heterogeneous substrates.

The two new characterized porphyranases clearly increase the substrate range that *Z. galactanivorans* can exploit as carbon source. In the case of porphyran approximately 2/3 of the polysaccharide are sulfated and in agars from other agarophytes significant proportions are sulfated. When L6S-G units are consecutively arranged, only porphyranases can bind, productively depolymerize the polysaccharide and liberate oligosaccharides.

Together with the previously described agarases, we describe today a characterized set of five GH16 enzymes which is the most comprehensive system of agarolytic enzymes reported to date (Figure 63). These enzymes could be used to produce new potentially interesting sulfated oligosaccharides for structure function studies. These enzymes could also be used for the characterization of agarophytic red algae. For example, an estimate of agar quality can be obtained within one hour by applying AgaD (agarose motifs), AgaB (agarose and hybrid motifs) and PorA (pure porphyran) on an algal extract, since the activity ratio directly relates to the amount of LA/L6S. Moreover this rapid estimate can be obtained without extensive oligosaccharide or polysaccharide purification or NMR analysis of the polysaccharide. This may be exploited for monitoring of the agar quality or for physiological studies in which environmental factors influencing agar composition in red algae are of interest.

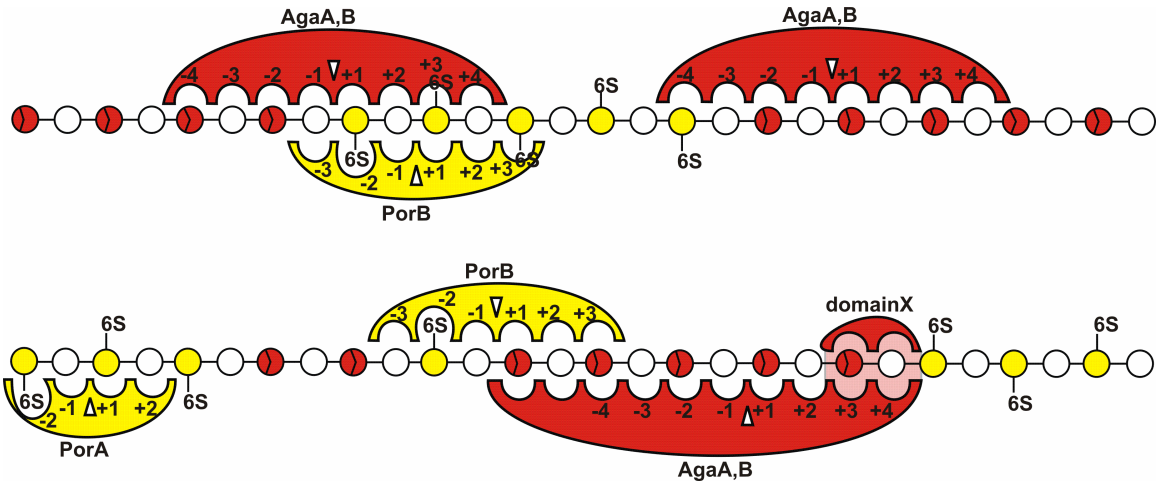


Figure 62: Model for the enzymatic degradation of complex agarocolloids by *Z. galactanivorans*.

VI.2 Screening for new marine glycoside hydrolases

In classical screening approaches when searching for glycoside hydrolase functions, the substrate is known and is used to screen for microbial activity. Today, by applying structural genomic approaches (in a way similar to the small medium throughput strategy that I applied here) on a selection of hypothetical proteins, we have convenient access to large numbers of recombinant enzymes with possible new functions. Nevertheless, the new bottleneck that appears is the functional characterization, a problem that is illustrated by the many structures deposited in the PDB without assigned function (http://www.jcsg.org/) or (http://www.pdb.org/). This highlights the need for new approaches, especially for activity assignation in marine glycoside hydrolases, since their substrates are different from terrestrial polysaccharides.

Two of the studied enzymes did not have an assigned activity (PorA and PorB). Due to the lack of commercial algal polysaccharides, therefore the enzymes were directly screened on none characterized algal extracts. By measuring released oligosaccharides with the reducing sugar assay high activity on *Porphyra* algae was found. As demonstrated here this approach was successful due to the thorough genome analysis and the rational selection of target genes, and therefore is possibly extendable to other enzymes. For instance, the enzyme Zg_3347

also showed activity on *Gracilaria* that still is waiting to be further characterized. When the reducing sugar assay is not sensitive enough, analytical methods such as fluorophore assisted oligosaccharide electrophoresis can be used to detect ng quantities of produced oligo or monosaccharide (Goubet et al. 2002). Moreover this technique is rapid and the necessary hardware is generally available in a molecular biology laboratory.

In the described approach, screening was applied only on the hot water extracted fraction from algal cell walls. It should therefore become even more powerful with advanced sequential extraction protocols, used for instance for the characterization of higher plant polysaccharides by microarrays, antibodies and CBMs (Moller et al. 2007; Obro et al. 2007). These protocols use different reagents to fractionate cell wall polysaccharides. By applying unknown enzymes on these polysaccharide fractions, one could narrow down the number of possible substrates and more easily pin down a new activity with subsequent analytical techniques such as MS, NMR, electrophoresis or HPLC.

VI.3 β-porphyranases discovery in marine bacteria

The surprising point about this novel identification of β-porphyranases is that they were not discovered much earlier since their substrate has been known for decades (at least 1961) (Turvey et al. 1961). One can argue that porphyran is only a marginally important polysaccharide compared to mass polysaccharides such as cellulose on land or chitin in the sea which implied that porphyranases were of minor interest. Nevertheless, porphyran is the worldwide most consumed natural algal dietary polysaccharide. Moreover, people have searched for porphyranases (Hatate et al. 1986) and claimed to have identified them (Aoki et al. 2006) but a porphyranase activity, such as the one described here, has never been proven before. The reason for this problem is the heterogeneous character of porphyran, which contains blocks of agarose and the amount of LA units in a consecutive polysaccharide chain can reach values up to 40%. These agarose blocks are recognized by agarases which therefore possess activity on this type of polysaccharide but they never cleave the β-1,4 linkages between two L6S-G motifs. What Aoki et al. did to find their "porphyranase" activity was

screening a functional clone library, from a marine bacterium collected from *Porphyra*, on an agar plate for hole digging clones. They found a clone which indeed digs a hole and they thereby cloned an agarase which shows activity on porphyran, but it was not a true porphyranase. That agarases cleave the non-substituted neutral blocks in porphyran has already been shown earlier (Turvey et al. 1967; Duckworth et al. 1969; Morrice et al. 1983).

The most important reason why porphyranases were not discovered earlier is the lack of commercial interest in non-exploited polysaccharides, such as porphyran. This is a general problem which obviously concerns all non biotechnologically used marine polysaccharides. This situation may hopefully change due to the current shift from land to sea in the biofuel research. For example ExxonMobil recently announced $ 600 million (US) funding for biofuel research on marine microalgae. Even though this funding is directed primarily at the production of oils from microalgae, such research opens new possibilities for applications with marine glycoside hydrolases in the bio-processing of these marine algae. This will hopefully lead to an increased interest in the marine glycoside hydrolases which could then help to tackle ecologically important questions and help fill up the "knowledge gap" covering marine polysaccharide-degrading enzymes.

VI.4 Seaweed polysaccharide degrading CAZymes in human gut bacteria

The surprising occurence of putative algal polysaccharide degrading CAZymes in human gut bacteria shows that such genes can be transmitted by horizontal gene transfer and that food associated bacteria influence genetic diversity in the gut ecosystem. The hypothesis that we propose is that seaweed associated bacteria were the original source of these new genes. Further we speculate that these genes occur in Japanese gut bacteria because the Japanese eat seaweeds on a daily basis. One main argument is that we found putative porphyranases and agarases in the genome of the bacterium *Bacteroides plebeius* but not in the genomes of gut bacteroidetes that were isolated from western individuals. We further analysed gut metagenome data and found putative porphyranase genes only in Japanese individuals and

not in americans. Nevertheless the problem with metagenome data is the limited sequencing depth in other words :"one scratches only the surface" (Bernard Henriessat pers. comm.). This means that American or European individuals may have bacteria with such genes but they can not be easily detected due to their low abundance. With the current experimental approaches proof for such a question likely will remain elusive. But with the new high throughput sequencing methods and very deep sequencing this question may be resolved. One could also argue that the genes described here, are very ancient genes which may have been common in humans. Therefore, one can not exclude the possibility that non Asian people contain such genes. Nevertheless, the genes described here, which were found in *B. plebeius*, were clearly derived from a transfer of genes as indicated by the adjacent mobility elements. The most propable explanation remains a HGT from marine to gut bacteria as a result of the consumption of seaweed.

VI.5 Marine glycoside hydrolases as tools to analyse marine POM and DOM

Important questions is there a resistant polysaccharide fraction in the sea which can act as a carbon sink and what is the exact composition of polysaccharides in the marine DOM and POM. New enzymes specific for marine polysaccharides, as well as the characterized glycoside hydrolases specific for polysaccharides from terrestrial ecosystems, may be useful to analyse POM and DOM to address such questions. The marine snow was, and is, extensively analysed for glycoside hydrolase activities from bacteria (Ziervogel et al. 2008). An inverse approach would be to use recombinant glycoside hydrolases to probe marine snow and concentrated DOM in order to understand the fine chemical structure of these important carbon pools. This may even be of broader interest, it is currently difficult to analyse their chemical composition by conventional approaches since biogenic material collected from sediment traps is too small in mass to chromatographically purify sufficient amounts for succeeding in characterization (for instance by NMR)(Hedges et al. 2001). The same problem appears concerning DOM, for which thousands of litres of seawater have to be

diafiltrated. The use of enzymes as analytical tools has two benefits, their sensitivity and their selectivity (if one uses non promiscuous well characterized enzymes). Therefore, enzymes together with new emerging technologies, glycan arrays for example, may find new applications in the characterization of marine DOM and POM.

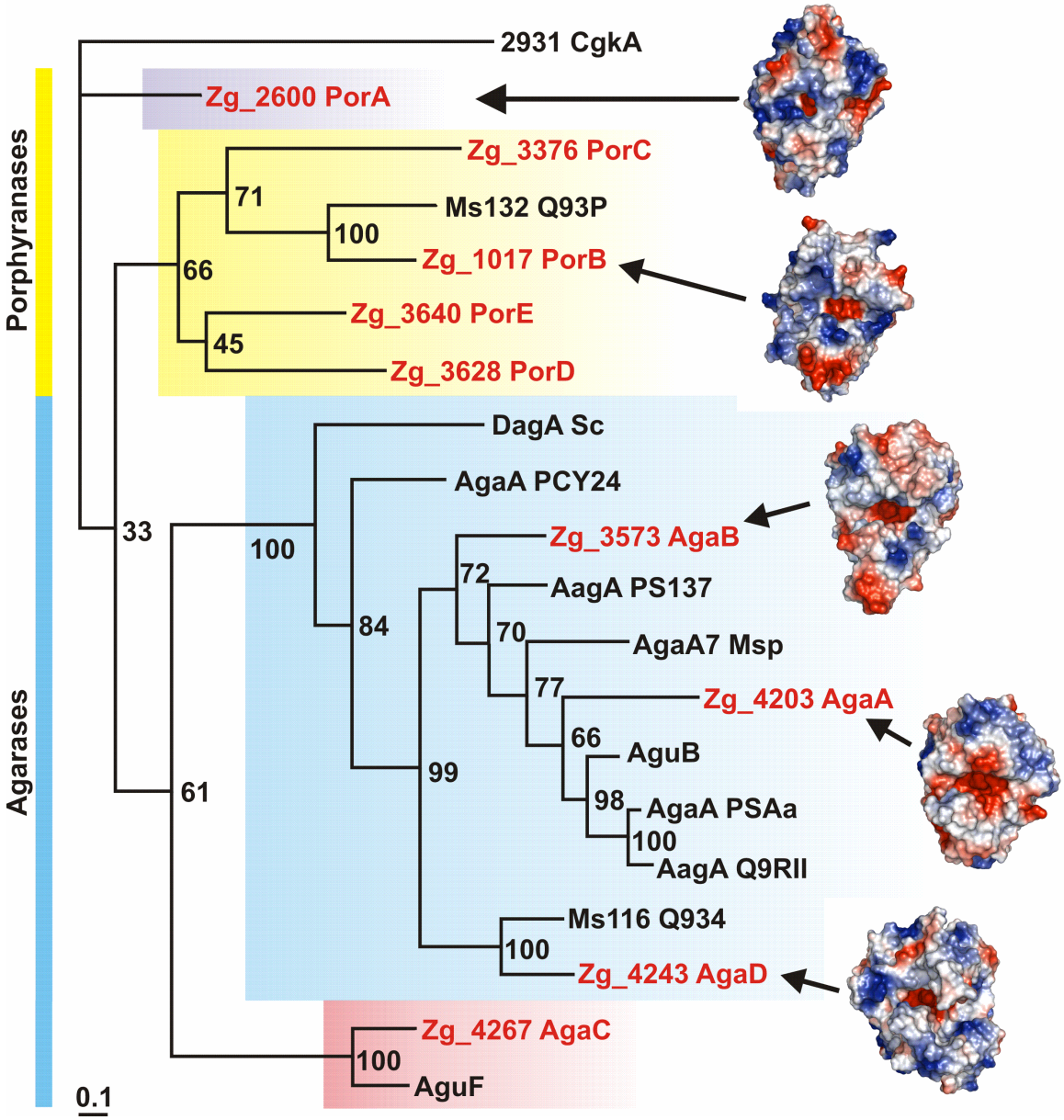


Figure 63: The GH16 agarolytic system of *Z. galactanivorans*. Sequences of the nine agarolytic GH16 enzymes from *Z. galactanivorans* (in red) are shown together with conserved agarolytic enzymes from other marine bacteria (in black). All crystal structures from agarolytic enzymes that have been solved so far are shown in an electrostatics surface representation in which the blue colour indicates basic and the red colour acidic residues. AgaA and AgaB from Allouch *et al.* 2003.

VII. References

Alban, S., N. Bourgougnon and G. Franz (1997). "Anticoagulant activity of an antiviral sulfated glucuronogalactan from *Schizymenia dubyi* (rhodophyta, gigartinales)." Thrombosis and Haemostasis: P2836-P2836.

Alderkamp, A. C., A. G. J. Buma and M. van Rijssel (2007). "The carbohydrates of *Phaeocystis* and their degradation in the microbial food web." Biogeochemistry **83**(1-3): 99-118.

Alldredge, A. L., U. Passow and B. E. Logan (1993). "The Abundance and Significance of a Class of Large, Transparent Organic Particles in the Ocean." Deep-Sea Research Part I-Oceanographic Research Papers **40**(6): 1131-1140.

Allouch, J., W. Helbert, B. Henrissat and M. Czjzek (2004). "Parallel substrate binding sites in a beta-agarase suggest a novel mode of action on double-helical agarose." Structure **12**: 623-632.

Allouch, J., M. Jam, W. Helbert, T. Barbeyron, B. Kloareg, B. Henrissat and M. Czjzek (2003). "The three-dimensional structures of two beta-agarases." Journal of Biological Chemistry **278**: 47171-47180.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Research **25**(17): 3389-3402.

Anderson, N. S. and D. A. Rees (1965). "Porphyran - a Polysaccharide with a Masked Repeating Structure." Journal of the Chemical Society(Nov): 5880-&.

Aoki, T., T. Araki and M. Kitamikado (1990). "Purification and characterization of a novel beta-agarase from *Vibrio* sp. AP-2." European Journal of Biochemistry **187**(2): 461-465.

Aoki, Y. and Y. Kamei (2006). "Preparation of recombinant polysaccharide-degrading enzymes from the marine bacterium, Pseudomonas sp ND137 for the production of protoplasts of Porphyra yezoensis." European Journal of Phycology **41**(3): 321-328.

Aquino, R. S., A. M. Landeira-Fernandez, A. P. Valente, L. R. Andrade and P. A. S. Mourao (2005). "Occurrence of sulfated galactans in marine angiosperms: evolutionary implications." Glycobiology **15**(1): 11-20.

Araki, T., M. Hayakawa, Z. Lu, S. Karita and T. Morishita (1998). "Purification and characterization of agarases from a marine bacterium, *Vibrio* sp. PO-303." Journal of Marine Biotechnology **6**(4): 260-265.

Arndt, E. R. and E. S. Stevens (1994). "A Conformational Study of Agarose by Vacuum Uv Cd." Biopolymers **34**(11): 1527-1534.

Arnott, S., A. Fulmer, W. E. Scott, I. C. Dea, R. Moorhouse and D. A. Rees (1974). "The agarose double helix and its function in agarose gel structure." Journal of Molecular Biology **90**(2): 269-284.

Azam, F. (1998). "Microbial control of oceanic carbon flux: The plot thickens." Science **280**(5364): 694-696.

Azam, F., T. Fenchel, J. G. Field, J. S. Gray, L. A. Meyerreil and F. Thingstad (1983). "The Ecological Role of Water-Column Microbes in the Sea." Marine Ecology-

Progress Series **10**(3): 257-263.
Azam, F. and F. Malfatti (2007). "Microbial structuring of marine ecosystems." Nature Reviews Microbiology **5**(10): 782-791.
Barbeyron, T., A. Gerard, P. Potin, B. Henrissat and B. Kloareg (1998). "The kappa-carrageenase of the marine bacterium *Cytophaga drobachiensis*. Structural and phylogenetic relationships within family-16 glycoside hydrolases." Molecular Biology and Evolution **15**(5): 528-537.
Barbeyron, T., S. L'Haridon, E. Corre, B. Kloareg and P. Potin (2001). "*Zobellia galactanovorans* gen. nov., sp. nov., a marine species of *Flavobacteriaceae* isolated from a red alga, and classification of." International Journal of Systematic and Evolutionary Microbiology **51**(Pt 3): 985-997.
Batey, J. F. and J. R. Turvey (1975). "Galactan Sulfate of Red Alga Polysiphonia-Lanosa." Carbohydrate Research **43**(1): 133-143.
Bauer, M., M. Kube, H. Teeling, M. Richter, T. Lombardot, E. Allers, C. A. Wurdemann, C. Quast, H. Kuhl, F. Knaust, D. Woebken, K. Bischof, M. Mussmann, J. V. Choudhuri, F. Meyer, R. Reinhardt, R. I. Amann and F. O. Glockner (2006). "Whole genome analysis of the marine *Bacteroidetes* '*Gramella forsetii*' reveals adaptations to degradation of polymeric organic matter." Environmental Microbiology **8**(12): 2201-2213.
Bayer, E. A., H. Chanzy, R. Lamed and Y. Shoham (1998). "Cellulose, cellulases and cellulosomes." Current Opinion in Structural Biology **8**(5): 548-557.
Belas, R. (1989). "Sequence analysis of the *agrA* gene encoding beta-agarase from *Pseudomonas atlantica*." Journal of Bacteriology **171**(1): 602-605.
Benner, R., J. D. Pakulski, M. Mccarthy, J. I. Hedges and P. G. Hatcher (1992). "Bulk Chemical Characteristics of Dissolved Organic-Matter in the Ocean." Science **255**(5051): 1561-1564.
Bergfors, T. (2003). "Seeds to crystals." Journal of Structural Biology **142**(1): 66-76.
Bibb, M. J., G. H. Jones, R. Joseph, M. J. Buttner and J. M. Ward (1987). "The agarase gene (dag A) of Streptomyces coelicolor A3(2): affinity purification and characterization of the cloned gene product." J Gen Microbiol **133**(8): 2089-96.
Bidle, K. D. and F. Azam (1999). "Accelerated dissolution of diatom silica by marine bacterial assemblages." Nature **397**(6719): 508-512.
Biersmith, A. and R. Benner (1998). "Carbohydrates in phytoplankton and freshly produced dissolved organic matter." Marine Chemistry **63**(1-2): 131-144.
Boraston, A. B., D. N. Bolam, H. J. Gilbert and G. J. Davies (2004). "Carbohydrate-binding modules: fine-tuning polysaccharide recognition." Biochemical Journal **382**(Pt 3): 769-781.
Buck, C. B., C. D. Thompson, J. N. Roberts, M. Muller, D. R. Lowy and J. T. Schiller (2006). "Carrageenan is a potent inhibitor of papillomavirus infection." Plos Pathogens **2**(7): 671-680.
Cantarel, B. L., P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard and B. Henrissat (2008). "The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics." Nucleic Acids Research: In the press.
Chin, W. C., M. V. Orellana and P. Verdugo (1998). "Spontaneous assembly of marine dissolved organic matter into polymer gels." Nature **391**(6667): 568-572.
Colin, S., E. Deniaud, M. Jam, V. Descamps, Y. Chevolot, N. Kervarec, J. C. Yvin, T. Barbeyron, G. Michel and B. Kloareg (2006). "Cloning and biochemical characterization of the fucanase FcnA: definition of a novel glycoside

hydrolase family specific for sulfated fucans." Glycobiology **16**(11): 1021-1032.
Collaborative Computational Project Number 4 (1994). "The CCP4 suite: programs for protein crystallography." Acta Crystallographica. Section D, Biological Crystallography **50**(Pt 5): 760-763.
Collins, B. K., S. J. Tomanicek, N. Lyamicheva, M. W. Kaiser and T. C. Mueser (2004). "A preliminary solubility screen used to improve crystallization trials: crystallization and preliminary X-ray structure determination of Aeropyrum pernix flap endonuclease-1." Acta Crystallographica Section D-Biological Crystallography **60**: 1674-1678.
Corpet, F. (1988). "Multiple sequence alignment with hierarchical clustering." Nucleic Acids Res **16**(22): 10881-90.
Cottrell, M. T. and D. L. Kirchman (2000). "Natural assemblages of marine proteobacteria and members of the Cytophaga-Flavobacter cluster consuming low- and high-molecular-weight dissolved organic matter." Applied and Environmental Microbiology **66**(4): 1692-1697.
Craigie, J. (1990). Cell Walls. Biology of the red algae. K. Cole and R. Sheath. Cambridge, Cambridge university Press**:** 221-257.
Craigie, J. and C. Leigh (1974). "carrageenans and agars." 109-133.
Cronshaw, J., A. Myers and R. D. Preston (1958). "A Chemical and Physical Investigation of the Cell Walls of Some Marine Algae." Biochimica Et Biophysica Acta **27**(1): 89-103.
Davies, G. J. and B. Henrissat (1995). "Structures and mechanisms of glycosyl hydrolases." Structure **3**(9): 853-859.
Davies, G. J., K. S. Wilson and B. Henrissat (1997). "Nomenclature for sugar-binding subsites in glycosyl hydrolases." Biochemical Journal **321**: 557-559.
Davis, T. A., F. Llanes, B. Volesky and A. Mucci (2003). "Metal selectivity of Sargassum spp. and their alginates in relation to their alpha-L-guluronic acid content and conformation." Environmental Science & Technology **37**(2): 261-267.
Davis, T. A., B. Volesky and A. Mucci (2003). "A review of the biochemistry of heavy metal biosorption by brown algae." Water Research **37**(18): 4311-4330.
Day, D. F. and W. Yaphe (1975). "Enzymatic hydrolysis of agar: purification and characterization of neoagarobiose hydrolase and p-nitrophenyl alpha-galactoside hydrolase." Canadian Journal of Microbiology **21**(10): 1512-1518.
Delong, E. F., D. G. Franks and A. L. Alldredge (1993). "Phylogenetic Diversity of Aggregate-Attached Vs Free-Living Marine Bacterial Assemblages." Limnology and Oceanography **38**(5): 924-934.
Descamps, V., S. Colin, M. Lahaye, M. Jam, C. Richard, P. Potin, T. Barbeyron, J. C. Yvin and B. Kloareg (2006). "Isolation and Culture of a Marine Bacterium Degrading the Sulfated Fucans from Marine Brown Algae." Marine Biotechnology **8**: 27-39.
Divne, C., J. Stahlberg, T. Reinikainen, L. Ruohonen, G. Pettersson, J. K. Knowles, T. T. Teeri and T. A. Jones (1994). "The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*." Science **265**(5171): 524-528.
Divne, C., J. Stahlberg, T. T. Teeri and T. A. Jones (1998). "High-resolution crystal structures reveal how a cellulose chain is bound in the 50 A long tunnel of cellobiohydrolase I from *Trichoderma reesei*." Journal of Molecular Biology **275**(2): 309-325.
Duckworth, M. and J. R. Turvey (1968). "The extracellular agarase from a *Cytophaga*

species." Biochemical Journal **109**(2): 6P.
Duckworth, M. and J. R. Turvey (1969). "The action of a bacterial agarase on agarose, porphyran and alkali-treated porphyran." Biochem. J. **113**(4): 687-692.
Duckworth, M. and J. R. Turvey (1969). "The action of a bacterial agarase on agarose, porphyran and alkali-treated porphyran." Biochemical Journal **113**(4): 687-692.
Duckworth, M. and J. R. Turvey (1969). "An extracellular agarase from a *Cytophaga* species." Biochem. J. **113**(1): 139-142.
Duckworth, M. and J. R. Turvey (1969). "The specificity of an agarase from a *Cytophaga* species." Biochem. J. **113**(4): 693-696.
Ekborg, N. A., J. M. Gonzalez, M. B. Howard, L. E. Taylor, S. W. Hutcheson and R. M. Weiner (2005). "*Saccharophagus degradans* gen. nov., sp. nov., a versatile marine degrader of complex polysaccharides." International Journal of Systematic Evolutionary Microbiology **55**(Pt 4): 1545-1549.
Ekborg, N. A., L. E. Taylor, A. G. Longmire, B. Henrissat, R. M. Weiner and S. W. Hutcheson (2006). "Genomic and proteomic analyses of the agarolytic system expressed by *Saccharophagus degradans* 2-40." Applied and Environmental Microbiology **72**(5): 3396-3405.
Elifantz, H., L. A. Waidner, V. K. Michelou, M. T. Cottrell and D. L. Kirchman (2008). "Diversity and abundance of glycosyl hydrolase family 5 in the North Atlantic Ocean." FEMS Microbiology Ecology **63**(3): 316-327.
Engel, A., S. Thoms, U. Riebesell, E. Rochelle-Newall and I. Zondervan (2004). "Polysaccharide aggregation as a potential sink of marine dissolved organic carbon." Nature **428**(6986): 929-932.
Fernandes, A. C., C. M. G. A. Fontes, H. J. Gilbert, G. P. Hazlewood, T. H. Fernandes and L. M. A. Ferreira (1999). "Homologous xylanases from Clostridium thermocellum: evidence for bi-functional activity, synergism between xylanase catalytic modules and the presence of xylan-binding domains in enzyme complexes." Biochemical Journal **342**: 105-110.
Ferreira, L. M. A., H. J. Gilbert, G. P. Hazlewood, T. H. Fernandes and C. M. G. A. Fontes (1997). "Molecular characterisation of a multi-domain cellulase from Cellvibrio mixtus." Faseb Journal **11**(9): A1036-A1036.
Field, C. B., M. J. Behrenfeld, J. T. Randerson and P. Falkowski (1998). "Primary production of the biosphere: integrating terrestrial and oceanic components." Science **281**(5374): 237-240.
Flament, D., T. Barbeyron, M. Jam, P. Potin, M. Czjzek, B. Kloareg and G. Michel (2007). "Alpha-agarases define a new family of glycoside hydrolases, distinct from beta-agarase families." Applied and Environmental Microbiology **73**(14): 4691-4694.
Ford, S. A. and E. D. T. Atkins (1989). "New X-ray diffraction results from agarose: extended single helix structures and implications for gelation mechanism." Biopolymers **28**: 1345-1365.
Frada, M., I. Probert, M. J. Allen, W. H. Wilson and C. de Vargas (2008). "The "Cheshire Cat" escape strategy of the coccolithophore Emiliania huxleyi in response to viral infection." Proceedings of the National Academy of Sciences of the United States of America **105**(41): 15944-15949.
Franz, G. and S. Alban (1996). "Structure activity relations of sulfated glucans in view of their anticoagulant effects." Abstracts of Papers of the American Chemical

Society **212**: 102-CELL.
Frei, E. and R. D. Preston (1961). "Variants in Structural Polysaccharides of Algal Cell Walls." Nature **192**(480): 939-&.
Fukuda, S., H. Saito, S. Nakaji, M. Yamada, N. Ebine, E. Tsushima, E. Oka, K. Kumeta, T. Tsukamoto and S. Tokunaga (2007). "Pattern of dietary fiber intake among the Japanese general population." European Journal of Clinical Nutrition **61**(1): 99-103.
Garcillan-Barcia, M. P., M. V. Francia and F. de la Cruz (2009). "The diversity of conjugative relaxases and its application in plasmid classification." Fems Microbiology Reviews **33**(3): 657-687.
Gattuso, J. P., M. Frankignoulle and R. Wollast (1998). "Carbon and carbonate metabolism in coastal aquatic ecosystems." Annual Review of Ecology and Systematics **29**: 405-434.
Gilbert, H. J., H. Stalbrand and H. Brumer (2008). "How the walls come crumbling down: recent structural biochemistry of plant polysaccharide degradation." Current Opinion in Plant Biology **11**(3): 338-348.
Glöckner, F. O., M. Kube, M. Bauer, H. Teeling, T. Lombardot, W. Ludwig, D. Gade, A. Beck, K. Borzym, K. Heitmann, R. Rabus, H. Schlesner, R. Amann and R. Reinhardt (2003). "Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1." Proceedings of the National Academy of Sciences of the United States of America **100**(14): 8298-8303.
Gloster, T. M., J. P. Turkenburg, J. R. Potts, B. Henrissat and G. J. Davies (2008). "Divergence of Catalytic Mechanism within a Glycosidase Family Provides Insight into Evolution of Carbohydrate Metabolism by Human Gut Flora." Chemistry & Biology **59**(10): 1058-1067.
Goubet, F., P. Jackson, M. J. Deery and P. Dupree (2002). "Polysaccharide analysis using carbohydrate gel electrophoresis: A method to study plant cell wall polysaccharides and polysaccharide hydrolases." Analytical Biochemistry **300**(1): 53-68.
Gouet, P., X. Robert and E. Courcelle (2003). "ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins." Nucleic Acids Res **31**(13): 3320-3.
Groleau, D. and W. Yaphe (1977). "Enzymatic hydrolysis of agar: purification and characterization of beta-neoagarotetraose hydrolase from *Pseudomonas atlantica*." Canadian Journal of Microbiology **23**(6): 672-679.
Groth, I., N. Grunewald and S. Alban (2009). "Pharmacological profiles of animal- and nonanimal-derived sulfated polysaccharides - comparison of unfractionated heparin, the semisynthetic glucan sulfate PS3, and the sulfated polysaccharide fraction isolated from Delesseria sanguinea." Glycobiology **19**(4): 408-417.
Guibet, M., S. Colin, T. Barbeyron, S. Genicot, B. Kloareg, G. Michel and W. Helbert (2007). "Degradation of lambda-carrageenan by *Pseudoalteromonas carrageenovora* lambda-carrageenase: a new family of glycoside hydrolases unrelated to kappa- and iota-carrageenases." Biochemical Journal **404**(1): 105-114.
Guindon, S. and O. Gascuel (2003). "A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood." Systematic Biology **52**(5): 696-704.
Ha, J. C., G. T. Kim, S. K. Kim, T. K. Oh, J. H. Yu and I. S. Kong (1997). "beta-Agarase

from *Pseudomonas* sp. W7: purification of the recombinant enzyme from *Escherichia coli* and the effects of salt on its activity." Biotechnology and Applied Biochemistry **26**: 1-6.

Hahn, M., O. Olsen, O. Politz, R. Borriss and U. Heinemann (1995). "Crystal structure and site-directed mutagenesis of *Bacillus macerans* endo-1,3-1,4-beta-glucanase." Journal of Biological Chemistry **270**(7): 3081-3088.

Hatada, Y., Y. Ohta and K. Horikoshi (2006). "Hyperproduction and application of alpha-agarase to enzymatic enhancement of antioxidant activity of porphyran." Journal of Agricultural and Food Chemistry **54**(26): 9895-9900.

Hatate, H., T. Aoki, T. Araki and M. Kitamikado (1986). "Preparation of Bacterial Enzymes Capable of Degrading the Cell-Wall of Red Alga Porphyra-Yezoensis." Bulletin of the Japanese Society of Scientific Fisheries **52**(3): 545-548.

Haug, A. and B. Larsen (1969). "Biosynthesis of Alginate . Epimerisation of D-Mannuronic to L-Guluronic Acid Residues in Polymer Chain." Biochimica Et Biophysica Acta **192**(3): 557-&.

Haug, A., B. Larsen and O. Smidsrod (1967). "Alkaline Degradation of Alginate." Acta Chemica Scandinavica **21**(10): 2859-&.

Hedges, J. I. (1992). "Global Biogeochemical Cycles - Progress and Problems." Marine Chemistry **39**(1-3): 67-93.

Hedges, J. I., J. A. Baldock, Y. Gelinas, C. Lee, M. Peterson and S. G. Wakeham (2001). "Evidence for non-selective preservation of organic matter in sinking marine particles." Nature **409**(6822): 801-804.

Hehemann, J. H., G. Correc, G. Michel, W. Helbert, T. Barbeyron and M. Czjzek (2009). Porphyranases et leur utilisation pour hydrolyser des polysaccharides. France.

Hehemann, J. H., L. Redecke, J. Murugaiyan, M. von Bergen, C. Betzel and R. Saborowski (2008). "Autoproteolytic stability of a trypsin from the marine crab Cancer pagurus." Biochemical and Biophysical Research Communications **370**(4): 566-571.

Helbert, W., P. Nyvall-Collen, G. Michel and M. Czjzek (2006). "Microscopic and molecular insights into heterogeneous phase degradation of agars and carrageenans by marine bacterial galactanases." Macromolecular Symposia **231**: 11-15.

Hemmingson, J. A., R. H. Furneaux and V. H. MurrayBrown (1996). "Biosynthesis of agar polysaccharides in Gracilaria chilensis Bird, McLachlan et Oliveira." Carbohydrate Research **287**: 101-115.

Henrissat, B. and G. Davies (1997). "Structural and sequence-based classification of glycoside hydrolases." Current Opinion in Structural Biology **7**(5): 637-44.

Henrissat, B. and G. Davies (1997). "Structural and sequence-based classification of glycoside hydrolases." Curr Opin Struct Biol **7**(5): 637-44.

Hosoda, A., M. Sakai and S. Kanazawa (2003). "Isolation and characterization of agar-degrading Paenibacillus spp. associated with the rhizosphere of spinach." Biosci Biotechnol Biochem **67**(5): 1048-55.

Inoue, N., N. Yamano, K. Sakata, K. Nagao, Y. Hama and T. Yanagita (2009). "The Sulfated Polysaccharide Porphyran Reduces Apolipoprotein B100 Secretion and Lipid Synthesis in HepG2 Cells." Bioscience Biotechnology and Biochemistry **73**(2): 447-449.

Ishihara, K., C. Oyamada, R. Matsushima, M. Murata and T. Muraoka (2005).

"Inhibitory effect of porphyran, prepared from dried "Nori", on contact hypersensitivity in mice." Bioscience Biotechnology and Biochemistry **69**(10): 1824-1830.

Ivanova, E. P., I. Y. Bakunina, O. I. Nedashkovskaya, N. M. Gorshkova, Y. V. Alexeeva, E. A. Zelepuga, T. N. Zvaygintseva, D. V. Nicolau and V. V. Mikhailov (2003). "Ecophysiological variabilities in ectohydrolytic enzyme activities of some Pseudoalteromonas species, P. citrea, P. issachenkonii, and P. nigrifaciens." Curr Microbiol **46**(1): 6-10.

Jam, M., D. Flament, J. Allouch, P. Potin, L. Thion, B. Kloareg, M. Czjzek, W. Helbert, G. Michel and T. Barbeyron (2005). "The endo-beta-agarases AgaA and AgaB from the marine bacterium *Zobellia galactanivorans*: two paralogue enzymes with different molecular organizations and catalytic behaviours." Biochemical Journal **385**: 703-713.

Kabsch, W. (1988). "Evaluation of Single-Crystal X-Ray-Diffraction Data from a Position-Sensitive Detector." Journal of Applied Crystallography **21**: 916-924.

Kang, N. Y., Y. L. Choi, Y. S. Cho, B. K. Kim, B. S. Jeon, J. Y. Cha, C. H. Kim and Y. C. Lee (2003). "Cloning, expression and characterization of a beta-agarase gene from a marine bacterium, Pseudomonas sp. SK38." Biotechnol Lett **25**(14): 1165-70.

Katoh, K., K. Misawa, K. Kuma and T. Miyata (2002). "MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform." Nucleic Acids Research **30**(14): 3059-3066.

Kazlowski, B., C. L. Pan and Y. T. Ko (2008). "Separation and quantification of neoagaro- and agaro-oligosaccharide products generated from agarose digestion by beta-agarase and HCl in liquid chromatography systems." Carbohydrate Research **343**(14): 2443-2450.

Keitel, T., O. Simon, R. Borriss and U. Heinemann (1993). "Molecular and active-site structure of a *Bacillus* 1,3-1,4-beta-glucanase." Proceedings of the National Academy of Sciences of the United States of America **90**(11): 5287-5291.

Kidby, D. K. and D. J. Davidson (1973). "A convenient ferricyanide estimation of reducing sugars in the nanomole range." Anal. Biochem. **55**(1): 321-325.

Kirchman, D. L. (2002). "The ecology of *Cytophaga-Flavobacteria* in aquatic environments." FEMS Microbiology Ecology **39**(2): 91-100.

Kitahara, M., M. Sakamoto, M. Ike, S. Sakata and Y. Benno (2005). "Bacteroides plebeius sp nov and Bacteroides coprocola sp nov., isolated from human faeces." International Journal of Systematic and Evolutionary Microbiology **55**: 2143-2147.

Kleywegt, G. J., J. Y. Zou, C. Divne, G. J. Davies, I. Sinning, J. Stahlberg, T. Reinikainen, M. Srisodsuk, T. T. Teeri and T. A. Jones (1997). "The crystal structure of the catalytic core domain of endoglucanase I from *Trichoderma reesei* at 3.6 Å resolution, and a comparison with related enzymes." Journal of Molecular Biology **272**(3): 383-397.

Kloareg, B. and R. Quatrano (1988). "Structure of the cell walls of marine algae and ecophysiological functions of the matrix polysaccharides." Oceanogr. Mar. Biol. Ann. Rev. **26**: 259-315.

Knutsen, S., D. Myslabodski, B. Larsen and A. Usov (1994). "A modified system of nomenclature for red algal galactans." Botanica Marina **37**: 163-169.

Koropatkin, N. M., E. C. Martens, J. I. Gordon and T. J. Smith (2008). "Starch catabolism by a prominent human gut symbiont is directed by the

recognition of amylose helices." Structure **16**(7): 1105-1115.
Koshland, D. E. (1953). "Stereochemistry and the mechanism of enzymatic reactions." Biol. Rev. Camb. Phylos. Soc. **28**: 416-436.
Kumar, S., K. Tamura and M. Nei (2004). "MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment." Briefings in Bioinformatics **5**(2): 150-163.
Lahaye, M. (2001). "Developments on gelling algal galactans, their structure and physico-chemistry." Journal of Applied Phycology **13**(2): 173-184.
Lahaye, M., W. Yaphe, M. T. P. Viet and C. Rochas (1989). "C-13-N-M-R Spectroscopic Investigation of Methylated and Charged Agarose Oligosaccharides and Polysaccharides." Carbohydrate Research **190**(2): 249-265.
Laine, R. A. (1994). "A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05 x 10(12) structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems." Glycobiology **4**(6): 759-67.
Leslie, A. G. W. (1992). "Recent changes to the MOSFLM package for processing film and image plate data." Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr. **26**:191.
Leslie, A. G. W. and H. R. Powell (2007). "Processing diffraction data with MOSFLM." Evolving Methods for Macromolecular Crystallography **245**: 41-51
Ley, R. E., C. A. Lozupone, M. Hamady, R. Knight and J. I. Gordon (2008). "Worlds within worlds: evolution of the vertebrate gut microbiota." Nature Reviews Microbiology **6**(10): 776-788.
Lozupone, C. A., M. Hamady, B. L. Cantarel, P. M. Coutinho, B. Henrissat, J. I. Gordon and R. Knight (2008). "The convergence of carbohydrate active gene repertoires in human gut microbes." Proceedings of the National Academy of Sciences of the United States of America **105**(39): 15076-15081.
Maciel, J. S., L. S. Chaves, B. W. S. Souza, D. I. A. Teixeira, A. L. P. Freitas, J. P. A. Feitosa and R. C. M. de Paula (2008). "Structural characterization of cold extracted fraction of soluble sulfated polysaccharide from red seaweed Gracilaria birdiae." Carbohydrate Polymers **71**(4): 559-565.
Mc Hugh, D. J. (2003). A guide to seaweed industry. FAO Fisheries Technical Paper No 441. FAO. Rome, Italy, FAO.
McCarthy, M., J. Hedges and R. Benner (1996). "Major biochemical composition of dissolved high molecular weight organic matter in seawater." Marine Chemistry **55**(3-4): 281-297.
Michel, G., L. Chantalat, E. Duee, T. Barbeyron, B. Henrissat, B. Kloareg and O. Dideberg (2001). "The kappa-carrageenase of *P. carrageenovora* features a tunnel-shaped active site: a novel insight in the evolution of Clan-B glycoside hydrolases." Structure **9**(6): 513-525.
Michel, G., L. Chantalat, E. Fanchon, B. Henrissat, B. Kloareg and O. Dideberg (2001). "The iota-carrageenase of *Alteromonas fortis*. A beta-helix fold-containing enzyme for the degradation of a highly polyanionic polysaccharide." Journal of Biological Chemistry **276**(43): 40202-40209.
Michel, G., P. Nyval-Collen, T. Barbeyron, M. Czjzek and W. Helbert (2006). "Bioconversion of red seaweed galactans: a focus on bacterial agarases and carrageenases." Applied Microbiology and Biotechnology **71**(1): 23-33.
Moller, I., I. Sorensen, A. J. Bernal, C. Blaukopf, K. Lee, J. Obro, F. Pettolino, A. Roberts,

J. D. Mikkelsen, J. P. Knox, A. Bacic and W. G. T. Willats (2007). "High-throughput mapping of cell-wall polymers within and between plants using novel microarrays." Plant Journal **50**(6): 1118-1128.

Morch, Y. A., S. Holtan, I. Donati, B. L. Strand and G. Skjak-Braek (2008). "Mechanical properties of C-5 epimerized alginates." Biomacromolecules **9**(9): 2360-2368.

Morrice, L. M., M. W. McLean, W. F. Long and F. B. Williamson (1983). "Beta-agarases I and II from *Pseudomonas atlantica*. Substrate specificities." European Journal of Biochemistry **137**(1-2): 149-154.

Morrice, L. M., M. W. Mclean, W. F. Long and F. B. Williamson (1983). "Porphyran Primary Structure - an Investigation Using Beta-Agarase-I from Pseudomonas-Atlantica and C-13-Nmr Spectroscopy." European Journal of Biochemistry **133**(3): 673-684.

Morrice, L. M., M. W. McLean, W. F. Long and F. B. Williamson (1983). "Porphyran primary structure. An investigation using beta-agarase I from Pseudomonas atlantica and 13C-NMR spectroscopy." Eur J Biochem **133**(3): 673-84.

Morrice, L. M., M. W. McLean, W. F. Long and F. B. Williamson (1983). "Porphyran primary structure. An investigation using beta-agarase I from Pseudomonas atlantica and 13C-NMR spectroscopy." European Journal of Biochemistry **133**(3): 673-84.

Morrice, L. M., M. W. Mclean, W. F. Long and F. B. Williamson (1984). "Porphyran Primary Structure." Hydrobiologia **116**(Sep): 572-575.

Morrice, L. M., M. W. McLean, F. B. Williamson and W. F. Long (1983). "beta-agarases I and II from *Pseudomonas atlantica*. Purifications and some properties." European Journal of Biochemistry **135**(3): 553-558.

Murano, E. (1995). "Chemical-Structure and Quality of Agars from Gracilaria." Journal of Applied Phycology **7**(3): 245-254.

Murshudov, G. N., A. A. Vagin and E. J. Dodson (1997). "Refinement of macromolecular structures by the maximum-likelihood method." Acta Crystallographica Section D-Biological Crystallography **53**: 240-255.

Myklestad, S. (1977). "Production of Carbohydrates by Marine Planktonic Diatoms .2. Influence of N-P Ratio in Growth Medium on Assimilation Ratio, Growth-Rate, and Production of Cellular and Extracellular Carbohydrates by Chaetoceros-Affinis Var Willei (Gran) Hustedt and Skeletonema-Costatum (Grev) Cleve." Journal of Experimental Marine Biology and Ecology **29**(2): 161-179.

Myklestad, S. M. (1995). "Release of Extracellular Products by Phytoplankton with Special Emphasis on Polysaccharides." Science of the Total Environment **165**(1-3): 155-164.

Nakai, K. and M. Kanehisa (1991). "Expert system for predicting protein localization sites in gram-negative bacteria." Proteins **11**(2): 95-110.

Nakamura, Y., T. Itoh, H. Matsuda and T. Gojobori (2004). "Biased biological functions of horizontally transferred genes in prokaryotic genomes." Nature Genetics **36**(7): 760-766.

Navaza, J. (2001). "Implementation of molecular replacement in AMoRe." Acta Crystallographica. Section D, Biological Crystallography **57**(Pt 10): 1367-1372.

Nisizawa, K., H. Noda, R. Kikuchi and T. Watanabe (1987). "The Main Seaweed Foods in Japan." Hydrobiologia **151**: 5-29.

Obro, J., T. Sorensen, I. Moller, M. Skjot, J. D. Mikkelsen and W. G. T. Willats (2007). "High-throughput microarray analysis of pectic polymers by enzymatic epitope deletion." Carbohydrate Polymers **70**(1): 77-81.

Ohta, Y., Y. Hatada, M. Miyazaki, Y. Nogi, S. Ito and K. Horikoshi (2005). "Purification and characterization of a novel alpha-agarase from a *Thalassomonas* sp." Current Microbiology **50**(4): 212-216.

Ohta, Y., Y. Hatada, Y. Nogi, Z. Li, S. Ito and K. Horikoshi (2004). "Cloning, expression, and characterization of a glycoside hydrolase family 86 beta-agarase from a deep-sea *Microbulbifer*-like isolate." Applied Microbiology and Biotechnology **66**(3): 266-275.

Pape, T. and T. R. Schneider (2004). "HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs." Journal of Applied Crystallography **37**: 843-844.

Passow, U. (2002). "Production of transparent exopolymer particles (TEP) by phyto- and bacterioplankton." Marine Ecology-Progress Series **236**: 1-12.

Paulsen, B. S. and S. Myklestad (1977). "Structural Studies of Reserve Polysaccharides in Diatoms." Journal of Phycology **13**: 52-52.

Percival, E. (1979). "Polysaccharides of Green, Red and Brown Seaweeds - Their Basic Structure, Biosynthesis and Function." British Phycological Journal **14**(2): 103-117.

Percival, E. and M. Young (1971). "Characterisation of Sucrose Lactate and Other Oligosaccharides Found in Cladophorales." Carbohydrate Research **20**(2): 217-&.

Perrakis, A., R. Morris and V. S. Lamzin (1999). "Automated protein model building combined with iterative structure refinement." Nature Structural Biology **6**(5): 458-463.

Pomin, V. H. and P. A. S. Mourao (2008). "Structure, biology, evolution, and medical importance of sulfated fucans and galactans." Glycobiology **18**(12): 1016-1027.

Potin, P., C. Richard, C. Rochas and B. Kloareg (1993). "Purification and characterization of the alpha-agarase from *Alteromonas agarlyticus* (Cataldi) comb. nov., strain GJ1B." European Journal of Biochemistry **214**(2): 599-607.

Potin, P., A. Sanseau, Y. Le Gall, C. Rochas and B. Kloareg (1991). "Purification and characterization of a new kappa-carrageenase from a marine *Cytophaga*-like bacterium." European Journal of Biochemistry **201**(1): 241-247.

Potterton, E., P. Briggs, M. Turkenburg and E. Dodson (2003). "A graphical user interface to the CCP4 program suite." Acta Crystallographica Section D-Biological Crystallography **59**: 1131-1137.

Potterton, L., S. McNicholas, E. Krissinel, J. Gruber, K. Cowtan, P. Emsley, G. N. Murshudov, S. Cohen, A. Perrakis and M. Noble (2004). "Developments in the CCP4 molecular-graphics project." Acta Crystallographica Section D-Biological Crystallography **60**: 2288-2294.

Rees, D. (1969). "Structure, conformation, and mechanism in the formation of polysaccharide gels and networks." Adv. Carbohydr. Chem. Biochem. **24**: 267-332.

Rees, D. A. (1961). "Enzymic synthesis of 3:6-anhydro-l-galactose within porphyran from l-galactose 6-sulphate units." Biochemical Journal **81**(2): 347-52.

Rees, D. A. and E. Conway (1962). "Structure and Biosynthesis of Porphyran - a Comparison of Some Samples." Biochemical Journal **84**(2): 411-&.

Rees, D. A. and E. Conway (1962). "The structure and biosynthesis of porphyran: a comparison of some samples." Biochemical Journal **84**: 411-6.

Rees, D. A., E. R. Morris, D. Thom and J. Madden (1982). The polysaccharides. G. O.

Aspinall. New York, USA, Academic Press. **1:** 195-290.
Reis, C. P., R. J. Neufeld, S. Vilela, A. J. Ribeiro and F. Veiga (2006). "Review and current status of emulsion/dispersion technology using an internal gelation process for the design of alginate particles." Journal of Microencapsulation **23**(3): 245-257.
Rochas, C., M. Lahaye and W. Yaphe (1986). "Sulfate Content of Carrageenan and Agar Determined by Infrared-Spectroscopy." Botanica Marina **29**(4): 335-340.
Sartoni, G., R. Urbani, P. Sist, D. Berto, C. Nuccio and M. Giani (2008). "Benthic mucilaginous aggregates in the Mediterranean Sea: Origin, chemical composition and polysaccharide characterization." Marine Chemistry **111**(3-4): 184-198.
Schneider, T. R. and G. M. Sheldrick (2002). "Substructure solution with SHELXD." Acta Crystallographica Section D-Biological Crystallography **58**: 1772-1779.
Schroeder, D. C., M. A. Jaffer and V. E. Coyne (2003). "Investigation of the role of a beta(1-4) agarase produced by *Pseudoalteromonas gracilis* B9 in eliciting disease symptoms in the red alga *Gracilaria gracilis*." Microbiology **149**(Pt 10): 2919-2929.
Shi, Y. M., G. W. Tyson and E. F. DeLong (2009). "Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column." Nature **459**(7244): 266-U154.
Shieh, W. Y. and W. D. Jean (1998). "*Alterococcus agarolyticus*, gen.nov., sp.nov., a halophilic thermophilic bacterium capable of agar degradation." Canadian Journal of Microbiology **44**(7): 637-645.
Sinnott, M. L. (1990). "Catalytic mechanisms of glycosyl transfer." Chemical Reviews **90**: 1171-1202.
Smith, D. C., M. Simon, A. L. Alldredge and F. Azam (1992). "Intense Hydrolytic Enzyme-Activity on Marine Aggregates and Implications for Rapid Particle Dissolution." Nature **359**(6391): 139-142.
Smith, S. V. (1981). "Marine Macrophytes as a Global Carbon Sink." Science **211**(4484): 838-840.
Sonnenburg, J. L., J. Xu, D. D. Leip, C. H. Chen, B. P. Westover, J. Weatherford, J. D. Buhler and J. I. Gordon (2005). "Glycan foraging in vivo by an intestine-adapted bacterial symbiont." Science **307**(5717): 1955-1959.
Starr, C. M., R. I. Masada, C. Hague, E. Skop and J. C. Klock (1996). "Fluorophore-assisted carbohydrate electrophoresis in the separation, analysis, and sequencing of carbohydrates." Journal of Chromatography A **720**(1-2): 295-321.
Stocker, R., J. R. Seymour, A. Samadani, D. E. Hunt and M. F. Polz (2008). "Rapid chemotactic response enables marine bacteria to exploit ephemeral microscale nutrient patches." Proceedings of the National Academy of Sciences of the United States of America **105**(11): 4209-4214.
Studier, F. W. (2005). "Protein production by auto-induction in high density shaking cultures." Protein Expression and Purification **41**(1): 207-234.
Sugano, Y., H. Kodama, I. Terada, Y. Yamazaki and M. Noma (1994). "Purification and characterization of a novel enzyme, alpha-agarooligosaccharide hydrolase (alpha-NAOS hydrolase), from a marine bacterium, Vibrio sp. strain JT0107." Journal of Bacteriology **176**: 6812-6818.
Sugano, Y., T. Matsumoto, H. Kodama and M. Noma (1993). "Cloning and sequencing of agaA, a unique agarase 0107 gene from a marine bacterium,

Vibrio sp. strain JT0107." Applied and Environmental Microbiology **59**(11): 3750-3756.

Sugano, Y., T. Matsumoto and M. Noma (1994). "Sequence analysis of the *agaB* gene encoding a new beta-agarase from *Vibrio* sp. strain JT0107." Biochimica Biophysica Acta **1218**(1): 105-108.

Sugano, Y., I. Terada, M. Arita, M. Noma and T. Matsumoto (1993). "Purification and characterization of a new agarase from a marine bacterium, *Vibrio* sp. strain JT0107." Applied and Environmental Microbiology **59**(5): 1549-1554.

Suttle, C. (2005). "The viriosphere: the greatest biological diversity on Earth and driver of global processes." Environmental Microbiology **7**(4): 481-482.

Swartz, M. N. and N. Gordon (1959). "Agarase from an Agar-Digesting Bacterium." Journal of Bacteriology **77**(4): 403-409.

Tempel, W., Z. J. Liu, P. S. Horanyi, L. Deng, D. Lee, M. G. Newton, J. P. Rose, H. Ashida, S. C. Li, Y. T. Li and B. C. Wang (2005). "Three-dimensional structure of GlcNAcalpha1-4Gal releasing endo-beta-galactosidase from *Clostridium perfringens*." Proteins - Structure Function and Bioinformatics **59**(1): 141-144.

Turnbaugh, P. J., M. Hamady, T. Yatsunenko, B. L. Cantarel, A. Duncan, R. E. Ley, M. L. Sogin, W. J. Jones, B. A. Roe, J. P. Affourtit, M. Egholm, B. Henrissat, A. C. Heath, R. Knight and J. I. Gordon (2009). "A core gut microbiome in obese and lean twins." Nature **457**(7228): 480-U7.

Turvey, J. R. and Christis.J (1967). "Enzymic Degradation of Porphyran." Biochemical Journal **105**(1): 317-&.

Turvey, J. R. and J. Christison (1967). "The hydrolysis of algal galactans by enzymes from a *Cytophaga* species." Biochemical Journal **105**: 311-321.

Turvey, J. R. and D. A. Rees (1961). "Isolation of L-Galactose-6-Sulphate from a Seaweed Polysaccharide." Nature **189**(476): 831-&.

Turvey, J. R. and E. L. Williams (1976). "Agar-Type Polysaccharide from Red Alga Ceramium-Rubrum." Carbohydrate Research **49**(Jul): 419-425.

Umani, S. F., P. Del Negro, C. Larato, C. De Vittor, M. Cabrini, M. Celio, C. Falconi, F. Tamberlich and F. Azam (2007). "Major inter-annual variations in microbial dynamics in the Gulf of Trieste (northern Adriatic Sea) and their ecosystem implications." Aquatic Microbial Ecology **46**(2): 163-175.

Usov, A., S. Yarotsky and A. Shashkov (1980). "^{13}C-NMR spectroscopy of red algal galactans." Biopolymers **19**: 977-990.

Usov, A. I. and L. I. Miroshnikova (1975). "Isolation of agarase from *Littorina mandshurica* by affinity chromatography on Biogel A." Carbohydrate Research **43**(1): 204-207.

Van der Meulen, H. J. and W. Harder (1975). "Production and characterization of the agarase of *Cytoplaga flevensis*." Antonie Van Leeuwenhoek **41**(4): 431-47.

Vattuone, M. A., E. A. de Flores and A. R. Sampietro (1975). "Isolation of neoagarobiose and neoagarotetraose from agarose digested by *Pseudomonas elongata*." Carbohydrate Research **39**(1): 164-167.

Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Y. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y. H. Rogers and H. O. Smith (2004). "Environmental genome shotgun sequencing of the Sargasso Sea." Science **304**(5667): 66-74.

Vera, J., R. Alvarez, E. Murano, J. C. Slebe and O. Leon (1998). "Identification of a marine agarolytic pseudoalteromonas isolate and characterization of its

extracellular agarase." Applied and Environmental Microbiology **64**(11): 4378-4383.

Verdugo, P., A. L. Alldredge, F. Azam, D. L. Kirchman, U. Passow and P. H. Santschi (2004). "The oceanic gel phase: a bridge in the DOM-POM continuum." Marine Chemistry **92**(1-4): 67-85.

Vocadlo, D. J. and G. J. Davies (2008). "Mechanistic insights into glycosidase chemistry." Current Opinion in Chemical Biology **12**(5): 539-555.

Voget, S., C. Leggewie, A. Uesbeck, C. Raasch, K. E. Jaeger and W. R. Streit (2003). "Prospecting for novel biocatalysts in a soil metagenome." Applied and Environmental Microbiology **69**(10): 6235-6242.

von Heijne, G. (1983). "Patterns of amino acids near signal-sequence cleavage sites." Eur J Biochem **133**(1): 17-21.

Weiner, R. M., L. E. Taylor, B. Henrissat, L. Hauser, M. Land, P. M. Coutinho, C. Rancurel, E. H. Saunders, A. G. Longmire, H. T. Zhang, E. A. Bayer, H. J. Gilbert, F. Larimer, I. B. Zhulin, N. A. Ekborg, R. Lamed, P. M. Richardson, I. Borovok and S. Hutcheson (2008). "Complete genome sequence of the complex carbohydrate-degrading marine bacterium, Saccharophagus degradans strain 2-40(T)." Plos Genetics **4**(5): -.

Weiss, M. S., U. Abele, J. Weckesser, W. Welte, E. Schiltz and G. E. Schulz (1991). "Molecular Architecture and Electrostatic Properties of a Bacterial Porin." Science **254**(5038): 1627-1630.

Withers, S. G. (2001). "Mechanisms of glycosyl transferases and hydrolases." Carbohydrate Polymers **44**(4): 325-337.

Wotton, R. S. (2004). "The ubiquity and many roles of exopolymers (EPS) in aquatic systems." Scientia Marina **68**: 13-21.

Xu, J., M. K. Bjursell, J. Himrod, S. Deng, L. K. Carmichael, H. C. Chiang, L. V. Hooper and J. I. Gordon (2003). "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis." Science **299**(5615): 2074-2076.

Xu, J., M. A. Mahowald, R. E. Ley, C. A. Lozupone, M. Hamady, E. C. Martens, B. Henrissat, P. M. Coutinho, P. Minx, P. Latreille, H. Cordum, A. Van Brunt, K. Kim, R. S. Fulton, L. A. Fulton, S. W. Clifton, R. K. Wilson, R. D. Knight and J. I. Gordon (2007). "Evolution of symbiotic bacteria in the distal human intestine." Plos Biology **5**(7): 1574-1586.

Zhang, Q. B., N. Li, X. G. Liu, Z. Q. Zhao, Z. Li and Z. H. Xu (2004). "The structure of a sulfated galactan from Porphyra haitanensis and its in vivo antioxidant activity." Carbohydrate Research **339**(1): 105-+.

Zhang, Q. B., H. M. Qi, T. T. Zhao, E. Deslandes, N. M. Ismaeli, F. Molloy and A. T. Critchley (2005). "Chemical characteristics of a polysaccharide from Porphyra capensis (Rhodophyta)." Carbohydrate Research **340**(15): 2447-2450.

Zhong, Z., A. Toukdarian, D. Helinski, V. Knauf, S. Sykes, J. E. Wilkinson, C. O'Bryne, T. Shea, C. DeLoughery and R. Caspi (2001). "Sequence analysis of a 101-kilobase plasmid required for agar degradation by a *Microscilla* isolate." Applied and Environmental Microbiology **67**(12): 5771-5779.

Ziervogel, K. and C. Arnosti (2008). "Polysaccharide hydrolysis in aggregates and free enzyme activity in aggregate-free seawater from the north-eastern Gulf of Mexico." Environmental Microbiology **10**(2): 289-299.