Insight into the Functionality of an Unusual Glycoside Hydrolase from Family 50

by

Kaleigh Giles
BSc, Brock University, 2011

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Biochemistry and Microbiology

Supervisory Committee

Insight into the Functionality of an Unusual Glycoside Hydrolase from Family 50

by

Kaleigh Giles
BSc, Brock University, 2011

Supervisory Committee

Dr. Alisdair B. Boraston, Department of Biochemistry and Microbiology
Supervisor

Dr. Martin J. Boulanger (Department of Biochemistry and Microbiology)
Departmental Member

Dr. Fraser Hof (Department of Chemistry)
Outside Member

Abstract

Supervisory Committee
Dr. Alisdair B. Boraston, Department of Biochemistry and Microbiology
Supervisor
Dr. Martin J. Boulanger, Department of Biochemistry and Microbiology
Departmental Member
Dr. Fraser Hof, Department of Chemistry
Outside Member

Agarose and porphyran are related galactans that are only found within red marine algae. As such, marine microorganisms have adapted to using these polysaccharides as carbon sources through the acquisition of unique Carbohydrate Active enZymes (CAZymes). A recent metagenome study of the microbiomes from a Japanese human population identified putative CAZymes in several bacterial species, including *Bacteroides plebeius* that have significant amino acid sequence similarity with those from marine bacteria. Analysis of one potential CAZyme from *B. plebeius* (*Bp*GH50) is described here. While displaying up to 30% sequence identity with β-agarases, *Bp*GH50 has no detectable agarase activity. Its crystal structure reveals that the topology of the active site is much different than previously characterized agarases, while containing the same core catalytic machinery. It is unclear whether the enzyme has *endo*- or *exo*-activity; the large binding 'groove' is typical of an *endo*-acting enzyme, while a loop at one end of the groove may provide a terminal pocket for the substrate, which is suggestive of *exo*-activity. Furthermore, the enzyme contains a basic pocket that may dock a sulphated substrate, like porphyran. While no quantifiable porphyran activity was observed, properties of the putative active site suggest that this unusual enzyme may be specific on an unusual substrate, such as a porphyran-agarose hybrid.

Table of Contents

List of Tables

List of Figures

Acknowledgments

First and foremost, I would like to thank my supervisor Dr. Alisdair Boraston for the incredible opportunity to work in his lab. His encouragement and guidance throughout my time here has been invaluable.

I also want to thank my committee members, Drs Martin Boulanger and Fraser Hof for their support and feedback on my progress.

I would also like to thank all the members of the Boraston lab for their assistance and comradery throughout these past two years. Many thanks go to Craig Robb and Dr. Benjamin Pluvinage for their guidance and help troubleshooting my experiments.

Thank you to Michelle Lee, Alex Fillo, Andrew Hettle and Crystal Pinto for your mutual support in the lab, during writing and through the trial by fire of BCMB 500.

Best of luck to you all in completing your own research goals!

Dedication

I would like to dedicate this work to my long time boyfriend, Jonathan Horgan, who has been there through the darker times and helped me to see the brighter side of things.

1.0 Introduction

1.1 General Polysaccharides

Polysaccharides are intrinsic to life on Earth. They are used for energy storage, plant cell walls, bacterial capsules, biofilms, and mammalian cell identification[37].Their wide spread use and diverse functionality make them the most abundant type of organic material on Earth. The key to their diversity is their ability to form multiple linkages with the same or different sugar types to form increasingly complex polymer chains. Just six sugar monomers can be organized in 1.05×10^{12} linear and branched forms (Laine, 1994; Thomas *et al*, 2011). This diversity in structure permits the development of unique three dimensional structures, which in turn bestow dynamic functional properties that are useful in countless roles and environments.

1.2 Plant Polysaccharides

Plant polysaccharides can comprise a significant amount of a plant's total biomass (Ishihara *et al*, 2005). In terrestrial plants, they provide two main functions: as an energy reserve and for structural support.

Starches constitute the bulk of energy reserve polysaccharides found in plants (Imberty *et al*, 1988). Starch is composed of two main components, the linear amylose and the branched amylopectin. Amylose is composed of repeated α- 1,4 D-glucose units (Hsien-Chih and Sarko 1978A and1978B). These linkages cause the polymer to bend around an axis to form a double helix structure with six monomers per turn, or a single helix structure if sequestering hydrophobic fatty acids or aromatic molecules (Cohen *et al*, 2008). This helical structure makes amylose much more compact than its counterpart amylopectin, which is the linear amylose with branches of α-1,6 linked D-glucose. The

branches interfere with the helix formation, giving it a more amorphous structure and thus rendering it more susceptible to enzymatic digestion.

Structural polysaccharides provide support against inter and intra osmotic pressure differences within plant cell walls, and in large plants also serve a load bearing function (Voet and Voet, 2004). Cellulose, the predominate component of these cell walls, is also linear polymer of D-glucose; however this time they are linked by β-1,4 glycosidic linkages (Voet and Voet, 2004). Individual chains of cellulose associate in sheets, and the crosslinking of these sheets together by hydrogen binding make it form a rigid polymer matrix. In large plants, the complexity of this matrix is augmented with lignin, a rigid phenolic polymer, which crosslinks cellulose to neighbouring hemicelluloses and pectins. Hemicelluloses are branched β-1,4 linked hexosyl sugars, including xyloglucan, glucomannan and galactomannan, while pectins contain 1,4 linked α-D-galacturonic acids (Voet and Voet, 2004). Since they are both composed of D-glucose, amylose and cellulose demonstrate the significance of orientation of the glycosyl linkage on structure.

1.3 Marine polysaccharides

While oceanic plants, such as macroalgae (seaweeds) contain some cellulose, the major structural polysaccharides include components not found in terrestrial plants (Percival, 1979; Hehemann *et al*, 2010 and 2012A). These include the use of L-sugars, 3,6-anhydro cycling of sugars and sulfated modifications as seen in agarose and carrageenans. Proportions of these unique sugars vary by genus, species and occasionally by cell type (Percival, 1979). These elements permit seaweeds to adapt to the constant movement of the ocean environment (such as tides and waves) through a more fluid support system by forming gels, in contrast to the rigid makeup of terrestrial plants, and sequester water to

buffer against dehydration when washed ashore (Percival, 1979; Hehemann *et al*, 2012A).

1.3.1 Macroalgae environmental impact and applications

While both terrestrial and aquatic plants contribute to the carbon cycle through the photosynthetic fixation of carbon dioxide (CO_2) from the atmosphere, as much as 50% of all carbon fixation occurs in the ocean (Azam and Malfatti, 2007). The turnover of this abundant plant biomass is facilitated by microorganisms, which utilize unique Carbohydrate Active enZymes (CAZymes) to break down and liberate this energy source. Additionally, the degradation of macroalgae facilitates the formation of particulate and dissolved organic matter, which then sinks, creating marine snow. Marine snow provides carbon, nitrogen and other key nutrients to heterobacteria within lower oceanic zones, and is later mineralized and sequestered on the ocean floor as inorganic carbon (Azam and Malfatti, 2007). The degradation pathways of plant polysaccharides, particularly marine polysaccharides, need to be better elucidated given the impacts of man-made augmentation of atmospheric CO_2 levels on the carbon cycle. Understanding these pathways will also facilitate the exploitation of macroalgae as a feedstock for the production of biofuels such as ethanol and butanol (Correc *et al*, 2011; Yun *et al* 2011). Seaweed is an attractive raw material since it does not require pesticides or fertilizers to grow quickly compared to seasonal feedstocks such as corn or soy (Gupta *et al*, 2013). It also does not require any land for farming so it would not divert land from food production.

While Green algae (*Chlorophyta*) cell walls are quite similar to terrestrial plants containing sulfated varieties of hemicelluloses and pectins, the low lignin content of

brown (*Phaeophyceae*) and red (*Rhodophyta*) macroalgae make them the better candiates for potential biofuel feed stocks (McCandless and Craigie, 1979; Percival 1979).

The majority of structural sugars from red macroalgae species are derived from galactans (Percival 1979, Michel *et al*, 2006), which allows them to have a reduced cell wall complexity compared to other algae. Red seaweed is a particularly attractive feedstock given its large carbohydrate content, which in *Porphyra* species, for example, approaches 50% of the total dried mass (MacArtain *et al*, 2007). The large sugar content coupled with the low wall complexity maximizes the potential yield with fewest CAZymes added, streamlining their degradation into fermentable sugar monomers by reducing the relative number of CAZymes required (Hehemann *et al*, 2012A; Gupta, *et al*, 2013; Kim *et al*, 2013).

Red algae galactan varieties can include some or all of the unique sugar elements mentioned previously. The three most relevant to this research are agarose, porphyran and carrageenan. These sugars all have a similar repeating heterodimer unit whose monomers are linked with alternating β-1,4 and α-1,3 linkages (Fu and Kim, 2010).

While there is no universally accepted nomenclature for these sugars, one approach proposed by Knutsen and colleagues in 1994, organizes these sugars into four main groups based on their repeating heterogalactan unit and the presence or absence of L-sugars and 3,6-anhydro rings, as noted in Figure 1(Knutsen *et al*, 1994; Chi *et al* 2012).

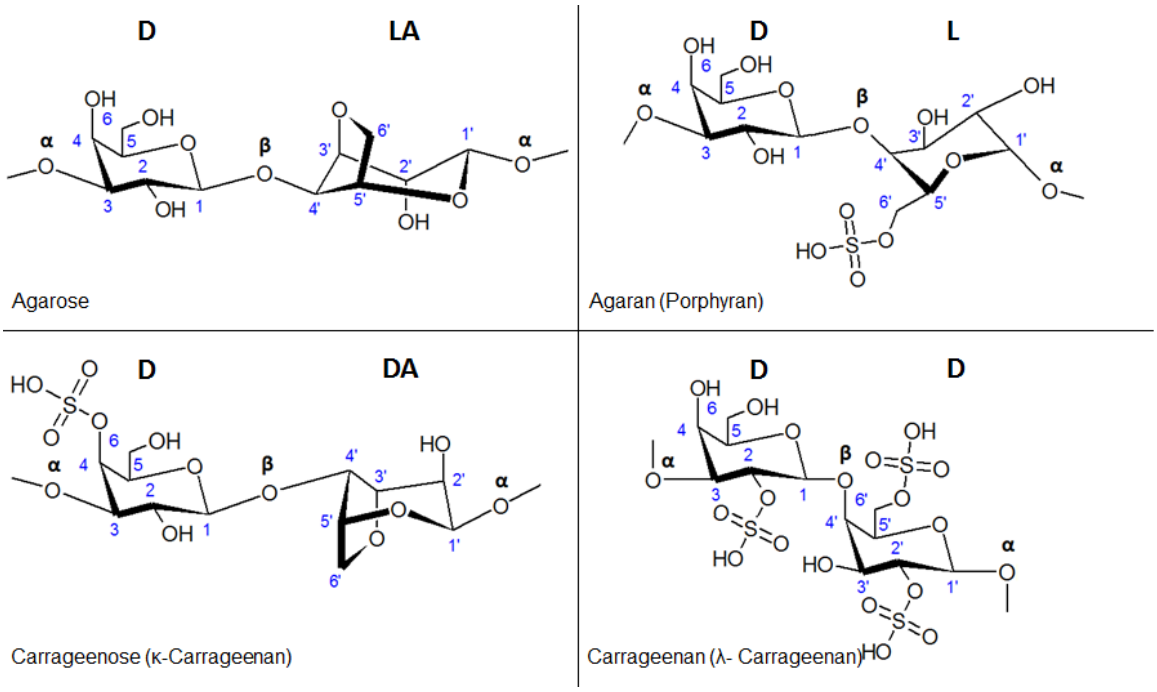


Figure 1: The organization of *Rhodophyta* structural galactans as proposed by Knutsen *et al* (1994). The classification is centered on the handedness of each of galactose (D or L) in the repeating dimer unit, and whether the dimers contain an 3,6-anhydro group (DA or LA). Examples of the four galactan types are observed here, their common names are in brackets if different from their classification. The numbers (N and N', in blue) constitute the numbered carbons of each hexamer. α and β depict the type of bond between the individual sugars and their repeating dimer units.

1.3.2 Agarose

Agar is the general name for galactans extracted from genera such as *Gelidium* and *Gracilaria* (Hehemann *et al*, 2010 and 2012A). Agarose is the predominant neutral fraction of agar; with the sulfated heterogeneous agaropectin impurities removed, it is used in research for agar plates and DNA gels, and in the food industry as a thickening/gelling agent used in marshmallows, processed cheese and icings (Fu and Kim, 2010; Correc *et al*, 2011). Agarose has a repeating unit of β-1,4-D- galactose and α-

1,3 linked 3,6-anhydro-L-galactose (LA, see Figure 2). While not a branched sugar, the agarose polymer will intertwine with a second strand to form a parallel double helix structure similar to amylose. These helices will in turn associate with other helices, forming a complex quaternary structure, manifesting as a gel (Pervical, 1979).

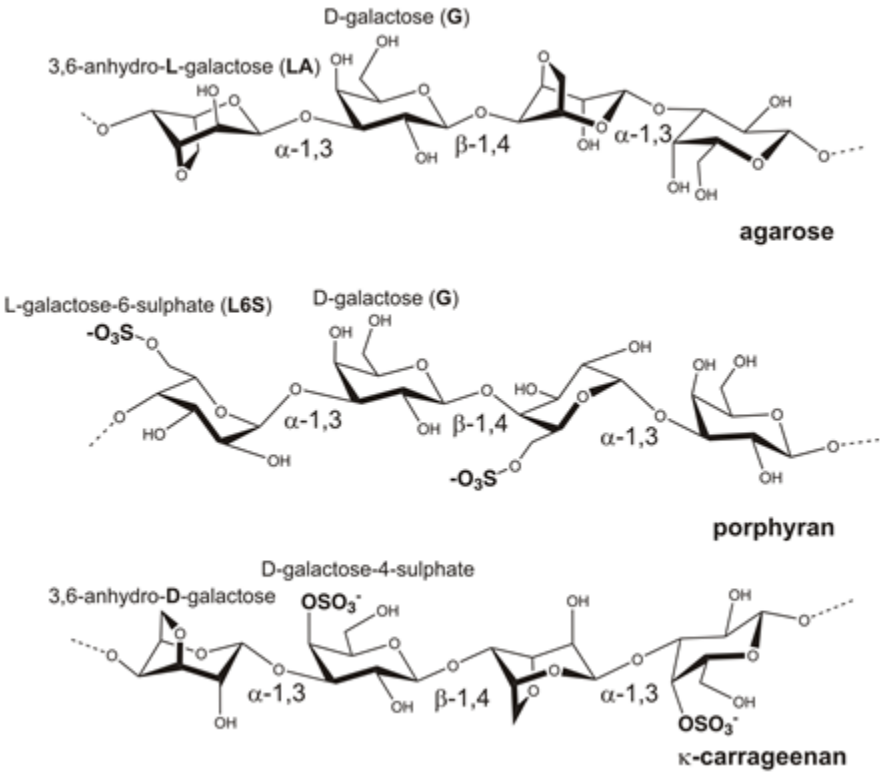


Figure 2: The general structure of agarose, porphyran and κ-carrageenan. [Adapted from Hehemann *et al* (2012B) © National Academy of Sciences of the United States of America]

1.3.3 Porphyran

Porphyran comprises approximately 40% of the total mass of seaweeds from the genus *Porphyra* (Hehemann *et al*, 2010 and 2012A), known commonly in Japan as nori. Nori is one of the most popular seaweeds in Japan and has been used for hundreds of years to make maki-sushi (Thomas *et al*, 2011; Ishihara *et al*, 2005). This makes porphyran one of the most widely consumed algal polysaccharides in Asian countries (Correc *et al*, 2011).

Porphryan is an agaran, so while it has the same D-Gal – L-Gal repeating unit as agarose, and is connected with the same α-1,3 and β- 1,4 linkages, there is a L-galactose-6 –sulfate instead of the 3,6-anhydro-L galactose (Hehemann *et al*, 2010). Furthermore, the repeating D-L unit is occasionally masked by the methylation of the D-Galactose (at the 6-OH position) (Anderson and Rees, 1965). Porphyran is the precursor sugar to agarose, and can be converted into a methylated agarose using galactose-6-sulfurylase, or through hot alkaline treatment (Correc *et al*, 2011; Rees, 1961). The presence of the sulfate group interrupts the tight interactions between helices, preventing solid gel formation, instead forming a viscous solution (Percival, 1979; Allouch *et al*, 2004). While pure porphyran does not have significant gelling properties, native porphyran contains interspersed agarose units (up to 30% depending on the species, Figure 3) to provide a gelling capacity that is inversely proportional to the relative amount of sulfation (Percival, 1979; Knutsen, *et al*, 1994; Correc *et al*, 2011).

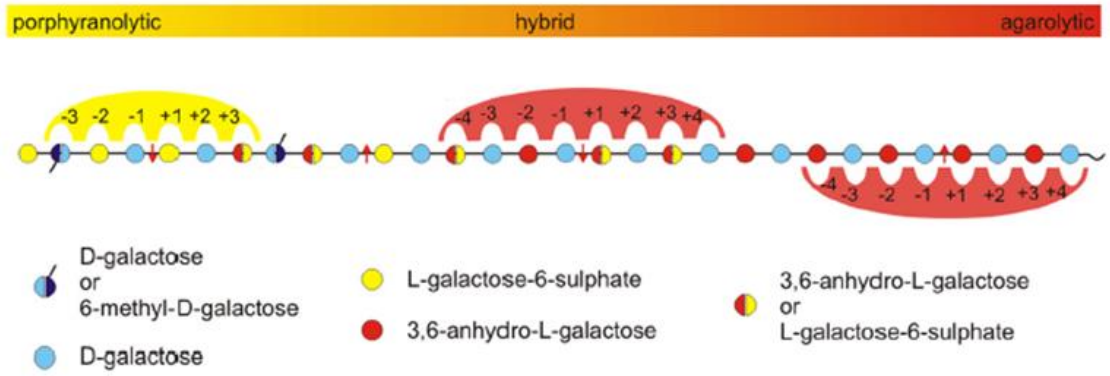


Figure 3: Hybridization of native porphyran. While the basic organization of the repeating porphyran unit is similar to agarose, it can be masked by occasional methylation of the D-Galactose. Unlike agarose polymer chains that tend to be homogeneous, native porphyran polymers can contain up to 30% agarose units, either in blocks (agarolytic porphyran) or alternating with porphyran units (hybrid porphyran), With this increased potential for complexity, the diversity of enzymes required to fully degrade native porphyran also

increases. [Originally published in JBC, by Hehemann *et al* (2012C) © the American Society for Biochemistry and Molecular Biology]

1.3.4 Carrageenan

Carrageenan and carrageenose are extracted from select red seaweed species, such as *Chondrus crispus* (Prajapati *et al*, 2014). The main unit for carrageenans or carrageenose is very similar to agarose, save for the use of only D-galactose sugars. Additionally, like porphyran, carrageenans can contain various sulfated linkages. The various types of carrageenan correspond to the different sulfated groups at C2, 5 and/or 6, resulting in a sulfate content of 22-38% by weight in commercially produced carrageenan (Prajapati *et al*, 2014; Van de Velde *et al*, 2002). The three most common types of carrageenan are iota, kappa and lambda. In accordance with the previous nomenclature, kappa and lambda would be considered carrageenose varieties because of their 3,6-anhydro-D-Galactose moieties (Knutsen *et al* 1994).

1.4 CAZymes

Given the plethora of complex polysaccharide varieties and their different roles, the four classes of CAZymes that are involved in their metabolism need to be just as specialized. For this reason, many different CAZymes are required for complete carbohydrate metabolism. As such, their genes constitute between 1-3% of the genomes of most organisms (Thomas *et al*, 2011).

Glycosyl transferases (GTs) are required for the synthesis of oligo- and polysaccharides through the formation of glycosidic bonds. This is done by the transfer of a sugar moiety from an activated sugar donor, to a sugar or non-sugar receptor (Coutinho, *et al*, 2003).

Glycoside hydrolases (GHs), polysaccharide lyases (PLs) and carbohydrate esterases (CEs) facilitate the degradation of the sugar polymers. CEs are involved in the deacylation of O- or N-linked saccharides (Lombard *et al*, 2014). While both GHs and PLs cleave O-linked glycosidic bonds, GHs will utilize hydrolysis to generate two alcohols, whereas PLs will use β- elimination to form an unsaturated product (Lombard *et al*, 2014; Abbott *et al*, 2010).

1.4.1 Glycoside Hydrolases (GHs)

Glycoside hydrolases are the most studied and most abundant of the CAZyme classes, being present within all three major kingdoms (Archaea, Prokaryota and Eukaryota) (Henrissat, 1991; Davies and Henrissat, 1995). They are utilized wherever polysaccharides are degraded, whether in soil, in the ocean or within animal gastrointestinal environments (Lombard *et al*, 2014).

GHs were originally organized by their substrate specificity, through the IUB Enzyme Nomenclature system (EC 3.2.1.x) developed in 1984. The first three numbers specified glycosyl linkages while the last number designates substrate and sometimes the molecular mechanism (Henrissat, 1991). This classification system helped to avoid the designation of trivial or ambiguous names to different GHs. However, it did not take into account the structural similarities (or lack thereof) between proteins with the same substrate designation.

As more GH sequences became available, it became apparent that sequence identity (and therefore structure) was a more efficient tool for GH organization. The current system of GH families was first implemented by Henrissat and his colleagues in 1991, when it became apparent that some protein sequences were more similar to proteins with

different substrates (Henrissat, 1991; Davies and Henrissat, 1995). Grouping the most similar sequences in potentially polyspecific families helps to suggest evolutionary divergence between proteins with different but structurally similar substrates. There are currently over 186 000 ORFs within 133 GH families catalogued on the online CAZyme database (www.cazy.org)(Lombard *et al*, 2014). Despite their lack of protein sequence similarity, these families are also organized into 14 larger 'clans' based on fold similarity. Folds from seven of the most populous clans are displayed in Figure 4, and include the β-helix, the β -jelly roll, 5 and 6 fold β-propellers and the $(\alpha / \beta)_8$ or TIM barrel motif (Lombard *et al*, 2014;Henrissat and Bairoch, 1996; Henrissat and Davies, 1997).

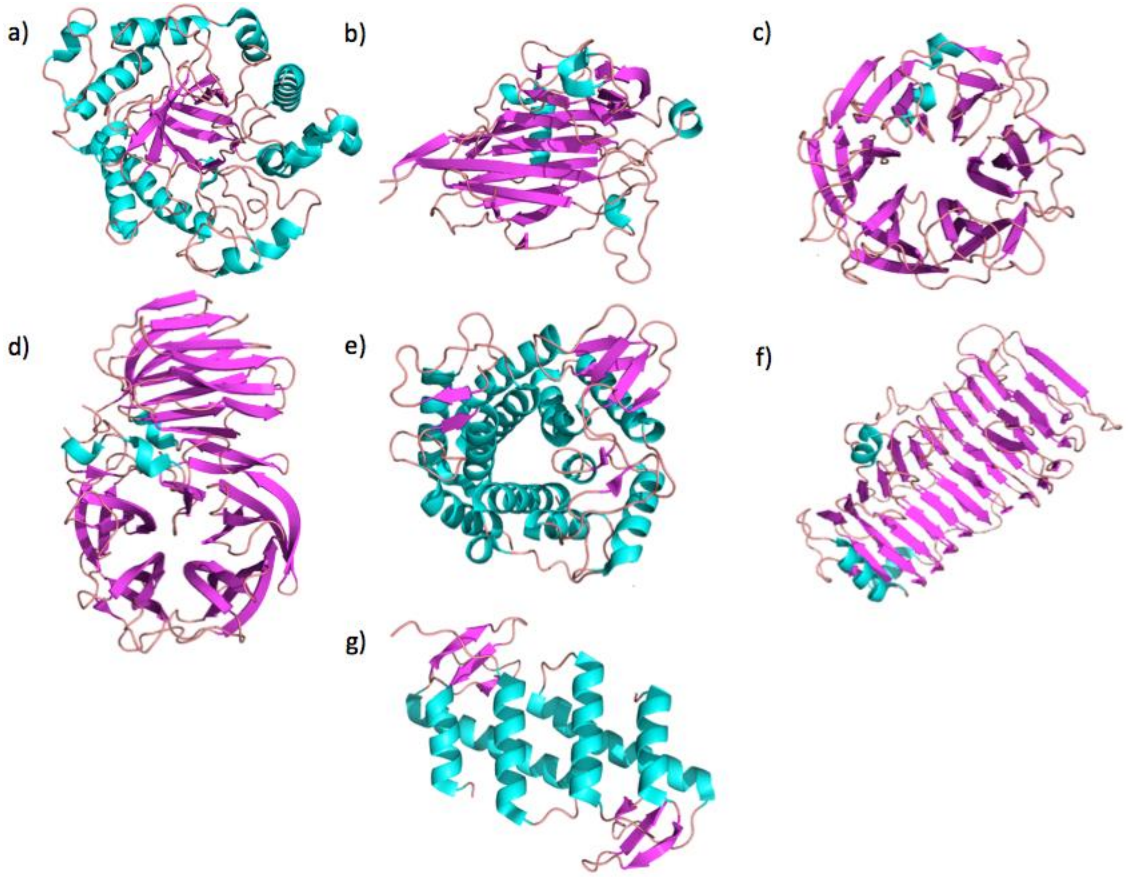


Figure 4: Main folds of glycoside hydrolases. All folds are displayed with β strands in magenta, and α helices in cyan. a) $(\alpha/\beta)_8$ barrel (or TIM barrel)motif from *Pyrococcus horikoshii* endocellulase (GH5) (pdb ID: 3QHO). b) β-jelly roll motif from *Bacteroides plebeius* β-porphyranase (GH16) (pdb ID: 4AWD). c) 6-fold β-propeller from *Peniciliium chrysogenum* α-L-arabinanase (GH93) (pdb ID: 3A71). d) 5 fold β-propeller from *Thermotoga maritima* β-fructofuranosidase (GH32) (pdb ID:1UYP). e) $(\alpha/\alpha)_6$ barrel from *Acetobacter xylinum endo*-β-1,4 glucanase (GH8) (pdb ID:1WZZ). f) β-helix from *Aspergillus aculeatus* rhamnogalacturonase (GH28) (pdb ID:1RMG). g) α+β motif from *Samonella enterica LT2* phage endolysin (GH46) (pdb ID:4EVX).

There are two main mechanisms for glycoside hydrolases that will result in either a one-step inversion or a two-step retention of stereochemistry at the anomeric carbon, both using an oxocarbenium transition state (Figure 5) (Davies *et al*, 1998). The active sites of inverting and retaining enzymes can appear very similar; both have two catalytic residues

on either side of the substrate binding pocket. Thus the main difference between the two is the distances between the residues. The average distance between catalytic residues in inverting enzymes tends to be much larger (9.0-9.5Å) than between those of retaining enzymes (4.8-5.3Å) (McCarter *et al*, 1994). The greater separation is thought to be due to the positioning of both water and substrate in between the residue, which is not an issue with the retaining mechanism (McCarter *et al*, 1994).

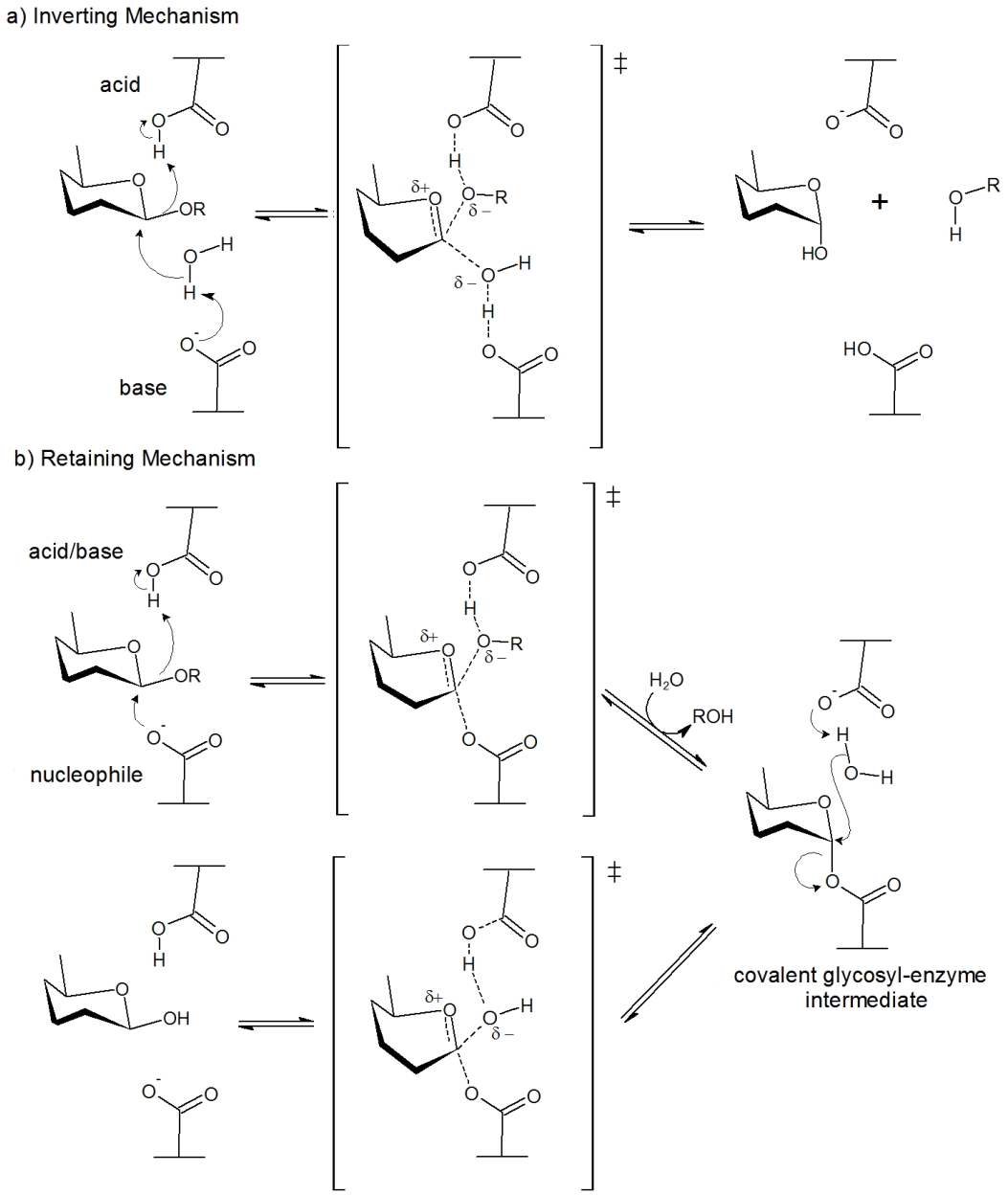


Figure 5: The inverting and retaining mechanisms of β-glycoside hydrolases. 5a) The one step inverting mechanism, the two carboxylates catalyze the direct substitution of the leaving group by water in a concerted fashion. 5b) Conversely, the retaining mechanism sees conservation of the original stereochemistry at the anomeric carbon. This is performed over a two step process through the formation of a covalent glycosyl-enzyme intermediate. Water then attacks the intermediate at the same place the original leaving group, thus retaining stereochemistry.

The active site topology of GHs also contributes to their functional capabilities. *Endo*-acting enzymes that cleave in the middle of long polysaccharide chains will tend to have large open active sites, such as a cleft or groove. This allows for protein interactions with multiple linked sugar monomers (Davies and Henrissat, 1995). Conversely, *exo*-acting enzymes will act at the ends of a polysaccharide chain, and tend to have smaller or more closed off active sites, such as a pocket or tunnel, which only allows a terminal chain end to enter (Davies and Henrissat, 1995).

There are also enzymes that appear to show both characteristics of *endo* and *exo* enzymes by their range of products. These enzymes exhibit processivity (Davies and Henrissat, 1995), where a product is cleaved, but the remainder of the substrate remains bound to the substrate, allowing for multiple actions on one long chain (Davies and Henrissat, 1995).

The topology of the active site will ultimately determine the number of sugars that will interact with the enzyme. The residues in the active site are organized to best interact with each sugar of the substrate, such that they can sufficiently bind their substrate and yet discriminate between other similar sugar chains the enzyme may encounter. From the identification of the catalytic residues, the scissile bond of the substrate can also be determined (Figure 6), which can predict the sizes of the enzyme products [by the number of sites of the glycone or non-reducing end (-n) and the aglycone or reducing (+n) ends] (Davies *et al*, 1997).

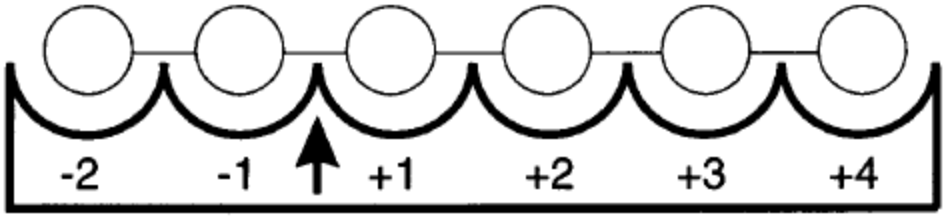


Figure 6: Visualization of CAZyme active site subsites with a hexasaccharide substrate. The negative values depict the non-reducing end of the substrate (or glycone), while the positive values depict the reducing end of the substrate (or aglycone), while the scissile bond is displayed with the arrow. Adapted from a figure originally published in Biochemical Journal by Davies *et al* (1997) © the Biochemical Society.

1.4.2 Agar acting CAZymes

CAZymes that degrade agar are found in 6 families and at least 3 clans thus far. Alpha (α-)agarases (EC 3.2.1.158) identified in families GH96 and GH117 hydrolyse α -1,3 linked sugars while β-agarases (EC 3.2.1.81) within GH16, GH50, GH86 and GH118 hydrolyse the β-1,4 linkages (Correc *et al*, 2012).

Due to their recent discovery in 2011, very few α-agarases have been characterized. Little information is currently known about GH96, but GH117 has structures for three family members as well as a proposed inverting mechanism (Hehemann *et al*, 2012A; Rebuffet, 2011). The characterized GH117 enzymes are all specific for agarose, and are all 1,3-α-3,6-anhydro-L-galactosidases, meaning they remove a single 3,6-anhydro-L-galactose monomer from α- neoagarobiose or longer substrate chains (Hehemann *et al*, 2012A; Rebuffet, 2011).

Compared to α-agarases, β–agarases are much better studied. β-agarases are quite diverse, not limited to a particular fold or even mechanism. The largest family, GH16 has

a conserved β-jelly roll fold, while the smaller GH50 and GH86 families utilize a TIM barrel fold. However, all three families employ a retaining hydrolysis mechanism, while the least studied family, GH118, is thought to use the inverting mechanism. (Lombard *et al*, 2014).

GH50 is a moderately sized family of 178 members, 21 of which have been characterized. Through recombinant protein The agarases in this family have been isolated from various marine bacteria such as *Pseudoalteromonas, Alteromonas, Agarivorans* and *Vibrio* species but also within the soil *(Streptomyces coelicolor)* (Lombard *et al*, 2014). All characterized GH50 proteins are β-agarases, and these tend to be *exo*-acting (producing neoagarobiose) but some can also produce larger products (neoagaro-hexaose and neoagaro-octaose) more typical of *endo*-acting enzymes (Sugano *et al*, 1993).

One such pseudo *endo*-acting enzyme Aga50D, from marine γ-proteobacterium *Saccharophagus degradans 2-40,* recently became the first GH50 structure published in 2013 (Pluvinage 2013). The protein contains a TIM barrel fold which contains the catalytic residues. These two Glu residues are sequestered within an active site tunnel (~25-30Å across), which allows for proposed processivity, even on gelled agarose. The main barrel is also fused to an auxiliary β- sandwich domain at the opening of the active site, which facilitates the binding of substrates with more than two aglycone subsites (more than +2). It is postulated that this auxillary domain may have originally been a distinct carbohydrate binding module (CBM) (Boraston *et al*, 2004; Pluvinage *et al*, 2013).

Because of the fundamental similarities between agarose and porphyran, it is no surprise that β-porphyranases have been found within the same GH families as agarases. While there are currently no β- porphyranases in the GH50 family, one has been characterized in GH86 and five have been characterized in GH16 (Lombard *et al*, 2014). The numerous porphyranases and agarases within GH16 allowed a comparison between the two groups to determine the structural differences. While several porphyranases and agarases can bind and degrade a hybrid substrate (one consisting of linked porphyran and agarose units), most are specific to their particular substrate and cannot accommodate 'other' units (an agarose unit in a porphyranase or a porphyran unit in an agarase) in subsites closest to the scissile bond. PorB, a GH16 *endo*-β-agarase from *Zobellia galactivorans* has been the only enzyme thus far that is promiscuous in the +1 and +2 subsites (Correc *et al*, 2011).

1.4.3 Marine Sugar Enzyme Pathways

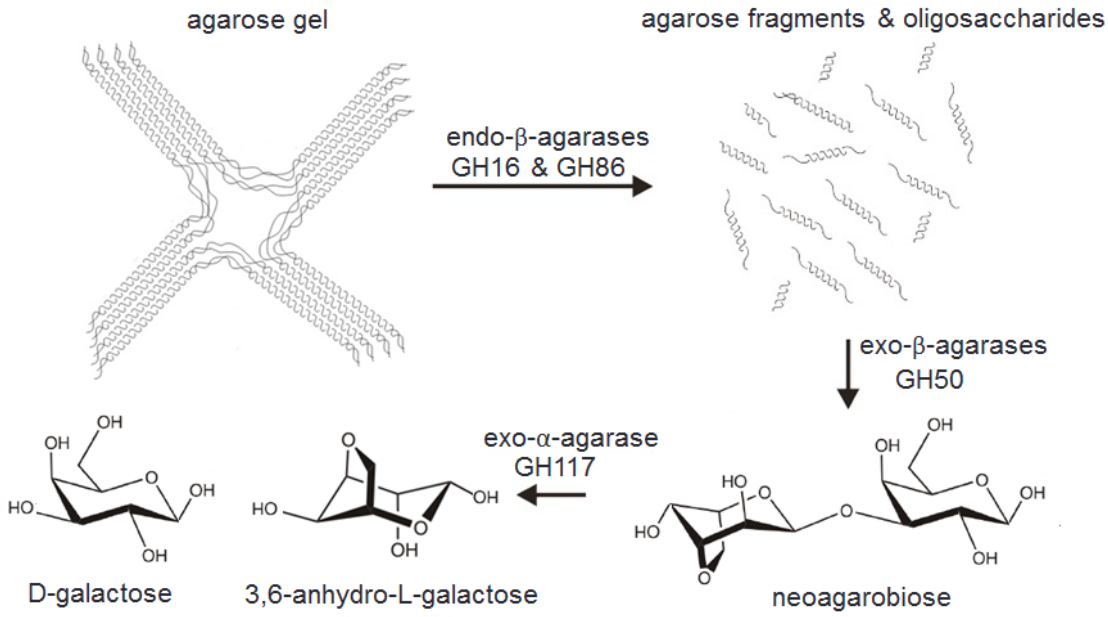


Figure 7: Visualization of the agarolytic degradation cascade of *Saccharophagus degradans* 2-40. [Adapted from a figure published in JBC, by Pluvinage *et al* (2013) © the American Society for Biochemistry and Molecular Biology]

In order to provide optimal conditions for industrial seaweed sugar degradation, we must understand how the heterotrophic bacteria that consume seaweed degrade these sugars. Because of the multiple bond types and variety of sugars within agarose and other agarans, it would be impossible for a single enzyme to completely degrade these complex chains into simpler monomer units. In order to have agarose or other agarans as a viable feedstock for biofuels, we need to first understand the roles of the multiple enzymes required for this complete degradation.

Agarose is the best studied marine polysaccharide, and a fair deal is known about the enzymatic cascades involved in agarose degradation in several systems, including *Saccharophagus degradans 2-40, Zobellia galactanivorans* and *Streptomyces coelicolorA3(2)*(Hehemann *et al*, 2010; Chi *et al*, 2012; Correc *et al*, 2011; Pluvinage *et*

al, 2013). Agarolytic pathways can be broken down into three main steps (Figure 5). In the first step *endo*-β-agarases degrade gelled agarose into oligosaccharides of varying lengths. Next, *exo*-β-agarases degrade these oligomers into neoagarobiose. This biose is finally cleaved by the α-L-galactosidase (all pathways thus far have used a GH117) in order to complete hydrolysis. *Endo*-β-agarases tend to be from GH16 or GH86, and *exo*-β-agarases tend to be from GH50 and GH16, while either α- or β-galactosidases have been from GH117 or GH2 (Pluvinage *et al*, 2013).

Native porphyran comprising of both porphyran and agarose blocks, is more complex than agarose, thus porphyran degradation cascades require additional enzymes to yield a complete breakdown to monomers (Michel *et al*, 2006; Chi *et al*, 2012). Our understanding of porphyran degradation lags behind that of agarose because, until recently, our research has been limited by the lack of porphyran specific enzymes. First discovered in 2010, a handful of β-porphyranases have now been characterized in *Z .galactivorans* and *B. plebeius*, and fittingly they reside in the same GH families as β-agarases (GH16 and GH86) (Correc *et al*, 2011; Hehemann *et al*, 2010 and 2012B). The analysis with *Z. galactivorans* porphyranases have demonstrated how porphyranases and sulfatases create products that can be funneled into the main agarolytic degradation system with subtle diversity in particular subsites compared to agarases in the same family. However, since so few porphyran degrading systems have been analyzed, all the enzyme players are not yet known.

Because of the localization of agarose and porphyran within *Rhodophyta* cell walls, the proliferation of their degradation pathways is largely limited to oceanic bacteria. However, recently a collection of agar and porphyran degrading loci have been

discovered with the gut microbiota of people of Japanese descent (Hehemann *et al*, 2010 and 2012B).

1.5 Human Gut Microbiota

The human body is limited in its ability to digest many of the things that we eat; our genome only codes for 97 CAZymes, and only 17 of these are thought to be related to digestion (Kaoutari *et al*, 2013). As such, the bulk of the catabolism of 'indigestible' starches, pectins, and other sugars falls to the human microbiome, particularly, the gut microbiota.

The human gut is home to hundreds of trillions of bacteria, which provides their host with up to 10% of their daily calories (Kau *et al* 2011, El Kaoutari *et al*, 2013) as well as additional vitamins and fatty acids, through the digestion and fermentation of 'indigestible' sugars in the host's diet. This mutualism provides additional nutrients to the host and promotes overall colon health (Thomas *et al*, 2011).

The complete physiological impact of the microbiota on human health is only beginning to be realized. The recent genomic sequencing of over 177 individual gut bacteria species is now making it possible to study their contributions of to human nutrition, as well as their roles in the development of chronic inflammatory diseases, such as Crohn's disease, and irritable bowel syndrome, as well as obesity and diabetes (Contemo *et al*, 2011; El Kaoutari *et al*, 2013). The microbiota may also protect the colon from pathogenic species, by stimulating immune responses through the activation of T-cells (Kau *et al*, 2011; Thomas *et al*, 2011).

The healthy human gut contains many diverse species; however, the majority (>90%) of these are from two main phyla, *Firmicutes* and *Bacteroidetes* (Sonnenburg *et al*, 2010; El Kaoutari *et al*, 2013). *Bacteroides* species are particularly adept at degrading polysaccharides because they use compact genetic cassettes known as Polysaccharide Utilization Loci (PULs). These PULs contain the genes necessary for the detection, transport, and degradation of specific polysaccharides (Hehemann *et al*, 2012B). These systems are not limited to the gut microbiota, and are present throughout *Bacteroides*, having been found in fresh water, marine and soil *Bacteroides* species as well (Martens *et al*, 2009). The development or maintenance of these PULs is ultimately determined by the continued abundance of their substrate.

1.5.1 Seaweed CAZymes within the Gut

Recently, a gut *Bacteroides* strain, *Bacteroides plebeius* DSM 17135 was identified containing a PUL that degraded red seaweed galactans (Hehemann *et al*, 2010). Many of the CAZymes localized within this PUL share significant sequence similarity with those found in marine bacteria (Hehemann *et al*, 2012B). Similar marine derived CAZymes have been found in the metagenomic analyses of Japanese (and Spanish) individuals, but are absent in the microbiota of North Americans (Thomas *et al*, 2011). Since seaweeds are a staple in Japanese cuisine (daily consumption around 14g /person/ day (Thomas *et al*, 2011), it is hypothesized the traditional non-sterile preparation of these seaweeds, when consumed, permitted contact between gut and marine bacteria. The gut microbiota then acquired these PULs via horizontal gene transfer (HGT) from the marine bacteria, and they were conserved because their genes provided access to a unique carbon source.

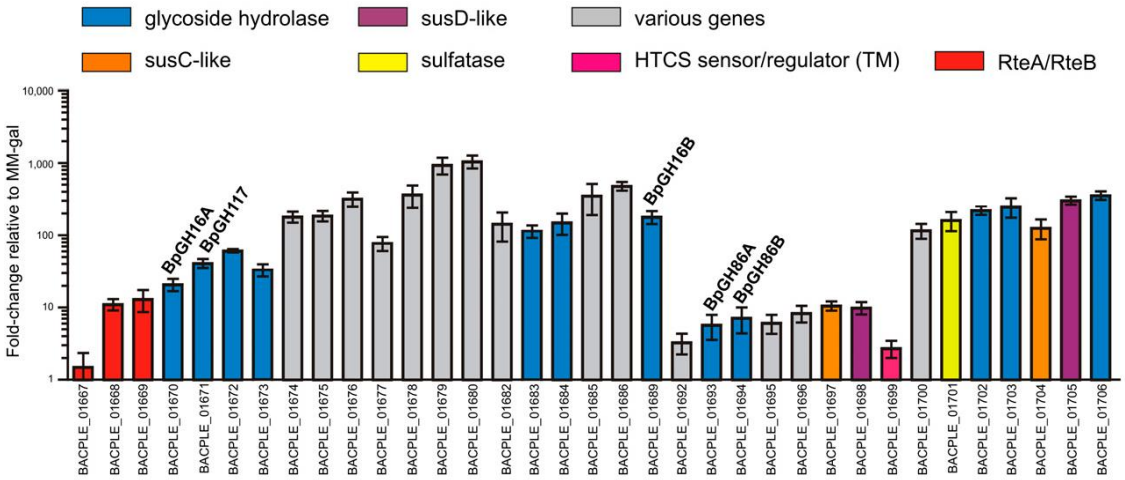


Figure 8: RT- PCR analysis demonstrating the upregulation of the PUL of *Bacteroides plebeius* DSM 17135 grown in the presence of porphyran (compared to D-Galactose). The enzymes that have been characterized have their names above their respective bar. (Adapted from Hehemann *et al*, (2012B), © National Academy of Sciences of the United States of America)

The *B. plebeius* PUL contains 40 genes that are all upregulated in the presence of porphyran and are required for *B. plebeius* porphyranase activity (Hehemann *et al*, 2012B). Eleven of these genes are conserved in marine *Bacteroides* species, and are interspersed with others conserved within gut *Bacteroides* (Hehemann *et al*, 2010 and 2012B). Our lab has characterized several of the genes within this PUL, including the *endo* β-agarase *Bp*GH16A, an α-L-galactosidase *Bp*GH117, two *endo* β-porphyranases (*Bp*GH86A, *Bp*GH16B), predicted β-galactosidases (*Bp*GH2A and B) and a sulfatase (Hehemann *et al*, 2010, 2012A and 2012B). Overall their respective enzymes constitute the main contributors of the *B.plebeius* porphyran enzymatic cascade. However, there are still select genes that have unknown functions.

1.6 Project Overview

*Bp*_01683 is a putative GH family 50 agarase, which has been previously shown to be upregulated in the presence of porphyran (Figure 8), along with nearly all enzymes from the *B.plebeius* PUL. Enzymes characterized from this PUL have demonstrated activity on either porphyran or agarose, which is consistent with the hybrid nature of native porphyran. These enzymes appear to work in concert in a porphyran degradation pathway; however, this pathway contains anticipated steps that have yet to be attributed to a particular enzyme.

The objective of this research is to elucidate the function of *Bp*_01683, in order to bring us closer to understanding the complete *B. plebeius* porphyran degradation pathway, the first from a gut derived bacteria. Given that other agarases have been previous characterized within this PUL, and that nearly all characterized GH50 proteins have been active on the β linkages of agarose, our hypothesis is that this protein is also a β-agarase (Lombard, 2014). Protein functionality will be determined by a two pronged approach. First, several biochemical assays (thin layer chromatography, reducing sugar assays, agar plate assay) will be performed to assess potential substrate binding and activity. Secondly, x-ray crystallography will be used to solve the protein structure and provide structural justification of the substrate analysis. Additionally, providing a second structure for the GH50 family will develop the working knowledge of this family and give an alternative GH50 model for other familial proteins.

2.0 Materials and Methods

2.1 Materials

All reagents and chemicals were purchased from Sigma unless otherwise specified. Full length agarose, both agarose A and low melting agarose are products of Biobasic. Native porphyran was isolated using the protocol described in Correc *et al* (2011), and was performed by Jan-Hendrick Hehemann, a previous postdoc of our lab. *P. umbilicalis* was collected in the intertidal zone of Perharidy Point (Roscoff, France). Specimens were extensively washed before being dried in an oven at 60˚C, and ground into a fine powder. The resulting powder was kept for 15 h 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 8 h. A clear solution was obtained through centrifugation and extensive filtration using diatomaceous earth and activated carbon. The sample was concentrated by rotary evaporation at 65 ˚C and the polysaccharides were precipitated by incubating with 4 volumes of pure methanol overnight at 4˚C. The precipitate was recovered by filtration and extensively washed with pure methanol and finally with acetone, prior to air drying.

Neoagaro-octaose, neoagaro-hexaose, neoagaro-tetraose and neoagarobiose were prepared through agarose hydrolysis as described in Pluvinage *et al* (2013). A 2% agarose solution in water was heated until dissolved and kept at 40˚C with *exo*-β-agarase, Aga50D from *Saccharophagus degradans* at a final concentration ~10 µg/mL overnight, shaking at 200 rpm. The enzyme was then heat inactivated, and denaturated protein aggregates were removed by centrifugation. The clarified solution was lyophilized and resuspended in water prior to filtration and size exclusion chromatography using Bio-gel

P2 resin (Bio-Rad) equilibrated with 50 mM ammonium carbonate buffer (pH 7.5). Elutions were analyzed by thin layer chromatography (TLC) and individual oligomers identified by comparison with previously prepared standards (neoagarobiose, neoagaro-tetraose, etc.). Fractions with pure oligosaccharides were pooled and lyophilized.

2.2 Cloning and Transformation

2.2.1 Polymerase Chain Reaction (PCR)

Two constructs of *Bp*_1683 were developed in tandem to optimize the chance of crystallization; one with an N-terminal 21 amino acid truncation corresponding to the removal of the putative signal peptide (designated as *Bp*GH50A, 22-523aa), and one with an N-terminal 65 amino acid truncation (*Bp*GH50B, 65-523aa). Each of the respected primers was designed for a pET28a vector as seen in Table 1. The PCR mixture contained the following reagents in their final concentration equivalents: 1x High Fidelity (HF) Phusion buffer (New England Biolabs), 200 µM dNTPs, 1.0 µM respective primers (For primers see Table 1), 100 ng template DNA (approximately 5 ng/uL stock), 1U of Phusion polymerase (New England Biolabs), and was brought up to volume with nuclease-free water. The reaction mixture was run in an Eppendorf PCR cycler using the following protocol, consistent with Phusion protocol specifications [Phusion extension time 1 kb/ 30 s]: initial denaturation at 98 °C for 30 s, 30 cycles of 10 s at 98 °C, 30 s at 36 °C, and 50 s at 72°C, followed by a final step of 10 min at 72 °C. The resulting PCR product was visualized on a 1% agarose gel containing Ethidum Bromide (EtBr), and the

nucleotide fragment of the correct size was purified using a PCR purification kit (Biobasic).

Table 1: PCR Primers for the *Bp*GH50 constructs into pET28a vector. Underline indicates restriction site added (NheI for Forward primers and XhoI for Reverse primers).

	Size of Constructs	Forward (5'-3')	Reverse (5'-3')
*Bp*GH50A (full length)	1503bp, 22-523aa	5' GAT CTA GCT AGC GAA GAC CCT CAG GAA GA 3'	5' GAT CTT CTC GAG TTA TTT GTC GAA ATA ATC TAT 3'
*Bp*GH50B (truncated)	1374bp, 65-523aa	5' GAT CTA GCT AGC CAA CTG CCT GTT CC 3'	(Reverse same as full length)

2.2.2 Digestion

The purified PCR insert and purified native pET28a vector were each incubated with 1x BSA, and 1x Buffer 4 (NEB), as well as 1.5U of restriction enzymes Nhe I and Xho I (New England Biolabs), brought up to volume with nuclease-free water. The mixtures were incubated at 37˚C for 2 h, and the restriction enzymes were denatured at 80˚C for 20 min. Both the digested insert and vector were either gel extracted or PCR purified.

2.2.3 Ligation

After purification, the vector and insert concentrations were estimated using their absorbance at 280 nm. Each ligation reaction mixture contains 1x ligase buffer, 1U DNA ligase (Sigma), 50 ng of vector and an ng amount of insert as determined by the equation below, to a final concentration of approximately 20 µL.

$$ng\ insert\ (for\ 3{:}1\ ratio) = \frac{50ng\ vector \times\ bp\ of\ insert\ \times\ 3}{bp\ of\ vector \times 1}$$

Example: For pET28a the native vector size is 5369 bp, and the size of the insert is dependent on the protein size (*Bp*GH50A: 1503 bp or *Bp*Gh50B: 1374 bp).

2.2.4 Transformation

Competent *Escherichia coli* BL21-DE3 cells (Invitrogen) were incubated on ice for 3-5 min before 10 µL of the completed ligation reaction was added. The cells were left to sit on ice for 15 additional min, before a heat shock at in a water bath at 42˚C for 1 min. The cells rested on ice for an additional minute, before 200 µL of Luria broth (LB) were added. The cells were incubated at 37˚C for 1 h, then plated on an LB plate supplemented with kanamycin (Kan, final concentration 1 mg/mL) and left at 37˚C overnight.

2.2.5 Colony PCR

If colonies were present on the LB+ Kan plates, the presence of the insert was detected using colony PCR. 1x Taq PCR (magnesium free) buffer was incubated with 200 µM dNTPs, 1.5 mM of magnesium chloride ($MgCl_2$), along with 1.0 µM of both forward and reverse (Forward 5'-TAA TAC GAC TCA CTA TAG G-3', Reverse 5' GCT AGT TAT TGC TCA GCG-3') T7 primers for pET28a. These were incubated with 1U of Taq polymerase (Invitrogen) brought up to volume with nuclease-free water, pipetted up and down to mix, then aliquoted into ten PCR tubes. Ten colonies from the plate were scratched and added to each of reaction mixtures, each time using a sterile p10 tip. The reactions were run in an Eppendorf cycler according to Taq polymerase protocol specifications: (Taq extension time 1 min/kb): initial denaturation at 94°C for 3 min, 30 cycles of 45 s at 94°C, 30 s at 45°C, and 2 min at 72°C, followed by a final step of 1 min

at 72°C. The completed reactions were then run on a 1% agarose gel containing EtBr, in order to detect the presence of the desired gene construct. The confirmed positive clones were confirmed by bi-directional DNA sequencing (Sequetech).

2.3 Protein Expression and Purification

Expression trials were performed in 2 mL cultures grown at 37˚C shaking at 200 rpm for approximately 6 h, before samples were split. Half were induced with isopropyl β-D-1-thiogalactopyranoside (IPTG, BioBasic, final concentration 0.5mM), and all samples were left at 37˚C overnight. The next day, all samples were pelleted before lysis using DNase (final concentration 2 µg) and BugBuster (Novagen). A final spin at 12000 rpm for 10 min using a table top centrifuge was performed before 2x Laemilli buffer (Sigma) was added to the resulting pellet and supernatant fractions and run on a 15% SDS-PAGE gel to assess the extent of the recombinant protein overexpression.

Large scale expressions were inoculated from 10 mL LB+Kan precultures inoculated from glycerol stock and incubated overnight at 37˚C. After the optical density reached 0.8-0.9, the cultures were cooled to 16˚C before the addition of IPTG (final concentration of 0.5 mM), then left overnight at 16˚C. The following day, cells were pelleted by centrifugation (6000 rpm for 15 min) and subsequently frozen at -20˚C until use.

Thawed cells were resuspended in sucrose solution (50 mM Tris-HCl, pH 8.0, 25% w/v sucrose) while stirring at room temperature. Lysozyme was added (10 mg), and the solution was allowed to stir for 10 min. Two volumes of deoxycholate solution [1% w/v deoxycholate, 1% w/v Triton X-100, 20 mM Tris pH 7.5, 100 mM sodium chloride, (NaCl)] were then added, incubating for another 10 min, before the addition of 5 mM $MgCl_2$ and 0.2 mg of DNase (Biobasic). After the solution was no longer viscous, it was

centrifuged at 15000 rpm for 45 min. The resulting supernatant was run over an Ni^{2+}-immobilized metal affinity column, first primed with binding buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl) and eluted with an imidazole gradient (0 -500 mM). 2x Laemilli buffer (Sigma) was added to elution samples before they were run by SDS-PAGE in order to identify the elution(s) the recombinant protein was in and to assess the relative purity. Elution fractions deemed sufficiently pure were pooled and concentrated on a stirred-ultrafiltration device (AMICON) with a 10 kDa molecular weight cut off membrane under pressurized nitrogen. Following concentration, the protein was run on a size exclusion column (GEhealthcare HiPrep 16/60 Sephacryl S-200 HR), equilibrated with running buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl), according to manufacturer's protocols.

2.3.1 Protein Concentration Determination

Protein concentration was determined from the absorbance at 280nm using the calculated molar extinction coefficient for the respective proteins used [Obtained through Expasy Server Analysis, Protparam (Gasteiger *et al*, 2005)].

GH50 (full length): Ext. coefficient 106120 M^{-1} cm^{-1}

GH50T (truncated): Ext. coefficient 101650 M^{-1} cm^{-1}

2.4 Agarose Gel Plate Assay

A 1.5% agarose solution was made using 1x PBS buffer. This solution was then heated to dissolve the agarose, and was poured into a petri dish, and left to solidify in antiseptic conditions.

Once cooled, *Bp*GH50A (full length wt, 90 μM) both active and boiled forms (boiled for 20 min at 90˚C), Aga50D (wt, kept at 4˚C, 60 μM) and Aga50D (wt, frozen at -80˚C, 240 μM) were all aliquoted onto the plate in 2 uL and 10 uL drops. The plate was then incubated at 37˚C overnight. A 5% Lugol's solution [5% (w/v) iodine, 10% (w/v) potassium iodide] was added to better visualize the deterioration of the gel matrix which is indicative of activity.

2.5 Thin Layer Chromatography

A 0.4% solution of substrate (neoagaro-tetraose, neoagaro-hexaose or neoagaro-octaose) was incubated with buffer (Final concentration 50 mM Tris-HCl, pH 8, citrate pH 6, or acetate pH 3), *Bp*GH50A (final concentration 15 μM), and cofactors if applicable [Final concentration 50 mM calcium chloride ($CaCl_2$), manganese chloride ($MnCl_2$) or magnesium chloride ($MgCl_2$)].

The reaction mixture was incubated for 1 h at 37˚C, or until completion. 6 μl of each reaction were spotted onto a silica gel-coated plate and allowed to dry completely. The plate was sealed in a glass container that was pre-equilibrated for 30 min with 20 mL solvent composed of a 3:1:1 mixture of butanol (Ana Chemia): 95% ethanol (Commercial Alcohols, Brampton ON): milliQ water (Millipore). The plate remained in the container until the solvent had reached ¾ of the way up the plate (approximately 1½ h). The plate was then removed and allowed to dry. The samples were visualized by dipping the plate in a solution of 5% sulfuric acid with 0.1% orcinol (and 94.9% ethanol), then baked at 110˚C for 10 min.

2.6 Reducing Sugar Assay

The assays were performed in 20 mM Tris-HCl, pH 8.0, 100 mM NaCl with 1% substrate [either low melting agarose (Biobasic), native porphyran (see materials) or κ-carrageenan (V-Labs)]. The substrate was incubated with enzyme(s) (including *Bp*GH50A where appropriate), all with a final concentration of 2 μM. Active colour reagent [0.5 M p-hydroxybenzoic acid hydrazide (PAHBAH), 0.01 M $CaCl_2$, 0.02 M trisodium citrate, and 0.3 M sodium hydroxide] was prepared according to Lever (1972), and created daily. The reaction mixtures were incubated at 37˚C for 2 h and 24 h in triplicate. Colour reagent (250 μL) was added to triplicate samples from each mixture, before being boiled at 100˚C for 10 min. The samples were transferred to a 96 well plate and allowed to cool to room temperature before their absorbance was taken at 410nm with a Molecular Devices SpectraMax M5 plate reader.

2.7 Crystallization, Structure Solution and Refinement

Both protein constructs crystallized but only the truncated form (*Bp*GH50B*)* was reproducible. The structure of this truncated form was solved by single isomorphous replacement with anomalous scattering (SIRAS) at 1.40 Å.

2.7.1 Full Length *Bp*GH50A

Crystals of Native *Bp*GH50A (Full length, 22-523aa) (40 mg/mL) were grown at 18˚C over three days using the vapor-phase diffusion technique from sitting drops in 25% (w/v) polyethylene glycol (PEG) 3350, 0.2 M sodium acetate. These drops were later optimized in hanging drops in 20% PEG, 0.2 M sodium acetate in a ratio of 1:1 protein to mother liquor. The crystals were cryoprotected with mother liquor supplemented with

25% (v/v) ethylene glycol before being flash frozen in liquid nitrogen. However, the proliferation of multiple diffraction patterns when screening made this crystal form unsuitable to produce a data set. A new crystal form of the same native protein was obtained after four months in 25% (w/v) PEG 3350, 0.1 M Bis-Tris, pH 6.5, from sitting drops in a drop ratio of 1:1 protein to mother liquor from the Index Screen (Hampton Research). The drop also contained a fungal contaminant, preventing crystal replication and optimization. These crystals were cryoprotected in the same manner as the previous crystals with mother liquor containing 25% (v/v) ethylene glycol. The first diffraction images of *Bp*GH50A were collected at the Stanford Synchrotron Radiation Laboratories (SSRL) on beamline 9-2 at 0.9792 Å. These images were able to produce a partial data set processed using iMosFLM (Battye *et al*, 2011) and SCALA (Evans, 2006) to 1.40 Å.

2.7.2 Truncated *Bp*GH50B

Crystals of Truncated *Bp*GH50B (65-523aa) (40mg/mL) were obtained from hanging drops in 17% (w/v) PEG 3350, 0.04M Bis-Tris pH 6.5 with a drop ratio of 1:2 protein to mother liquor. An iodide derivative dataset of the *Bp*GH50B crystals was obtained by soaking the crystals in mother liquor supplemented with 1M sodium iodide (NaI) for five minutes, before cryoprotecting in 75% NaI & mother liquor, 25% (v/v) ethylene glycol and being flash cooled in the cryostream.

Diffraction images of *Bp*GH50B were collected with a Rigaku R-AXIS 4++ area detector coupled to a MM-002 X-ray generator with Osmic 'blue' optics and an Oxford Cryostream 700. The iodide derivative data was then processed using Crystal Clear/d*trek (Pflugrath, 1999).

Analysis of the iodide derivative data using ShelX_cde (Sheldrick, 2010) identified 20 iodide sites within the crystal packing (ten within each monomer in the asymmetric unit). This data was combined with the anomalous scattering data from the previously collected *Bp*GH50A high quality data set. Once the two datasets were unified and scaled [using CAD and SCALEIT (Howell and Smith, 1992)], they were processed using SIRAS in ShelX_cde (Sheldrick, 2010). The initial model was built with ARP/WARP (Langer *et al*, 2008) before completion with COOT (Emsley and Cowtan, 2004) and refinement in REFMAC (Murshudov *et al*, 2011).Water molecules were added using REFMAC and inspected visually prior to deposition. In each data set, five percent of observations were flagged as free and used to verify refinement procedures. The statistics for the final model (designated as *Bp*GH50), are present in Table 2. Figures 14, 15 and17 were created using PyMol (The PyMOL Molecular Graphics System, Version 1.6.0 Schrödinger, LLC.).

In order to estimate the active site residues that would be involved in substrate binding, three models were created based on the alignment of the *Bp*GH50 structure with an Aga50D-neoagaro-octaose complex. All alignments and overlays were performed in Coot (Emsley and Cowtan, 2004).

The *Bp*GH50 neoagaro-octose model was created using the alignment was performed between *Bp*GH50 and an Aga50D-neoagaro-octaose complex (4BQ5) (RMSD: 2.41Å, 360 Cαs aligned). This alignment placed the agarose directly in the putative active site of *Bp*GH50. The Aga50D protein structure was then removed, leaving only the *Bp*GH50 protein structure and sugar model.

The neoporphyran-tetraose model was created using the previous *Bp*GH50 agarose model. A neoporphyran-tetraose oligomer was obtained from a complex with a β-porphyranase (*Bp*GH86A, 4AW7). The tetraose was manually overlaid with the neoagaro-octaose sugar monomers corresponding to the -2,-1, +1 and +2 subsites of *Bp*GH50, with special attention given to align the porphyran D-Gal with the agarose D-Gal at the -1 subsite. The model was adjusted at the +1 and +2 subsites, causing a breakage in the tetraose at the scissile bond. This was necessary to prevent direct steric interference between the *Bp*GH50 surface model and the porphyran sugar at these sites. The agarose substrate was then selectively removed in PyMol for the final picture.

The hybrid substrate model (containing both porphyran and agarose units) was obtained by combining the two previous models. From the porphyran model, the porphyran monomers in subsites +1 and +2 were selectively removed in PyMol and their equivalent agarose units made visible once again.

Table 2: X-ray data collection and refinement statistics for *Bp*GH50, where () is the outer shell data.

Data Collection	*Bp*GH50A High Resolution Data Set (Native)	*Bp*GH50B Data Set (Iodide Derivative)
Space Group	$P2_1$	$P2_1$
Unit Cell		
a,b,c (Å)	73.10 87.37 74.67	73.59 88.01 74.57
A,β,γ (˚)	90 105.5 90	90 105.2 90
Wavelength	0.9792	1.5418
Resolution (Å)	39.80-1.40 (1.42-1.40)	19.76-2.30 (2.38-2.30)
Rmerge (%)	0.056 (0.470)	0.138 (0.433)
Completeness (%)	99.3 (91.0)	99.2(98.2)
No. of Total Reflections	1057352	710297
No. of Unique Reflections	176279	40551
Redundancy	6.0 (4.4)	17.5 (17.5)
I/sigma(I)	39.4 (3.7)	18.4 (7.3)
Anomalous completeness	99.2 (89.3)	98.9 (97.4)
Anomalous multiplicity	3.0 (2.2)	8.7 (7.9)
Refinement		
Resolution	1.40	
Rwork/Rfree (%)	0.16/0.18	
No. Atoms		
Protein Chain A	3674	
Protein Chain B	3628	
Water	1091	
B-factors		
Protein Chain A	14.0	
Protein Chain B	16.8	
Water	29.5	
RMS deviations		
Bond Lengths(Å)	0.019	
Bond Angles (˚)	1.870	
Ramachandran		
Preferred (%)	876 (97.3)	
Generously Allowed (%)	18 (2.0)	
Outliers (%)	6 (0.7)	

3.0 Results

3.1 Bioinformatics

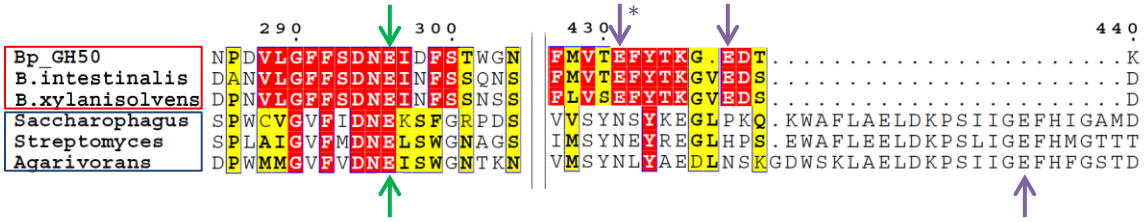


Figure 9: Partial sequence alignment of *Bp*GH50 with other GH50 proteins (full sequence alignment in Appendix D). *Bp*GH50 was aligned with uncharacterized proteins from gut *Bacteroides intestinalis* (NCBI refsequence WP_022392986, 98% coverage and 45% identity) and *Bacteroides xylanisolvens* (Bxy_10650, 97% coverage and 47% identity), and proteins from marine sources; Aga50D from *Saccharophagus degradans* (PDB ID 4QB5, 67% coverage and 28% identity), Sco3487, a β-agarase from *Strepomyces coelicolor A3(2)* (45% coverage and 46% identity) (Temuujin *et al*, 2012), and β- agarase A from *Agarivorans sp. QM38* (83% coverage and 25% identity) (Lee *et al*, 2006). Sequences and statistics extracted from pBLAST analysis (Altschul *et al*, 1997) unless otherwise indicated, alignment by SALIGN and ESPript (Braberg *et al*, 2012; Gouet *et al*, 2003). Catalytic residues of Aga50D (*S. degradans*) are highlighted at the bottom with arrows (green for acid/base, purple for nucleophile). Potential catalytic residues for *Bp*GH50 (E431 and E438) are highlighted by the arrows above (green for acid/base, purple for nucleophiles). The purple asterisk depicts the *Bp*GH50 nucleophile that was ultimately determined to be structurally conserved with the nucleophile of Aga50D.

The GH50 family is comprised primarily of β-agarases (Lombard *et al*, 2014). The protein from the gene BACPLE_01683, is hereby known as *Bp*GH50, due to its significant shared amino acid sequence identity with other GH50 members (up to 28%) from *Saccharophagus degradans*, *Streptomyces coelicolor* and *Agarivorans sp. QM38* (Figure 9). This shared identity coupled with the location of BACPLE_01683 within the *B. plebeius* PUL, which primarily degrades porphyran but also contains previously

characterized agarases (Hehemann *et al*, 2012A and 2012B), suggests that *Bp*GH50 may be a β-agarase.

Of the aligned agarases in Figure 9, the *exo*-β-agarase from *S. degradans*, Aga50D is currently the only GH50 that has been structurally characterized. The catalytic residues of Aga50D are shown in Figure 9 (bottom green and purple arrows). Interestingly, only one of these residues, the putative acid/base (E296) is conserved within *Bp*GH50 (top green arrow). The nucleophile of Aga50D is contained on 23 residue stretch completely absent in the gut *Bacteroides* proteins that were used in the alignment. However, there are two Glu residues (Figure 9, top purple arrows) that are within a strongly conserved section among the gut *Bacteroides*, 431–**E**FYTKGX**E**D-439, just before this absent stretch. One of these residues may provide the second residue involved in catalysis, but this needed to be confirmed using additional methods.

3.2 Activity Assays

3.2.1 Agarose Plate Assay

Bioinformatics has predicted that *Bp*GH50 is an agarase since it is in the same GH family that contains almost exclusively agarases. Thus, in order to test for agarase activity a series of activity assays were performed. The agar plate assay was initially used to test agarase activity on a gelled agarose matrix (Figure 10).

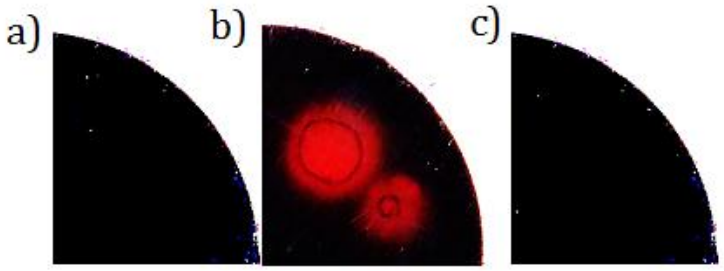


Figure 10: Gelled agarose plate assay to determine *Bp*GH50A activity on full length agarose. Purified *Bp*GH50A and Aga50D protein were added in 10 µL and 2 µL drops

directly to a 1.5% agarose plate and incubated for 18 h at 37˚C. a) Boiled *Bp*GH50A served as a negative control, while established agarase, Aga50D was the positive control (b). The native *Bp*GH50A is present in c).

The *Bp*GH50A showed no appearance of indents in the agarose gel after an 18h incubation (Figure 10c), and similarly, staining with the Lugol's Solution did not reveal any areas of agarose matrix breakdown, indicating that no digestion had occurred, which was consistent with the inactivated protein control. In contrast, the positive control Aga50D *exo*-β-agarase demonstrated activity as both indents in the agar, and the appearance of lighter reddish circles surrounding the two aliquots on the plate when Lugol's solution was added, observed in Figure 10b). This suggests that *Bp*GH50A is not active on full length agarose. However, the complex structure which gives rise to the agarose gel matrix may inhibit *Bp*GH50A from properly acting on the substrate (Percival, 1979). As such, *Bp*GH50A may be active on smaller oligomers of agarose once they are liberated from the gel.

3.2.2 Thin Layer Chromatography

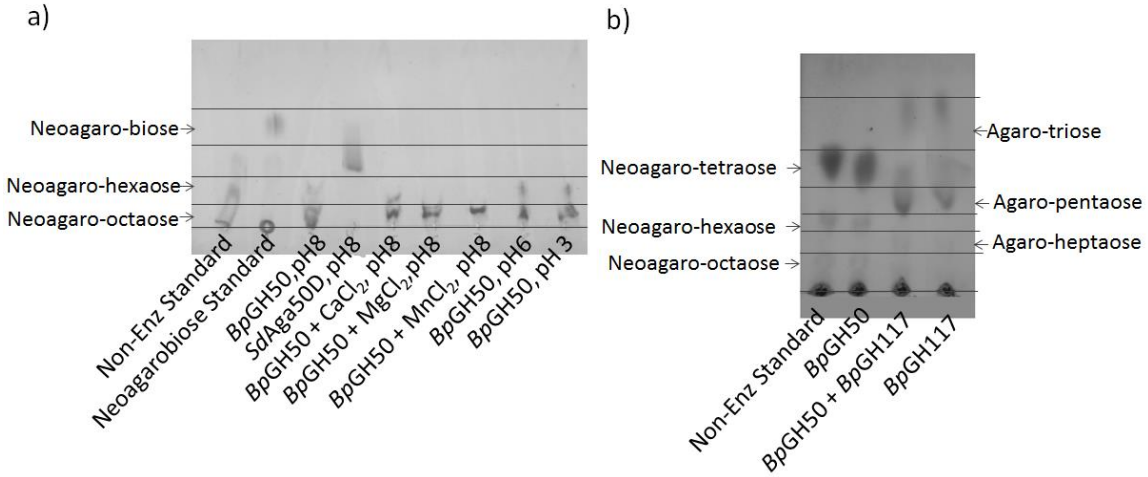


Figure 11: Thin layer chromatography of short length agarose (eight sugars or fewer) after incubation with *Bp*GH50A. a) *Bp*GH50A was incubated with neoagaro-octaose [with a

neoagaro-hexaose impurity, see Non-Enzymatic (Non-Enz) Standard] in conditions varying in pH and presence of three common cofactors. b) *Bp*GH50A was incubated with a mix of neoagaro-octaose, hexaose and tetraose [previously digested by Aga50D, see Non-Enzymatic (Non-Enz) Standard], in the presence and absence of *Bp*GH117, an α-3,6-anhydro-L-galactosidase.

Thin layer chromatography was used in order to explore the hypothesis that *Bp*GH50A was active on smaller oligomers of agarose. Oligomers of neoagarotetraose, neoagarohexaose and neoagaroocatose were incubated with *Bp*GH50A (Figure 9a; hexaose and octaose; 9b) tetraose, hexaose and octaose); however, no digestion was detected, as the reaction mixture spots migrated the same distances as the oligomer controls (Figure 11a). This is compared to the positive control, the Aga50D which in Figure 11a shows no hexaose or octaose bands, only a smear of smaller products closer to neoagarobiose in size. Separate *Bp*GH50A reaction mixtures were supplemented with several divalent cations (Ca^{2+}, Mn^{2+}, Mg^{2+}) that can be sequestered within gelled agarose (within the *Rhodophyta* cell wall) (Percival, 1979). It was thought that one of these cations could serve as a potential cofactor for the protein. However, in each case, the bands that appeared were the same as the controls. Three buffers of different pHs (pH 3.0, pH 6.0, pH 8.0) were also tested to determine if *Bp*GH50A was only active within a certain pH threshold. These results were not indicative of activity. Finally, *Bp*GH50A was incubated with *Bp*GH117, an α-neoagarobiase from the same PUL, in order to provide *Bp*GH50A additional smaller oligomers with D-Galactose at the non-reducing end (Figure 9b). The addition of the *Bp*GH117 did modify the profile of the sugars as compared to the non-enzymatic control; however, the bands were consistent with the *Bp*GH117 positive control. This data implies that *Bp*GH50A is not degrading any of the

agarose oligomers that were provided, and when combined with the previous assay, strongly suggests that *Bp*GH50A is likely not active on agarose.

3.4 Reducing Sugar Assay

The lack of agarose activity of *Bp*GH50A, despite the amino acid sequence similarities to agarases, lead us to the expansion of our potential substrate scope. To this end, other galactan types were tested using a reducing sugar assay. The reducing sugar assay uses hydrazine which readily reacts with the reducing ends of carbohydrates in solution, producing a yellow colour. If a protein is active on a select substrate, more reducing ends will be present and will give a more intense colour.

Kappa-carrageenan and porphyran were chosen for this assay because both have similar structural characteristics to agarose. Porphyran being an agaran, has the same D-L sugar repeating unit, while κ-carrageenan is a carrageenose, with a D-Gal for the first sugar and an anhydro-galactose on the second.

Additionally, previously classified enzymes from within the *B. plebeius* PUL [and one from *Bacteroides uniformis,* another gut derived *endo*-β-agarase *which* produces the same products as *Bp*GH16A from the B. plebeius PUL (Ben Pluvinage, Personal Communication)] were also added to increase the variety of substrate sizes for agarose and porphyran, all of which *Bp*GH50 may encounter *in vivo*.

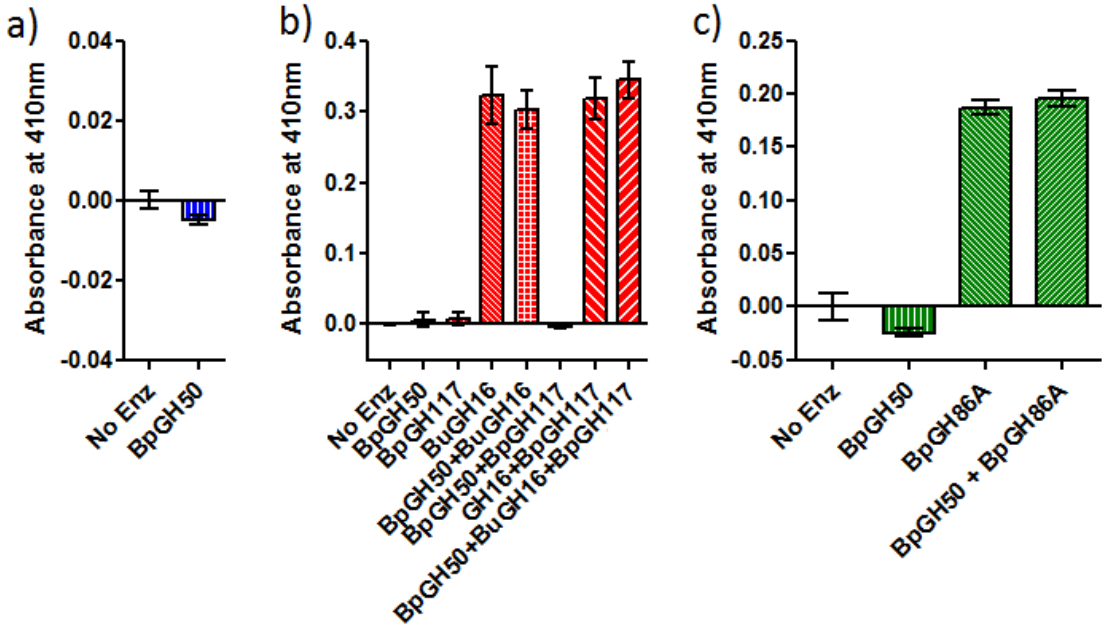


Figure 12: Sugar reducing assays of *Bp*GH50A after incubation with agarose, porphyran and κ-carrageenan oligomers. Assay performed on κ-carrageenan [a), in blue] and agarose [b), in red]. All data is normalized with respect to the non-enzymatic (No Enz) reaction. All reactions in a) and b) were performed in triplicate. c) Reducing sugar assay to dispute GH50A activity on porphyran (in green) in the presence of *Bp*GH86A. All samples in b) were performed in triplicate, and triplicates of each were taken, making nine samples total per reaction type.

Alone, *Bp*GH50A was not active on any of the available substrates, because, when incubated with each substrate, the reactions containing *Bp*GH50A did not demonstrate any additional absorbance over non-enzymatic control (Figure 12a, 12b, and 12c). This result is consistent with the previous agarose assays. *Bp*GH117 also did not show any significant absorbance changes (despite activity as observed in TLC, Figure 11b); however, this was expected given how the protein liberates individual 3,6-anhydro-L-galactose units from the main chain. This anhydro-galactose, while theoretically able to rearrange to form a reducing end, has not been observed in practice within our lab. The

3,6-anhydro group is thought to interfere with the reducing end rearrangement, and thus it does not form, preventing an increase in absorbance within this assay (Craig Robb, personal communication).

Of the enzymes tested, only two proteins; the *Bu*GH16 and the *Bp*GH86A were able to break down their respective substrates (agarose and porphyran respectively), creating many reducing ends, which resulted in an increased absorbance when reacted with the hydrazine.

When coupled with these *endo* acting enzymes, *Bp*GH50A did not provide any distinguishable increase in activity on agarose or porphyran. Similarly, the addition of *Bp*GH50 did not noticeably affect absorbance levels when coupled with the *Bp*GH117. When *Bp*GH50A was coupled with both the *Bp*GH117 and the *Bu*GH16 agarase, there was a slight but insignificant increase in absorbance. Overall this suggests that *Bp*GH50A is not active on the products of agarose or porphyran degradation.

3.3 Crystals and Structure

The previous biochemical results suggest a lack of agarase activity. We used X-ray crystallography to solve the structure of *Bp*GH50, to understand the active site architecture so that it may provide insight into the natural substrate of this protein.

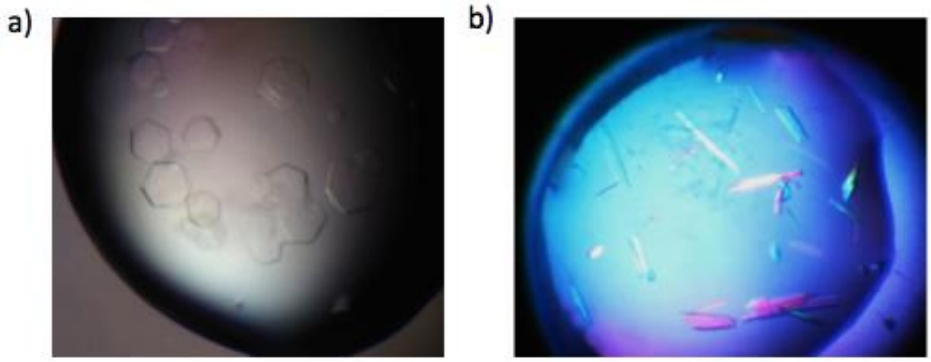


Figure 13: Crystal isoforms of *Bp*GH50. a) Optimized hexagonal plates in 0.2 M sodium acetate, 20% w/v PEG 3350. b) Crystals in 0.1 M Bis-Tris, pH 6.5, 25% w/v PEG 3350.

The full length *Bp*GH50A protein construct (22-523aa) crystallized after three days incubation and optimized in 0.2 M sodium acetate, and 20% (w/v) PEG 3350. The crystals were hexagonal plates (Figure 13a) which were very reproducible, but prone to stacking. When obtaining diffraction data, the images produced contained multiple patterns that could not be differentiated from one another. Thus, these crystals could not be used to collect a complete dataset.

Fortunately, after four months a different crystal form was found in a condition of 0.1M Bis-Tris pH 6.5, 25% PEG 3350. This crystal form (Figure 13b) had a space group of $P2_{1,}$ and from which a high resolution dataset was obtained. However, these crystals were unreproducible given the presence of a fungal contaminant, but the second *Bp*GH50 construct (*Bp*GH50B, 65-523aa) was able to crystallize in the same crystal condition with minimal difficulty. Using the *Bp*GH50B construct, an iodide derivative was obtained by soaking the crystal in mother liquor supplemented with 1 M NaI. Using the derivative and the high resolution data set, the structure was solved to a resolution of 1.40 Å using SIRAS (see 2.7.1 for complete processing information, and Table 2 for all collection, processing and refinement data).

3.3.1 *Bp*GH50 Structure

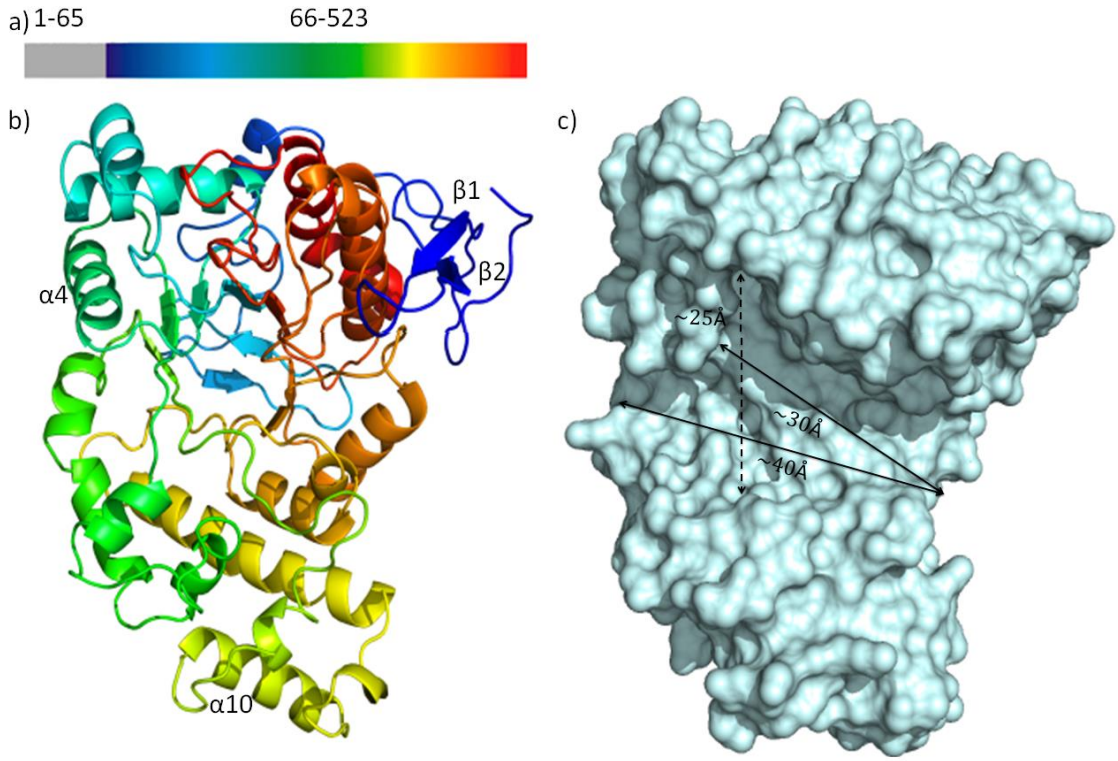


Figure 14: Ribbon and surface structure models of *Bp*GH50. a) The primary structure of *Bp*GH50 consists of an N-terminal signal peptide (1-24) followed by a segment that interfered with crystallization (25-65). The remaining residues make up a truncated N-terminal domain of unknown function and the TIM barrel of the GH50. b) A cartoon model of *Bp*GH50, in a rainbow gradient from N-terminus (blue) to C-terminus (red). The labels of select secondary structures are added for descriptive purposes. c) A surface model of *Bp*GH50, which demonstrates the large wide groove across the width of the protein (~40 Å) and contains the putative catalytic residues. A smaller trajectory along this groove (~30 Å) is blocked at one end by a loop. The width of the groove at its widest point (~25 Å) is measured along the dotted line. Structures generated in PyMol.

*Bp*GH50 was crystallized with two molecules in the asymmetric unit. In one molecule, Chain A, residues 65-523 could be traced, and in Chain B, residues 68-523 could be traced, both with no gaps present. The protein contains ten β-strands and fifteen

α-helices, with eight of each forming the main TIM barrel fold, consistent with other GH50 proteins (best observed in Figure 14b). There are also extra loops and secondary structure that deviates the protein from a perfect TIM barrel. The most apparent is the collection of loops and two additional beta strands that constitute the most N-terminal section of the protein sequence. N-terminal domains are not uncommon among TIM barrel containing enzymes, as observed in other GH50 enzymes (Aga50D) as well as GH5 and GH86 enzymes (*Bp*GH86A)(Hehemann *et al*, 2012B; Pluvinage *et al*, 2013). The N-terminal domain of *Bp*GH50 is comparatively quite small (only ~60 residues) compared to Aga50D's 210 residues), and may indeed be truncated at the N-terminus, which was necessary for consistent crystallization.

Overall, the protein is globular, with a large groove running across its width (approximately ~40 Å, Figure 14c). This groove is very wide (~25Å at its widest point), and is also partly blocked on one side, leading to a small open-faced pocket (the length of this smaller size of groove is ~30 Å).

3.3.2 Comparison with Aga50D

Aga50D from *S. degradans* is the most similar (and only) GH50 for which there is a structure available[36]. A global alignment of the proteins was performed [Figure 15, global root-mean-square deviation (RMSD) 2.41Å, 360 Cαs aligned] to confirm the putative catalytic residues of *Bp*GH50 and also to assess the proteins' differences in substrate preference.

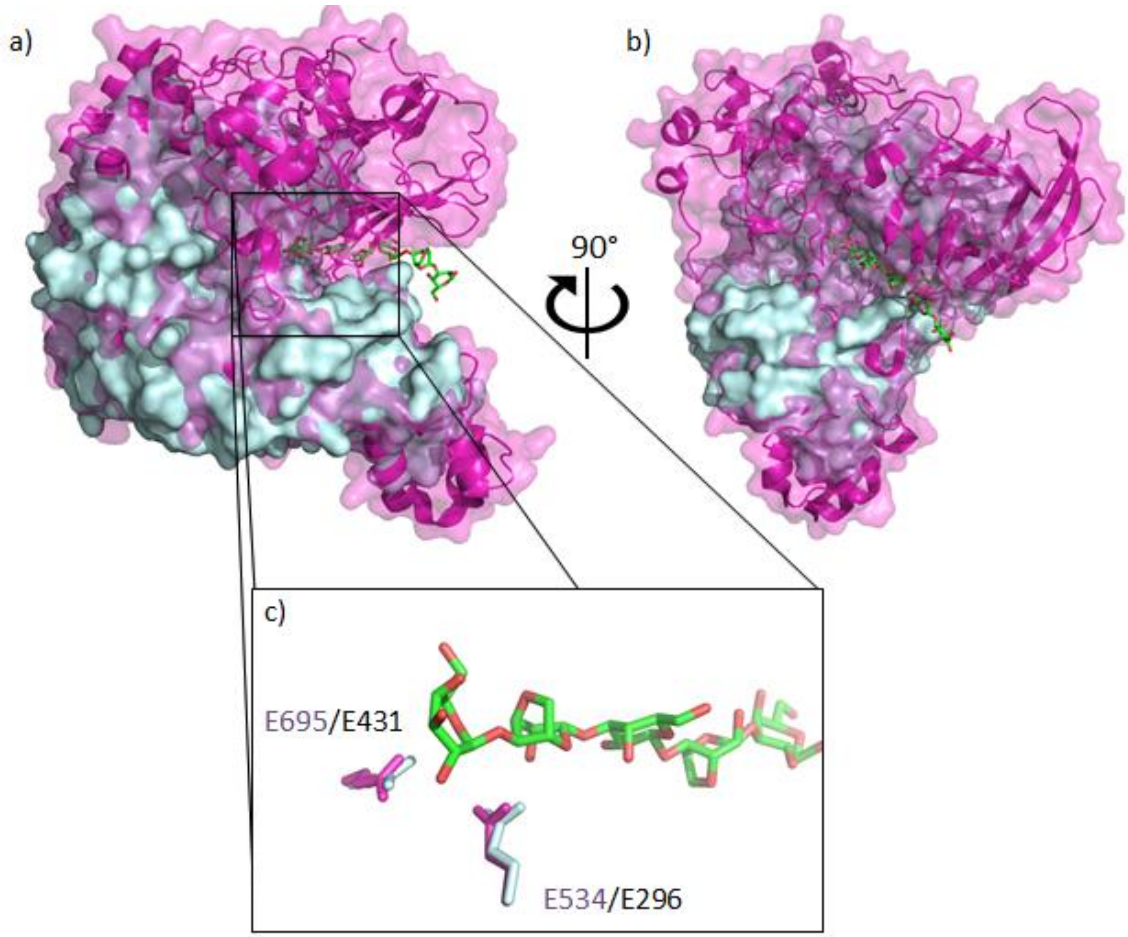


Figure 15: Global alignment of *Bp*GH50 (cyan) with Aga50D (purple) complexed with neoagaro-octaose (green) (PDB ID: 4BQ5, Global RMSD 2.41Å, 360 Cαs aligned). a) b) Surface model of *Bp*GH50 with cartoon model of Aga50D (with transparent surface), highlighting the extra loops present in Aga50D. The complexed octaose substrate (stick model), demonstrates the active site location in Aga50D (and by extension, *Bp*GH50) c) The *Bp*GH50 putative catalytic residues (cyan, labels in black) with their corresponding Aga50D equivalents (residues and labels in purple) are displayed relative to the ocatose substrate. Images generated in PyMol.

By number of amino acids, Aga50D is approximately 40% larger than that of *Bp*GH50 (747 amino acids vs 459 amino acids), and as such the structural comparison reveals a significant number of loops are not present in *Bp*GH50 (Figure 15a and b). Using *Bp*GH50 as the benchmark (Figure 14b), the largest sections of these extra loops are

present around the β1 and β2 strands, extending across to the α4 helix, reaching the far side of the *Bp*GH50 groove. A large part of this section in Aga50D is taken up by a Carbohydrate Binding Module (CBM)-like domain that is fused to the main TIM Barrel, which despite its similar location, has no sequential similarities to the *Bp*GH50 N-terminal domain (using BLAST sequential comparison of the first 120 residues of *Bp*GH50 to the first 170 residues of Aga50D). A second smaller section of two α-helices also extends away from the TIM Barrel around the α10 helix, an expansion of the similar *Bp*GH50 shape (all helix and strands are in *Bp*GH50). The most distinct difference in Aga50D is the lack of the pronounced *Bp*GH50 groove.

The active site of Aga50D consists of a large pocket with a small opening on the opposite side of the main opening which may give the site a 'tunnel'-like quality. This 30Å pocket is approximately the same shape and occupies the same space as the *Bp*GH50 groove up until the partial blockage (see 30 Å block in Figure 14b). The exception is that this pocket in Aga50D is completely covered by several bridging loops not present in *Bp*GH50. By comparison, *Bp*GH50 is very open, with only two sides of the groove available for substrate binding. Interestingly, the side pocket offshoot of the groove blockage is not retained in Aga50D, which may hint at a different substrate of choice for *Bp*GH50. Given the lack of other grooves or pockets that may serve as alternative active site location, and the conservation of this particular section with Aga50D, this is the most likely location for the putative *Bp*GH50 active site.

This was supported by the localization of the putative catalytic residues within this section of the groove. Based on the sequence alignment (Figure 9), it was suspected that

these residues were E296 and either E431 or E438. The overlay of *Bp*GH50 with Aga50D (Figure 15C) indeed confirms that E296 and E431 are the catalytic residues of *Bp*GH50 as they are nearly superimposable with those of Aga50D (Figure 15C, green and purple arrows at bottom).

4.0 Discussion

4.1 Bioinformatics

The marine origin of the *B. plebeius* porphyran PUL has been previously established (Hehemann *et al*, 2012B), stemming from the realization that the closest homologs of many PUL CAZymes are found in marine organisms. *Bp*GH50 is not one of these CAZymes, as it shares much more sequence identity with uncharacterized proteins from other gut *Bacteroides*, *B.intestinalis* (45% identity) and *B. xylanisolvens* (47% identity). Agars occupy a unique dietary niche, and gut agarases and porphyranases indeed have been limited to the microbiota of individuals that regularly consume seaweeds (to date those of Japanese and Spanish descent) (Thomas *et al*, 2011). While *B.intestinalis* like *B.plebeius* was isolated from the Japanese microbiota (Abu Dakir *et al*, 2006), *B. xylanisolvens* was obtained from the French microbiota (Chassard *et al*, 2008). Since French cuisine does not contain significant amounts of seaweed (certainly less than the ~14g/person/day consumed in Japan) (Thomas *et al*, 2011; Hehemann *et al*, 2012B), agarose or agarans would not be regularly encountered by the French microbiota, suggesting that the *B. xylanisolvens* protein is not an agarase. The significant shared identity among all three of these *Bacteroides* proteins instead suggests they are derived from an ancestral gut *Bacteroides* protein rather than being of marine origin, effectively dividing the GH50 family into at least two subsections. This is supported by cladogram analysis (Figure 16) which also displays a distinct separation with an 88% bootstrap value between GH50s similar to *Bp*GH50, and those similar to Aga50D.

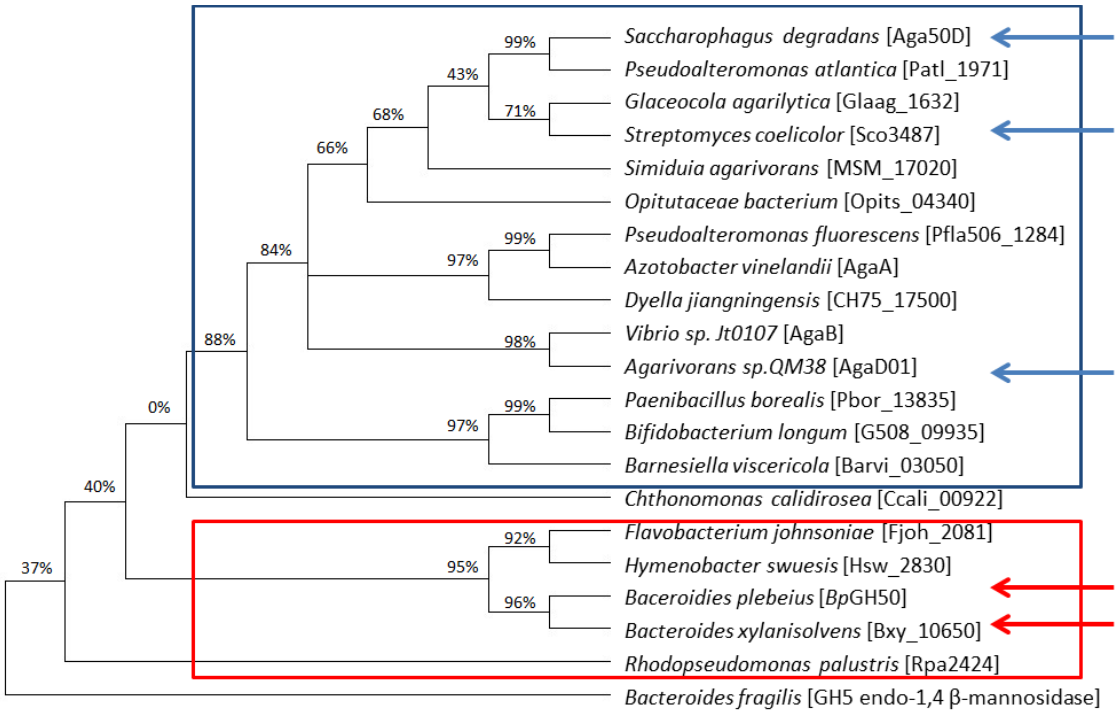


Figure 16: Cladogram of 20 aligned GH50 protein sequences. The protein sequence of *Bp*GH50 was combined with 20 randomly selected GH50 proteins sequences (and one GH5 protein sequence that served as an outgroup). The sequence names are displayed [protein name if characterized, gene name if not] alongside their respective organisms. which we re aligned through SALIGN (Braberg *et al*, 2012) and processed through PhyML (Guindon *et al*, 2010) in order produce a cladogram of the sequences via Nearest Neighbour Interchange (NNI) clustering, visualized using Geneious (Kearne *et al*, 2012). Bootstrap values (%) observed show the branching likelihood after 100 repetitions. The red box specifies the GH50s closest in sequence to gut derived GH50 (such as *Bp*GH50), while the blue box indicates those closer to more marine derived GH50s (such as Aga50D). Since the bootstrap value for the *C. calidirosea* GH50 was so poor (0%), it was not included in either grouping. The red and blue arrows indicate the proteins that were present in the alignment in Figure 9.

4.2 Overlay Model with agarose and porphyran oligomers

4.2.1 Agarose

To provide a more in depth analysis the lack of agarose activity observed in the biochemical assays, we overlaid the crystal structure of neoagaro-octaose derived from an Aga50D complex, on top of our *Bp*GH50 structure to provide a guide to what theoretical interactions may occur. In reality, the sugars bonds within this rigid model (Figure 17b) and those derived from it (Figure 17c,d) would be much more flexible and would likely produce more interactions that those shown.

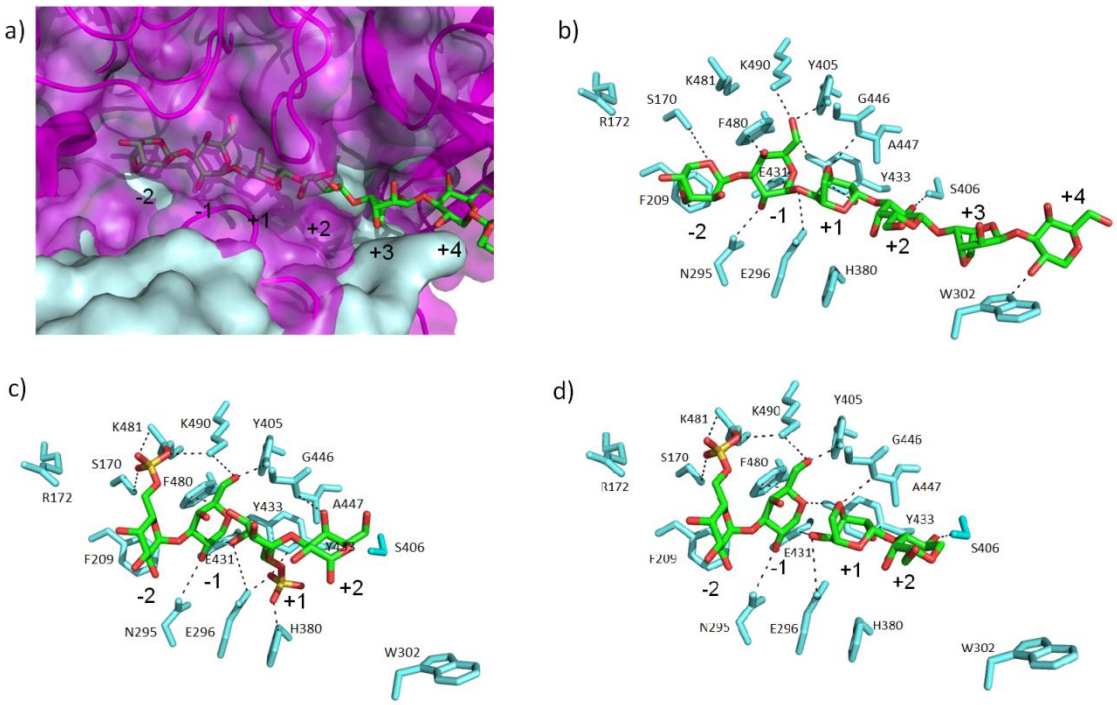


Figure 17: Stick model overlay of agarose and porphyran oligomers present in native porphyran within the *Bp*GH50 active site. Close up of the active site from the global alignment observed in Figure 15 (Aga50D, pdb ID: 4QB5). b) Overlay of neoagaro-octaose obtained from the Aga50D complex, with only *Bp*GH50 active site residues visible. Only 6 of the 8 sugars are displayed, the other two (corresponding to subsites +5, and +6) did not associate with any *Bp*GH50 residues and were thus omitted. c) Overlay of neoporphyra-tetraose (isolated from the aglycone region of a mixed agaro-porphyran substrate originally

crystallized with *Bp*GH86A, pdb ID: 4AW7). d) Combination of the two overlays, (-2 and -1 from porphyran oligomer, and +1 and +2 from the agarose oligomer) forming a putative hybrid substrate, the most likely substrate candidate. In all examples, the dashes between *Bp*GH50 residues (in cyan) and the ligand (in green for carbon/red for oxygen/gold for sulfur), correspond to potential interactions, as the distances between the two points are <5.0 Å in length. All images generated in PyMol.

With the neoagaro-octaose in the active site, the substrate rests along the smaller of the two grooves (~30 Å) (Figure 14c, 17a). This means that the two loops at the end of this groove would prevent the cleavage of products larger than three monomers by steric inference, suggesting that the protein may have *exo*-activity instead of the *endo*-activity proposed by the larger ~40 Å groove.

*Bp*GH50 has multiple residues in a close enough proximity to interact (which for the purposes of this analysis are distances <5 Å) with at least 3 sugar subsites of the agarose (the -2,-1 and +1 subsites, see Figure 16). There are also single potential interactions present with the +2 and +4 D-galactoses (with S406 and W302). The Trp residue resides at the very entrance to the groove and appears as if it may facilitate docking of longer sugars, but lack of extensive interactions at these sites may suggest that the protein may prefer smaller substrates of 3-4 units, since a longer chain may not have enough possible interactions to sufficiently bind.

The neoagaro-octaose substrate appears to fit the *Bp*GH50 best at the -1 subsite, since this subsite is involved with six of the ten hydrogen bonds previously described (if catalytic residues are included). The majority of the residues that interact with the agarose at this site in Aga50D are conserved in *Bp*GH50, namely the catalytic residues E296 and

E431, but also those that interact with the hydroxyl groups of the D-Gal such as N295, as well as form sugar-aromatic interactions with D-Gal with F480. The residues K490 is slightly further away from the centre of the active site but may help to anchor the D-Gal in place, providing a similar function to that of E757 in Aga50D.

The +1 subsite has 3 hydrogen bond interactions including both the catalytic residues, and one aromatic interaction with Y405. Interestingly, this remaining hydrogen bond is situated between the 3,6-anhydro ring oxygen and the backbone between G446 and A447 rather than a specific residue. There is a histidine residue (H380) on the opposite site of the main chain interaction, and while it is too far away (~6.9 Å) to interact with an 3,6-anhydro group, it may potentially provide some support to a sulfated sugar.

The most interesting findings concern the residues surrounding the theoretical -2 subsite. While there is an F209 residue that would form sugar-aromatic interactions, and a single interaction with S170, there are no other residues nearby to make hydrogen bonds with the LA sugar. However, there is a pocket off centre from the binding groove. Two basic residues (R172 and K481) and the previously mentioned serine form this pocket, which is not conserved within Aga50D, but is sequentially conserved within other gut derived *Bacteroides* GH50s (See Appendix D). These basic residues appear to be too far away (6-7 Å) to interact with the anhydro group of the agarose.

Overall, there does not appear to be residues that would sterically interfere with agarose binding, but the limited number of apparent subsites suggests a small substrate (consisting of 3-4 linked monomers). Aga50D has the benefit of a tunnel-like active site, so that it can interact with agarose from all sides. With a wide groove, *Bp*GH50 is limited to two sides, as the groove is much wider than the substrate, so the agarose can only

interact with one side of the groove at a time. This limited capacity for substrate interactions coupled with lack of interaction with the basic pocket at the -2 subsite may contribute to the lack of agarose binding observed in the biochemical analysis, as the interactions observed in this model may not be sufficient to anchor the substrate to the binding site.

4.2.2 Porphyran

Since the basic pocket in the -2 subsite is too far from a neutral agarose to be properly utilized, we instead looked at porphyran as a substrate contender. Since porphyran has an acidic 6'-sulfate group within its repeating unit, it may extend far enough at the -2 subsite to interact with this basic pocket.

A neoporphyan-tetraose substrate obtained from the complex of a β-porphyranase from *Z. galactivorans* (pdb 3ILF) (Hehemann *et al*, 2012C). In order to dock the tetraose and properly overlay it in approximately the same space as the agarose, it needed to be broken at the putative site of cleavage. When the tetraose was overlaid at the -1 subsite, it was observed that the L6S sugar in the -2 subsite had its sulfate group in the proper orientation to associate with the basic pocket. Furthermore, two of the three residues thought to interact with the sulfate were <4 Å. This suggests that the natural substrate of the *Bp*GH50 may be sulfated in order to associate with this pocket.

The analysis of the +1 subsite however, appears to be less conclusive. When oriented in the same manner as the agarose, the sulfate of L-Gal is close enough to H380 (2.7 Å) to suggest an interaction. However, this also puts the sulfate in the proximity of the putative nucleophile E296 (2.1 Å away). In the catalytically active enzyme, this residue would be charged, and would likely repel the like-charged sulfate. Furthermore, the backbone of

G446-A447, which previously showed an interaction with the 3,6-anhydro group of the agarose, is now distanced too far away from any oxygen of the L6S to suggest an interaction, which may indicate that a 3,6-anhydro group many fit better in this subsite.

From these findings, it appears as if a porphyran oligomer best fits in the aglycone (-2 and -1) subsites, while an agarose oligomer would fit better in the glycone (+1 and +2) subsites. Given that native porphyran contains both agarose and porphyran units, it is reasonable to suggest that *Bp*GH50 may require a small hybridized substrate. Considering that the different substrate types would be present on opposite ends of the scissile bond, *Bp*GH50 may act to separate the porphyran and agarose units from each other.

4.3 Role of *Bp*GH50 in the *B. plebeius* PUL enzymatic pathway

Given the information gleaned from the structure, *Bp*GH50 may be specific to accommodate an agaro-porphyan hybrid substrate. Native porphyran contains blocks of both agarose and porphyran repeating units, and thus for effective degradation, the separation of these two sugar types must be performed by CAZymes that can accommodate both types. The presence of hybrid substrate enzymes is already apparent in the porphyranolytic system of *Zobellia galactanivorans* (Hehemann *et al*, 2012C), although no such hybrid enzymes have been yet been found within the *B. plebeius* PUL.

Compared to agarolytic cascades, a porphyran degrading cascade is considerably more complex as it requires additional enzymes such as β-porphyranases and sulfatases to break down the sugar into usable monomers (Michel *et al*, 2006; Chi *et al*, 2012). The augmentation in complexity leads to the production of many different possible substrates/ products. Both porphyranases and agarases often have subsites that can accommodate

both L6S and LA units; however these are often distanced from the critical -1,-2 and +1, +2 subsites. One *Z. galactivorans* agarase has been found to have +1 and +2 subsites that are promiscuous (Correc *et al*, 2011), which enables it to degrade the bond between agarose and porphyran units, increasing the degradation efficiency of the entire agarolytic system by liberating more agarose units to be degraded. Understanding all the roles of porphyran and agarose degradation is important, and we are still uncovering new types of enzymes; while β-agarases have been known for decades , α-agarases have only recently been discovered (Hehemann *et al*, 2012A).

This research suggests that *Bp*GH50 would perform a similar function as the 'hybrid' agarase; cleavage at the interface between agarose and porphyran units, in order to liberate the each unit type so that they can be further degraded by CAZymes specific to only one substrate type. But unlike the aforementioned hybrid GH16, *Bp*GH50 may digest the smaller products of *endo* enzymes. This would justify its upregulation in the presence of porphyran (Hehemann *et al*, 2012B), but may also help to explain why it may have been a late addition to the PUL. Since its main purpose is predicted to improve overall system efficiency, the system could function without such an enzyme, but may result in the production of small products that would be otherwise indigestible to downstream CAZymes (α- and β-Galactosidases) because of their hybrid topology.

If the activity of *Bp*GH50 can be confirmed, it will be the first porphyran degrading enzyme within the GH50 family. Each β-agarase GH family would also be polyspecific for porphyran, which will help us to further observe the subtle differences between the two functions (Chi *et al*, 2012; Hehemann *et al*, 2010; Van de Velde *et al*, 2002).

Moreover, the gut and soil *Bacteroides* proteins that are homologous to *Bp*GH50, hint at further undiscovered GH50 functionalities.

4.4 Future Work/Conclusions

The objective of this project was to determine the substrate of *Bp*GH50 and deduce where it would fit within the *Bacteroides plebeius* enzymatic degradation pathway for porphyran. While activity analysis indicated no activity on pure agarose or porphyran, structural analysis suggests that *Bp*GH50 degrades the β-linkage of small agarose/porphyran hybrid oligomers. This protein hints at the variety of enzymes required for the complex agaran degradation of the PUL of *B.plebeius* and paints a more complete porphyranolytic degradation picture.

Because the scarcity of a natural hybrid substrate would make the activity particularly difficult to map, the next logical step would be to create the hybrid molecule (whether through a controlled enzyme degradation or synthetically) then soak in mutant GH50T crystals to confirm our hypothesis via a substrate complex.

The understanding of complex agarolytic degradation systems is critical if we are to utilize red seaweed as a potential biofuel source (Chi *et al*, 2012). *Bacteroides plebeius DSM* 17135 provides a unique opportunity to study a specifically porphyran degrading PUL. Until recently, only one other porphyran degradation system (*Z. galactivorans*) has been characterized (Correc *et al*, 2011, Hehemann *et al*, 2012C). Porphyranolytic degradation systems have an advantage of agarolytic systems because they can degrade both agarose and porphyran, exponentially increasing the variety of species that could be used for a biofuels venture. Furthermore, *Bp*GH50 could be an asset even in an agarolytic

system, where porphyran concentrations are low, since it would maximize the agarose turnover by liberating agarose units or their respective monomers from otherwise indigestible porphyran components.

Given the significant shared sequence identity, it would also be fruitful to characterize other homologues of *Bp*GH50 from other gut and soil *Bacteroides* species, in order to determine the ancestor of this enzyme. Because of its widespread distribution, it is unlikely that this progenitor enzyme was an agarase. Given the conservation of the basic pocket among the gut *Bacteroides* surveyed (Appendix D), the protein may instead be involved in the degradation of charged (terrestrial) galactans or galactose containing polymers more plentiful within the human diet, much like the non-agarase members of the polyspecific GH16 family (Lombard *et al*, 2014). This would develop our understanding of these GH50s and further explain substrate divergence within the GH50 family.

Bibliography

Abbott, D.W., Gilbert, H.J., and Boraston, A.B. (2010). The Active Site of Oligogalacturonate Lyase Provides Unique Insights into Cytoplasmic Oligogalacturonate β-Elimination. J. Biol. Chem. *285*, 39029–39038.

Allouch, J., Helbert, W., Henrissat, B., and Czjzek, M. (2004). Parallel Substrate Binding Sites in a β-Agarase Suggest a Novel Mode of Action on Double-Helical Agarose. Structure *12*, 623–632.

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

Anderson, N.S., and Rees, D.A. (1965). 1104. Porphyran: a polysaccharide with a masked repeating structure. J. Chem. Soc. 5880–5887.

Azam, F., and Malfatti, F. (2007). Microbial structuring of marine ecosystems. Nat Rev Micro *5*, 782–791.

Bakir, M.A., Kitahara, M., Sakamoto, M., Matsumoto, M., and Benno, Y. (2006). *Bacteroides intestinalis* sp. nov., isolated from human faeces. Int. J. Syst. Evol. Microbiol. *56*, 151–154.

Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G.W. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallogr. D Biol. Crystallogr. *67*, 271–281.

Boraston, A.B., Bolam, D.N., Gilbert, H.J., and Davies, G.J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem. J. *382*, 769–781.

Braberg, H., Webb, B.M., Tjioe, E., Pieper, U., Sali, A., and Madhusudhan, M.S. (2012). SALIGN: a web server for alignment of multiple protein sequences and structures. Bioinformatics *28*, 2072–2073.

Chassard, C., Delmas, E., Lawson, P.A., and Bernalier-Donadille, A. (2008). *Bacteroides xylanisolvens* sp. nov., a xylan-degrading bacterium isolated from human faeces. Int. J. Syst. Evol. Microbiol. *58*, 1008–1013.

Chi, W.-J., Chang, Y.-K., and Hong, S.-K. (2012). Agar degradation by microorganisms and agar-degrading enzymes. Appl. Microbiol. Biotechnol. *94*, 917–930.

Cohen, R., Orlova, Y., Kovalev, M., Ungar, Y., and Shimoni, E. (2008). Structural and functional properties of amylose complexes with genistein. J. Agric. Food Chem. *56*, 4212–4218.

Conterno, L., Fava, F., Viola, R., and Tuohy, K.M. (2011). Obesity and the gut microbiota: does up-regulating colonic fermentation protect against obesity and metabolic disease? Genes Nutr *6*, 241–260.

Correc, G., Hehemann, J.-H., Czjzek, M., and Helbert, W. (2011). Structural analysis of the degradation products of porphyran digested by *Zobellia galactanivorans* β-porphyranase A. Carbohydrate Polymers *83*, 277–283.

Coutinho, P.M., Deleury, E., Davies, G.J., and Henrissat, B. (2003). An Evolving Hierarchical Family Classification for Glycosyltransferases. Journal of Molecular Biology *328*, 307–317.

Davies, G., and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. Structure *3*, 853–859.

Davies, G.J., Wilson, K.S., and Henrissat, B. (1997). Nomenclature for sugar-binding subsites in glycosyl hydrolases. Biochem J *321*, 557–559.

Davies, G.J., Mackenzie, L., Varrot, A., Dauter, M., Brzozowski, A.M., Schülein, M., and Withers, S.G. (1998). Snapshots along an enzymatic reaction coordinate: analysis of a retaining beta-glycoside hydrolase. Biochemistry *37*, 11707–11713.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Evans, P. (2006). Scaling and assessment of data quality. Acta Crystallogr. D Biol. Crystallogr. *62*, 72–82.

Fu, X.T., and Kim, S.M. (2010). Agarase: review of major sources, categories, purification method, enzyme characteristics and applications. Mar Drugs *8*, 200–218.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocols Handbook, J.M. Walker, ed. (Humana Press), pp. 571–607.

Gouet, P., Robert, X., and Courcelle, E. (2003). ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res. *31*, 3320–3323.

Guindon, S. Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst. Biol. *59*, 307-321.

Gupta, V., Trivedi, N., Kumar, M., Reddy, C.R.K., and Jha, B. (2013). Purification and characterization of exo-β-agarase from an endophytic marine bacterium and its catalytic potential in bioconversion of red algal cell wall polysaccharides into galactans. Biomass and Bioenergy *49*, 290–298.

Hehemann, J.-H., Correc, G., Barbeyron, T., Helbert, W., Czjzek, M., and Michel, G. (2010). Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. Nature *464*, 908–912.

Hehemann, J.-H., Smyth, L., Yadav, A., Vocadlo, D.J., and Boraston, A.B. (2012A). Analysis of keystone enzyme in Agar hydrolysis provides insight into the degradation (of a polysaccharide from) red seaweeds. J. Biol. Chem. *287*, 13985–13995.

Hehemann, J.-H., Kelly, A.G., Pudlo, N.A., Martens, E.C., and Boraston, A.B. (2012B). Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. Proc. Natl. Acad. Sci. U.S.A. *109*, 19786–19791.

Hehemann, J.-H., Correc, G., Thomas, F., Bernard, T., Barbeyron, T., Jam, M., Helbert, W., Michel, G., and Czjzek, M. (2012C). Biochemical and structural characterization of the complex agarolytic enzyme system from the marine bacterium *Zobellia galactanivorans*. J. Biol. Chem. *287*, 30571–30584.

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. *280 (Pt 2)*, 309–316.

Henrissat, B., and Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. Biochem. J. *316 (Pt 2)*, 695–696.

Henrissat, B., and Davies, G. (1997). Structural and sequence-based classification of glycoside hydrolases. Current Opinion in Structural Biology *7*, 637–644.

Howell, L., and Smith,D. (1992) Normal Probability Analysis. J.Appl. Cryst. 25: 81-86

Hsien-Chih, H.W., and Sarko, A. (1978A). The double-helical molecular structure of crystalline a-amylose. Carbohydrate Research *61*, 27–40.

Hsein-Chih, H.W., and Sarko, A. (1978B). The double-helical molecular structure of crystalline b-amylose. Carbohydrate Research *61*, 7–25.

Imberty, A., Chanzy, H., Pérez, S., Bulèon, A., and Tran, V. (1988). The double-helical nature of the crystalline part of A-starch. Journal of Molecular Biology *201*, 365–378.

Ishihara, K., Oyamada, C., Matsushima, R., Murata, M., and Muraoka, T. (2005). Inhibitory effect of porphyran, prepared from dried “Nori”, on contact hypersensitivity in mice. Biosci. Biotechnol. Biochem. *69*, 1824–1830.

Kaoutari, A.E., Armougom, F., Gordon, J.I., Raoult, D., and Henrissat, B. (2013). The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. Nat Rev Micro *11*, 497–504.

Kau, A.L., Ahern, P.P., Griffin, N.W., Goodman, A.L., and Gordon, J.I. (2011). Human nutrition, the gut microbiome and the immune system. Nature *474*, 327–336.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., and Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647-1649

.Kim, H.T., Lee, S., Lee, D., Kim, H.-S., Bang, W.-G., Kim, K.H., and Choi, I.-G. (2010). Overexpression and molecular characterization of Aga50D from *Saccharophagus degradans* 2-40: an exo-type beta-agarase producing neoagarobiose. Appl. Microbiol. Biotechnol. *86*, 227–234.

Kim, H.T., Yun, E.J., Wang, D., Chung, J.H., Choi, I.-G., and Kim, K.H. (2013). High temperature and low acid pretreatment and agarase treatment of agarose for the production of sugar and ethanol from red seaweed biomass. Bioresour. Technol. *136*, 582–587.

Knutsen, S.H., Myslabodski, D.E., Larsen, B., and Usov, A.I. (1994). A Modified System of Nomenclature for Red Algal Galactans. Botanica Marina *37*, 163–170.

Laine, R.A. (1994). Invited Commentary: A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05 × 1012 structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. Glycobiology *4*, 759–767.

Langer, G., Cohen, S.X., Lamzin, V.S., and Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat Protoc *3*, 1171–1179.

Lee, D.-G., Park, G.-T., Kim, N.Y., Lee, E.-J., Jang, M.K., Shin, Y.G., Park, G.-S., Kim, T.-M., Lee, J.-H., Lee, J.-H., et al. (2006). Cloning, expression, and characterization of a glycoside hydrolase family 50 beta-agarase from a marine *Agarivorans* isolate. Biotechnol. Lett. *28*, 1925–1932.

Lever, M. (1972). A new reaction for colorimetric determination of carbohydrates. Anal. Biochem. *47*, 273–279.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. *42*, D490–495.

MacArtain, P., Gill, C.I.R., Brooks, M., Campbell, R., and Rowland, I.R. (2007). Nutritional value of edible seaweeds. Nutr. Rev. *65*, 535–543.

Martens, E.C., Koropatkin, N.M., Smith, T.J., and Gordon, J.I. (2009) Complex Glycan Catabolism by the Human Gut Microbiota: The Bacteroidetes Sus-like Paradigm. J. Biol. Chem. *284*, 24673-24677.

McCandless, E.L., and Craigie, J.S. (1979). Sulfated Polysaccharides in Red and Brown Algae. Annual Review of Plant Physiology *30*, 41–53.

McCarter, J.D., and Withers, S.G. (1994). Mechanisms of enzymatic glycoside hydrolysis. Curr. Opin. Struct. Biol. *4*, 885–892.

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

Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. D Biol. Crystallogr. *67*, 355–367.

Percival, E. (1979). The polysaccharides of green, red and brown seaweeds: Their basic structure, biosynthesis and function. British Phycological Journal *14*, 103–117.

Persin, Z., Stana-Kleinschek, K., Foster, T.J., van Dam, J.E.G., Boeriu, C.G., and Navard, P. (2011). Challenges and opportunities in polysaccharides research and technology: The EPNOE views for the next decade in the areas of materials, food and health care. Carbohydrate Polymers *84*, 22–32.

Pflugrath, J. W. (1999). The finer things in X-ray diffraction data collection. *Acta Crystallogr D Biol Crystallogr 55*, 1718-25.

Pluvinage, B., Hehemann, J.-H., and Boraston, A.B. (2013). Substrate recognition and hydrolysis by a family 50 exo-β-agarase, Aga50D, from the marine bacterium *Saccharophagus degradans*. J. Biol. Chem. *288*, 28078–28088.

Prajapati, V.D., Maheriya, P.M., Jani, G.K., and Solanki, H.K. (2014). Carrageenan: A natural seaweed polysaccharide and its applications. Carbohydr Polym *105C*, 97–112.

Rebuffet, E., Groisillier, A., Thompson, A., Jeudy, A., Barbeyron, T., Czjzek, M., and Michel, G. (2011). Discovery and structural characterization of a novel glycosidase family of marine origin. Environ. Microbiol. *13*, 1253–1270.

Rees, D.A. (1961). Enzymic synthesis of 3:6-anhydro-l-galactose within porphyran from l-galactose 6-sulphate units. Biochem J *81*, 347–352.

Sheldrick, G.M. (2010). Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. Acta Crystallogr. D Biol. Crystallogr. *66*, 479–485.

Sonnenburg, E.D.,Zheng, H., Joglekar, P., Higginbottom,S.K., Firbank, S.J., Bolam, D.N., and Sonnenburg, J.L. (2010) Specificity of polysaccharide use in intestinal Bacteroides species determines diet-induced microbiota alterations. Cell *141*, 1241-1252.

Sugano, Y., Terada, I., Arita, M., Noma, M., and Matsumoto, T. (1993). Purification and characterization of a new agarase from a marine bacterium, *Vibrio* sp. strain JT0107. Appl. Environ. Microbiol. *59*, 1549–1554.

Temuujin, U., Chi, W.-J., Chang, Y.-K., and Hong, S.-K. (2012). Identification and biochemical characterization of Sco3487 from *Streptomyces coelicolor* A3(2), an exo- and endo-type β-agarase-producing neoagarobiose. J. Bacteriol. *194*, 142–149.

Thomas, F., Hehemann, J.-H., Rebuffet, E., Czjzek, M., and Michel, G. (2011). Environmental and gut Bacteroidetes: the food connection. Front Microbiol *2*, 93.

Van de Velde, F., Knutsen, S.H., Usov, A.I., Rollema, H.S., and Cerezo, A.S. (2002). 1H and 13C high resolution NMR spectroscopy of carrageenans: application in research and industry. Trends in Food Science & Technology *13*, 73–92.

Voet, D., and Voet, J.G. (2004). Biochemistry (John Wiley & Sons) 362-366

Yun, E.J., Shin, M.H., Yoon, J.-J., Kim, Y.J., Choi, I.-G., and Kim, K.H. (2011). Production of 3,6-anhydro-l-galactose from agarose by agarolytic enzymes of *Saccharophagus degradans* 2-40. Process Biochemistry *46*, 88–93.

Appendix

Appendix A:
Copyright Policy for Journal of Biological Chemistry

ASBMB

American Society for Biochemistry and Molecular Biology

11200 Rockville Pike
Suite 302
Rockville, Maryland 20852

August 19, 2011

To whom it may concern,

It is the policy of the American Society for Biochemistry and Molecular Biology to allow reuse of any material published in its journals (the Journal of Biological Chemistry, Molecular & Cellular Proteomics and the Journal of Lipid Research) in a thesis or dissertation at no cost and with no explicit permission needed. Please see our copyright permissions page on the journal site for more information.

Best wishes,

Sarah Crespi

American Society for Biochemistry and Molecular Biology
11200 Rockville Pike, Rockville, MD
Suite 302
240-283-6616
JBC | MCP | JLR

Appendix B:
Copyright Policy for Biochemical Journal

How do I find out about rights and permissions?

Rights

For terms and conditions for online usage of journals published by Portland Press Ltd, please visit the following websites:

- Biochemical Journal
- Clinical Science
- Bioscience Reports
- Biochemical Society Transactions
- Essays in Biochemistry
- Biochemical Society Symposia

Licensing enquiries

Permissions

The guidelines below refer to articles published in the following journals, published by Portland Press Limited:

Biochemical Journal	*Cell Signalling Biology*
Biochemical Society Symposia	*Clinical Science*
Biochemical Society Transactions	*Essays in Biochemistry*
Bioscience Reports	*The Biochemist*

Authors of articles published in the above journals

Authors do NOT usually need to contact Portland Press Limited to request permission to reuse their own material, as long as the material is properly credited to the original publication. It is usual to provide the citation of the original publication thus:

"This research was originally published in Journal Name. Author(s), Title. Journal Name. Year; Volume: pp-pp © copyright holder"

The copyright holder for each publication is as follows:

© the Biochemical Society	**© Portland Press Limited**
Biochemical Journal	*Cell Signalling Biology*
Biochemical Society Symposia	
Biochemical Society Transactions	
Bioscience Reports	
Clinical Science	
Essays in Biochemistry	
The Biochemist	

Provided the original publication of the article, or portion of the article, is properly cited, Authors retain the following non-exclusive rights:

1. To reproduce their article in whole or in part in any printed volume (book or thesis) of which they are the Author or Editor
2. To reproduce their article in whole or in part at the Author's current academic institution for teaching purposes
3. To reuse figures, tables, illustrations or photos from the article in commercial or non-commercial works created by them
4. To post a copy of the Immediate Publication (i.e. the Accepted Manuscript*) at the Author's Institutional Repository, 6 months after publication, provided that this is linked to the article on the journal website (e.g. through the DOI).

In addition, authors of Opt2Pay articles, or authors of articles published in our open access journal *Bioscience Reports*, may post the Version of Record** to their Institutional Repository. Portland Press Limited will deposit the Version of Record in PubMed Central on behalf of the author, where applicable.

Requests for non-commercial use

Other parties wishing to use reuse an article in whole or in part for educational purposes (e.g. in lectures or tutorials) may do so at no cost providing the original source is attributed, as outlined above.

You may also reproduce an article in whole or in part in your thesis at no cost providing the original source is attributed.

If you wish to copy and distribute an article in whole for teaching (e.g. in a course pack), please contact your librarian who will advise you on the various clearance options available.

STM Permission Guidelines

Portland Press Limited is a signatory of the **STM Permission Guidelines**, which aims to reduce the administration involved in clearing permissions.

Publishers who are also signatories to the STM Permission Guidelines may re-use work published in Portland Press Limited journals, without contacting us, as long as the work is appropriately used and attributed as stated in the Guidelines.

All other usage requests, including Books

For all other requests please download and complete the **permission request form** and send by email or fax to:

(e) **permissions@portlandpress.com**
(f) +44 (0) 20 7685 2469

---

*Accepted Manuscript – (also known as Immediate Publication) - the version of the article that has been accepted for publication and includes the revisions made following Peer Review.

**Version of Record - the version of the article, after processes such as copyediting, proof corrections, layout and typesetting have been applied.

Please see the **NISO/ALPSP guidelines** for a full description of these terms.

ISSN Print: 0264-6021
ISSN Electronic: 1470-8728

Published by Portland Press
on behalf of
the Biochemical Society

Portland Press Limited
Third floor
Charles Darwin House
12 Roger Street
London WC1N 2JU
Tel: +44(0) 20 7685 2410
Fax: +44(0) 20 7685 2469
Email: editorial@portlandpress.com

Appendix C:
Copyright Policy for the Proceedings of the National Academy of Sciences of the United States of America

Rights and Permissions

Beginning with articles submitted in Volume 106 (2009) the author(s) retains copyright to individual articles, and the National Academy of Sciences of the United States of America retains an exclusive license to publish these articles and holds copyright to the collective work. Volumes 90–105 (1993–2008) copyright © by the National Academy of Sciences. Volumes 1–89 (1915–1992), the author(s) retains copyright to individual articles, and the National Academy of Sciences holds copyright to the collective work.

The PNAS listing on the Sherpa RoMEO publisher copyright policies & self-archiving detail pages can be found here.

Requests for Permission to Reprint

Requests for permission should be made in writing. For the fastest response time, please send your request via e-mail to PNASPermissions@nas.edu. If necessary, requests may be faxed to 202-334-2739 or mailed to:

PNAS Permissions Editor
500 Fifth Street, NW
NAS 340
Washington, DC 20001 USA

Anyone may, without requesting permission, use original figures or tables published in PNAS for noncommercial and educational use (i.e., in a review article, in a book that is not for sale) provided that the original source and the applicable copyright notice are cited.

For permission to reprint material in volumes 1–89 (1915–1992), requests should be addressed to the original authors, who hold the copyright. The full journal reference must be cited.

For permission to reprint material in volumes 90–present (1993–present), requests must be sent via e-mail, fax, or mail and include the following information about the original material:

1. Your full name, affiliation, and title
2. Your complete mailing address, phone number, fax number, and e-mail address
3. PNAS volume number, issue number, and issue date
4. PNAS article title
5. PNAS authors' names
6. Page numbers of items to be reprinted
7. Figure/table number or portion of text to be reprinted

Also include the following information about the intended use of the material:

Also include the following information about the intended use of the material:

1. Title of work in which PNAS material will appear
2. Authors/editors of work
3. Publisher of work
4. Retail price of work
5. Number of copies of work to be produced
6. Intended audience
7. Whether work is for nonprofit or commercial use

PNAS authors need not obtain permission for the following cases: (1) to use their original figures or tables in their future works; (2) to make copies of their papers for their own personal use, including classroom use, or for the personal use of colleagues, provided those copies are not for sale and are not distributed in a systematic way; (3) to include their papers as part of their dissertations; or (4) to use all or part of their articles in printed compilations of their own works. Citation of the original source must be included and copies must include the applicable copyright notice of the original report.

Authors whose work will be reused should be notified. PNAS cannot supply original artwork. Use of PNAS material must not imply any endorsement by PNAS or the National Academy of Sciences. The full journal reference must be cited and, for articles published in Volumes 90–105 (1993–2008), "Copyright (copyright year) National Academy of Sciences, USA."

Requests for Permission to Photocopy

For permission to photocopy beyond that permitted by Section 107 or 108 of the US Copyright Law, contact:

Copyright Clearance Center
222 Rosewood Drive
Danvers, MA 01923 USA
Phone: 1-978-750-8400
Fax: 1-978-750-4770
E-mail: info@copyright.com

Authorization to photocopy items for the internal or personal use of specific clients is granted by The National Academy of Sciences provided that the proper fee is paid directly to CCC.

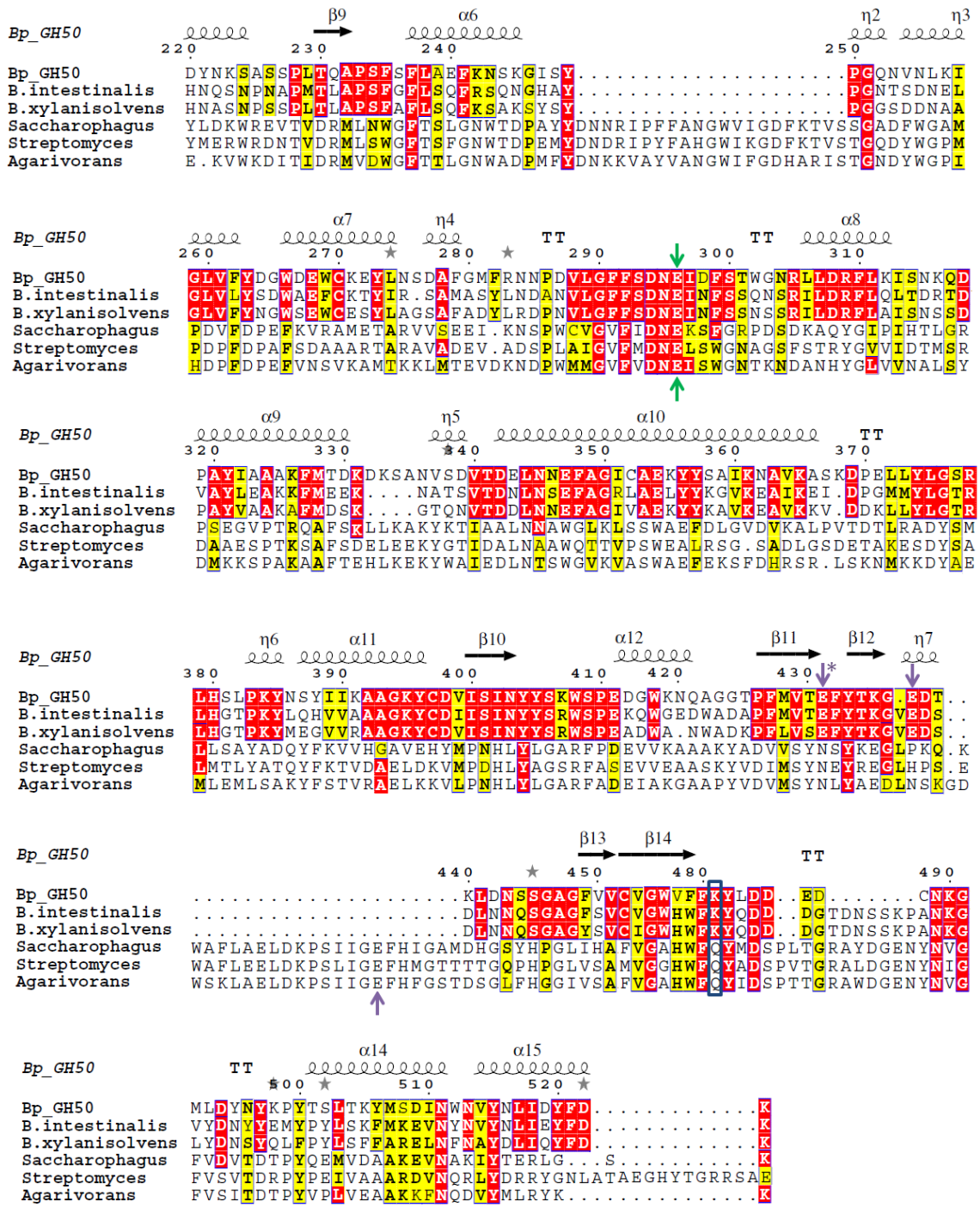
[08/14]

Appendix D:
Complete Sequence Alignment of *Bp*GH50 with other GH50 proteins.

***Bp*GH50 was aligned with uncharacterized proteins from gut *Bacteroides intestinalis* (NCBI ref sequence WP_022392986, 98% coverage and 45% identity) and *Bacteroides xylanisolvens* (Bxy_10650, 97% coverage and 47% identity), and proteins from marine sources; Aga50D from *Saccharophagus degradans* (PDB ID 4QB5, 67% coverage and 28% identity), Sco3487, a β-agarase from *Strepomyces coelicolor A3(2)* (45% coverage and 46% identity) (Temuujin *et al*, 2012), and β- agarase A from *Agarivorans sp. QM38* (83% coverage and 25% identity) (Lee *et al*, 2006). Sequences and statistics extracted from pBLAST analysis (Altschul *et al*, 1997) unless otherwise indicated, alignment by SALIGN and ESPript (Braberg *et al*, 2012; Gouet *et al*, 2003). Catalytic residues of Aga50D (*S. degradans*) are highlighted at the bottom with arrows (green for acid/base, purple for nucleophile). Potential catalytic residues for *Bp*GH50 (E296, E431, and E438) are highlighted by the arrows above (green for acid/base, purple for nucleophiles). The purple asterisk depicts the *Bp*GH50 nucleophile that was ultimately determined to be structurally conserved with the nucleophile of Aga50D. The blue boxes highlight the conservation of 'basic pocket' residues (S170, R172, and K481).**

```
Bp_GH50
                  1                                                  10        20
Bp_GH50           M.................................KYSTIITKTICLSVPFFALSCEDPQE
B.intestinalis    M.................................KIKSFYITASFACLFMLTSCVDYEVK
B.xylanisolvens   M.................................KKLFILGTFLFISSIPMVSCTDDDDK
Saccharophagus    M...........LFDF............ENDQVPSNIHFLNARASIETYTGINGEPSKGL
Streptomyces      M...........TVHK............RACTTPPPRASRSFRVRWPVLIAAACAGLVLA
Agarivorans       MVEVMKFTKNKIAALLSLTLLGVYGCGSTPSSSDAEGAVEDVGGTIPDFESAAFFKKVKK

Bp_GH50
                   30
Bp_GH50           EFSYIPPEV.................................................
B.intestinalis    DPNFMPPDV.................................................
B.xylanisolvens   DPNFMPPDI.................................................
Saccharophagus    KLAMQSKQHSYTGLAIVPEQPWDWSEFT..............SASLYFDIVSVGDHSTQF
Streptomyces      TTSPPAVAAGAHDLGDETMLYDFQDGLVPAEVGPY.......QAKTTIVGRGDKKLRVDF
Agarivorans       DHAKAEVVSDQGVTSGSSALKVNFDSVSEANKFKYWPNVKVHPDSGFWNWNAKGSLSLDI

Bp_GH50

Bp_GH50           ..........................................................
B.intestinalis    ..........................................................
B.xylanisolvens   ..........................................................
Saccharophagus    YLDVTDQNGAVFTR................SIDIPVGKMQSYYAKLSGHDLEVPD.....
Streptomyces      QARKNYYSSFSVRPEPVWNWSAEESESLGIAMELTNPSDRSVQLTIDLESSTGVA.TRSV
Agarivorans       TNPTDSPANIILKLADNVGVMGSGDNQLNYAVNVPAGETVPVEMLFNGTKRKLDGYWGGE
```

```
Bp_GH50                                                                                  40
Bp_GH50          ..................................................IYSQPGN
B.intestinalis   ..................................................VLDEGD.
B.xylanisolvens  ..................................................VMGGGDV
Saccharophagus   ............SGDVN...........DLNLASGLRSNPPTWTS.DDRQFVWMWGVKNL
Streptomyces     NVPAGGGGTYYFDVDSP...........ALHRDTGLRADPSWLADKDVTSAVWMWGSKET
Agarivorans      KIMLPNIVEFQIFVQGPMDAQTVIIDNFNLVDATGDFIEASGQEVKVSGPIPTVASITSF

                                                β1
Bp_GH50                                         →  TT
                          50         60         70
Bp_GH50          VEDYNTPGDNAGDNDDDDVELQLPVPGPAETYPNS.......................
B.intestinalis   DDEIIEGLPTPGEMQAYSPSLLGKPYRPIKVKYSS.......................
B.xylanisolvens  ESEYPEDLPAPGASVMYTPSLNANMYRPISVKYSS.......................
Saccharophagus   DLSGIAKISLSVQSAMHDKTVIIDNIRIQPNPPQDEN....................
Streptomyces     DTSRISQLNFYVAGLLHDRSVIVDDIRVVRDAPADPD.....................
Agarivorans      DEGQPTFVAFDRSAAATVTELKTDMGGLLAVKLAATNAYPNITFKAPQPWDWSEYGDFSL

                 β2      β3          TT      β4      η1      TT        α1        TT
Bp_GH50                  →           →       →       ⌇       →         ⌇         →
                 80         90                 100        110       120       130
Bp_GH50          ...TKQYQPIIVEYAEKPDKAFIEAKTRILPYLVGYEQQTKTQDEYLQSVNKYGSYAKGQ
B.intestinalis   ...............QFPPVASWTEANTRIVAYMGEYKPSIKTESDYKAITNKYGSLTTGA
B.xylanisolvens  ...............AYPPISSWKTENTRIIAYMDGYKPAIKTLKAYQESVNKYGSSTTLP
Saccharophagus   .............FLVGLVDEFGQNAKVDYKGKIHSLEELHAARDVELAELD.GKPMPSR
Streptomyces     .............YLKGLVDAFGQNNKVDYKGKVSRTSEILRQRAAEAKDLRRHPVPEDR
Agarivorans      AFDLESKTDEPLQLFVRVDDAENENWGGTANGVVDSMSSYVTLAPGDDGTFYLPLGQTGS

                                         β5      β6      β7
Bp_GH50                                  →       →       →                 ⌇....
                                         TT      TT
                          140        ★ 150      160        170
Bp_GH50          K...............FKATGRFRVEKNSNGRSWIVDPEGYPYYVRGIASFRMDGNS....
B.intestinalis   K...............QQATGRFYVKK.VNGRWWIIDPEGYPHYERSVTSLRYGSSS....
B.xylanisolvens  K...............QAATGRFYTKK.IDGRWWLVDPEGCLHLERSATSLRKGTSS....
Saccharophagus   SKFGGWLAGP.....KLKATGYFRTEKINGKWMLVDPEGYPYFATGLDIIRLSNSSTMTG
Streptomyces     SRYGGWLNGP.....RLEATGNFRVEKYQGRWTLVDPDGYLFFSTGIDNARMFDSPTTTG
Agarivorans      QIVSGMRAEPPKKSYNAQAISYGWGEKSLDTSNIVSFQLYLQNPTKDAEFNIKSVRLIPN

Bp_GH50          ..............................................................

Bp_GH50          ..............................................................
B.intestinalis   ..............................................................
B.xylanisolvens  ..............................................................
Saccharophagus   YDYDQATVAQRSADDVTPEDSK.......................................
Streptomyces     YDFDHDAIQELPPPSLTAGGPE........................................
Agarivorans      IDADATRYEGLIDQYGQFTGSEWPKKISEDEELETMGKLAKMSLKSTSQMPGRSIYGGWA

Bp_GH50          ..............................................................

Bp_GH50          ..............................................................
B.intestinalis   ..............................................................
B.xylanisolvens  ..............................................................
Saccharophagus   ...............................................GLMAVSEKSFATR
Streptomyces     ...............................................DLNRVQKSALPTR
Agarivorans      DGPKLKGTGFFRTEKVDGKWSLVDPQGNLFFATGVDNIRMDDTVTITGHDFADKDKRSGK

                                α3          α4                  β8          α5
Bp_GH50          ................⌇           ⌇                   →           ⌇
                                180         190        200      210
Bp_GH50          ................SAFGKLYSSVDDWVAKSQKQFSEIGFHSVCAFGKEEGDKAVN
B.intestinalis   ..............RNKEAWNKRFGNDNMWLSKTQSELASIGFHGTGAFCTNTYSKIQT
B.xylanisolvens  ..............RNKTAWNSRFGTDEKWLSITQRELSEIGFHGTGAFCTGTYSLIQT
Saccharophagus   HLASPTRAAMFNWLPDYDHPLANHYNYRRSAHSGPLKRGEAYSFYSANLERKYGETYPGS
Streptomyces     TKMSETRADLFSKLPKYRTRAGEGFGYAPDTLAGPVAQGETYSFYKANVARKYPGSN...
Agarivorans      EVASEVRRSMFTWLPEDDDVLAENYDYANWVHSGALKKGEVFSFYGANLQRKYGGTFSEA
```


Bp_GH50
β9
α6
η2
η3
220 230 240 250
Bp_GH50 DYNKSASSPLTQAPSFSFLAEFKNSKGISY.....PGQNVNLKI
B.intestinalis HNQSNPNAPMTLAPSFGFLSQFRSQNGHAY.....PGNTSDNEL
B.xylanisolvens HNASNPSSPLTLAPSFAFLSQFKSAKSYSY.....PGGSDDNAA
Saccharophagus YLDKWREVTVDRMLNWGFTSLGNWTDPAYYDNNRIPFFANGWVIGDFKTVSSGADFWGAM
Streptomyces YMERWRDNTVDRMLSWGFTSFGNWTDPEMYDNDRIPYFAHGWIKGDFKTVSTGQDYWGPM
Agarivorans E.KVWKDITIDRMVDWGFTTLGNWADPMFYDNKKVAYVANGWIFGDHARISTGNDYWGPI
α7
η4
TT
TT
α8
260 270 280 290 300 310
Bp_GH50 GLVFYDGWDEWCKEYLNSDAFGMFRNNPDVLGFFSDNEIDFSTWGNRLLDRFLKISNKQD
B.intestinalis GLVLYSDWAEFCKTYIR.SAMASYLNDANVLGFFSDNEINFSSQNSRILDRFLQLTDRTD
B.xylanisolvens GLVFYNGWSEWCESYLAGSAFADYLRDPNVLGFFSDNEINFSSNSSRILDRFLAISNSSD
Saccharophagus PDVFDPEFKVRAMETARVVSEEI.KNSPWCVGVFIDNEKSFGRPDSDKAQYGIPIHTLGR
Streptomyces PDPFDPAFSDAAARTARAVADEV.ADSPLAIGVFMDNELSWGNAGSFSTRYGVVIDTMSR
Agarivorans HDPFDPEFVNSVKAMTKKLMTEVDKNDPWMMGVFVDNEISWGNTKNDANHYGLVVNALSY
α9
η5
α10
TT
320 330 340 350 360 370
Bp_GH50 PAYIAAAKFMTDKDKSANVSDVTDELNNEFAGICAEKYYSAIKNAVKASKDPELLYLGSR
B.intestinalis VAYLEAKKFMEEK....NATSVTDNLNSEFAGRLAELYYKGVKEAIKEI.DPGMMYLGTR
B.xylanisolvens PAYVAAKAFMDSK....GTQNVTDDLNNEFAGIVAEKYYKAVKEAVKKV.DDKLLYLGTR
Saccharophagus PSEGVPTRQAFSKLLKAKYKTIAALNNAWGLKLSSWAEFDLGVDVKALPVTDTLRADYSM
Streptomyces DAAESPTKSAFSDELEEKYGTIDALNAAWQTTVPSWEALRSG.SADLGSDETAKESDYSA
Agarivorans DMKKSPAKAAFTEHLKEKYWAIEDLNTSWGVKVASWAEFEKSFDHRSR.LSKNMKKDYAE
η6
α11
β10
α12
β11
β12
η7
380 390 400 410 420 430
Bp_GH50 LHSLPKYNSYIIKAAGKYCDVISINYYSKWSPEDGWKNQAGGTPFMVTEFYTKG.EDT..
B.intestinalis LHGTPKYLQHVVAAAGKYCDIISINYYSRWSPEKQWGEDWADAPFMVTEFYTKGVEDS..
B.xylanisolvens LHGTPKYMEGVVRAAGKYCDVISINYYSRWSPEADWA.NWADKPFLVSEFYTKGVEDS..
Saccharophagus LLSAYADQYFKVVHGAVEHYMPNHLYLGARFPDEVVKAAAKYADVVSYNSYKEGLPKQ.K
Streptomyces LMTLYATQYFKTVDAELDKVMPDHLYAGSRFASEVVEAASKYVDIMSYNEYREGLHPS.E
Agarivorans MLEMLSAKYFSTVRAELKKVLPNHLYLGARFADEIAKGAAPYVDVMSYNLYAEDLNSKGD
β13
β14
TT
440 450 480 490
Bp_GH50 .....KLDNSSGAGFVVCVGWVFFKYLDD..ED.....CNKG
B.intestinalis .....DLNNQSGAGFSVCVGWHWFKYQDD..DGTDNSSKPANKG
B.xylanisolvens .....DLNNQSGAGYSVCIGWHWFKYQDD..DGTDNSSKPANKG
Saccharophagus WAFLAELDKPSIIGEFHIGAMDHGSYHPGLIHAFVGAHWFQYMDSPLTGRAYDGENYNVG
Streptomyces WAFLEELDKPSLIGEFHMGTTTTGQPHPGLVSAMVGGHWFQYADSPVTGRALDGENYNIG
Agarivorans WSKLAELDKPSIIGEFHFGSTDSGLFHGGIVSAFVGAHWFQYIDSPTTGRAWDGENYNVG
α14
α15
TT
500 510 520
Bp_GH50 MLDYNYKPYTSLTKYMSDINWNVYNLIDYFD.....K
B.intestinalis VYDNYYEMYPYLSKFMKEVNYNVYNLIEYFD.....K
B.xylanisolvens LYDNSYQLFPYLSFFARELNFNAYDLIQYFD.....K
Saccharophagus FVDVTDTPYQEMVDAAKEVNAKIYTERLG...S.....K
Streptomyces FVSVTDRPYPEIVAAARDVNQRLYDRRYGNLATAEGHYTGRRSAE
Agarivorans FVSITDTPYVPLVEAAKKFNQDVYMLRYK.....K