

# Analysis of Keystone Enzyme in Agar Hydrolysis Provides Insight into the Degradation of a Polysaccharide from Red Seaweeds<sup>\*[5]</sup>

Received for publication, January 23, 2012, and in revised form, February 29, 2012. Published, JBC Papers in Press, March 5, 2012, DOI 10.1074/jbc.M112.345645

Jan-Hendrik Hehemann<sup>†1</sup>, Leo Smyth<sup>‡</sup>, Anuj Yadav<sup>§</sup>, David J. Vocadlo<sup>§2</sup>, and Alisdair B. Boraston<sup>‡3</sup>

From the <sup>†</sup>Department of Biochemistry and Microbiology, University of Victoria, P. O. Box 3055 STN CSC, Victoria, British Columbia V8W 3P6, Canada and <sup>§</sup>Department of Chemistry, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia V5A 1S6, Canada

**Background:** The catalytic mechanism and substrate recognition required for cleavage of the  $\alpha$ -linkage in agarose are unclear.

**Results:** Structural analysis of a family 117 glycoside hydrolase details substrate recognition and supports an inverting mechanism.

**Conclusion:** GH117 enzymes use substrate distortion and an unusual general acid for catalysis.

**Significance:** Microbes may utilize alternate strategies to catalyze the degradation of polysaccharides with unique structural characteristics.

Agars are abundant polysaccharides from marine red algae, and their chemical structure consists of alternating D-galactose and 3,6-anhydro-L-galactose residues, the latter of which are presumed to make the polymer recalcitrant to degradation by most terrestrial bacteria. Here we study a family 117 glycoside hydrolase (*BpGH117*) encoded within a recently discovered locus from the human gut bacterium *Bacteroides plebeius*. Consistent with this locus being involved in agarocolloid degradation, we show that *BpGH117* is an exo-acting 3,6-anhydro- $\alpha$ -(1,3)-L-galactosidase that removes the 3,6-anhydrogalactose from the non-reducing end of neoagar-oligosaccharides. A Michaelis complex of *BpGH117* with neoagarobiose reveals the distortion of the constrained 3,6-anhydro-L-galactose into a conformation that favors catalysis. Furthermore, this complex, supported by analysis of site-directed mutants, provides evidence for an organization of the active site and positioning of the catalytic residues that are consistent with an inverting mechanism of catalysis and suggests that a histidine residue acts as the general acid. This latter feature differs from the vast majority of glycoside hydrolases, which use a carboxylic acid, highlighting the alternative strategies that enzymes may utilize in catalyzing the cleavage of glycosidic bonds.

Algae are the plants of the ocean; however, they produce oligo- and polysaccharides that are structurally and chemically distinct from those found on land. Although microbial degradation of these algal carbohydrates is of ecological importance and of increasing interest for biofuel applications (1), how bacteria depolymerize and metabolize algal carbohydrates remains unclear in many cases. Like the cell walls of terrestrial plants, the cell wall and storage polysaccharides of marine macroalgae comprise a massive sink of biomass and photosynthetically fixed carbon (2). Unlike the cell walls of terrestrial plants, however, the cell walls of macroalgae either lack lignin or are much less lignified. Furthermore, their matrix polysaccharides form flexible gels instead of crystalline polysaccharides like those of terrestrial plants, and this confers important properties to macroalgae (3). First, it permits dehydration and rehydration of the seaweed cell wall and thereby protects against desiccation occurring in the intertidal zone. Second, it enables the macroalgae to remain flexible enough to withstand the physical challenge imposed by this particular environment (*e.g.* waves and tides). Agars are the dominant gel-forming cell wall matrix polysaccharide (70%) in red seaweeds (agarophytes) and comprise a class of heterogeneous polymers having various degrees of sulfation and other substitutions including O-methylation. One of the major constituents of the cell wall of agarophytes, however, is the neutral and linear polysaccharide termed agarose (4, 5).

Agarose is made of neoagarobiose motifs in which 3,6-anhydro-L-galactose is  $\alpha$ -1,3-linked to D-galactose; the resulting disaccharide is linked by  $\beta$ -1,4-glycosidic bonds to form the linear polymer (see Fig. 1A). Agarose forms a helical structure stabilized by intra- and intermolecular hydrogen bonds to generate a high gel strength polymer. The 3,6-anhydro-L-galactose, which is a bridged bicyclic system containing fused five- and six-membered rings, is critical for helix formation. Polysaccharides that lack this modification, like porphyran, which is a sulfated agarocolloid related to agarose, form viscous solutions

\* This work was supported in part by a Natural Sciences and Engineering Research Council of Canada discovery grant.

[5] This article contains supplemental Figs. 1 and 2 and Table 1.

The atomic coordinates and structure factors (codes 4ak5, 4ak6, and 4ak7) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

<sup>1</sup> Supported by a European Molecular Biology Organization long term fellowship.

<sup>2</sup> Canada Research Chair in Chemical Glycobiology and a Michael Smith Foundation for Health Research (MSFHR) Career Scholar.

<sup>3</sup> Canada Research Chair in Molecular Interactions and an MSFHR Career Scholar. To whom correspondence should be addressed. Tel.: 250-472-4168; Fax: 250-721-8855; E-mail: boraston@uvic.ca.

This is an open access article under the [CC BY](https://creativecommons.org/licenses/by/4.0/) license.

## Analysis of $\alpha$ -Agarase from *B. plebeius*

instead of flexible gels (6). The high gel strength of agarose and its resistance to degradation by many microbes have supported its wide utilization in microbiological, molecular biological, and food applications (7). However, marine bacteria have evolved systems of metabolic enzymes to enable utilization of this abundant marine polymer as both a carbon and energy source (4). These metabolic pathways are initiated by specific carbohydrate-active enzymes (CAZymes)<sup>4</sup> (8) that are responsible for the initial depolymerization steps.

The largest class of carbohydrate-degrading CAZymes are the glycoside hydrolases, which hydrolyze glycosidic bonds found within high molecular weight carbohydrates to produce mono-, di-, and short oligosaccharides. These enzymes are ubiquitous throughout all domains of life. Currently, 130 families of GHs are classified in the CAZy database (Carbohydrate-Active Enzyme database) (8). Enzymes within one family are related by amino acid sequence and thus share common protein folds and catalytic mechanisms (8). In general, glycoside hydrolases catalyze hydrolysis of the glycosidic bond using either a two-step double displacement mechanism resulting in retention of stereochemistry at the anomeric carbon or a one-step, single displacement mechanism that leads to a product with inverted stereochemistry at the anomeric center (9).

The highly polymerized nature of agarose, the presence of both  $\alpha$ - and  $\beta$ -bonds in its backbone, and its composition of two different monosaccharide units create a challenge for its degradation by CAZymes. As such, both endo-acting (cleaving internally in the backbone) and exo-acting (acting only on the chain ends) CAZymes that cleave the  $\beta$ -1,4-glycosidic bond of agarose (Fig. 1A) have been described; these belong to glycoside hydrolase families 16, 50, 86, and 118. CAZymes that cleave the  $\alpha$ -1,3-glycosidic bond presently belong to families 96 and 117.  $\beta$ -Agarases are relatively well studied, and much mechanistic and structural insight into the cleavage of the  $\beta$ -bond in agarose has been gained (4, 10, 11); however, comparatively little is known about the  $\alpha$ -agarases (12).

Family GH96 comprises a unique but very small family of endo- $\alpha$ -agarases. In contrast, family GH117 is substantially larger and more widely distributed in nature, suggesting a more general role of this enzyme family in agarose metabolism. Members of this family are known to cleave the  $\alpha$ -1,3-glycosidic bond in neoagarose oligosaccharides to release 3,6-anhydro-L-galactose. This family was created based on the identification of the neoagarobiose hydrolase from *Saccharophagus degradans* (13), whereas the first structural and bioinformatic description for family GH117 was provided for Zg4663 (ZgAhgA) from the marine flavobacterium *Zobellia galactanivorans* (14). Proteins belonging to GH117 are the only known enzymes that release monosaccharides from agaro-oligosaccharides; thus, they are of critical importance for polysaccharide utilization in agarolytic bacteria. Most GH117 enzymes are found in bacteria associated with seaweed, coral, and the intestines of fish, which is consistent with the abundance of agarose

**TABLE 1**  
Oligonucleotide primers used for gene amplification and mutagenesis

Primer name	Sequence
BpGH117F	5'-CATATGGCTAGCAATACAGACTATTGCCG-3'
BpGH117R	5'-GTGTGTCTCGAGTTATTTACGTTCTTTTGATTCA-3'
D90NF	5'-GTAGTAAGAAGAAACCCGAGTGC AATGCTG-3'
D90NR	5'-CAGCATTGCACCTCGGGTTCTTCTTACTAC-3'
E167QF	5'-CTGTTTTCACTCCTAAGGTAATGGAATGG-3'
E167QR	5'-CCATTCATTACCTTAGGAGTGAAAACAG-3'
D245NF	5'-CATAAGGTACACAACCCGTGTATAATACC-3'
D245NR	5'-GGTATTATACACGGGTTGTGTACCTTATG-3'
E303QF	5'-CAAACAGTGGGCATCAAGTCTGCGTCTGG-3'
E303QR	5'-CCAGACGCAGACTTGATGCCACTGTTTG-3'
H302QF	5'-CAAACAGTGGGCAAGAAGTCTGCGTCTGG-3'
H302QR	5'-CCAGACGCAGACTTCTTGCCACTGTTTG-3'
H302EF	5'-CAAACAGTGGGGAAGAAGTCTGCGTCTGG-3'
H302ER	5'-CCAGACGCAGACTTCTTCCCCTGTTTG-3'

in the marine environment, but surprisingly these enzymes are also found, albeit with lower frequency, in bacteria from soil, the oral cavity, and the human intestine (14, 15).

The x-ray crystal structures of two GH117 enzymes, ZgAhgA from *Z. galactanivorans* (14) and SdGH117 from *S. degradans* (13), revealed the structural similarity between GH117 and GH43 enzymes (16). In particular, conservation of active site residues has led to the postulation that the GH117 enzymes share an inverting catalytic mechanism with the GH43 enzymes (14). Significant alterations in the active sites of the two enzyme families, however, have kept predictions about the general mechanism of catalysis and the identity of the catalytic residues in GH117 speculative. Here we present a structural and mutational analysis of the GH117 enzyme from *Bacteroides plebeius* called BpGH117. The x-ray crystallographic analysis of an inactive mutant of the enzyme in complex with unhydrolyzed substrate reveals the complex suite of residues involved in recognizing substrate and hydrolyzing the glycosidic bond and provides strong support for the proposed inverting hydrolytic mechanism. This analysis reveals distortion of the constrained 3,6-anhydro-L-galactose residue to a conformation conducive to hydrolysis of the glycosidic bond. In the proposed inverting catalytic mechanism, a histidine residue is suggested to act as the general acid, which would be the first observation of its kind for a glycoside hydrolase and would add to the recent example of a histidine acting as the acid/base catalytic residue for the retaining catalytic mechanism used by the *N*-acetylglucosaminidase from *Bacillus subtilis* (17).

## EXPERIMENTAL PROCEDURES

**Gene Cloning, Protein Production, and Protein Purification**—The portion of the gene encoding BpGH117 (locus tag Bacple\_1671) without the signal peptide, which was identified by the SignalP server (18), was amplified by polymerase chain reaction from *B. plebeius* genomic DNA and comprised amino acid residues 22–402 (see Table 1 for primer sequences). The DNA that served as template for the PCR was extracted from *B. plebeius* obtained from DMSZ GmbH. The PCR product was cloned into pET-28 using standard directional cloning procedures using the restriction enzymes NheI and XhoI. The sequence was verified by bidirectional sequencing. The encoded protein, referred to as BpGH117, contained an N-terminal 6-histidine tag separated from the domain of interest by a

<sup>4</sup> The abbreviations used are: CAZymes, carbohydrate-active enzymes; GH, glycoside hydrolase; Bis-Tris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol; G residue, D-galactose standard; LA residue, 3,6-anhydro-L-galactose.

thrombin cleavage site. This cloned gene construct was used as a template for site-directed mutagenesis using the QuikChange site-directed mutagenesis kit (Stratagene catalogue number 200518). The standard protocol was used with the primers listed in Table 1. All mutations were verified by bidirectional DNA sequencing.

The protein production and purification protocol was identical for *BpGH117* and mutants. Plasmids were transformed into chemically competent *Escherichia coli* BL21(DE3) Star cells (Invitrogen). For protein expression, 5 ml of LB medium supplemented with 50 mg/ml kanamycin was inoculated with a single bacterial colony and incubated at 30 °C for 8 h. One milliliter was used to inoculate 1 liter of autoinduction medium (19), and the cultures were incubated with shaking for 5 days at 20 °C.

Cells were harvested by centrifugation and resuspended in 25 ml of sucrose solution (25% w/v), 50 mM Tris-HCl, pH 8.0. Ten milligrams of lysozyme was added to the suspension, and stirring was continued for 10 min. 50 ml of a deoxycholate solution containing 1% deoxycholate (w/v), 1% Triton X-100 (v/v), 20 mM Tris-HCl, pH 7.5, 100 mM NaCl was added with continued stirring for 10 min. The solution was adjusted to 5 mM MgCl<sub>2</sub>, and 2 mg of DNase was added. Cellular debris was separated by centrifugation at 15,000 rpm for 45 min at 4 °C.

The supernatant was applied onto a 2-ml Ni<sup>2+</sup>-charged nitrilotriacetic acid-Sepharose resin (GE Healthcare), which was equilibrated with binding buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl). Proteins were eluted with a linear gradient of imidazole from 0 to 500 mM in binding buffer. Fractions were analyzed by SDS-PAGE, and fractions containing protein were pooled and dialyzed overnight into 20 mM Tris-HCl, pH 8.0. After dialysis, the proteins were further purified by anion exchange chromatography on a Bio-Rad UNO Q6 column. Proteins were eluted in 20 mM Tris-HCl, pH 8.0 with a linear gradient to 1 M NaCl; protein-containing fractions were concentrated in a stirred cell ultrafiltration unit using a 5-kDa molecular mass cutoff membrane. Following concentration, a final polishing step of size exclusion chromatography was performed on a Sephacryl 16/60 Hiprep S-200 column using 20 mM Tris-HCl, pH 8.0. The proteins were concentrated in a stirred ultrafiltration unit prior to crystallization and activity measurements. Selenomethionine-containing protein was produced in *E. coli* B834(DE3) cells in minimal autoinduction medium that lacked yeast extract and tryptone and was supplemented with selenomethionine (40 mg/liter) and the SelenoMet Nutrient Mix (5.1 g/liter), both from Athena Enzyme Systems.

**Crystallization and Data Collection**—Crystallization experiments were performed at 18 °C by the vapor diffusion method using both hanging drops and sitting drops. Crystals of *BpGH117* and mutants were obtained in conditions comprising polyethylene glycol 3350 at 15–25% (w/v) with 100 mM ammonium acetate (Sigma), 100 mM Bis-Tris, pH 5.5 (Sigma). Prior to crystallization, mother liquor and protein solution were mixed and centrifuged for 30 min at 12,000 × *g* and 4 °C to remove precipitated protein; crystallization experiments were set up with the supernatant. The selenomethionine derivative of *BpGH117* was crystallized in 15% PEG 3350 with 100 mM ammonium acetate, 100 mM Bis-Tris, pH 5.5 by streak seeding

with microseeds of native crystals. A complex with neoagarobiose was obtained by co-crystallizing *BpGH117E303Q* (at 10 mg/ml) with 2 mM neoagarobiose in 0.1 M HEPES, pH 7.5, 0.15 M ammonium acetate, 17–21% PEG 3350.

The crystals were cryoprotected by soaking in crystallization solution supplemented with 20% ethylene glycol or glycerol. The crystals were flash frozen in liquid nitrogen, and data collection was carried out on either BL9-2 at the Stanford Synchrotron Radiation Laboratories (Stanford, CA) or the CMCF1 beam line at the Canadian Light Source (Saskatoon, Canada). Data were processed using MOSFLM (20) and SCALA (21) as part of the CCP4 suite (22).

**Structure Solution and Model Refinement**—The structure of *BpGH117* was solved by the single wavelength anomalous dispersion method using an x-ray wavelength optimized for the selenium peak determined by a fluorescence scan ( $\lambda = 0.97911$ ). SHELXC/D was used to determine the selenium substructure, which comprised 20 of the possible 22 total seleniums present in the two *BpGH117* monomers in the asymmetric unit (23). Phasing was performed with SHARP, which resulted in figures of merit of 0.43 and 0.17 for acentric and centric reflections, respectively, and a phasing power of 1.5 over the full resolution range of the data set (24). The 2-fold non-crystallographic symmetry operators were determined using PROFESS followed by phase improvement by solvent flattening (with an estimated solvent content of 44%) and non-crystallographic symmetry averaging using DM (22). These phases were sufficient for automatic model building of a nearly complete model using ARP/wARP (25). This model was corrected and completed by refinement with REFMAC (26) against the high resolution native data set followed by model building using Coot (27). The completed model was used as a starting point to solve the structures of the neoagarobiose complex and mutant proteins. In all cases, water molecules were added using Coot Find Waters and inspected visually prior to deposition. Five percent of the observations were flagged as “free” and used to monitor refinement procedures (28). Model validation was performed with MolProbity (29). All data collection and model statistics are given in Table 2. Coordinates and structure factors have been deposited in the Protein Data Bank with the following accession codes: native *BpGH117*, 4ak5; *BpGH117H302E* mutant, 4ak6; and *BpGH117E303Q* mutant in complex with neoagarobiose, 4ak7. Figures were prepared using PyMOL Molecular Viewer.

**Activity Assays**—Detection of enzyme activity by TLC was performed as described previously (30). The activity of the *BpGH117* mutants was compared with that of the native protein by a time course digestion assay with an enzyme concentration of 0.5  $\mu$ M and neoagarobiose at a concentration of 0.5% (1.54 mM) in 50 mM NaPO<sub>4</sub>, pH 7.0. Aliquots were taken from each reaction at 5-min intervals and added to an equal volume of 0.1 M NaOH to stop the reaction. The samples were subsequently diluted into sterile distilled water to a final sugar concentration of 20 mM. The reactant and product sugars were then separated by high performance anion exchange chromatography with pulsed amperometric detection using a Dionex ICS 3000 HPLC equipped with an ASI 100 automated sample injector and an ED50 electrochemical detector (Dionex) with a

# Analysis of $\alpha$ -Agarase from *B. plebeius*

**TABLE 2**

Data collection and refinement statistics for BpGH117

r.m.s., root mean square; N/A, not applicable.

	BpGH117 Se-Met	BpGH117 native	BpGH117E303Q + neoagarbiose	BpGH117H302E
<b>Data collection</b>				
Beamline	SSRL BL9-2	SSRL BL9-2	CLS 08ID-1	SSRL BL9-2
Wavelength (Å)	0.97911	0.95369	0.97949	1.0
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Cell dimensions ( <i>a, b, c</i> ) (Å)	83.96, 92.97, 105.15	83.99, 93.00, 105.54	83.90, 93.55, 104.96	85.26, 96.55, 103.44
Resolution (Å)	40.0–2.40 (2.53–2.40) <sup>a</sup>	37.96–1.70 (1.79–1.70)	40.0–1.80 (1.90–1.80)	48.28–1.90 (2.0–1.90)
<i>R</i> <sub>merge</sub>	0.142 (0.455)	0.072 (0.374)	0.086 (0.255)	0.065 (0.347)
<i>I</i> / $\sigma$ <i>I</i>	14.7 (5.5)	17.0 (5.1)	14.7 (5.5)	16.1 (3.8)
Completeness (%)	100.0 (99.9)	99.8 (100.0)	95.0 (73.2)	99.8 (99.4)
Redundancy	14.3 (14.1)	7.4 (7.4)	6.7 (4.6)	6.2 (5.8)
<b>Refinement</b>				
Resolution (Å)		1.70	1.80	1.90
No. reflections		86,666	77,119	88,357
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>		0.17/0.21	0.16/0.21	0.22/0.25
No. atoms				
Protein		2,941 (A); 2,948 (B)	2,943 (A); 2,979 (B)	2,925 (A); 2,861 (B)
Ligand <sup>b</sup>		8	22 (sug_A); 22 (sug_B)	
Water		674	784	408
B-factors				
Protein		18.2 (A); 17.4 (B)	14.4 (A); 14.4 (B)	32.4 (A); 29.9 (B)
Ligand <sup>b</sup>		N/A	13.3 (sug_A); 14.9 (sug_B)	N/A
Water		24.5	22.5	33.2
r.m.s. deviations				
Bond lengths (Å)		0.0103	0.0145	0.0072
Bond angles (°)		1.1214	1.3494	0.9180
Ramachandran				
Preferred (%)		95.8	96.3	93.8
Generously allowed (%)		3.7	3.0	5.3
Outliers (%)		0.5	0.7	1.0

<sup>a</sup> Values in parentheses are for the highest resolution bin.

<sup>b</sup> Refers to carbohydrates and carbohydrate derivatives associated with the A and B protein chains in the asymmetric unit.

gold working electrode and an Ag/AgCl reference electrode. Products were analyzed using a CarboPac PA-20 column set (analytical plus guard column). The running buffer comprised 0.1 M NaOH, and elution of the sugars was achieved using a linear gradient of 0–300 mM sodium acetate in 0.1 M NaOH.

**Bioinformatics**—Homologues of BpGH117 were extracted from GenBank<sup>TM</sup> through Phylogeny.fr (31). The calculation of conservation scores was carried out with ConSurf (32) using 33 amino acid sequences, and the conservation was shown on the BpGH117 structure. The phylogenetic tree was constructed with MEGA5 (33) using the same data set that was used for the ConSurf calculation. The phylogenetic tree was calculated with the maximum likelihood method. A model for a member of Clade C was created using the Phyre server (34) with BpGH117 and the sequence of a homologue from Clade C, Zg3615 from *Z. galactanivorans*, which shares 40% sequence identity with BpGH117.

## RESULTS

**BpGH117 Is  $\alpha$ -1,3-3,6-Anhydro-L-galactosidase**—BpGH117 was overproduced by heterologous expression in *E. coli* and purified to homogeneity with three chromatographic steps. The activity of this recombinant protein on neoagaro-oligosaccharides was examined using thin layer chromatography to visualize the reaction products. Two products, one of which migrated faster than the D-galactose standard (G residue), appeared from degradation of neoagarbiose (LA-G~), neoagarotetraose (LA-G-LA-G~), and neoagarohexaose (LA-G-LA-G-LA-G~) (Fig. 1B). This new high mobility product showed the typical color (not shown) and ratio-to-front (*R*<sub>f</sub>) of 3,6-anhydro-L-galactose (LA residue) as observed with the res-

orcinol staining procedure (30). The lower running bands could therefore be interpreted as G~, G-LA-G~, and G-LA-G-LA-G~ products. This product pattern is identical to the reaction profiles observed for ZgAhgA (14) and the *S. degradans* enzyme described by Suzuki *et al.* (35), indicating that BpGH117 possesses the same activity as these other two GH117 enzymes and therefore cleaves the  $\alpha$ -1,3-glycosidic bond to release 3,6-anhydro-L-galactose from the non-reducing end of neoagaro-oligosaccharides.

**General Structural Features of BpGH117**—The x-ray crystal structure of BpGH117, determined to 1.7-Å resolution, revealed a dimer in the asymmetric unit (Fig. 2, A and B). Residues 33–401 for one molecule (Chain A) and residues 30–402 for the second molecule in the dimer (Chain B) could be continuously modeled into the electron density. Each monomer folds into a globular 5-fold  $\beta$ -propeller comprising five  $\beta$ -sheets arranged radially around a central axis (Fig. 2, A and B). The  $\beta$ -sheets contain three to four antiparallel  $\beta$ -strands. BpGH117 forms a dimer with a total molecular interface of 8700 Å<sup>2</sup> of buried surface area, ~7500 Å<sup>2</sup> of which is contributed through interactions of N- and C-terminal extensions (Fig. 2, A and B). The structures of ZgAhgA and BpGH117, superposed as their dimers, display a root mean square deviation of 1.0 Å (36) over 708 aligned residues. This is consistent with their high amino acid sequence identity (>60%) and also reveals both the high structural identity of the monomers and the organization of the dimer. Similar to ZgAhgA, the N-terminal 28 residues of BpGH117 (residues 33–59) comprise a helix-turn-helix domain that reaches over to interact with the other monomer in the dimer. Thus, the main function of the helix-turn-helix domain

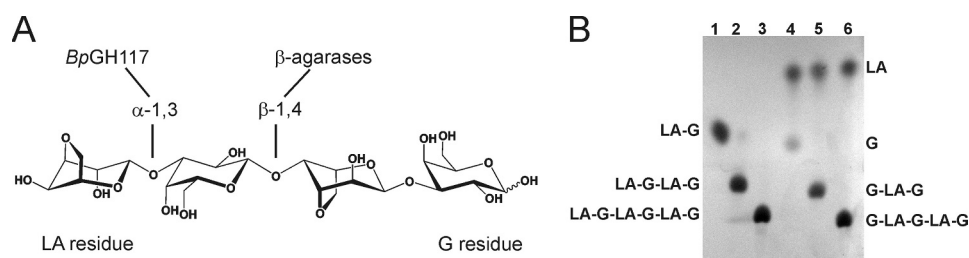


FIGURE 1. ***BpGH117* shows neoagarobiose hydrolase activity.** *A*, neoagaro-oligosaccharides, here represented by neoagarotetraose, as substrates for  $\beta$ -agarases and  $\alpha$ -agarases (i.e. GH117). The LA residue at the non-reducing end is degraded to LA and G-LA-G by *BpGH117*. *B*, TLC analysis of neoagarose oligosaccharide degradation by *BpGH117*. *Lanes 1, 2, and 3* show the untreated standards, whereas *lanes 4, 5, and 6* show the products resulting from treatment of the di-, tetra-, and hexasaccharide standards with *BpGH117*.

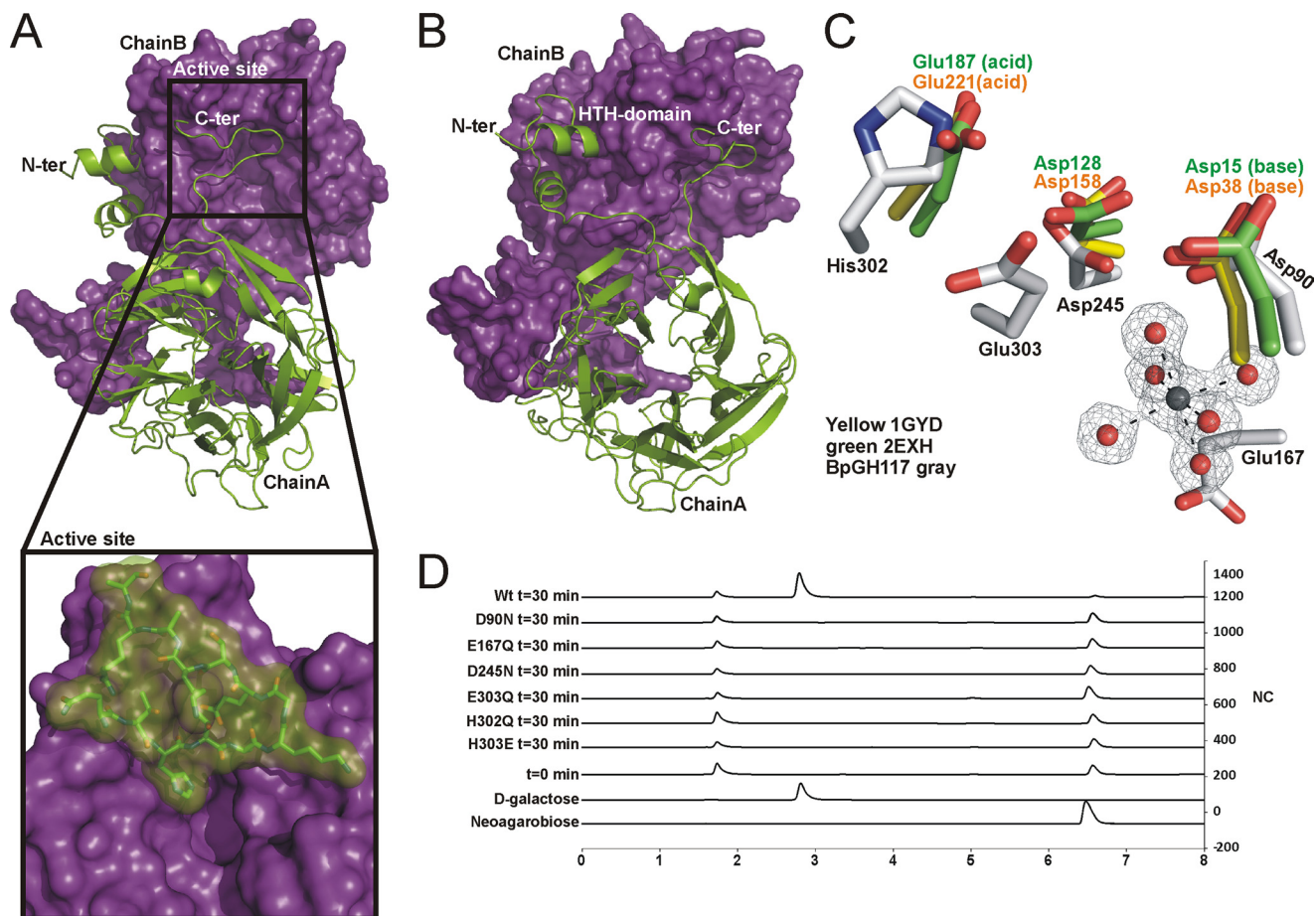


FIGURE 2. ***BpGH117* forms dimer in which both monomers contribute to architecture of active site.** *A* and *B*, the structure of the *BpGH117* dimer. Chain A (green) is shown in schematic representation, and Chain B (purple) is shown as a solvent-accessible surface representation. *A* and the inset show the intimate association of the C-terminal extensions with the neighboring monomer. *B* shows a slightly rotated view of the dimer to show both the N- and C-terminal interactions as well as the  $\beta$ -propeller fold of the monomer. *C*, comparison of the active site of *BpGH117* (gray) with the catalytic residues in a GH43 arabinase (Protein Data Bank code 1gyd; yellow) and xylanase (Protein Data Bank code 2exh; green) and the proximity of the active site to the magnesium/water cluster. *D*, high performance anion exchange chromatography with pulsed amperometric detection analysis of neoagarobiose hydrolysis by *BpGH117* and various active site mutants. The bottom two traces represent relevant standards, whereas the third trace from the bottom shows a representative time 0 sample before substrate is hydrolyzed. Additional traces show the chromatograms for samples taken after 30 min of enzyme treatment. HTH, helix-turn-helix. NC, nanocoulombs.

in both enzymes is suggested to be in stabilizing the dimer by formation of extensive interactions between the two monomers. However, a notable difference between *BpGH117* and *ZgAhgA* is that a 12-residue segment at the C terminus of *BpGH117* (residues 390–401) was found to be ordered and formed a loop structure that, like the N-terminal helix-turn-helix, interacts with the surface of the adjacent monomer in the dimer, thus increasing the contact surface between the monomers and likely stabilizing its quaternary structure. A similar set

of interactions was observed for the dimer of the GH117 enzyme *SdGH117* from *S. degradans* (37).

The five blades of the *BpGH117*  $\beta$ -propeller fold around a deep central cavity. In structurally related glycoside hydrolases from Clans F (GH43 and GH62) and J (GH32 and GH68), a similarly positioned pocket harbors the active site (16, 38, 39). In *BpGH117* and *ZgAhgA*, this central cavity houses a water/metal cluster with the metal being completely coordinated by water molecules in its inner coordination sphere (Fig. 2C)

## Analysis of $\alpha$ -Agarase from *B. plebeius*

(supplemental Fig. 1). On the basis of the bond lengths (2.05–2.19 Å), B-factors, and coordination geometry, the metal atom bound to the *Bp*GH117 monomers was modeled as a  $Mg^{2+}$  ion. This choice of metal ion was further supported by the absence of any detectable anomalous difference peaks for the atom in data sets collected with x-ray wavelengths of 0.95–0.98 or 1.5419 Å (unreported data sets were ultimately superseded by higher resolution synchrotron data). The metal binding site is absolutely conserved. In the case of *Zg*AhgA, however, the ion was modeled as  $Zn^{2+}$ ; a fluorescence scan of *Bp*GH117 crystals at the zinc edge produced no signal, supporting the identity of the *Bp*GH117 bound ion as something other than  $Zn^{2+}$ , namely  $Mg^{2+}$ . Likewise, this metal site is relatively well conserved with the structurally related GH43 enzymes like the  $\alpha$ -L-arabinase Arb43A from *Cellvibrio japonicus* of GH43 (Protein Data Bank code 1gyd) in which it was modeled as a chloride ion (16). This and the appearance of  $Zn^{2+}$  in *Zg*AhgA imply a degree of plasticity in the use of ions by these structurally similar and functionally related enzymes.

The coordination sphere of water molecules surrounding the  $Mg^{2+}$  adopts a near-perfect octahedral bipyramidal geometry, all of which is accommodated in a cavity of relatively large volume within the protein. This cavity is in part formed by the unique structural features of proline residues. Three central  $\beta$ -strands forming the wall of the metal cavity are interrupted by prolines (Pro<sup>91</sup>, Pro<sup>166</sup>, and Pro<sup>246</sup>), one located in each of these strands (supplemental Fig. 1), and these residues are highly conserved in GH117. These prolines lead to a bulge in the three  $\beta$ -strand structures, creating a larger space in the center of the  $\beta$ -propeller where the metal/water cluster is located. The cluster is locked in this cavity by the carboxyl groups of Glu<sup>167</sup> from one side and by Asp<sup>245</sup> and Asp<sup>90</sup> on the active site face and by a number of main chain carbonyl oxygen atoms. In addition to hosting the bound metal, the central cavity also connects the putative active site with the distal side of the protein through a solvent channel running along the rotational axis of the  $\beta$ -propeller leading to a connected water network through the entire protein (supplemental Fig. 1), resembling *Zg*AhgA in this regard. Although it is clear that this metal ion is important to the catalytic activity of the enzymes in this family, its role in catalysis remains unknown.

The active site of *Bp*GH117 is a pronounced pocket, which is typical for exo-acting glycoside hydrolases. To determine the catalytic residues and infer the catalytic mechanism, we structurally aligned *Bp*GH117 with the catalytic domain of the  $\beta$ -xyylanase XynB3 from *Geobacillus stearothermophilus* (39) and with the  $\alpha$ -L-arabinase Arb43A from *C. japonicus*; the catalytic residues for both of these enzymes have been described previously (16) (Fig. 2C). In both of these inverting enzymes, the general base has been predicted as Asp<sup>38</sup> in Arb43A and Asp<sup>15</sup> in Xyn3B. In *Bp*GH117, the residue Asp<sup>90</sup> aligns well with these two residues. The general acid has been assigned to Glu<sup>221</sup> in Arb43A and to Glu<sup>187</sup> in Xyn3B, whereas a histidine, His<sup>302</sup>, is located at this structural position in *Bp*GH117. We found that His<sup>302</sup> has a strained geometry and is an outlier in the Ramachandran plot, which suggested a possible catalytic function. However, being an unusual candidate for a catalytic residue in glycoside hydrolases, it was the adjacent residue Glu<sup>303</sup>, which

is not conserved in GH43, that has been proposed as a catalytic residue in GH117 (14). Thus, there remains ambiguity as to the identity of the catalytic residues in GH117 and the catalytic mechanism that this family uses. To provide insight into this, we systematically mutated residues in the active site.

**Mutation of Active Site Residues**—The predicted catalytic residues Glu<sup>303</sup>, Asp<sup>90</sup>, and Asp<sup>245</sup> were mutated to Gln, Asn, and Asn, respectively. Additionally, Glu<sup>167</sup>, which is located behind the metal/water cluster, was mutated to Gln, and His<sup>302</sup>, which is located in place of the general acid in GH43, was mutated to both Glu and Gln. All of these mutants crystallized under the same conditions as the wild-type enzyme and eluted without apparent signs of aggregation during the gel filtration purification (not shown), indicating the fidelity of their folds, but they all lacked activity under conditions where the wild-type enzyme hydrolyzed >95% of the neoagarobiose substrate (Fig. 2D). Although this did not resolve the ambiguity in the identity of the catalytic residues, it did indicate that all of the targeted residues are important to catalysis. This analysis also provided a tool to probe the molecular basis of the enzyme-substrate interaction and the more specific roles of these enzymes by obtaining the inactive E303Q mutant, which could be crystallized with neoagarobiose.

**Structure of *Bp*GH117 in Complex with Neoagarobiose**—GH117E303Q was co-crystallized with an excess of neoagarobiose under the same conditions and in the same space group as the native protein. Clear electron density for fully occupied neoagarobiose molecules was found in both active sites of the dimer, allowing unambiguous modeling of the carbohydrate (Fig. 3A). The disaccharide was found with an intact glycosidic linkage between the LA residue in the –1 subsite and the G residue in the +1 subsite (40). An extended network of hydrogen bonds and hydrophobic interactions is made between the neoagarobiose molecule and the enzyme active site with particular interactions in each of the subsites providing specificity for the sugar subunits (Fig. 3B) (41).

The  $\beta$ -face of the G residue in the +1 subsite sits on a hydrophobic platform formed by Trp<sup>127</sup> and Phe<sup>125</sup>. The side chains of Arg<sup>273</sup> and Glu<sup>262</sup> approach the  $\alpha$ -face of the G residue and make hydrogen bonds with the C4-OH. Remarkably, the C terminus of the adjacent monomer in the dimer, in addition to contributing to the dimer interface, plays a role in contouring the +1 subsite (Fig. 3B). The C terminus from the adjacent monomer folds over the active site, narrowing the opening and introducing His<sup>392</sup> into the cleft. His<sup>392</sup> forms hydrogen bonds with both the endocyclic oxygen and the C6-OH of the G residue in the +1 subsite. Although the C-terminal region was disordered in the *Zg*AhgA, His<sup>392</sup> is conserved in GH117 enzymes from marine Bacteroidetes, suggesting that His<sup>392</sup> may have a conserved functional role in substrate recognition. C1-OH of the G residue occupying the +1 subsite is solvent-exposed, consistent with the ability of *Bp*GH117 to remove LA from the non-reducing end of longer agarose oligosaccharides. Notably, in this complex, His<sup>302</sup> was found to be suitably located so that its Ne2 is positioned at a distance of 2.6 Å from the glycosidic oxygen and appropriately oriented to make a hydrogen bond (Fig. 3A). Furthermore, Asp<sup>320</sup> interacts closely (2.6 Å) with the Ne1 of His<sup>302</sup>, suggesting these two residues

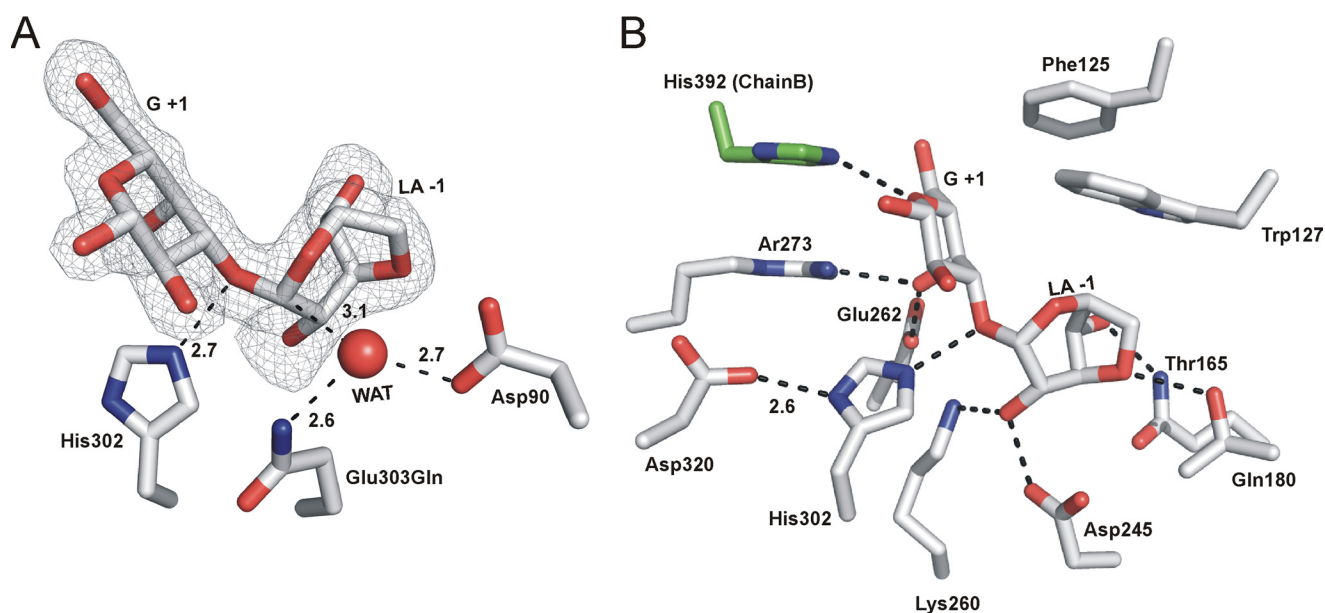


FIGURE 3. **Structural features of neogarobiose recognition in BpGH117.** *A*, neogarobiose trapped in the active site of the E303Q mutant of BpGH117. The electron density of the neogarobiose is shown in gray mesh as a maximum likelihood/ $\sigma_a$ -weighted  $F_o - F_c$  map contoured at  $3\sigma$  ( $0.19e/\text{\AA}^3$ ). The three putative catalytic residues are shown in stick representation. *B*, residues forming hydrogen bonds and hydrophobic interactions with neogarobiose are shown in stick representations. Hydrogen bonds are shown as dashed lines. WAT, water.

may act in tandem to effect general acid catalysis and relay a proton from Asp<sup>320</sup> to the leaving group (Fig. 3B) (17).

The LA residue is fully enclosed in the  $-1$  subsite. Hydrogen bonds are made between C4-OH and Gln<sup>180</sup>, the 3,6-anhydro bridge oxygen to Thr<sup>165</sup>, and C2-OH to both Lys<sup>260</sup> and Asp<sup>245</sup> (Fig. 3B). In addition, the C4-OH makes a water-mediated interaction with a Cl<sup>-</sup> atom, which is structurally conserved in ZgAhgA (14). Trp<sup>127</sup> contributes to forming one wall of this subsite; however, this residue does not interact in the classical way of protein sugar recognition where the aromatic residue forms a platform for the face of a sugar in a sandwich-like manner. Here Trp<sup>127</sup> is oriented toward the side of the LA residue with the endocyclic ring oxygen (Fig. 3B).

## DISCUSSION

**GH117 Is Keystone Enzyme in Agarocolloid Degradation—**Agarolytic bacteria with sequenced genomes contain genes encoding varied numbers and combinations of GH86, GH16, and GH50 enzymes, all of which act on the  $\beta$ -1,4-glycosidic bond of agarose. For example, *Z. galactanivorans* utilizes mainly GH16  $\beta$ -agarases (42) to degrade agar, whereas the estuarine bacterium *S. degradans* utilizes a mixture of GH86, GH50, and GH16 (43). All of these organisms also contain at least one gene encoding a conserved GH117 (supplemental Table 1), raising the question whether GH117 enzymes are the predominant enzymes to break the  $\alpha$ -glycosidic bond in agarose or whether there are alternative endo-acting  $\alpha$ -agarases that may play a significant role in agar degradation. The only other known examples with such an activity are  $\alpha$ -agarases of family GH96; however, none of the currently available microbial genomes that are annotated in CAZy contain such an enzyme (supplemental Table 1). Compared with families GH16, GH50, GH86, GH117, and GH118, enzymes in family GH96 are extremely rare with only four sequences in CAZy (8). Both

endo-acting  $\alpha$ - and  $\beta$ -glycoside hydrolases are identified by agar lysis around agarolytic bacteria, which has been the classical method for functional screening and agarase discovery (44, 45), identifying both  $\alpha$ - and  $\beta$ -activities. Thus, screening bias does not account for the discrepancy in total numbers of highly abundant endo- $\beta$ -agarases versus extremely rare endo- $\alpha$ -agarases, strongly suggesting that the latter are actually rare in nature. Because all agarolytic bacteria appear to contain at least one conserved GH117  $\alpha$ -agarase, it seems at present that this family is the primary evolutionary solution to breaking the  $\alpha$ -glycosidic bond in agarose. Thus, this suggests that family GH117 proteins are keystone enzymes in agarose metabolism.

**Relating Sequence to Specificity in GH117—**Using Blast Explorer (Phylogeny.fr), we extracted 35 sequences with over 35% identity and over 75% query coverage to BpGH117. Examination of these family GH117 sequences, in light of the complex structure, reveals near invariance for catalytic and sugar binding residues His<sup>302</sup>, Glu<sup>303</sup>, Asp<sup>90</sup>, Trp<sup>127</sup> (except *Cyclobacter marinum*), Asp<sup>245</sup>, Gln<sup>180</sup> (except *C. marinum*), Thr<sup>165</sup>, and His<sup>244</sup> (except *Lentisphaera araneosa* and *C. marinum*). When calculating the conservation of residues in this data set with ConSurf, a small degree of variation has been found for residues around the G residue in subsite +1 (Fig. 4A). The residues in subsite  $-1$  are invariant, showing that recognition of the LA residue is conserved within these agarases like GH117. Thus, with the exception of *C. marinum* and *L. araneosa*, we can confidently predict these proteins to be active on agarose oligosaccharides for cleavage of the  $\alpha$ -1,3-glycosidic bond. The phylogenetic tree initially shown by Rebuffet *et al.* (14) and an updated version with additional sequences presented here show that three separate Clades (A, B, and C) are formed within GH117. Apart from the highly conserved subsites +1 and  $-1$ , the highest sequence variation in the active site of GH117 is

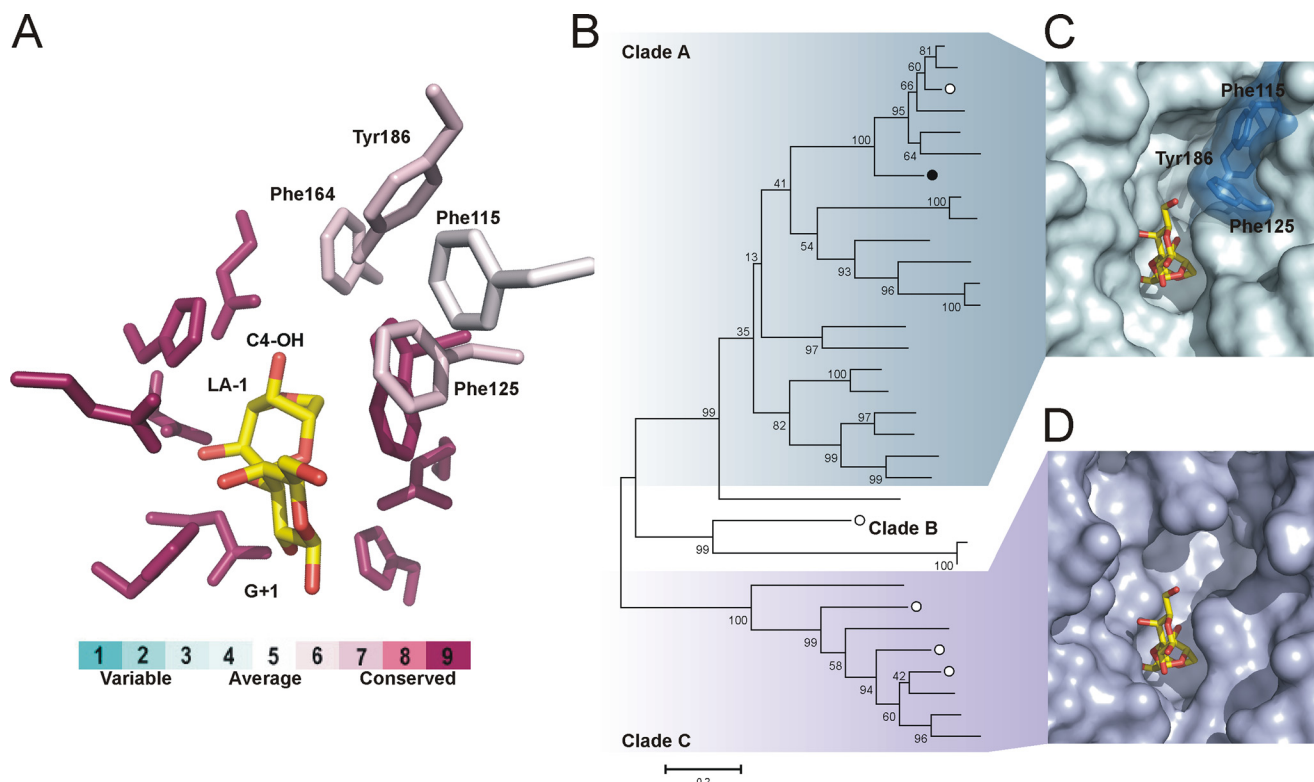


FIGURE 4. Residues for agarose recognition are highly conserved throughout sequences belonging to glycoside hydrolase family 117. *A*, the conservation of active site residues (shown as sticks) in GH117 as calculated with ConSurf (see "Experimental Procedures"). The residues that interact with neoagarobiose in *BpGH117* are largely conserved in GH117. A cluster of bulky hydrophobic residues that is located at the non-reducing end of the LA residue is less conserved in GH117; these are labeled. *B*, the phylogenetic tree of GH117 as proposed by Rebuffet *et al.* (14) and updated here with additional sequences. *BpGH117* is indicated by a black circle, and GH117 from *Z. galactanivorans* is indicated by white circles. The clustering into clades is indicated. (A more detailed phylogenetic tree showing the identity of GH117 members is given in supplemental Fig. 2.) *C*, the *BpGH117* active site is shown with the neoagarobiose complex and the aromatic residues that close the minus (–) subsite at the non-reducing end of the LA residue. *BpGH117* belongs to Clade A of the phylogenetic tree, the sequences of which all share the cluster of bulky aromatic side chains at the non-reducing end of the LA residue and thus contain one minus (–) subsite. *D*, surface representation of a homology model of *Zg3615* with neoagarobiose modeled into the active site.

found in Clade C. We found that the structural variation in Clade C is concentrated in proximity to the C4-OH of the LA residue in the –1 subsite (Fig. 4A). These residues are bulky aromatic amino acids in Clade A (*i.e.* Phe<sup>115</sup>, Phe<sup>125</sup>, Phe<sup>162</sup>, and Tyr<sup>186</sup> in *BpGH117*) but are substituted with smaller amino acids, often serines and alanines, in Clade C. This leaves considerably more space at the –1 subsite of Clade C enzymes, and we therefore hypothesize that sequences belonging to Clade C may have additional subsites, possibly allowing accommodation of longer substrates. We created a model (34) based on the *BpGH117* structure using a homologue from Clade C, *Zg3615* from *Z. galactanivorans*, which shares 40% sequence identity with *BpGH117*. The model showed structurally aligned residues for neoagarobiose recognition and an extended space, possibly for an additional –2 subsite. This may allow binding of a G residue in addition to the LA residue bound in subsite –1. A system containing additional GH117 enzymes with an extra minus (–) subsite would be advantageous for agarolytic bacteria because characterized enzymes of Clade A are inactive on their own reaction products, *i.e.* G-LA-G~, due to the new non-reducing end (G residue), which blocks re-entry into the active site. These products, however, may be a substrate for and further degraded by a GH117 enzyme of Clade C, allowing synergistic degradation of agarose oligosaccharides. This may explain the occurrence of multiple GH117 enzymes in *Z. galac-*

*tanivorans* and other agarolytic microbes. In conclusion, the unique features encountered in this class of enzymes reflect the unusual glycochemical character of its substrate. Once microbes evolved an efficient solution, the high constraints for the distortion of the 3,6-anhydro-L-galactose may have acted as a selection pressure, explaining the conservation observed in the active site of GH117 enzymes.

*Substrate Distortion and Catalytic Mechanism in GH117*—Quantum mechanical calculations predict that the 3,6-anhydro bond constrains the 3,6-anhydrofuranose ring in either a <sup>1</sup>C<sub>4</sub> chair or B<sub>1,4</sub> boat conformation with an energy barrier of 10 kcal mol<sup>–1</sup> to other conformations (46). Consistent with this latter point, only the <sup>1</sup>C<sub>4</sub> chair has been experimentally reported in the x-ray crystal structures of enzymes and carbohydrate binding modules in complex with neoagarose-oligosaccharides (42, 47). Because the LA residue preferentially adopts a <sup>1</sup>C<sub>4</sub> chair rather than a <sup>4</sup>C<sub>1</sub> chair, the  $\alpha$ -1,3-glycosidic bond in agarose mimics the conformation of a  $\beta$ -glycosidic bond. The enzyme-catalyzed hydrolysis of  $\beta$ -glycosidic bonds is facilitated by distortion of the substrate to position the leaving group in a pseudoaxial orientation, thereby placing the lone pair on the endocyclic oxygen antiperiplanar to the scissile bond (48). This distortion relieves unfavorable 1,3-diaxial interactions involving the incoming nucleophile and permits in-line attack of the nucleophile at the anomeric center. The conformational



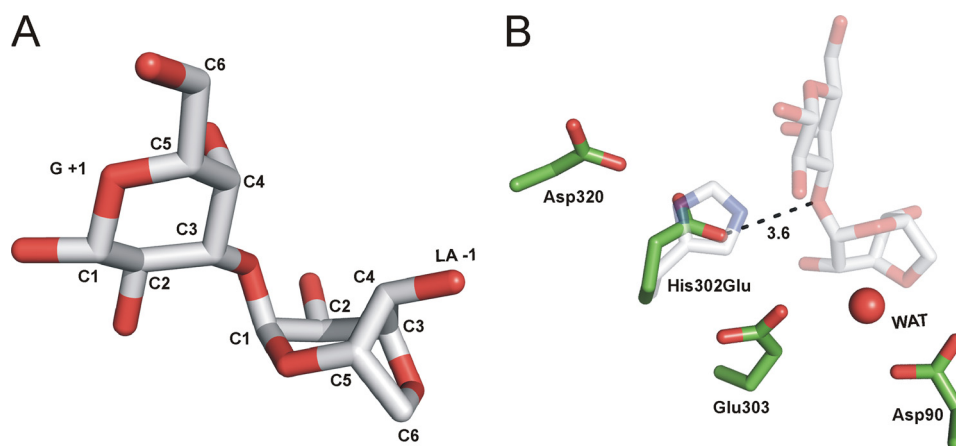


FIGURE 5. **Distortion of substrate and features of catalytic residues.** *A*, the neogagarobiose molecule shows distortion of the 3,6-anhydro-L-galactose into the  $B_{1,4}$  conformation. *B*, the catalytic residues in the mutant H302E are shown in *green* and are superposed with the E303Q mutant in complex with neogagarobiose (*faint gray*). The H302E mutant is inactive likely due to the longer distance between its carboxyl group and the glycosidic bond (3.6 Å) compared with the native His<sup>302</sup> residue (2.7 Å). *WAT*, water.

change in the substrate also brings the glycosidic oxygen in position for efficient general acid catalysis by His<sup>302</sup>. Thus, one would predict that distortion of the terminal LA residue occurs during hydrolysis of the  $\alpha$ -1,3-glycosidic bond by GH117 enzymes.

Indeed, the Michaelis complex formed from *Bp*GH117 and neogagarobiose reveals that the LA residue in the  $-1$  subsite adopts a  $B_{1,4}$  conformation (Fig. 5A), suggesting a reaction coordinate involving distortion of the free substrate from a  ${}^1C_4$  conformation to a less favorable  $B_{1,4}$ . The resulting substrate conformation enables efficient engagement of the catalytic machinery.

Based on partial conservation of active site residues with GH43 glycoside hydrolases, the GH117 enzymes were postulated to use an inverting mechanism. Glycoside hydrolases that catalyze with an inverting catalytic mechanism use general acid and base catalysis from two residues, typically either aspartate or glutamate. The general base enhances the nucleophilicity of a suitably positioned water molecule, enabling it to attack the anomeric carbon. In *Bp*GH117, the pseudoaxial orientation of the leaving group arising from substrate distortion enables this in-line attack of a molecule of water, one positioned only 3.14 Å distant from C1 of the LA residue in the  $-1$  subsite, on the  $\sigma^*$  orbital at the anomeric carbon (9, 49, 50) (Fig. 3A). This water interacts with the side chain carboxylates of both Asp<sup>90</sup> and Glu<sup>303</sup>. Although Asp<sup>90</sup> is well positioned for an in-line activation of the catalytic water molecule, the Glu<sup>303</sup> residue may modulate the electrostatic environment and tailor the  $pK_a$  of the general base Asp<sup>90</sup>. This is supported by mutation of either residue to the corresponding amide side chains, both of which inactivated the enzyme. Of these two residues, only Asp<sup>90</sup> is structurally conserved with the general base in the GH43 enzymes Arb43A and XynB3, generating further support for this residue as the more likely candidate for the general base catalytic residue.

The distorted conformation of the substrate (Figs. 3A and 5A) positions the glycosidic bond such that it could accept a proton from His<sup>302</sup>, which based on its distance from and orientation to the glycosidic bond oxygen is the only candidate for

a general acid. His<sup>302</sup> is structurally conserved with the position of Glu<sup>187</sup> in the GH43 XynB3 (Protein Data Bank code 2exh) and with Glu<sup>221</sup> in Arb43A (Protein Data Bank code 1gyd) (Fig. 2B). In support of the catalytic role of His<sup>302</sup>, substitutions of this amino acid with glutamate or glutamine abolished the activity of the enzyme. Although glutamine is unable to act as a general acid, glutamate, however, is a common general acid found within glycoside hydrolases. Given the structural conservation with GH43 and the superposition of His<sup>302</sup> with the general acid of GH43, we expect that glutamate could have conferred activity to this H302Q variant. To analyze why H302E lacked activity, we crystallized the protein together with the substrate neogagarobiose. The structure of H302E lacks a bound neogagarobiose molecule, but nevertheless, its comparison with the Michaelis complex reveals that the side chain of Glu<sup>302</sup> is too distant from the glycosidic oxygen to serve as a general base (3.48 Å) (Fig. 5B). Furthermore, whereas His<sup>302</sup> interacts with Asp<sup>320</sup> (2.6 Å) in the native enzyme (Fig. 3B), mostly likely acting in tandem as a proton relay system, the Glu<sup>302</sup> of the H302E mutant makes no interactions with Asp<sup>320</sup>, resulting in the greater mobility of the Glu<sup>302</sup> side chain as judged by its elevated B-factors.

The structural features of the active site revealed by the complex of *Bp*GH117 with neogagarobiose are therefore consistent with the assignment of an inverting catalytic mechanism. This mechanism displays the common feature of a carboxylate side chain, in this case Asp<sup>90</sup>, acting as a general base to activate the catalytic water (Fig. 6). An unusual feature of this mechanism, however, is the utilization of a histidine, His<sup>302</sup>, as the general acid (Fig. 6). Indeed, the Asp<sup>320</sup>-His<sup>302</sup> dyad of *Bp*GH117 is similar to the Asp<sup>232</sup>-His<sup>324</sup> catalytic dyad observed for family 3 glycoside hydrolase *Bs*NagZ from *B. subtilis* (17). A notable difference between the mechanisms of these two enzymes, however, is that *Bs*NagZ uses a retaining mechanism, and therefore the histidine residue acts as both acid and base, whereas the histidine in *Bp*GH117 acts only as an acid in the proposed inverting mechanism of catalysis for this enzyme.

GH117 enzymes are the only known enzymes that release monosaccharides from agar, the major cell wall polysaccharide

## Analysis of $\alpha$ -Agarase from *B. plebeius*

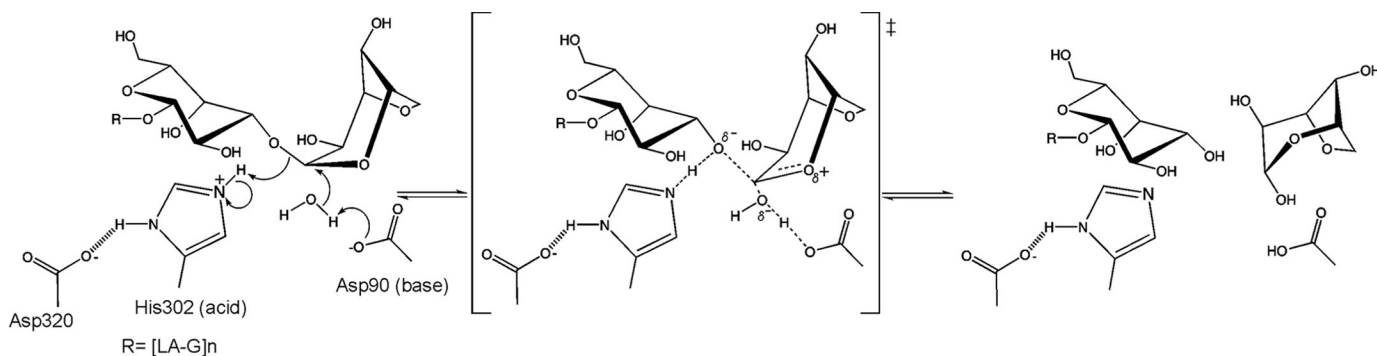


FIGURE 6. **Proposed mechanism of neogaro-oligosaccharide hydrolysis by GH117 enzymes.** The reaction occurs via a single displacement mechanism showing the oxocarbenium-like transition state. His<sup>302</sup> acts as the general acid to facilitate cleavage of the glycosidic bond, whereas Asp<sup>90</sup> acts as a general base to activate a water molecule for nucleophilic attack on the C1 carbon of the 3,6-anhydro-L-galactose.

of agarophytes. Thus, the structural and mechanistic data presented add to the increasing understanding of how bacteria degrade seaweed polysaccharides, insight that is much needed for a refined understanding of the ocean carbon cycle. Furthermore, the unusual use of a histidine as a catalytic residue in this family of enzyme highlights the alternative mechanistic strategies that glycoside hydrolases can use to catalyze the hydrolysis of glycosidic bonds.

*Acknowledgments*—We thank the Stanford Synchrotron Radiation Laboratories beamline 9-2 staff and also the Canadian Light Source beamline CMCF1 staff. The Canadian Light Source is supported by the Natural Sciences and Engineering Research Council of Canada, the National Research Council Canada, the Canadian Institutes of Health Research, the Province of Saskatchewan, Western Economic Diversification Canada, and the University of Saskatchewan.

## REFERENCES

- Wargacki, A. J., Leonard, E., Win, M. N., Regitsky, D. D., Santos, C. N., Kim, P. B., Cooper, S. R., Raisner, R. M., Herman, A., Sivitz, A. B., Lakshmanaswamy, A., Kashiwayama, Y., Baker, D., and Yoshikuni, Y. (2012) An engineered microbial platform for direct biofuel production from brown macroalgae. *Science* **335**, 308–313
- Smith, S. V. (1981) Marine macrophytes as a global carbon sink. *Science* **211**, 838–840
- Martone, P. T., Estevez, J. M., Lu, F., Ruel, K., Denny, M. W., Somerville, C., and Ralph, J. (2009) Discovery of lignin in seaweed reveals convergent evolution of cell-wall architecture. *Curr. Biol.* **19**, 169–175
- 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
- Rees, D. A. (1961) Enzymic synthesis of 3:6-anhydro-L-galactose within porphyrin from L-galactose 6-sulphate units. *Biochem. J.* **81**, 347–352
- Correc, G., Hehemann, J. H., Czjzek, M., and Helbert, W. (2011) Structural analysis of the degradation products of porphyrin digested by *Z. galactanivorans*  $\beta$ -porphyrinase *Carbohydr. Polym.* **83**, 277–283
- Hitchens, A. P., and Leikind, M. C. (1939) The introduction of agar-agar into bacteriology. *J. Bacteriol.* **37**, 485–493
- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* **37**, D233–D238
- Davies, G., and Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases. *Structure* **3**, 853–859
- Allouch, J., Jam, M., Helbert, W., Barbeyron, T., Kloareg, B., Henrissat, B., and Czjzek, M. (2003) The three-dimensional structures of two  $\beta$ -agarases. *J. Biol. Chem.* **278**, 47171–47180
- Jam, M., Flament, D., Allouch, J., Potin, P., Thion, L., Kloareg, B., Czjzek, M., Helbert, W., Michel, G., and Barbeyron, T. (2005) The endo- $\beta$ -agarases AgaA and AgaB from the marine bacterium *Zobellia galactanivorans*: two paralogue enzymes with different molecular organizations and catalytic behaviours. *Biochem. J.* **385**, 703–713
- Flament, D., Barbeyron, T., Jam, M., Potin, P., Czjzek, M., Kloareg, B., and Michel, G. (2007)  $\alpha$ -Agarases define a new family of glycoside hydrolases, distinct from  $\beta$ -agarase families. *Appl. Environ. Microbiol.* **73**, 4691–4694
- Lee, S., Lee, J. Y., Ha, S. C., Jung, J., Shin, D. H., Kim, K. H., and Choi, I. G. (2009) Crystallization and preliminary x-ray analysis of neogaro-oligosaccharide hydrolyase from *Saccharophagus degradans* 2-40. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **65**, 1299–1301
- 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
- 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
- Nurizzo, D., Turkenburg, J. P., Charnock, S. J., Roberts, S. M., Dodson, E. J., McKie, V. A., Taylor, E. J., Gilbert, H. J., and Davies, G. J. (2002) *Cellvibrio japonicus*  $\alpha$ -L-arabinanase 43A has a novel five-blade  $\beta$ -propeller fold. *Nat. Struct. Biol.* **9**, 665–668
- Litzinger, S., Fischer, S., Polzer, P., Diederichs, K., Welte, W., and Mayer, C. (2010) Structural and kinetic analysis of *Bacillus subtilis* N-acetylglucosaminidase reveals a unique Asp-His dyad mechanism. *J. Biol. Chem.* **285**, 35675–35684
- Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783–795
- Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234
- Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 271–281
- Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82
- Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763
- Sheldrick, G. M. (2008) A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122
- Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007) Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230
- 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
- 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
27. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
  28. Brünger, A. T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475
  29. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383
  30. Duckworth, M., and Yaphe, W. (1970) Thin layer chromatography analysis of enzymic hydrolysates of agar. *J. Chromatogr. A* **49**, 482–487
  31. Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J. F., Guindon, S., Lefort, V., Lescot, M., Claverie, J. M., and Gascuel, O. (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36**, W465–W469
  32. Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38**, W529–W533
  33. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739
  34. Kelley, L. A., and Sternberg, M. J. (2009) Protein structure prediction on the web: a case study using the Phyre server. *Nat. Protoc.* **4**, 363–371
  35. Suzuki, H., Sawai, Y., Suzuki, T., and Kawai, K. (2002) Purification and characterization of an extracellular  $\alpha$ -neogaroooligosaccharide hydrolase from *Bacillus* sp. MK03. *J. Biosci. Bioeng.* **93**, 456–463
  36. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2256–2268
  37. Ha, S. C., Lee, S., Lee, J., Kim, H. T., Ko, H. J., Kim, K. H., and Choi, I. G. (2011) Crystal structure of a key enzyme in the agarolytic pathway,  $\alpha$ -neogaroobiose hydrolase from *Saccharophagus degradans* 2-40. *Biochem. Biophys. Res. Commun.* **412**, 238–244
  38. Alberto, F., Bignon, C., Sulzenbacher, G., Henrissat, B., and Czjzek, M. (2004) The three-dimensional structure of invertase ( $\beta$ -fructosidase) from *Thermotoga maritima* reveals a bimodular arrangement and an evolutionary relationship between retaining and inverting glycosidases. *J. Biol. Chem.* **279**, 18903–18910
  39. Brüx, C., Ben-David, A., Shallom-Shezifi, D., Leon, M., Niefind, K., Shoham, G., Shoham, Y., and Schomburg, D. (2006) The structure of an inverting GH43  $\beta$ -xylosidase from *Geobacillus stearothermophilus* with its substrate reveals the role of the three catalytic residues. *J. Mol. Biol.* **359**, 97–109
  40. Davies, G. J., Wilson, K. S., and Henrissat, B. (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem. J.* **321**, 557–559
  41. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127–134
  42. Allouch, J., Helbert, W., Henrissat, B., and Czjzek, M. (2004) Parallel substrate binding sites in a  $\beta$ -agarase suggest a novel mode of action on double-helical agarose. *Structure* **12**, 623–632
  43. Ekborg, N. A., Taylor, L. E., Longmire, A. G., Henrissat, B., Weiner, R. M., and Hutcheson, S. W. (2006) Genomic and proteomic analyses of the agarolytic system expressed by *Saccharophagus degradans* 2-40. *Appl. Environ. Microbiol.* **72**, 3396–3405
  44. Potin, P., Richard, C., Rochas, C., and Kloareg, B. (1993) Purification and characterization of the  $\alpha$ -agarase from *Alteromonas agaralyticus* (Cataldi) comb. nov., strain GJ1B. *Eur. J. Biochem.* **214**, 599–607
  45. Barbeyron, T., L'Haridon, S., Corre, E., Kloareg, B., and Potin, P. (2001) *Zobellia galactanovorans* gen. nov., sp. nov., a marine species of Flavobacteriaceae isolated from a red alga, and classification of [*Cytophaga*] *uliginosa* (ZoBell and Upham 1944) Reichenbach 1989 as *Zobellia uliginosa* gen. nov., comb. nov. *Int. J. Syst. Evol. Microbiol.* **51**, 985–997
  46. Navarro, D. A., and Stortz, C. A. (2005) Modeling ring puckering in strained systems: application to 3,6-anhydroglycosides. *Carbohydr. Res.* **340**, 2030–2038
  47. Henshaw, J., Horne-Bitschy, A., van Bueren, A. L., Money, V. A., Bolam, D. N., Czjzek, M., Ekborg, N. A., Weiner, R. M., Hutcheson, S. W., Davies, G. J., Boraston, A. B., and Gilbert, H. J. (2006) Family 6 carbohydrate binding modules in  $\beta$ -agarases display exquisite selectivity for the non-reducing termini of agarose chains. *J. Biol. Chem.* **281**, 17099–17107
  48. Zechel, D. L., and Withers, S. G. (2000) Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Acc. Chem. Res.* **33**, 11–18
  49. Vocadlo, D. J., and Davies, G. J. (2008) Mechanistic insights into glycosidase chemistry. *Curr. Opin. Chem. Biol.* **12**, 539–555
  50. Sulzenbacher, G., Driguez, H., Henrissat, B., Schülein, M., and Davies, G. J. (1996) Structure of the *Fusarium oxysporum* endoglucanase I with a non-hydrolyzable substrate analogue: substrate distortion gives rise to the preferred axial orientation for the leaving group. *Biochemistry* **35**, 15280–15287