The structure of a family 110 glycoside hydrolase provides insight into the hydrolysis of α-1,3-galactosidic linkages in λ-carrageenan and blood group antigens

Received for publication, September 16, 2020, and in revised form, October 23, 2020. Published, Papers in Press, October 30, 2020, DOI 10.1074/jbc.RA120.015776

Bailey E. McGuire1,‡, Andrew G. Hettle1,‡, Chelsea Vickers1,‡, Dustin T. King2, David J. Vocadlo2,3, and Alisdair B. Boraston1,*

From the 1Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia, Canada, and the Departments of 2Molecular Biology and Biochemistry, and 3Chemistry, Simon Fraser University, Burnaby, British Columbia, Canada

Edited by Gerald W. Hart

α-Linked galactose is a common carbohydrate motif in nature that is processed by a variety of glycoside hydrolases from different families. Terminal Galα1–3Gal motifs are found as a defining feature of different blood group and tissue antigens, as well as the building block of the marine algal galactan λ-carrageenan. The blood group B antigen and linear α-Gal epitope can be processed by glycoside hydrolases in family GH110, whereas the presence of genes encoding GH110 enzymes in polysaccharide utilization loci from marine bacteria suggests a role in processing λ-carrageenan. However, the structure–function relationships underpinning the α-1,3-galactosidase activity within family GH110 remain unknown. Here we focus on a GH110 enzyme (PdGH110B) from the carrageenolytic marine bacterium Pseudoalteromonas distincta U2A. We showed that the enzyme was active on Galα1–3Gal but not the blood group B antigen. X-ray crystal structures in complex with galactose and unhydrolyzed Galα1–3Gal revealed the parallel β-helix fold of the enzyme and the structural basis of its inverting catalytic mechanism. Moreover, an examination of the active site reveals likely adaptations that allow accommodation of fucose in blood group B active GH110 enzymes or, in the case of PdGH110, accommodation of the sulfate groups found on λ-carrageenan. Overall, this work provides insight into the first member of a predominantly marine clade of GH110 enzymes while also illuminating the structural basis of α-1,3-galactoside processing by the family as a whole.

The ABH glycan antigens define the ABO blood types, the appropriate matching of which is a key consideration in blood transfusions and organ transplantations. The O-blood group, defined by the smaller H antigen, is considered a universal donor. Accordingly, enzymatic conversion of the more elaborate and immunogenic A/B antigens to the H antigen provides an attractive route to avoid antigen mismatching and create a readily supply of universal donor blood. A campaign to identify enzymes that could hydrolyze terminal α-1,3-linked N-acetyl D-galactosamine and/or α-galactose from the A and B antigens, respectively, and thereby provide a set of biocatalytic tools to perform this antigen switching revealed a set of enzymes with specific B antigen exo-α-1,3-galactosidase activity (1, 2). These enzymes were the founding members of GH110 (glycoside hydrolase 110) family, of which all currently characterized members are exo-α-1,3-galactosidases that are able to hydrolyze the B antigen glycan (Galα1–3(Fucα1–2)Gal-R) (1–3). The GH110 family presently has over 400 classified members in the Carbohydrate-Active Enzyme Database (4). These putative enzymes are encoded by genes found in the genomes of a variety of environmental and host-adapted bacteria. Most recently, genes encoding proteins that are classified by amino acid sequence identity into GH110 have been identified in polysaccharide utilization loci (PULs) from human gut microbiome bacteria and marine bacteria; these PULs are postulated to target λ-carrageenan, which is a polysaccharide found in marine algae (5-7).

Most of the photosynthetically fixed carbon present on Earth is in land plants with the amount of photosynthetically fixed carbon in the oceans (i.e. present in microalgae and macroalgae (seaweed)) estimated to be only ~1/200 that of terrestrial plant biomass. However, the annual turnover rate (total mass/time) of photosynthetically fixed carbon in the oceans is roughly equal to that on land, indicating a highly dynamic process with a normalized rate of recycling that is ~2 orders of magnitude greater than on land (8). This is a remarkable biotransformation that occurs in the marine environment, and it is largely made possible by the metabolic capabilities of marine microbes, who return the carbon locked in algal storage and structural polysaccharides to the global carbon cycle (9). The dedicated metabolic pathways, which include numerous glycoside hydrolase, deployed by marine microbes to perform this biotransformation are uniquely adapted to the distinctive chemical compositions of marine algal polysaccharides. However, the identification and molecular details of several pathways that target major classes of marine polysaccharides, such as λ-carrageenan, remain to be uncovered. Understanding these pathways is key to, for example, the development of complete biogeochemical models of the global carbon cycle (10), comprehending the rise and fall of algal blooms (11), unlocking farmable seaweed biomass feedstocks for the generation of biofuels or other high-value products (12), and even identifying and engineering the metabolic capabilities of the human gut microbiome (7, 13, 14).
Carrageenans are a family of abundant marine polysaccharides comprising unbranched, high-molecular-mass sulfated galactans. They are typically found in the cell walls of marine red macroalgae, where they can make up to 50% of the dry mass. The backbone comprises β-d-galactose with alternating α-1,3- and β-1,4-linkages. The occurrence and specific patterns of sulfate esters on the free hydroxyl groups of the galactose backbone, along with the presence or absence of 3,6-anhydro-β-d-galactose, gives rise to numerous different carrageenan families (15–17). λ-Carrageenan is made of necroarabiose motifs in which β-d-galactose-2,6-sulfate is α-1,3-linked to β-d-galactose-2-sulfate, and this disaccharide is joined by β-1,4-glycosidic linkages forming a linear λ-carrageenan polymer that is distinct from other carrageenans by its lack of 3,6-anhydro-β-d-galactose. Presently, how microbes process λ-carrageenan is poorly understood with only endo-acting β-1,4-λ-carrageenases from Pseudoalteromonas carrageenovora $^{9T}$ (18, 19) having been identified. To date, no other enzymes having activity consistent with λ-carrageenan processing have been experimentally identified, including the distinct absence of identified enzymes that are active on the α-1,3-linkages. We postulate that this is an activity performed by the GH110 enzymes found in λ-carrageenan PULs.

Toward testing this hypothesis, we characterized the structure and function of PdGH110B. This enzyme is encoded by a gene we identified in the recently reported genome of Pseudoalteromonas carrageenovora $^{9T}$ (6). PdGH110B has ~25% amino acid sequence identity with the characterized GH110 enzymes from Bacteroides sp. (1, 2) and ~94% amino acid sequence identity with a putative GH110 enzyme present in the P. carrageenovora $^{9T}$ PUL that is proposed to target λ-carrageenan (5). Here we demonstrate that PdGH110B is an α-galactosidase that can hydrolyze the Galα1-3Gal disaccharide, but not the blood group B-trisaccharide Galα1-3(Fucα1-2)Gal, thus distinguishing it from previously characterized GH110 enzymes. The structural determination of PdGH110B in complex with enzyme substrate and products revealed the parallel β-helix fold of GH110 enzymes and a −1 subsite (20) that accommodates an unmodified galactose residue. An analysis of the structure points to a potential +1 subsite that is key to accommodating sulfate modifications, as present in λ-carrageenan, or fucose, a defining constituent of the blood group B antigen glycan. Overall, the results are consistent with PdGH110B representing the founding member of a GH110 subfamily that exo-hydrolytically processes terminal α-linked galactose residues in λ-carrageenan while providing general insight into the specificity of the family as a whole.

Results

**Identification of a GH110 in Pseudoalteromonas carrageenovora sp. U2A**

U2A was isolated from the marine environment for its capacity to grow on macroalgal polysaccharides, including carrageenan, as previously described (6). We identified two adjacent genes (locus tags EU511_08545 and EU511_08540) encoding proteins with 30% sequence identity to one another and ~25% sequence identity to previously characterized GH110 enzymes. This pair of proteins displayed 98 and 94% amino acid sequence identity to two putative orthologous GH110 enzymes in P. carrageenovora $^{9T}$ that are encoded by adjacent genes in a presumed λ-carrageenan PUL (5). A gene truncation of EU511_08540, encoding a protein we refer to as PdGH110B, lacking the predicted signal peptide was overproduced and purified, followed by qualitative assessment of activity on the synthetic substrates pNP-α-D-galactopyranoside and pNP-β-D-galactopyranoside. Recombinant PdGH110B showed activity only on pNP-α-D-galactopyranoside and a pH optimum of ~5.6 (Fig. S1A).

We further tested the activity of PdGH110B on more natural substrates using Galα1-3Gal (αG2) and Galβ1–4Gal (βG2), which represent the basic unmodified motifs present in λ-carrageenan, by quantifying galactose release. PdGH110B released β-galactose when incubated with αG2, whereas there was no activity on βG2 (Fig. 1A). The $K_m$ and $k_{cat}$ for αG2 were 5.9 ± 1.1 mM and 18.3 ± 0.002 s$^{-1}$, respectively (Fig. 1B). Given the activity of other GH110 enzymes on the blood group B glycan, we tested PdGH110B using TLC but could not detect any activity (Fig. S1B).

**The overall structure of PdGH110B α-1,3-galactosidase in complex with D-galactose**

An initial preliminary structure of PdGH110B was determined by single-wavelength anomalous dispersion using a cadmium derivative. This initial model was used to solve the structure of PdGH110B in complex with a D-galactose monosaccharide to 2.35 Å resolution. The crystal structure of PdGH110B in complex with D-galactose revealed two chains in the asymmetric unit. Residues 25–238/241–616 for one molecule and residues 21–186/238–241–616 for the second molecule were modeled, with the missing residues residing in loop regions. The noncrystallographic dimer shows no evidence of being stable; however, each monomer in the asymmetric unit forms a crystallographic dimer (Fig. S2A). The crystallographic dimers (Fig. 2A) have a total molecular interface, determined by PISA (Proteins, Interfaces, Structures and Assemblies) (21) analysis, of 2300 Å$^2$, predicting a stable dimeric state. Dynamic light scattering analysis of PdGH110B in solution at protein concentrations of 0.18, 0.37, and 0.73 mg/ml yielded a molecular mass of 136.3 ± 9.4 kDa. The expected molecular mass of the PdGH110B monomer is 67.9 kDa, resulting in an expected molecular mass of 135 kDa for a dimer. This indicates that PdGH110B adopts a dimeric quaternary structure, which is most likely the biologically relevant assembly.

The overall fold of PdGH110B is that of a right-handed parallel β-helix of 11 complete turns. A structural homology search using the DALI server (22) identified a fold most similar to those of a GH87 α-1,3-glucanase from Bacillus circulans (23), as well as two epimerases, AlgE4 and AlgE6, from Azotobacter vinelandii (16) (PDB code 5LW3). This core β-helix is surrounded by two small β-barrel domains (domains I and II) that contribute to the residues involved in dimerization of PdGH110B (Fig. 2B). The α-helix of domain II from the
adjacent monomer folds over and along the wall of the active site, which was identified by a bound D-galactose, with several amino acid side chains protruding into the cleft that contains the active site pocket (Fig. 2C).

The bound D-galactose monosaccharide was identified by clear electron density (Fig. S2B) found in the central region of the β-helix domain of both active sites of the dimer. The modeled monosaccharide occupied a pocket that sequesters the monosaccharide in a fashion that is typical for glycoside hydrolases that are exo-acting on the nonreducing end of glycans (Fig. 2D). Specifically, Asp-344 is located within hydrogen bonding distance of the C1-OH, where the scissile bond of an intact substrate would be (Fig. 2E), indicating that this residue is a likely candidate to play the catalytically essential role of general acid. To trap an intact substrate complex of the enzyme, we targeted this residue to generate an inactive D344N mutant, which indeed lacked activity on pNP-α-galactopyranoside.

Structure of PdGH110B in complex with α-G2

Crystals of the PdGH110B D344N mutant were soaked with an excess of αG2, and the structure was determined to 2.20 Å resolution.

![Figure 1. Activity of PdGH110B on α-1,3- and β-1,4-galactobiose. A, galactose release curve showing release of galactose units by PdGH110B when incubated with αG2 (closed squares) and βG2 (closed triangles), along with a D-galactose standard (closed circles). B, kinetic analysis of PdGH110B activity on αG2 at 25 °C using a galactose release kit to measure units of galactose released. In both panels, error bars show standard deviations of measurements made in triplicate.

![Figure 2. Structural features of PdGH110B in complex with D-galactose. A, the structure of the PdGH110B dimer. Monomer A (purple) is shown in cartoon representation, and monomer B (gray) is shown as a solvent-accessible surface representation. B, cartoon representation showing the different structural domains of PdGH110B, the β-helix domain (purple), domain I (gray), and domain II (dark gray). C, the association of the domain II α-helix extending into the neighboring chain’s active site. The coloring is the same as in A. D, the active site pocket, shown as a solvent-accessible surface in gray, sequesters the galactose residue. The O1, which would be engaged in a glycosidic linkage, is indicated. E, the interaction of the galactose residue with the likely catalytic acid residue. Interacting side chains are shown in purple, and the water molecule is shown as a red sphere. In all A–E, the D-galactose monosaccharide is represented as yellow sticks.](18428 - J. Biol. Chem. (2020) 295(52) 18426–18435)
resolution. The refined structure revealed four monomers of PdGH110B D344N in the asymmetric unit. The monomers were organized as two noncrystallographic dimers with identical arrangements to the crystallographic dimer observed in the PdGH110B D-galactose complex, supporting the concept that the dimer is a stable quaternary structure. Clear electron density for the αG2 disaccharide was found in each monomer active site (Fig. S3) with the intact glycosidic linkage spanning the –1 subsite and +1 subsites and the catalytic machinery (Fig. 3A).

There is an extended network of hydrogen bonds, as well as a single interacting aromatic residue, made between the αG2 molecule and the enzyme active site (Fig. 3A). The C2–C6 portion of the α-face of the D-galactose unit in the –1 subsite sits on a hydrophobic platform created by Trp-486, interacting through CH–π interactions often employed by CAZymes (24). The remainder of the –1 subsite is created by an extensive hydrogen bond network comprising interactions between Asn-85, Glu-488, Asn-348, and Glu-480 and the C3-OH, C4-OH, and C6-OH hydroxyl groups. Arg-265 coordinates both C2-OH and C3-OH, and Arg-453 interacts with C6-OH and the endocyclic oxygen (Fig. 3A). The +1 subsite is formed exclusively by the positively charged side chains of Arg-451, Arg-208, and through a water coordinated by Lys-207. The electrostatic potential of the active site surface indicates a generally acidic –1 subsite but a +1 subsite and neighboring surfaces that are basic (Fig. 3B).

GH110 enzymes were previously shown to operate through use of a single displacement, or inverting, catalytic mechanism (1). We confirmed this for PdGH110B by using 1H NMR to monitor the initial release of the β-anomer of D-galactose from pNP-α-D-galactopyranoside, which indicates inversion of the anomeric configuration of C1 involved in the glycosidic bond (Fig. 4A). The architecture of the PdGH110B catalytic center is also consistent with an inverting catalytic mechanism (25, 26) (Fig. 4B). In the αG2 complex with the mutant enzyme, Asn-344, which would be Asp-344 in the WT enzyme, is 3.1 Å from the glycosidic oxygen and appropriately positioned to act as a general acid. A water molecule that sits 3.5 Å beneath C1 of the D-galactose residue in the –1 subsite is suitably positioned to be activated as a nucleophile by Asp-321 and/or Asp-345 (Fig. 4B).

Discussion

The GH110 family was initially identified by examination of members that specifically removed the immunodominant α-1,3-linked galactose residues of blood group B antigen. Subsequent characterization of additional GH110 enzymes revealed some to be less stringent by possessing the ability to process the linear α-Gal epitope, as well as the blood group B antigen. In contrast to the majority of other α-galactosidases, which employ a retaining catalytic mechanism (i.e. GH4, GH27, GH31, GH36, GH57, and GH97 (27–32)), family GH110 was shown to utilize an inverting mechanism for this hydrolytic activity. Here, through examination of PdGH110B, which originates from a marine bacterium and has relatively low amino acid sequence identity to previously characterized GH110 members, we also demonstrated α-Gal activity, although it differed from other GH110 enzymes by its inability to hydrolyze the blood group B antigen. The structure of PdGH110B revealed the molecular basis of the inverting mechanism utilized by the family. Notably, the core parallel β-helix fold and catalytic machinery of PdGH110 is conserved within GH families 28, 49, and 87 (Fig. 4C) (33–35). GH28 and GH49 are classified into glycoside hydrolase clan GH-N. Given that the fold and catalytic machinery of these founding families of the clan are conserved with GH87 and GH110, we note that it is likely that the latter two GH families also belong to clan GH-N.

A phylogenetic tree constructed from 334 GH110 amino acid sequences, including PdGH110B and another from U2A (PdGH110A), displays several distinct clades (Fig. 5A and Fig. S4). Notably, the sequences originating primarily from marine microbes form their own clade that branches off directly from the origin, hinting at evolution toward potentially distinct functions. Indeed, the P. carrageenovora 9T orthologue of PdGH110B (~94% amino acid sequence identity) resides in a locus proposed to target the highly sulfated marine galactan λ-carrageenan (5).

The genes neighboring that encoding PdGH110B in U2A show similarly high amino acid sequence identity to components of the putative P. carrageenovora 9T λ-carrageenan PUL. This led to the hypothesis that PdGH110B (and its orthologue) would target λ-carrageenan. The activity of PdGH110B on αG2, and the molecular recognition thereof, is consistent with exo-α-1,3-galactosidase activity on the nonreducing ends of λ-carrageenan. In
contrast, because of its \(^1\)C\(_4\) chair conformation, the 3,6-anhydro-D-galactose residue found at the nonreducing termini of \(\kappa\)-and \(\iota\)-neocarrageenoligosaccharides would not be accommodated in the active site of PdGH110B. Notably, the architecture of the \(-1\) subsite leaves no room for a sulfate modification on either C2 or C6 of the galactose residue, indicating that PdGH110B requires an unmodified D-galactose at the nonreducing end of its substrate. This is consistent with our inability to observe activity for this enzyme on sulfated \(\lambda\)-neocarrageenoligosaccharides (not shown) and suggests that preprocessing of \(\lambda\)-carrageenan by as-yet-unidentified sulfatases would be necessary prior to the action of PdGH110B.

An examination of amino acid conservation within the \(-1\) subsite of the representative GH110 sequences reveals this subsite to be remarkably conserved across the family (Fig. 5B), including among the characterized enzymes (Fig. S5), suggesting that \(\text{exo-}\alpha\)-galactosidase activity is a general feature of the family. This observation points to features of the \(\text{exo}\)-subsite as being critical determinants of substrate selectivity. In the case of PdGH110B, a series of specific interactions, particularly those mediated by Arg-451, select for a galacto-configured sugar in the \(+1\) subsite. In \(\lambda\)-carrageenan, the galactose residue in this subsite would likely bear a 2-sulfate group. The structure of PdGH110B in complex with \(\alpha\)G2 shows a small pocket, which we will refer to as a \(+1\) subsite, with the appropriate size and charge to accommodate a sulfate modification (Fig. 3B). Although PdGH110B was active on \(\alpha\)G2 (i.e. lacking sulfates) the high \(K_m\) (\(~5\) mM) suggests nonoptimal recognition of this substrate; 2-sulfation of the galactose in the \(+1\) subsite may be expected to improve binding of the substrate. Thus, we suggest that the features of the PdGH110B active site are entirely consistent with \(\lambda\)-carrageenan recognition.

The \(+1\) subsite has one conserved residue, Arg-451, which makes specific interactions with the galactose residue in the \(+1\) subsite that are likely maintained throughout the family (Figs. 3A and 5B). The other residue in this subsite that makes a direct interaction, Arg-208, is not well-conserved (Fig. 5B). Nevertheless, the noteworthy conservation of the \(-1\) subsite and conservation of Arg-452 in the \(+1\) subsite, which makes directional interactions specific to a galacto-configured monosaccharide, reveal a mechanism that is shared by the GH110 family to recognize a \(\text{Gal}\alpha1–3\text{Gal}\) motif. Lys-207 makes potential water-mediated hydrogen bond to the galactose in the \(+1\) subsite; however, it appears more relevant to the \(+1\) subsite of PdGH110B, which is not well-conserved among the GH110 family (Fig. 5, B and C). In particular, Lys-195 (contributed by...
the other monomer of the dimer) and Lys-207 on either side of the +1′ subsite close it off this subsite in PdGH110B making it of insufficient size to accommodate the 2-fucosyl residue of the blood group B antigen. These residues are not well-conserved among the family as a whole (Fig. 5C) nor in the characterized enzymes (Fig. S5). In the characterized enzymes, Lys-195 resides in a variable region, which in PdGH110B comprises the “finger” that extends between the two monomers of the dimer (Fig. 2C and Fig. S5), although Lys-207 is typically an amino acid with a smaller sidechain (Fig. S5). These alterations seen other GH110 enzymes likely open up the +1′ subsite, allowing it to accommodate the 2-fucosyl residue of the blood group B antigen. Furthermore, these structural changes, along with substitutions of the poorly conserved Arg-208, would likely reduce the basic nature of the +1′ subsite, making it less suitable for accommodating an anionic sulfate residue and more appropriate for a neutral fucosyl residue.

To date, all characterized examples of GH110 enzymes display α-1,3-galactosidase activity. These enzymes partition into three classes of activity: enzymes that are specific for the blood group B antigen, those that are active on both the blood group B antigen and linear α-1,3-galactose epitopes, and now a class represented by PdGH110B that is selective for linear α-1,3–linked galactose. PdGH110B notably sorts into a clade that comprises proteins from primarily marine microbes. This observation likely reflects that enzymes in this clade have adapted to enable biological processing of λ-carrageenan. Indeed, the other entries in this clade that are not from marine microbes are typically from human gut bacteria, which themselves may also be able to target λ-carrageenan present in the diet. Further examination of this marine clade of GH110 enzymes may ultimately reveal additional adaptations that confer the ability to accommodate, or even specifically recognize, the sulfate modifications on λ-carrageenan.

**Experimental procedures**

**Materials**

All reagents, chemicals and other carbohydrates were purchased from Sigma unless otherwise specified.

**Cloning and mutagenesis**

The gene fragment encoding GH110B without predicted signal peptides (amino acids 24–620) was amplified from *Pseu-
dodalteromonas distincta* U2A genomic DNA using the oligonucleotide primers 5′-CTA GCT AGC AAT GAT AAA GTG ATA GAT G-3′ (PdGH110B_fwd) and 5′-CCG CTC GAG TTA GTT TTT AGC TCT TTT AT-3′ (PdGH110B_rev). The product was ligated into pET28a between the NheI and XhoI restriction sites (underlined in sequences).

SaGal110A from *Streptomyces avermitilis*; arm 5, BbAgaBb from *Bifidobacterium bifidum*; and arm 6, BfGal110B from *B. fragilis*. The shaded region indicates the “marine” clade. See also Fig. S4 and Table S1 for complete annotation of the tree. B, the conservation of active site residues of GH110 mapped onto the structure of PdGH110B by ConSurf analysis (46). C, conservation of the active site shown as accessible surface representation. The color scheme representing degree of residue conservation in B and C is shown in B. Active site subsites and sugar residues are labeled in green.
Structural analysis of GH110

All DNA amplifications were done using CloneAmp HiFi PCR premix. Mutations were created by site-directed mutagenesis (QuikChange site-directed mutagenesis kit) using the primers 5′-CTC CAT GGA TGT TAG TAG CAT CAT TCT TTT GAC TTT CGA ATA AGT TAT C-3′ and 5′-GAT AAC TTA TTC GAA AGT CAA AAG AAT GAT GCT ACT AAC ATC CAT GGA G-3′. All constructs were sequence confirmed by bidirectional sequencing.

Protein expression and purification

The expression plasmid of the PdGH110B was transformed into Escherichia coli BL21 (DE3) Star and grown in 2-liter cultures of LB broth containing 50 μg ml⁻¹ kanamycin sulfate at 37 °C with agitation at 180 rpm until cell density reached an A₆₀₀ of ~0.5, at which time the temperature was dropped to 16 °C, and recombinant protein production was induced with a final concentration of 0.5 mM isopropyl β-D-thiogalactopyranoside and allowed to incubate for an additional 16 h.

The cultures were then centrifuged at 6300 × g for 10 min. Harvested cells were chemically lysed by resuspension in 35% (w/v) sucrose, 1% (w/v) deoxycholate, 1% Triton X-100, 500 mM NaCl, 10% glycerol, 20 mM Tris (pH 8.0), 10 mg lysozyme, and 0.2 μg ml⁻¹ DNase. The resulting lysates were clarified by centrifugation at 16,500 × g for 30 min.

PdGH110B protein was purified by applying the clarified lysate supernatant to a nickel-affinity chromatography column (IMAC) in 20 mM Tris (pH 8.0) and further purified using a HiPrep 16/60 Sephacryl S-300 HR size-exclusion chromatography column in 20 mM Tris (pH 8.0) and 500 mM NaCl.

pH studies

The reactions with synthetic substrates were set up in triplicate using a 96-well plate and incubated at 25 °C in the dark for 1 h. The reactions (100 μl) contained 50 mM McIlvaine buffer (pH 2.9–9.6), 1 mM 1,4-dithio-D-threitol in binding buffer with 7.5% glycerol, 1 mM pNP-α-D-galactopyranoside or pNP-β-D-galactopyranoside in D₂O, and 1 μM recombinant PdGH110B. The reactions were stopped with an equal volume of 100 mM NaCl in this mother liquor containing 1 mM CdCl₂ prior to data collection. A 10 mM D-galactose control reaction in NMR buffer was also monitored as a standard. The reactions were measured before the addition of PdGH110B and after subsequent incubations at the 5-min, 15-min, 30-min, 1-h, 2-h, and 24-h time points using a Bruker Avance II 500 MHz NMR spectrometer and a 5-mm TXI inverse probe. The data were processed using the MestReNova software package.

Phylogenetic analysis

The sequences from the glycoside hydrolase family 110 were retrieved. Of the 417 sequences, only 332 had entries in the NCBI database. A multiple sequence alignment was performed using Clustal Omega (36). The evolutionary relationships of PdGH110A, PdGH110B, and the other GH110 sequences was inferred by maximum likelihood method based on the Jones–Taylor–Thornton matrix-based model using FastTree (37). The phylogenetic tree was visualized and annotated using the iTol web tool.

Crystallization, diffraction data collection, and processing

All crystals were grown at 18 °C by hanging-drop vapor diffusion with 1:1 ratios of crystallization solution and protein. A native PdGH110B crystal grown in 0.5 M NaI, 2% Tacsimate, 0.1 M HEPES (pH 7.5), and 17% PEG 3350 was soaked for 10 min in this mother liquor containing 1 mM CdCl₂ prior to data collection. To obtain the product complex, a PdGH110B crystal grown 0.1 M HEPES (pH 7.5), 9% PEG 6000, and 5.5% 2-methyl-2,4-pentanediol was soaked in mother liquor containing excess D-galactose to prior to data collection. To obtain a substrate complex, PdGH110B_D344N was crystallized in 1 M NaI, 2% Tacsimate, 0.1 M HEPES (pH 7.5), and 17% PEG 3350 and then soaked in this mother liquor containing excess α-D-galacto-biose prior to data collection. 25% (v/v) ethylene glycol was used as the cryo-protectant for all crystals, with the exception of the D-galactose-soaked crystal. The 2-methyl-2,4-pentanediol content of the D-galactose-soaked crystal was increased from 5.5% (mother liquor concentration) to 20% for cryo-protection.

Diffraction data were collected on a instrument comprising a Pilatus 200K 2D detector coupled to a MicroMax-007HF X-ray generator with a VarimaxTM-HF ArcSec confocal optical system and an Oxford Cryostream 800. The data were...
integrated, scaled, and merged using HKL2000. The data processing statistics are shown in Table 1.

**Structure solution and refinement**

The structure of PdGH110B was determined by the single-wavelength anomalous dispersion method using the cadmium derivative. Initial phases were determined using the SHARP/autoSHARP pipeline (38). Phases were improved using PARROT (39) to perform density modification and noncrystallographic averaging. An initial model comprising 95% completeness was constructed by autobuilding using ARP/wARP (40). The most complete monomer from this initial model was used as a molecular replacement model to determine the structure of the PdGH110B galactose complex, which was finished by manual building with COOT (41) and refinement with REFMAC (42). A monomer from this model was then used to solve the structure of the PdGH110B_D234N structure in complex with α-G2 by molecular replacement using PHASER (43) and the same building and refinement procedures.

For all structures, the addition of water molecules was performed in COOT with FINDWATERS and manually checked after refinement. In all data sets, refinement procedures were monitored by flagging 5% of all observations as “free” (44). Model validation was performed with MOLPROBITY (45). The model refinement statistics are shown in Table 1.

**Data availability**

The atomic coordinates for the two crystal structures reported here have been deposited in the Research Collaboratory for Structural Bioinformatics Protein Data Bank under accession codes 7JW4 and 7JWF. All other data are available in the article and in the supporting information.

**Acknowledgments**—We thank Edward Meier for technical assistance.

**Author contributions**—B. E. M., A. G. H., C. V., and A. B. B. formal analysis; B. E. M. and D. T. K. investigation; B. E. M. methodology; B. E. M., A. G. H., C. V., D. J. V., and A. B. B. writing-review and editing; A. G. H. validation; A. G. H. and A. B. B. writing-original draft; D. J. V. and A. B. B. supervision; A. B. B. conceptualization; A. B. B. funding acquisition; A. B. B. project administration.

**Funding and additional information**—This work was supported by Discovery Grant FRN 04355 from the Natural Sciences and Engineering Research Council of Canada.

**Conflict of interest**—The authors declare that they have no conflicts of interest with the contents of this article.

**Abbreviations**—The abbreviations used are: GH, glycoside hydrolase; PUL, polysaccharide utilization locus; Fuc, fucose; pNP, para-

---

**Table 1**

X-ray data collection and structure refinement statistics

|                      | PdGH110B Cadmium derivative | PdGH110B Galactose complex | PdGH110B_D234N α-1,3-Galactobiose (α-G2) complex |
|----------------------|----------------------------|----------------------------|-----------------------------------------------|
| **Data collection**  |                            |                            |                                               |
| Beamline             | In-house                   | In-house                   | In-house                                      |
| Wavelength           | 1.514                      | 1.514                      | 1.514                                         |
| Space group           | P21                        | C2                         | P21                                          |
| Cell dimensions      |                            |                            |                                               |
| a                    | 96.70                      | 169.79                     | 95.82                                         |
| b                    | 124.95                     | 128.09                     | 124.76                                        |
| c                    | 142.34                     | 100.14                     | 142.42                                        |
| β                    | 94.47                      | 123.15                     | 93.86                                         |
| Resolution (Å)       | 25.00–2.00 (2.03–2.00)     | 30.00–2.35 (2.39–2.35)     | 30.00–2.20 (2.24–2.20)                        |
| \( R_{merge} \)      | 0.166 (0.914)              | 0.127 (0.916)              | 0.142 (0.590)                                |
| \( R_{merge} \)      | 0.040 (0.337)              | 0.071 (0.327)              | 0.096 (0.421)                                |
| \( CC_{1/2} \)       | 0.997 (0.782)              | 0.989 (0.888)              | 0.991 (0.833)                                |
| \(<I/\sigma(I)\)>    | 17.9 (2.3)                 | 10.4 (2.3)                 | 7.1 (2.3)                                    |
| Completeness (%)     | 99.9 (99.7)                | 99.6 (98.5)                | 98.6 (84.6)                                  |
| Redundancy           | 17.5 (8.0)                 | 4.0 (4.0)                  | 3.6 (3.4)                                    |
| No. of reflections   | 3,964,261                  | 289,894                    | 630,568                                      |
| No. of unique reflections | 89,189                  | 17,427                     | 174,470                                      |
| **Refinement**       |                            |                            |                                               |
| Resolution (Å)       | 2.35                       | 2.20                       |                                               |
| \( R_{work} / R_{free} \) | 0.185/0.231               | 0.205/0.238               |                                               |
| No. of atoms         |                            |                            |                                               |
| Protein              | 4663 (A), 4640 (B)         | 4634 (A), 4613 (B), 4625 (C), 4603 (D) |                                               |
| Ligand               | 24 (GLA), 12 (GAL)         | 92 (α-G2)                  |                                               |
| Water                | 901                        | 1661                       |                                               |
| B-factors            |                            |                            |                                               |
| Protein              | 33.9 (A), 35.1 (B)         | 33.7 (A), 35.2 (B), 35.2 (C), 33.5 (D) |                                               |
| Ligand               | 28.2 (GLA), 54.0 (GAL)     | 30.0 (α-G2)                |                                               |
| Water                | 40.7                       | 38.5                       |                                               |
| Root-mean-square deviation |                     |                            |                                               |
| Bond lengths (Å)     | 0.007                      | 0.002                      |                                               |
| Bond angles (°)      | 0.873                      | 0.473                      |                                               |
| Ramachandran (%)     | 95.2                       | 94.5                       |                                               |
| Preferred            | 4.8                        | 5.5                        |                                               |
| Allowed              | 0.0                        | 0.2                        |                                               |
| Disallowed           |                            |                            |                                               |

J. Biol. Chem. (2020) 295(52) 18426–18435 18433
Structural analysis of GH110

nitrophenol; αG2, Gaia1-3Gal; βG2, Galβ1-4Gal; PDB, Protein Data Bank.

References

1. Liu, Q. P., Yuan, H., Bennett, E. P., Leyer, S. B., Nudelman, E., Spence, J., Pietz, G., Saunders, K., White, T., Olsson, M. L., Henrisatt, B., Sulzenbacher, G., and Clausen, H. (2008) Identification of a GH110 subfamily of α1,3-galactosidases: novel enzymes for removal of the αG1al xenotransplantation antigen. J. Biol. Chem. 283, 8545–8554 CrossRef Medline

2. Liu, Q. P., Sulzenbacher, G., Yuan, H., Bennett, E. P., Pietz, G., Saunders, K., Spence, J., Nudelman, E., Leyer, S. B., White, T., Neeve, J. F., Lane, W. S., Bourne, Y., Olsson, M. L., Henrisatt, B., et al. (2007) Bacterial glycosidases for the production of universal red blood cells. Nat. Biotechnol. 25, 454–464 CrossRef Medline

3. Wakinaka, T., Kiyohara, M., Kurihara, S., Hirata, A., Chaiwangsi, T., Ohnuma, T., Fukumizo, T., Katayama, T., Ashida, H., and Yamamoto, K. (2013) Bifidobacterial α-galactosidase with unique carbohydrate-binding module specifically acts on blood group B antigen. Glycobiology 23, 232–240 CrossRef Medline

4. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrisatt, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 42, D490–D495 CrossRef Medline

5. Gotet, A., Barbeyron, T., Matard-Mann, M., Magdelenat, G., Vallenet, D., Duchaud, E., and Michel, G. (2018) Evolutionary evidence of algal polysaccharide degradation acquisition by Pseudoalteromonas carrageenovora 9T adapted to macroalgal nicks. Front. Microbiol. 9, 2740 CrossRef Medline

6. Hettle, A. G., Hobbs, J. K., Plunvice, B., Vickers, C., Abe, K. T., Salama-Alber, O., McGuire, B. E., Hehemann, J. H., Hui, J. P. M., Berrue, F., Banskota, A., Zhang, J., Botos, E. M., Van Hamme, J., and Boraston, A. B. (2019) Insights into the α/β-carrageenan metabolism pathway of some marine Pseudoalteromonas species. Commun. Biol. 2, 474 CrossRef Medline

7. Pudlo, N. A., Pereira, G. V., Parnami, J., Cid, M., Adams, R., Martin, D., Bolam, D. N., Becher, D., and Schmidt, T. M. (2020) Extensive transfer of genes for edible seaweed digestion from marine to human gut bacteria. bioRxiv CrossRef

8. Field, C. B., Behrenfeld, M. J., Randerson, J. T., and Falkowski, P. (1998) Primary production of the biosphere: integrating terrestrial and oceanic components. Science 281, 237–240 CrossRef Medline

9. Hehemann, J. H., Boraston, A. B., and Czjzek, M. (2014) A sweet new wave: structures and mechanisms of enzymes that digest polysaccharides. Curr. Opin. Struct. Biol. 28, 77–86 CrossRef Medline

10. Falkowski, P. G., Barber, R. T., and Smetacek, V. (1998) Biogeochemical controls and feedbacks on ocean primary production. Science 281, 200–206 CrossRef Medline

11. Teeling, H., Fuchs, B. M., Bennke, C. M., Krüger, K., Chafee, M., Kappelmann, L., Reintjes, G., Waldmann, J., Quast, C., Glocker, F. O., Lucas, J., Wichels, A., Gerdts, G., Wiltshire, K. H., and Amann, R. I. (2016) Recurring evidence of algal polysaccharide degradation acquisition by Pseudoalteromonas carrageenovora 9T adapted to macroalgal nicks. Front. Microbiol. 9, 2740 CrossRef Medline

12. Oht, Y., and Hatada, Y. (2006) A novel enzyme, α-carrageenase, isolated from a deep-sea bacterium. J. Biochem. 140, 475–481 CrossRef Medline

13. Guibert, M., Colin, S., Barbeyron, T., Genicot, S., Kloareg, B., Michel, G., and Helbert, W. (2007) Degradation of α-carrageenan by Pseudoalteromonas carrageenovora α-carrageenase: a new family of glycoside hydrolases unrelated to kappa- and iota-carrageenases. Biochem. J. 404, 105–114 CrossRef Medline

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

15. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 774–797 CrossRef Medline

16. Usow, A. I. (1998) Structural analysis of red seaweed galactans of agar and carrageenan groups. Food Hydrocoll. 12, 301–308 CrossRef

17. Usow, A. (2011) Polysaccharides of the red algae. Adv. Carbohydr. Chem. Biochem. 65, 115–179 CrossRef Medline

18. Oht, Y., and Hatada, Y. (2006) A novel enzyme, α-carrageenase, isolated from a deep-sea bacterium. J. Biochem. 140, 475–481 CrossRef Medline

19. Spyrov, V. (2017) CH/β interactions in carbohydrate recognition. Molecules 22, 1038–1012 CrossRef Medline

20. Davies, G., and Henrisatt, B. (1995) Structures and mechanisms of glycosyl hydrolases. Structure 3, 853–859 CrossRef Medline

21. Vodcolado, J. D., and Davies, G. J. (2008) Mechanistic insights into glycosidase chemistry. Curr. Opin. Chem. Biol. 12, 539–555 CrossRef Medline

22. Yip, V. L. Y., Varrot, A., Davies, G. J., Rajan, S. S., Yang, X., Thompson, J., Anderson, W. F., and Withers, S. G. (2004) An unusual mechanism of glycoside hydrolysis involving redox and elimination steps by a family 4 β-glycosidase from Thermotoga maritima. J. Am. Chem. Soc. 126, 8354–8355 CrossRef Medline

23. Frandsen, T. P., and Svensson, B. (1998) Plant α-galactosidases of the glycoside hydrolase family 31: molecular properties, substrate specificity, reaction mechanism, and comparison with family members of different origin. Plant Mol. Biol. 37, 1–13 CrossRef Medline

24. Comfort, D. A., Bobrov, K. S., Ivan, D. R., Shabalin, K. A., Harris, J. M., and Helbert, W. (2007) Degradation of α-carrageenan by Pseudoalteromonas carrageenovora α-carrageenase: a new family of glycoside hydrolases unrelated to kappa- and iota-carrageenases. Biochem. J. 404, 105–114 CrossRef Medline

25. Okuyama, M., Kitamura, M., Hondo, H., Kang, M. S., Mori, H., Kimura, A., Tanaka, I., and Yao, M. (2009) Catalytic mechanism of retaining α-galactosidase belonging to glycoside hydrolase family 97. J. Mol. Biol. 392, 1232–1241 CrossRef Medline

26. Mizuno, M., Koidz, A., Yamamura, A., Akeboshi, H., Yoshiha, H., Kami- tori, S., Sakano, Y., Nishikawa, A., and Tonozuka, T. (2008) Crystal structure of Aspergillus niger isopullanase, a member of glycoside hydrolase family 49. J. Mol. Biol. 376, 210–220 CrossRef Medline

27. Itoh, T., Intuy, R., Suyotha, W., Hayashi, J., Yano, S., Makabe, K., Wakayama, M., and Hibi, T. (2020) Structural insights into substrate recognition and catalysis by glycoside hydrolase family 87 α-1,3-glucanase from Paenibacillus glycanylicicus FH11. FEBS J. 287, 2524–2543 CrossRef Medline

28. Abbott, D. W., and Boraston, A. B. (2007) The structural basis for exopolysaccharide degradation and catabolism by family 28 glycoside hydrolases. J. Mol. Biol. 368, 1215–1222 CrossRef Medline

18434 J. Biol. Chem. (2020) 295(52) 18426–18435

ASMB
36. Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D., and Lopez, R. (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 47, W636–W641 CrossRef Medline

37. Price, M. N., Dehal, P. S., and Arkin, A. P. (2010) FastTree 2: approximately maximum-likelihood trees for large alignments. *PLoS One* 5, e9490 CrossRef

38. Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007) Automated structure solution with autoSHARP. *Methods Mol. Biol.* 364, 215–230 CrossRef Medline

39. Cowtan, K. (2010) Recent developments in classical density modification. *Acta Crystallogr. D Biol. Crystallogr.* 66, 470–478 CrossRef Medline

40. 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 CrossRef Medline

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

42. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* 53, 240–255 CrossRef Medline

43. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674 CrossRef Medline

44. Brünger, A. T. (1992) Free R-value: a novel statistical quantity for assessing the accuracy of crystal-structures. *Nature* 355, 472–475 CrossRef Medline

45. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 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, 375–383 CrossRef Medline

46. Goldenberg, O., Erez, E., Nimrod, G., and Ben-Tal, N. (2009) The Consurf-DB: pre-calculated evolutionary conservation profiles of protein structures. *Nucleic Acids Res.* 37, 323–327 CrossRef Medline