Biochemical and structural investigation of two paralogous glycoside hydrolases from Zobellia galactanivorans: novel insights into the evolution, dimerization plasticity and catalytic mechanism of the GH117 family.

E Ficko-Blean, D Duffieux, É Rebuffet, R Larocque, A Groisillier, G Michel, M Czjzek

Biochemical and structural investigation of two paralogous glycoside hydrolases from *Zobellia galactanivorans*: novel insights into the evolution, dimerization plasticity and catalytic mechanism of the GH117 family

Elizabeth Ficko-Blean[1,2*], Delphine Duffieux[1,2], Étienne Rebuffet[1,2], Robert Larocque[1,2], Agnes Groisillier[1,2], Gurvan Michel[1,2*], Mirjam Czjzek[1,2*]

[1]Sorbonne Universités, UPMC Univ Paris 06, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, F-29688, Roscoff cedex, Bretagne, France

[2]CNRS, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, F-29688, Roscoff cedex, Bretagne, France

*Running title: *Enzymatic degradation of red seaweed galactans*

*To whom correspondence should be addressed: Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, Bretagne, France. Tel.: 33-298-29-23-30; Fax: 33-298-29-23-24; E-mails: efickoblean@sb-roscoff.fr; gurvan@sb-roscoff.fr; czjzek@sb-roscoff.fr

Keywords: glycoside hydrolase; GH117; inverting mechanism; red algae; agarose

ABSTRACT

The family 117 glycoside hydrolase enzymes (GH117s) have exo-α-1,3-(3,6-anhydro)-L-galactosidase activity, removing terminal non-reducing α-1,3-linked 3,6-anhydro-L-galactose residues from their red algal neoagarose substrate. These enzymes have previously been phylogenetically divided into clades and only the Clade A enzymes have been experimentally studied to date. Investigation into two GH117 enzymes, Zg3615 and Zg3597, produced by the marine bacterium *Zobellia galactanivorans*, reveals structural, biochemical and further phylogenetic diversity between clades. A product complex with the unusual β-3,6-anhydro-L-galactose residue sheds light on the inverting catalytic mechanism of the GH117 enzymes as well as into the structure of this unique sugar produced by hydrolysis of the agarophyte red algal cell wall.

Red seaweeds (Rhodophyta) are ancient photosynthetic eukaryotes and are considered as the first organisms to have developed advanced multicellularity (1). These red macroalgae are important primary

producers to the marine ecosystem and a large proportion of their organic biomass is recycled through the food chain. This biomass is composed of approximately 50% polysaccharides, mainly starch and cell wall components (2). Red seaweeds contain either agars or carrageenans, both of which are sulfated galactans, and are thus referred to as agarophytes (e.g. *Gracillaria*, *Gelidium*, *Porphyra* spp.) or carrageenophytes (e.g. *Chondrus*, *Kappaphycus*, *Eucheuma* spp.), respectively. Sulfated galactans contribute to the flexibility and hydration of the macroalgae which is important given the extreme environmental forces exerted by the ocean on the organism (3). These polysaccharides consist of a linear backbone of galactose residues linked by alternating β-1,4 and α-1,3 glycosidic bonds. A specific feature of red algal agars is the presence of L-sugars, which may in addition contain 3,6-anhydro-bridges, while the equivalent unit is a D-sugar in carrageenans. Thus, the nomenclature of the α-linked galactose units refers to their configuration, with the abbreviations LA and L6S referring to L-sugars, or DA and D6S that refer to D-sugars, in agars or carrageenans, respectively. In red algal polysaccharides the β-linked residues are always classical galactose units in the D configuration (G monomer), which can also contain different degrees of sulfatations.

Both agarophyte and carrageenophyte algae synthesize an unusual carbohydrate moiety, the α-1,3-linked bicyclic 3,6-anhydrogalactose (LA monomer in agars, DA monomer in carrageenans), which they have incorporated into their cell wall galactans. The synthesis of 3,6-anhydrogalactose is unique to red algae and imparts important physiochemical properties to the cell wall. These bicyclic sugars are also crucial for the gelling properties of agars and carrageenans which are widely used as food ingredients (4). The formation of 3,6-anhydrogalactose moieties is catalyzed at the polymer level by galactose-6-sulfurylases. These unique enzymes convert galactose 6-sulfate (L6S or D6S monomer) into 3,6-anhydrogalactose, releasing a free sulfate ion (5,6). However the knowledge on the biosynthesis of sulfated galactans remains very limited and the first global view of the carbohydrate metabolism of red seaweeds has been only recently provided by the genome sequencing of the carrageenophyte *Chondrus crispus* (7,8).

The regular structure of the agar backbone is often modified with methyl groups and variations in acidic side groups such as sulfate or pyruvate (2,4). The main repeating disaccharides are the agarobiose (G-β-1,4-LA) and the porphyranobiose (G-β-1,4-L6S) and these moieties are found in various amounts in the agar chains of most agarophytes. Complex algal polysaccharides constitute a crucial carbon source for numerous marine bacteria. These microorganisms produce several glycoside hydrolases (GHs) from different amino acid sequence based CAZyme (Carbohydrate-Active Enzyme) families (9) in order to efficiently degrade this unusual polysaccharide. Notably, different endo-hydrolases catalyze the initial steps of agar degradation. Endo-α-agarases (EC 3.2.1.158) cleave the α-1,3 linkage between two agarobiose units and belong to a single family, the GH96 family (10). In contrast, the β-agarases (EC 3.2.1.81) which hydrolyze the β-1,4 bond between two neoagarobiose units (LA-α-1,3-G) fall into various GH families: GH16, GH50, GH86 and GH118 (see for review, (11)). Finally the β-porphyranases which cleave the β-1,4 bond between two neoporphyranobiose units (L6S-α-1,3-G) constitute new subfamilies within the families GH16 (12) and GH86 (13).

Recently, the characterization of the enzyme Zg4663 as an α-1,3-(3,6-anhydro)-L-galactosidase (ZgAhgA) has defined a new GH family which includes agar-specific enzymes, the GH117 family (14). ZgAhgA specifically removes 3,6-anhydro-L-galactose at the non-reducing end of oligo-agars released by β-agarases and thus is involved in the terminal steps of agar catabolism.

The marine flavobacterium *Zobellia galactanivorans* is a model for the biodegradation of most algal polysaccharides (11). Particularly, it can assimilate sulfated galactans from red seaweeds. *Z. galactanivorans* possesses one GH16 kappa-carrageenase (15) and three GH82 iota-carrageenases (16) (17). This bacterium has also evolved a complex agarolytic system composed of four GH16 β-agarases and five GH16 β-porphyranases (18). Several of these enzymes have been already biochemically and structurally characterized: the β-agarases ZgAgaA, ZgAgaB (19-21) and ZgAgaD (18) and the β-porphyranases ZgPorA and ZgPorB (12,18).

Z. galactanivorans possesses five GH117 enzymes, including the above mentioned ZgAhgA (Zg4663). The GH117 family has been divided into three phylogenetic clades (A, B, and C), ZgAhgA belonging to Clade A (14). This enzyme adopts a five-bladed β-propeller fold and forms a dimer in solution by swapping of a small N-terminal domain. The activity of this enzyme is also cation-dependent and a unique cation-binding site is found next to the putative catalytic residues. Two other Clade A GH117s have been characterized since then: SdGH117 from *Saccharophagus degradans* (22) and BpGH117 from *Bacteroides plebeius* (23). These two enzymes are also α-1,3-(3,6-anhydro)-L-galactosidases and form dimers by swapping of the conserved N-terminal domain but also of a C-terminal extension absent in ZgAhgA. The structures of SdGH117 and BpGH117 were solved in complex with a D-galactose at subsite +1 (22) and neoagarobiose at subsites +1 and -1 (23), respectively. A site-directed mutagenesis analysis identified the potential catalytic residues, and notably a conserved histidine was proposed to act as a general acid (23). To date, only clade A enzymes have been structurally and functionally characterized and for the two clade B enzymes little is known. Based on sequence comparison within the active site region of the Clade C enzymes which suggested replacement of bulky hydrophobic amino acids with smaller amino acids such as serine and alanine, Hehemann and coworkers have also proposed that there is an additional -2 subsite within the Clade C enzymes, possibly for accommodating longer oligo-agars (23).

Besides ZgAhgA, *Z. galactanivorans* has one Clade B GH117 (Zg185) and three Clade C GH117 enzymes (Zg205, Zg3597 and Zg3615). These proteins are relatively distant from ZgAhgA (35, 37 and 41% sequence identity for Zg205, Zg3597 and Zg3615, respectively), suggesting a potential functional diversity and/or synergy. New phyologenetic analysis reveals that within the *Z. galactanivorans* previously defined Clade C orthologues (14) there are significant differences between enzymes; this divergence results in Clade C enzymes which might not have the same specificities. With the growth in the genomic sequences available, and the structural and biochemical characterization done on the GH117

enzymes to date, we have new phylogenetic evidence that there are at least 6 clades within this highly unusual family.

In order to investigate the fine differences within the clades we chose to study two of the GH117 (previously defined) Clade C enzymes, Zg3615 (now Clade E) and Zg3597 (now Clade D). The structure of Zg3597 was determined in its native form and Zg3615 was determined in complex with its product, β-3,6-anhydro-L-galactose (Figure 1). This has important implications on the mechanism of catalysis and it is the first time, to the best of our knowledge, that the structure of this unusual sugar has been clearly elucidated.

EXPERIMENTAL PROCEDURES

All materials were obtained from Sigma-Aldrich unless otherwise stated.

Cloning- PCR was used to amplify the gene fragment from *Z. galactanivorans* genomic DNA encoding the loci for Zg3615 and Zg3597 without their predicted signal peptide. SignaIP 4.1 (24) predicts a secretion signal peptide cleavage site between amino acids 20 and 21 for Zg3615 and between amino acids 24 and 25 for Zg3597. Based on comparison with the structure of ZgAhgA (pdb id 3P2N) (14) and secondary structure predictions, nucleotides 76-1248 corresponding to amino acids 26-416 of the Zg*3615* gene, and nucleotides 79-1299 corresponding to amino acids 27-433 of the Zg*3597* gene, were cloned into the PFO4 vector which encodes an N-terminal hexahistidine tag (25). The following primers and restriction enzyme sites were used to clone Zg3615: forward primer 5'-TTTTTTGTCGACCAGCCCGAAGGGTTTCCGTTCA-3' including SalI (going into XhoI of pFO4) and reverse primer 5'-AAAAAACTGCAGTTATTTAGGTTGTTTTTCGATGCGTTC-3' including a PstI site (going into NsiI site of pFO4). The following primers and restriction enzyme sites were used to clone Zg3597: forward primer 5'-GGGGGGGGATCCTGTTCCGAACAAAAAACGGAAACCA-3' with a

BamHI site and reverse primer 5'-CCCCCCGAATTCTTATCGTCCTTCGGGACTTAGGTTT-3' with an EcoRI site.

Protein production and purification- Plasmids were transformed into competent *Escherichia coli* strain BL21 (DE3) and grown shaking in 1L LB medium containing 100 µg/mL ampicillin at 37^{0}C to an optical density of ~1.0 whereupon the temperature was lowered to 20^{0}C for one hour followed by overnight induction of protein production with the addition of IPTG to 0.5 mM. Cells were harvested the following morning by centrifugation at 3000 rpm for 30 minutes. A chemical lysis procedure was used to lyse the cells. Briefly, cells were resuspended in 20 mL of 50 mM Tris pH 8.0 with 25 % sucrose and 10 µg of lysozyme and left stirring at 4^{0}C for 20 minutes. Next, 40 mL of 1% deoxycholate, 1% triton X-100, 20 mM Tris pH 7.5 and 100 mM NaCl was added and incubated stirring at 4^{0}C for 10 minutes.

Finally, $MgCl_2$ was added to 5 mM and 0.1 mg of DNase and the lysis reaction was incubated at room temperature for 20 minutes, until the viscosity decreased. The lysis solution was clarified at 13000 rpm over 45 minutes at 4 ^{0}C. Supernatant containing the soluble proteins of interest was run on a 5 mL GE His Trap HP column. Briefly, using an ÄCTA FPLC, the IMAC column was loaded with 0.45 μM filtered supernatant in Buffer A (20 mM Tris pH 8.0, 100 mM NaCl), washed with Buffer A + 1% Buffer B (20 mM Tris pH 8.0, 100 mM NaCl, and 1M imidazole) and eluted using a gradient of 1%-100% Buffer B. Fractions containing the protein of interest were pooled and concentrated using an Amicon stir cell concentrator (MWCO 3 kDa). This was followed by size exclusion chromatography on a Superdex S200 column in 20 mM Tris pH 8.0 and 100 mM NaCl. Again, fractions containing the protein of interest were pooled and concentrated using an Amicon stir cell concentrator (MWCO 3 kDa). Proteins were assessed at >95% purity using SDS-PAGE. The concentration was calculated using A280 and the extinction coefficient calculated by ExPASy ProtParam tool ($\varepsilon_{0.1\%}$ of 2.1 for Zg3615 and 2.0 for Zg3597). The final concentration for Zg3615 was 18.0 mg/mL and 10.0 mg/mL for Zg3597.

Thin Layer Chromatography, TLC- Enzymatic reactions were performed in 0.5 % oligo- or polysaccharide, 100 mM Tris pH 7.5, 0.1 mM $ZnSO_4$, 100 mM NaCl, and 25 mM $CaCl_2$ and 0.1 μg/μL enzyme. Neoagarooligosaccharides were purified as described previously (20). 2 μL of the reaction volume was spotted onto Merck TLC Silica Gel 60 plates and developed in 3x butanol, 1x ethanol and 1x water. The plates were dried then stained in 50 mL of 10% H_2SO_4 in ethanol added to 25 mL of 0.2%

solution of naphthoresorcinol (1,3-dihydroxynaphthalene) in ethanol just before using. Plates were again dried and incubated at 60 ^{0}C until spots appeared (approximately 45 minutes).

Flurophore Assisted Chromatography Electrophoresis, FACE- Enzymatic reactions were performed as above. 50 µg of oligo- or polysaccharide reaction volume were dried down using a speed vacuum. Labelling was done by adding 2 µL of ANTS (8-Aminonaphthalene-1,3,6-Trisulfonic Acid) in DMSO to the dried sample followed by the addition of 5 µL of freshly made sodium cyanoborohydride in DMSO. The samples were mixed well, centrifuged and left at 37 ^{0}C in the dark overnight. 20µL of 20 % glycerol was then added as a gel loading aid. 8 µL of sample was loaded on a 27% polyacrylamide gel. Running buffer used was 0.192 M glycine and 0.025 M Tris pH 8.5. Gels were migrated at 200V for approximately 3 hours and then visualized under UV light.

Small Angle X-ray Scattering, SAXS- SAXS experiments were carried out on the beamline X33 at the DORIS-III storage ring at DESY (Deutsches Elektronen-Synchrotron), Hamburg, Germany. The wavelength λ was 0.15 nm. The sample-to-detector distance was set at 2.7 m, resulting in scattering vectors q ranging from s = 0.06 nm^{-1} to 6 nm^{-1}. All experiments were performed at 16°C on a series of concentrations ranging from 2.0 to 7.9 mg/mL for the heterologously expressed and purified enzyme Zg3597. The scattering vector is defined as $q=4p/\lambda \sin\theta$, where 2θ is the scattering angle. The detector used was a 2D Photon counting Pilatus 1M-W pixel X-ray detector and 10 successive frames of 10 second

exposure times were recorded for each sample. Each frame was carefully checked for possible bubble formation or radiation induced aggregation. If such effects were not observed, the individual frames were averaged. A 4.8 mg/ml solution of BSA was measured as a reference and for calibration procedures. Up to five different concentrations of the protein sample were measured to test for consistency and eventually detect and eliminate concentration dependent effects. Background scattering was quantified before and after each protein measurement by measuring the corresponding buffer, which is then subtracted from the protein patterns using the Primus suite (26).

The radius of gyration, Rg, of each measurement was derived by the Guinier approximation: $I(q)=I(0)\exp(-q^2Rg\ 2/3)$ up to $qRg<1.0$. The radii of gyration Rg, calculated for the different protein concentrations, displayed no evident concentration dependence. The program GNOM (27) was used to compute the pair-distance distribution functions P(r). This approach gives the maximum dimension of the macromolecule Dmax and offers an alternative calculation of Rg, which is based on the entire scattering curve (see Table 1).

The overall shapes of all assemblies were restored from the experimental data by using the *ab initio* modelling programs GASBOR and DAMMIF (28,29). The scattering profiles were used up to $qmax=3\ nm^{-1}$. Nine low-resolution models obtained from different runs were compared using the program DAMAVER (30) to give an estimate of the reproducibility of the results inferred from the *ab initio* shape calculation and to construct the average model representing the general structural features of all the reconstruction filtered by DAMFILT. The atomic structures of the individual modules were then positioned into the low resolution envelopes using the program SUPCOMB (31). For the Zg3597 dimer obtained from the crystal structure the theoretical SAXS profile, the Rg and the corresponding fit (represented by χ^2 values) to the experimental data were calculated using the program CRYSOL (32).

Crystallization-Initial hits were obtained with sitting drop vapour diffusion and the Cartesian Dispensing System (Genomic Solution) using four different commercial screens: JCSG+, PEGs suite,

PEGs II and the PACT suite. Hits were optimized using hanging drop vapour diffusion. The optimized condition for Zg3615 was: 0.1 M sodium acetate pH 5.0, 25% PEG 3350, in a 1:1 ratio of protein (at 10 mg/mL) to mother liquor. Soaks were done on the native protein crystals using 10 mM neoagarotetraose made up in mother liquor and mixed 1:1 with the crystal drop. The optimized crystallization conditions for Zg3597 were: 16% glycerol, 0.15 M ammonium sulfate, and 20 % PEG 4000 in a 1:1 ratio with protein at 5 and 2.5 mg/mL and 14% PEG 3350 and 75 mM sodium acetate with a ratio of 2:1 protein (at 7.5 mg/mL) to mother liquor. Crystals were flash frozen in liquid nitrogen using a cryoprotectant of 20 % glycerol or 20 % ethylene glycol for Zg3597 and 30% ethylene glycol for Zg3615. Crystals for both proteins grew at 18 ^{0}C.

X-ray Crystallography Data Collection and Processing- Diffraction data were collected for Zg3615 on the PROXIMA 1 line at SOLEIL (Source optimisée de lumière d'énergie intermédiaire du LURE, Paris). Diffraction data for Zg3597 were collected on the ID14-4 beamline at the ESRF (European Synchrotron Radiation Facility, Grenoble). The data was processed using MOSFLM (33) and scaled using SCALA (34). The structures were solved by molecular replacement using the one-to-one threading tool from PHYRE2 (35) to build a model (100% confidence) using monomer A from the RSCB deposited GH117 Zg4663 pdb 3P2N (14) and removing any non-conserved amino acids from the MOLREP (36) input pdb. A library for the co-crystallized sugar was created using ProDrg from the CCP4 suite of programs (37). Data was refined using REFMAC5 (38) from the CCP4 suite of programs. Waters were added using the COOT:FIND WATERS option. Successive rounds of model building and refinement were done using COOT (39) and REFMAC5. All data collection statistics are summarized in Table 2.

Ramachandran statistics were calculated using the program Rampage from the CCP4 suite of programs (40). Zg3597 has been given the pdb id 4U6B and for Zg3615 the pdb id is 4U6D.

Phylogeny-New sequences of GH117 have been identified using BlastP at the Genbank database. These new sequences have been added to the previous dataset of GH117 sequences (14) and all sequences were aligned using MAFFT with the iterative refinement method and the scoring matrix Blosum62 (41). The MAFFT alignment of these sequences was manually refined using Bioedit (© Tom Hall), on the basis of the superposition of the five available crystal structures of GH117 (ZgAhgA (Zg4663): PDB code, 3P2N; Zg3597: 4U6B; Zg3615: 4U6D; SdGH117: 3R4Y and BpGH117: 4AK5). This refined alignment allowed calculating model tests and maximum likelihood trees with MEGA version 6.0.6 (42). Tree reliability was tested by bootstrap using 100 resamplings of the dataset. The trees were displayed with MEGA 6.0.6 (42). Accession numbers of the selected sequence homologues are available in Supplementary Table 1.

RESULTS

Activity Assays- Two representatives (Zg3615 and Zg3597) from the recently discovered GH117 family of enzymes were chosen for investigation. Zg3615 and Zg3597 activities were tested on agarose and a variety of neoagaro-oligosaccharides. The enzyme activities were also screened in combination with one another and with Zg4663 (ZgAhgA,(14)), in case the activity of one enzyme is dependent on exposing the substrate by another. Zg3615 exo-α-1,3-(3,6-anhydro)-L-galactosidase activity was initially demonstrated using TLC which showed the release of terminal 3,6-anhydro-L-galactose from neoagarotetraose, neoagarohexaose, and neoagarooctaose (not shown). Enzymatic activity was also tested on agarose; however, there was no detectable difference between the enzymatic digested and undigested reactions, consistent with exo-cleavage as there are not sufficient terminal 3,6-anhydro-L-galactose residues for detection by TLC. In order to release detectable levels of 3,6-anhydro-L-galactose the activity

of a β-agarase would also be required. Thus, by TLC the activity of Zg3615 is indistinguishable from that of Zg4663. Therefore, FACE was employed to further investigate any potential differences in activity between the clades. Enzyme reactions using neoagarobiose/neoagarotetraose, neoagarotetraose/neoagarohexaose (Figure 2), and agarose (not shown) as substrates were migrated on FACE gels. Both Zg3615 and Zg4663 were active on neoagarooligosaccharides; however, there was no discernible difference between Zg3615 and Zg4663 activities on these unsubstituted oligo-agars. The enzymatic release of one 3,6-anhydro-L-galactose residue resulted in a corresponding electrophoretic shift in migration from neoagarotetraose and neoagarohexaose to 3,6-anhydro-L-galactose, agarotriose and agaropentaose (Figure 2, Lanes 1-7). The cleavage of neoagarobiose and neoagarotetraose resulted in the production of agarotriose and the monosaccharides D-galactose and 3,6-anhydro-L-galactose (Figure 2, Lanes 8-14). The enzymatic digests on agarose polysaccharide showed no visible bands, again consistent with the exo-activity within this family. No noticeable differences were detected between Zg3615 reactions alone or in combination with Zg3597 or Zg4663. In contrast, and though thorough screening of activity of Zg3597 on potential agarose substrates was undertaken, it was not possible to discern activity on any of the oligo- or polysaccharides tested, suggesting a novel and as yet unidentified activity within the family.

Overall homodimer structural features- The X-ray crystal structure of Zg3615 was determined to 1.7 Å resolution and the structure of Zg3597 to 2.3 Å. The globular Zg3615 and Zg3597 GH117 monomers share the 5-fold β-propeller fold with the other members of the family and each β-sheet has 3 or 4 anti-parallel β-strands. The GH117 enzymes form a structural homodimer, previously shown to be mediated by both the N- and the C-terminal domains (SdGH117, BpGH117) or just the N-terminal domain (Zg4663) (Figure 3). The structures of Zg3615 and Zg3597 reveal differences in their N- and the C-terminal domains. Zg3615 has a helix-random coil N-terminal motif (amino acids 28-61) which interacts with its other monomer through multiple surface hydrophobic residues. The N-terminal is in a structurally similar position to the N-terminal helix, followed by a turn, followed by a helix from Zg4663 (amino acids

38-71) and BpGH117 (amino acids 33-65). This N-terminal domain is not visible in the Zg3597 crystal structure. Zg3615 has a helix, followed by a turn, followed by a helix (amino acids 386-402) within its C-terminal domain which provides considerable hydrophobic interactions with its partner monomer. This C-terminal domain is not conserved in the sequence of Zg4663 and its C-terminus is reduced to a short loop after the last β-strand, whereas BpGH117 and SdGH117 have a large interacting surface area contributed by their random coil C-termini (amino acids 386-401 and amino acids 350-371). Interestingly, while spatially the C-terminal domains of the Clade A enzymes are approximately conserved within the clade, the C-terminal domain is not spatially conserved in Zg3615 or Zg3597. The Zg3597 C-terminal domain consists only of one helix motif (amino acids 408-425) which is spatially conserved with the Zg3615 C-terminal end. All contributions to the termini-mediated dimerization of Zg3597 come from the C-terminal and this has resulted in a different dimer configuration for Zg3597 when compared to the other GH117 enzymes.

Using the webservice PDBePISA, Proteins, Interfaces, Structures and Assemblies (https://www.ebi.ac.uk/pdbe/pisa/) the overall interaction interface within the Zg3615 dimer in the asymmetric unit was determined to be ~4100 $Å^2$, this gave a Complex Formation Significance Score (CSS) of 0.73 indicating it is essential for complex formation. To this, the N-terminal contributes approximately ~1900 $Å^2$ of surface area and the C-terminal contributes ~1750 $Å^2$ of surface area. Zg3597 has a total interaction interface with its monomeric partner of ~1820 $Å^2$ giving a CSS score of 0.25, indicating it's relevant for complex formation. Zg3597 is a tetramer in the asymmetric unit, though the surface interaction area (~740 $Å^2$) that is not mediated by the N- or C-terminal domains appears to be the result of crystal packing only (CSS=0).

Active Site Architecture- Both Zg3615 and Zg3597 enzymes have a shallow cavity typical of exo-acting enzymes. The proposed catalytic residues are conserved (Figure 4), in Zg3615 these are His284, proposed to be the catalytic acid, Asp89, proposed to be the general base, and Glu285, proposed to modulate the pKa of the nearby general base (23). Though sequentially conserved (Figure 4), the

proposed catalytic acid (His306) from Zg3597 is shifted compared to Zg3615 (Figure 5 A) and to the structures of the other enzymes. This is a rather pronounced positional change, putting it well out of hydrogen bonding position with the glycosidic oxygen seen in the BpGH117 complex (approximately 6.9 Å away compared to between 2.7-3.6 Å for the other enzymes).

Soaks were done on the native Zg3615 crystals with neoagarotetraose. Examination into the active sites of the homodimer pair revealed an unusual electron density in monomer A. We first attempted to fit α-3,6-anhydro-L-galactose into the density but our attempts were unsuccessful; however, β-3,6-anhydro-L-galactose fit the density well in an unconstrained manner (Figure 5 B). This enzyme-product complex provides a novel glimpse into the structure of the unique β-3,6-anhydro-L-galactose molecule. The bicyclic sugar moiety sits at the base of the pocket in the -1 subsite in a $^4C_1$ conformation. Its anomeric C1 hydroxyl appears in a 'flattened' conformation at the anomeric centre. The $^4C_1$ conformation is stabilized by several hydrogen bonding interactions with the C1 hydroxyl provided by Zg3615 Arg88, Asp89, and Glu285. Arg88 also hydrogen bonds to the 6 membered ring's endocyclic oxygen (Figure 5 A). The axial C2 hydroxyl interacts via hydrogen bonds to Lys246 and Asp231. Gln181 and Asp192 are within hydrogen bonding distance of the equatorial C4 hydroxyl group and Thr166 provides a hydrogen bond to the 5-membered ring's cyclic oxygen residue. Finally, a remarkable difference is seen when comparing the two monomers in the Zg3615 homodimer. The loop containing Asp 192 participates in coordinating the β-3,6-anhydro-L-galactose residue in the active site of monomer A of Zg3615; however, in monomer B this loop is shifted considerably away from the active site. The contribution of Asp192 to the active site of monomer A is probably due to interactions with the carbohydrate moiety. Hydrophobic contributions with the product complex are provided by the His230 imidazole face which interacts with the equatorial C4 arm of β-3,6-anhydro-L-galactose. Trp128 provides a hydrophobic platform for the sugar's C5 and C6 atoms which form part of the 5-membered sugar ring and the arm of Arg88 also provides hydrophobic contributions to this ring component.

The C-terminal domains of Zg3597 and Zg3615 lie distal to the active site and do not contribute directly to the +1 subsite. This is in direct contrast with two of the Clade A enzymes, SdGH117 and BpGH117, who's C-termini do participate in the active site substructure and interact with the galactose component in the +1 subsite (BpGH117, His392; SdGH117, Ser358 and Tyr359). The last residues of the Clade A's Zg4663 C-terminal domain are disordered in the native, non-complexed crystal structure (His400-Gln408), though its structural placement is similar to the other Clade A enzymes.

Structural analysis of the enzymes has revealed several important amino acid substitution differences between Zg3597 and Zg3615 which may affect both active site substructure and N-terminal dimerization. Firstly, Asp250 from Zg3615 forms a salt bridge with Arg58 from the N-terminal of the second monomer. Zg3597 has Asn268 at this site, effectively destroying the electrostatic interaction. In the structure of Zg3615, the phenyl ring of Phe32 inserts between two β-strands of the other monomer and sits 3.75 Å from the sidechain of Tyr227. Zg3597 replaces these two residues with a histidine residue (His48- not visible in the crystal structure) and Asp245, suggesting this interaction is abrogated in Zg3597.

A loop comprising Ile322-Tyr329 is flipped in Zg3597 (Figure 6 A and B) which sits just beside the turn holding His306. One reason for this flip may be the hydrophobic Tyr329 which would otherwise be surface located. In the +1 subsite, Arg257 from Zg3615 is spatially, though not sequentially, conserved with a conserved Clade A arginine that interacts with the C4 hydroxyl group of galactose. Arg257 sits against the imidazole ring of His284 (Figure 6 C), the putative catalytic acid. Zg3597 replaces Arg257 with Val279, a much smaller non-polar residue, leaving a pocket at the C4 site, possibly for accommodating a substituent on the galactose C4 hydroxyl group. His306, the Zg3597 putative catalytic acid, is able to shift into the space left by the absence of an arginine residue at this site (e.g. Arg257 in Zg3615). The consequence of the loop flip and the replacement of Arg257 (Zg3615) with Val279 (Zg3597) is that the Zg3597 His306 lays sandwiched hydrophobically between Val279 and Asp324 (Figure 6 B). In the other structures, this aspartic acid residue is within hydrogen bonding distance to His306. The change in position of the Zg3597 His306 residue represents a rather large deviation for a

proposed catalytic acid. Though an activity has yet to be described for Zg3597, it is reasonably possible that this protein has a divergent specificity/function.

Within the -1 subsite there are several rearrangements of bulky hydrophobic groups which again result in spatial, though not sequential, conservation of amino acid residues. For example, the BpGH117 Phe164 residue is spatially conserved with the Zg3615 Phe183 and Zg3597 Phe201 residues. The residue BpGH117 Tyr186 is not spatially conserved in Zg3597 or Zg3615, however, the functionality of the Tyr hydroxyl group is roughly conserved through Zg3615 Ser187 and Zg3597 Ser205. BpGH117 Phe115 is replaced by Ala116 in Zg3615 and by a more conservative Trp131 in Zg3597 which is disordered at the enzyme surface. The bulky BpGH117 Phe125 residue is replaced by the smaller Ala26 for Zg3615 and the polar Thr144 in the case of Zg3597. Zg3615's C4 hydroxyl interacting residue Asp192 is conserved in the Zg3597 crystal structure (Asp210); however, it is shifted considerably leaving a more open active site entrance at the enzyme surface. This residue is not found in any of the Clade A enzyme structures. Both Zg3615 and Zg3597 enzymes have more open and accessible active site entrances when compared to BpGH117 and SdGH117 (though not Zg4663), Zg3597 in particular.

Metal binding site- With the exception of SdGH117, divalent metal ions have been modelled into a conserved metal binding pocket adjacent to the active site. For BpGH117 a Mg^{2+} ion was modelled and for Zg4663 a Zn^{2+} ion. A Ca^{2+} molecule has been modelled for Zg3615 and Zg3597 at this same site which is supported by 7 organized water molecules in the solvation shell (43) and the similarity of the B factors between the metal atoms and their neighboring water molecules. Furthermore, the distances between the water atoms and the Ca^{2+} molecule in Zg3615 (monomer A) range from 2.32 Å to 2.60 Å and in Zg3597 (monomer A) from 2.18 Å to 2.58 Å, these values are consistent with the expected distribution range (44). A water molecule has been modelled in precisely the same site as the metal atoms in the SdGH117 crystal structure. Examination of the F_o-F_c maps for SdGH117 at this site reveal significant density above 5 sigma ($0.317e^-/Å^3$), consistent with the molecule being modelled incorrectly as a water. Furthermore, the distorted octahedral geometry of 6 ligands surrounding a central atom is incompatible with the central

atom being a highly coordinated water molecule (43). Regardless, this metal binding site appears to have some degree of plasticity and is conserved throughout the GH117 structures solved to date.

SAXS- The solution structure of Zg3597 was determined by small angle X-ray scattering. We used SAXS analyses to address the oligomeric state and conformational plasticity of the Zg3597 enzyme in solution (Figure 7 D). The slope and intensity at zero angle of the experimental scattering data allowed to calculate Rg values and estimate the molecular weight in solution (Table 1) that were coherent with the protein taking a non-spherical, bimodular form. Several forms were calculated to fit the experimental curves and compared using DAMAVER. The overall shapes were rather similar, although indicative of variations, as indicated by a mean normalized spatial discrepancy value (NSD) of 1.20±0.01. The best envelopes that fitted the experimental curve (χ^2 = 1.6) are obtained when imposing a 2-fold symmetry for a dimeric form of the enzyme. These data also allowed determination of an Rg of 34.6 Å. Similar values were obtained using several distinct calculation methods (Table 1). The distance distribution function P(r) displayed a biphasic pattern, also indicating an elongated shape with a maximum diameter (Dmax) of 100±2 Å (Figure 7 D). The molecular mass determined from the scattering intensity extrapolated to zero scattering angle indicated a molecular mass (Mw) of approximately 109 kDa (Table 1), in close agreement with Zg3597 existing as a homodimer in solution, which is also observed by size-exclusion chromatography, DLS and in the crystal structure (data not shown). The Mw determined by SAXS also compares well to the theoretical Mw of a homodimer (~97 kDa), as calculated from the protein sequence. A theoretical scattering curve of the dimer that exists in the crystal structure of Zg3597 was calculated using CRYSOL. Although the deduced scattering profile from the crystal structure displayed a similar shape as the experimental scattering profile (data not shown) and roughly an Rg value (31.9 Å) in the same order of magnitude as the experimental scattering curve, the χ^2 value of 3.4 describing the fit between experimental and theoretical profiles suggests structural differences between the solution and crystallographic structures. In agreement with this notion, we found that when the crystallographic structure was docked onto the SAXS envelope, an additional volume likely corresponding to the catalytic

amino-terminal domains was present in the SAXS envelope (Figure 7 A, B and C). This may reflect conformational flexibility of the amino-terminus region with respect to the catalytic domains, having a different position in solution and in the crystal structure. The closer contact with the catalytic domain in the solution structure might be mediated by weak protein-protein interactions (Figure 7 C). In addition, the SAXS envelope calculated by GASBOR (Figure 7 A) was of sufficient resolution to show the position of the crater like active site at the heart of the β -propeller. The calculation of an approximate dihedral angle between the orientations of these two depressions in the SAXS envelopes appeared to be closer to perpendicular at roughly 100° (Figure 7 A), whereas this angle estimated from the crystal structure is about 130° (Figure 7 C). This would imply a rotational rearrangement of the two monomers, most probably due to constraints in the crystal structure that might disfavor the establishment of the interactions of the amino-terminal domains with the adjacent catalytic domains in the swapped geometry.

Phylogenetic Analysis- In the initial phylogenetic analysis of GH117 family three clades were defined: Clade A contained most of the sequences, while Clades B and C only comprised few sequences and essentially paralogous proteins from *Z. galactanivorans* (Clade B: Zg185, clade C: Zg205, Zg3597, and Zg3165 (14). Since then, numerous genomes of marine bacteria have been sequenced and novel GH117 sequences have been identified, particularly in the initial Clades B and C. Therefore, we have generated an updated phylogenetic tree (Figure 8 and Supplementary Table 1) in order to revise the definition of the GH117 subfamilies. In the initial tree (14), Clade B only contained two sequences, Zg185 and RB13146 from *Rhodopirellula baltica*. In the revised tree, these sequences are much more distant: Zg185 groups with a new sequence (RspSWK7GH) with a weak statistical support (51%), while RB13146 clusters together with 10 new sequences, forming a solid clade (bootstrap: 100%). To be consistent with the previous nomenclature, Zg185 remains in the subfamily referred to as Clade B, while RB13146 and its closely related sequences are defined as the new Clade F. Similarly, the former Clade C has significantly expanded and changed. The protein LNTAR_17733 from *Lentisphaera araneosa* was at the root of the former Clade C and is still an isolated outgroup. In contrast, the remaining sequences of the former Clade

C have now diverged into two solid clades (bootstrap values of 98% and 90% respectively): (i) the first containing Zg3597 and 5 new sequences and (ii) the second encompassing 28 sequences, including Zg205 and Zg3615. The difference of enzymatic activity between Zg3597 (still unknown) and Zg3615 (exo-α-1,3-(3,6-anhydro)-L-galactosidase) is consistent with the existence of two distinct subfamilies. Consequently, one can expect that the outgroup sequence LNTAR_17733 may display an activity different from Zg3597 and Zg3615. Altogether, we propose that the clade composed of LNTAR_17733 retains the name Clade C, while the clades of Zg3597 and of Zg3615 are referred to as Clade D and Clade E, respectively.

DISCUSSION

Z. galactanivorans has evolved a complex GH117 family- Sulfated galactans from red algae are unique polysaccharides with a complex chemical structure. To assimilate these cell wall polymers, *Z. galactanivorans* has evolved a diverse enzymatic system, mainly based on GH16 enzymes including four β-agarases (18,20), five β-porphyranases (12) and one κ-carrageenase (15). Alongside the GH16 family, *Z. galactanivorans* possesses a multigenic family of five GH117 genes, localized in different clusters containing sulfatases and additional carbohydrate-related genes (14) and one could expect a similar functional diversity for this family. The first characterized GH117 enzyme, Zg4663, is an α-1,3-(3,6-anhydro)-L-galactosidase and thus has been named ZgAhgA (14). In the present work, we have shown that, despite the sequence divergence, Zg3615 has an activity similar to that of Zg4663, at least on unsubstituted agarose oligosaccharides (Figure 2). Therefore, Zg3615 and its corresponding gene will be referred to as ZgAhgB and *ahgB*, respectively. In contrast, Zg3597 is inactive on oligo-agars. Thus, Zg3597 is different from ZgAhgA and ZgAhgB, although its exact activity remains to be determined. The complexity of the GH117 family is confirmed by our revised phylogenetic tree (Figure 8). While the former Clades B and C initially contained very few sequences (14), these subfamilies are largely expanded

due to the GH117 sequences identified in the numerous genomes of marine bacteria sequenced the last three years. Our phylogenetic analysis has led to the division of the former clade B into a new Clade B and Clade F, and of the former Clade C into a new Clade C, Clade D and Clade E. ZgAhgB (Zg3615) aligns into the new Clade E and Zg3597 aligns into the new clade D, consistent with the differences in dimerization conformation, active site structure and function. Although Clade E is solidly supported (bootstrap value: 90%), this clade is still quite large and contains two paralogous enzymes from Z. *galactanivorans*, Zg3615 (ZgAhgB) and Zg205. Depending on the biochemical characterization of additional Clade E enzymes, this clade may be further divided in the future to reflect the difference in enzymatic specificity.

Implications on catalysis- The GH117 family shares distant sequence similarity with families GH43 and GH32. The first crystal structure, Zg4663, revealed a zinc ion octahedrally coordinated by water molecules located at the bottom of the active site pocket (43). The divalent cation was proposed to have a role in the activation of the catalytic water molecule to provoke hydrolysis. This ion-binding site is conserved in the structures of all the enzymes solved to date, though the ion type differs. The BpGH117 Michaelis substrate complex trapped the α-1,3-linked 3,6-anhydro-L-galactose moiety in a higher energy $B_{1,4}$ boat conformation (Figure 9 B) which aligns the anomeric centre for nucleophilic attack and demonstrates how the enzymes alter the sugar conformation to favor the oxocarbenium ion transition state (23). Based on crystallographic and mutational evidence for BpGH117, Hehemann *et al* proposed Asp90 (Figure 9 B) acts as the general base the adjacent Glu303 as a pKa modulator (23). The positioning of the BpGH117 active site residues suggests an inverting mechanism (45) and unexpectedly revealed a histidine residue positioned to act as the putative catalytic acid. This distinguishes this family from the vast majority of glycoside hydrolases which use a carboxylic acid moiety, though unusual glycoside hydrolase mechanisms are not unheard of (46). The three proposed catalytic residues, histidine, aspartic acid, and glutamic acid, are conserved throughout the GH117 family, with the exception of two Clade F (Figure 8)

enzymes (GjGH117-1 and RRSWK_05156) which have an alanine instead of a glutamate in their predicted active sites.

The ZgAhgB $^4C_1$ β-3,6-anhydro-L-galactose chair complex provides a snapshot of the last step of the conformational itinerary. The chair conformer of the β-sugar closely mimics the $^4C_1$ conformation of its relaxed α-anomer counterpart (19,47,48) implying this is not a high energy intermediate (Figure 9 A and C). The unusual positioning of the C1 hydroxyl in a 'flattened' conformation is likely due in part to repulsive steric forces contributed by the rigid constraints of the five-membered 3,6-anhydro ring and the axial C2 hydroxyl group.

In solution, the 3,6-anhydrogalactose reducing end exists in its aldehyde and hydrated aldehyde forms and, to date, has not been experimentally demonstrated to undergo mutarotation or to exist in α- or β-anomeric form in solution (49,50). If GH117 enzymes were to proceed via a retaining mechanism the product would be an α-sugar, thus, only α-conformers or open chain forms would be expected to be found. Attempts to model the α-$^4C_1$ conformer into the active site revealed there are no residues within H-bonding distance to the α-C1 hydroxyl group except for Glu285 which is within 2.0 Å (too close for a hydrogen bond) and at the wrong angle for a hydrogen bond. Sterically, the $^4C_1$ α-conformation at this site is unlikely to be tolerated without a significant rearrangement of Glu285, a residue that is spatially conserved in the structures of all the GH117 enzymes. Only the $B_{1,4}$ conformation of α-1,3-linked 3,6-anhydro-L-galactose is likely to be accommodated, as seen in the BpGH117 Michaelis complex which is a snapshot right before catalysis (23). Therefore, the trapped β-3,6-anhydro-L-galactose provides substantial crystallographic evidence supporting a mechanism of stereochemistry inversion at the anomeric centre as has been previously suggested (23).

Most of the interactions with the 3,6-anhydro-L-galactose residue are conserved between the BpGH117 $B_{1,4}$ α-3,6-anhydro-L-galactose residue and the ZgAhgB's β-anomeric $^4C_1$ complex. The exceptions in ZgAhgB are Asp192 which hydrogen bonds to the C4 hydroxyl group of β-3,6-anhydro-L-

galactose and three amino acids (Asp89, Glu285, Arg88) which provide hydrogen-bonds to the β-anomeric hydroxyl group (Figure 5 A). The three residues which interact with the β-anomeric hydroxyl are conserved in all the GH117 structures, whereas the C4 hydroxyl binding Asp192 is found only in Zg3597. An ordered water (HOH 2092) is found 3.1 Å from the BpGH117 $B_{1,4}$ α-anomeric carbon (Figure 9 B). BpGH117 Asp90 is positioned for in-line activation of the water molecule for direct attack on the $B_{1,4}$ anomeric carbon. The BpGH117 water molecule is coordinated by the same conserved residues as the ZgAhgB $^4C_1$ β-anomer's C1 hydroxyl. The interactions at this key site likely stabilize the oxocarbenium ion transition state for the one step single displacement inverting mechanism. An inverting mechanism is further supported by the distance of 8.5 Å between the putative catalytic acid and base in the ZgAhgB crystal structure (51).

ZgAhgB *and Zg3597 have novel modes of dimerization*- Both ZgAhgB and Zg3597 have novel modes of dimerization compared to the Clade A enzymes. Zg3597 is unusual among the GH117 enzymes as it is missing its predicted N-terminal dimerization domain in the crystal structure. Interestingly, this domain appears to be present in solution, as shown by the SAXS experiments, most likely indicating that this domain displays significant flexibility. This is not unlikely considering the flexibility demonstrated by the ZgAhgB loop which contributes Asp192 to the active site in the sugar product complex. In the Zg3597 crystal structure dimerization is mediated only by the C-terminal, shifting its relative orientation compared to ZgAhgB and the Clade A GH117 enzymes. ZgAhgB also dimerizes in a shifted state relative to the Clade A enzymes, indicating there is some plasticity in dimerization conformations; however, the C-terminal domains of ZgAhgB and Zg3597 overlap in a similar structural position (Figure 3 B and D).

Zg*3597 has a significantly modified active site*- There are several possible scenarios that might arise from the divergent Zg3597 active site sub-structure. First, that the substrate is significantly different for this enzyme and thus His306 is shifted to accommodate the substrate. Second, that a substrate stabilized rearrangement occurs upon binding substrate, possibly implicating the N-terminal in dimerization as, for example, observed in the solution structure. Third, it is possible that this protein is a

non-catalytic protein such as seen in the "inactivated" GH18 chitin binding proteins (52,53). Finally, that the crystal structure of Zg3597 is in a non-native form induced by crystallization, although the SAXS analysis does reveal that the protein exists as a dimer in solution (Figure 7 A). In our actual working hypothesis we assume that the dimeric form of Zg3597 is the biological active one, and the observed flexibility of the amino-terminal domain in Zg3597 is associated to its catalytic function, but since the natural substrate is not known yet we are not able to verify this assumption.

GH117 novel subsite structure- ZgAhgB and Zg3597 in particular, have a more open entrance to their active sites when compared to their Clade A counterparts. This may confer an advantage when presented with the diverse decorated sugars found in the agar polygalactan chain. Though it has been postulated that ZgAhgB has an extra -2 subsite (23) we find that it is unlikely, based on this structural analysis. The ZgAhgB C4 hydroxyl-binding Asp192 impedes the presence of a -2 subsite without significant structural reorganization; however, Zg3597's Asp210 is shifted considerably (Figure 5 A) leaving a more open pocket and some accessibility to the 3,6-anhydro-L-galactose C4 hydroxyl group, thus, we cannot discount the possibility of a shallow, surface located, -2 subsite in Zg3597. Sequence analysis reveals that all the new Clade F enzymes have a glycine in place of the tryptophan within their active sites and the glutamine, which coordinates the C4 hydroxyl group, is replaced by a threonine. These Clade F enzymes are certainly candidates for a -2 subsite.

The absence of activity for the Clade D Zg3597 on neoagaro-oligosaccharides and the predicted novel active sites of the Clade F GH117 enzymes suggest further research is required in order to understand the structure and function of the GH117 enzymes. The red algal agar substrate is truly unique and classic methods of determining catalytic mechanism are challenging, thus, much remains to be discovered on the mechanism of GH117 catalysis.

The reported structures of enzymes and CBMs in complex with agarose derivatives, with the exception of BpGH117 (23), have trapped the 3,6-anhydro-L-galactose in a relaxed α-$^{4}C_{1}$ conformation

(19,47,48). This is the first time the β-anomer $^4C_1$ conformation has been demonstrated experimentally and it sheds important insight into the mechanism of the GH117 enzymes as well as into the structure of these unique bicyclic sugars.

REFERENCES

1. Butterfield, N. (2000) Bangiomorpha pubescens n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. *Paleobiology* **26**, 386-404
2. Popper, Z. A., Michel, G., Herve, C., Domozych, D. S., Willats, W. G., Tuohy, M. G., Kloareg, B., and Stengel, D. B. (2011) Evolution and diversity of plant cell walls: from algae to flowering plants. *Annual review of plant biology* **62**, 567-590
3. Kropf, D. L., Kloareg, B., and Quatrano, R. S. (1988) Cell wall is required for fixation of the embryonic axis in Fucus zygotes. *Science* **239**, 187-190
4. Rees, D. A. (1969) Structure, conformation, and mechanism in the formation of polysaccharide gels and networks. *Advances in carbohydrate chemistry and biochemistry* **24**, 267-332
5. Lawson, C. J., and Rees, D. A. (1970) An enzyme for the metabolic control of polysaccharide conformation and function. *Nature* **227**, 392-393
6. Genicot-Joncour, S., Poinas, A., Richard, O., Potin, P., Rudolph, B., Kloareg, B., and Helbert, W. (2009) The cyclization of the 3,6-anhydro-galactose ring of iota-carrageenan is catalyzed by two D-galactose-2,6-sulfurylases in the red alga *Chondrus crispus*. *Plant physiology* **151**, 1609-1616
7. Collen, J., Porcel, B., Carre, W., Ball, S. G., Chaparro, C., Tonon, T., Barbeyron, T., Michel, G., Noel, B., Valentin, K., Elias, M., Artiguenave, F., Arun, A., Aury, J. M., Barbosa-Neto, J. F., Bothwell, J. H., Bouget, F. Y., Brillet, L., Cabello-Hurtado, F., Capella-Gutierrez, S., Charrier, B., Cladiere, L., Cock, J. M., Coelho, S. M., Colleoni, C., Czjzek, M., Da Silva, C., Delage, L., Denoeud, F., Deschamps, P., Dittami, S. M., Gabaldon, T., Gachon, C. M., Groisillier, A., Herve, C., Jabbari, K., Katinka, M., Kloareg, B., Kowalczyk, N., Labadie, K., Leblanc, C., Lopez, P. J., McLachlan, D. H., Meslet-Cladiere, L., Moustafa, A., Nehr, Z., Nyvall Collen, P., Panaud, O., Partensky, F., Poulain, J., Rensing, S. A., Rousvoal, S., Samson, G., Symeonidi, A., Weissenbach, J., Zambounis, A., Wincker, P., and Boyen, C. (2013) Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 5247-5252
8. Jonas Collén, M. L. C., James Craigie, Elizabeth Ficko-Blean, Cécile Hervé, Stacy A. Krueger-Hadfield, Catherine Leblanc, Gurvan Michel, Philippe Potin, Thierry Tonon, Catherine Boyen. (2014) *Chondrus crispus* – A Present and Historical Model Organism for Red Seaweeds. in *Advances in Botanical Research*, Elsevier Ltd. pp 53-89
9. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2008) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic acids research*
10. 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. *Applied and environmental microbiology* **73**, 4691-4694

11. Martin, M., Portetelle, D., Michel, G., and Vandenbol, M. (2014) Microorganisms living on macroalgae: diversity, interactions, and biotechnological applications. *Applied microbiology and biotechnology* **98**, 2917-2935
12. 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
13. Hehemann, J. H., Kelly, A. G., Pudlo, N. A., Martens, E. C., and Boraston, A. B. (2012) Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 19786-19791
14. 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. *Environmental microbiology* **13**, 1253-1270
15. Barbeyron, T., Gerard, A., Potin, P., Henrissat, B., and Kloareg, B. (1998) The kappa-carrageenase of the marine bacterium Cytophaga drobachiensis. Structural and phylogenetic relationships within family-16 glycoside hydrolases. *Molecular biology and evolution* **15**, 528-537
16. Barbeyron, T., Michel, G., Potin, P., Henrissat, B., and Kloareg, B. (2000) iota-Carrageenases constitute a novel family of glycoside hydrolases, unrelated to that of kappa-carrageenases. *The Journal of biological chemistry* **275**, 35499-35505
17. Rebuffet, E., Barbeyron, T., Jeudy, A., Jam, M., Czjzek, M., and Michel, G. (2010) Identification of catalytic residues and mechanistic analysis of family GH82 iota-carrageenases. *Biochemistry* **49**, 7590-7599
18. Hehemann, J. H., Correc, G., Thomas, F., Bernard, T., Barbeyron, T., Jam, M., Helbert, W., Michel, G., and Czjzek, M. (2012) Biochemical and structural characterization of the complex agarolytic enzyme system from the marine bacterium *Zobellia galactanivorans*. *The Journal of biological chemistry*
19. 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
20. 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. *The Biochemical journal* **385**, 703-713
21. Allouch, J., Jam, M., Helbert, W., Barbeyron, T., Kloareg, B., Henrissat, B., and Czjzek, M. (2003) The three-dimensional structures of two beta-agarases. *The Journal of biological chemistry* **278**, 47171-47180
22. 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-neoagarobiose hydrolase from *Saccharophagus degradans* 2-40. *Biochemical and biophysical research communications* **412**, 238-244
23. Hehemann, J. H., Smyth, L., Yadav, A., Vocadlo, D. J., and Boraston, A. B. (2012) Analysis of Keystone Enzyme in Agar Hydrolysis Provides Insight into the Degradation of a Polysaccharide from Red Seaweeds. *Journal of Biological Chemistry* **287**, 13985-13995
24. Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**, 785-786
25. Groisillier, A., Herve, C., Jeudy, A., Rebuffet, E., Pluchon, P. F., Chevolot, Y., Flament, D., Geslin, C., Morgado, I. M., Power, D., Branno, M., Moreau, H., Michel, G., Boyen, C., and Czjzek, M.

(2010) MARINE-EXPRESS: taking advantage of high throughput cloning and expression strategies for the post-genomic analysis of marine organisms. *Microbial cell factories* **9**, 45
26. P. V. Konarev, V. V. V., A. V. Sokolova, M. H. J. Koch and D. I. Svergun. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr* **36**, 1277-1282
27. Svergun, D. I. (1992) Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *J Appl Crystallogr* **25**, 495-503
28. Svergun, D. F. a. D. I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Crystallogr* **42**, 342-346
29. Svergun, D. I., Petoukhov, M. V. & Koch, M. H. (2001). *Biophys. J.*, 2946-2953
30. Volkov, V. V. S., D. I. (2003) Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Crystallogr*, 860-864
31. Kozin, M. B., and Svergun, D. I. (2001) Automated matching of high- and low-resolution structural models. *J Appl Crystallogr* **34**, 33-41
32. Svergun, D., Barberato, C., and Koch, M. H. J. (1995) CRYSOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* **28**, 768-773
33. Powell, H. R. (1999) The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta crystallographica. Section D, Biological crystallography* **55**, 1690-1695
34. Evans, P. (2006) Scaling and assessment of data quality. *Acta crystallographica. Section D, Biological crystallography* **62**, 72-82
35. Lawrence Kelley, B. J. Phyre2: Protein Homology/analogY Recognition Engine V 2.0. Structural Bioinformatics Group, Imperial College, London
36. Vagin, A., and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J Appl Crystallogr* **30**, 1022-1025
37. (1994) The CCP4 suite: programs for protein crystallography. *Acta crystallographica. Section D, Biological crystallography* **50**, 760-763
38. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta crystallographica. Section D, Biological crystallography* **53**, 240-255
39. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography* **60**, 2126-2132
40. Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* **50**, 437-450
41. Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research* **30**, 3059-3066
42. Kumar, S., Tamura, K., and Nei, M. (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Briefings in bioinformatics* **5**, 150-163
43. Perez, C., Muckle, M. T., Zaleski, D. P., Seifert, N. A., Temelso, B., Shields, G. C., Kisiel, Z., and Pate, B. H. (2012) Structures of cage, prism, and book isomers of water hexamer from broadband rotational spectroscopy. *Science* **336**, 897-901
44. Zheng, H., Chruszcz, M., Lasota, P., Lebioda, L., and Minor, W. (2008) Data mining of metal ion environments present in protein structures. *Journal of inorganic biochemistry* **102**, 1765-1776
45. Zechel, D. L., and Withers, S. G. (2000) Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Accounts of chemical research* **33**, 11-18

46. Jongkees, S. A. K., and Withers, S. G. (2014) Unusual Enzymatic Glycoside Cleavage Mechanisms. *Accounts of chemical research* **47**, 226-235
47. Pluvinage, B., Hehemann, J. H., and Boraston, A. B. (2013) Substrate Recognition and Hydrolysis by a Family 50 exo-beta-Agarase, Aga50D, from the Marine Bacterium *Saccharophagus degradans*. *Journal of Biological Chemistry* **288**, 28078-28088
48. 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. *Journal of Biological Chemistry* **281**, 17099-17107
49. Rochas, C., Potin, P., and Kloareg, B. (1994) Nmr Spectroscopic Investigation of Agarose Oligomers Produced by an Alpha-Agarase. *Carbohydrate research* **253**, 69-77
50. Ducatti, D. R. B., Colodi, F. G., Goncalves, A. G., Duarte, M. E. R., and Noseda, M. D. (2011) Production of agaro- and carra-oligosaccharides by partial acid hydrolysis of galactans. *Rev Bras Farmacogn* **21**, 296-304
51. Rye, C. S., and Withers, S. G. (2000) Glycosidase mechanisms. *Current opinion in chemical biology* **4**, 573-580
52. Hennig, M., Jansonius, J. N., Vanscheltinga, A. C. T., Dijkstra, B. W., and Schlesier, B. (1995) Crystal-Structure of Concanavalin-B at 1.65 Angstrom Resolution - an Inactivated Chitinase from Seeds of Canavalia-Ensiformis. *Journal of molecular biology* **254**, 237-246
53. Bleau, G., Massicotte, F., Merlen, Y., and Boisvert, C. (1999) Mammalian chitinase-like proteins. *Exs* **87**, 211-221
54. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research* **42**, W320-324
55. Read, R. J. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallographica Section A* **A42**, 140-149

FOOTNOTES

EFB was funded by a post-doctoral fellowship supported by the Région Bretagne through the program 'Algevol' with reference SAD_Obex_EMBRC 12010152. ER benefited from a PhD fellowship by the Région Bretagne through the program 211-B2-9/ARED 091539 'Iotase3D' and MC is thankful for support by the French Centre National de Recherches Scientifiques. This work also benefited from the support of the French Government through the National Research Agency with regards to an investment expenditure program IDEALG with reference ANR-10-BTBR-04. We are indebted to the European Synchrotron Research Facilities (Grenoble, France) for regular access to X-ray beamlines and to all local contacts for their support during data collection at the MX beamlines ID23-1 and ID29. We also thank the German synchrotron at DESY (Hamburg) for access to the SAXS beamline X33, and especially Clement Blanchet (EMBL-Hamburg) for valuable help during the SAXS data collection.

FIGURE LEGENDS

Table 1. Data evaluation of experimental SAXS curves at different concentrations of Zg3597.

Table 2. X-ray crystallography data collection statistics.

Figure 1. Chemical drawing of $^{4}C_{1}$ β-3,6-anhydro-L-galactose.

Figure 2. FACE on enzymatic digests of neoagarooligosaccharide. In lanes 1-7 digests were done on a mixture of neoagarotetraose and neoagarohexaose. In lanes 8-14 digests were done on a mixture of neoagarobiose and neoagarotetraose. Lane 1. Zg3615, Lane 2. Zg3597, Lane 3. Zg4663, Lane 4. Zg3597, Zg4663, Lane 5. Zg3615, Zg4663 Lane 6. Zg3615, Zg3597, Lane 7. No enzyme, Lane 8. Zg3615, Lane 9. Zg3597, Lane 10. Zg4663, Lane 11. Zg3597, Zg4663, Lane 12. Zg3615, Zg4663 Lane 13. Zg3615, Zg3597, Lane 14. No enzyme.

Figure 3. Dimer conformations of the GH117 enzymes: A. ZgAhgA (Zg4663) with Zn^{2+}, B. ZgAhgB (Zg3615) with Ca^{2+}, C. SdGH117 (no metal ion), D. Zg3597 with Ca^{2+} and E. BpGH117 with Mg^{2+}. The N and C termini have been labelled and highlighted.

Figure 4. ESPript 3.0 (54) generated figure of the sequence alignment between the five GH117 enzymes from *Z. galactanivorans*. Active site residues from the -1 subsite of Zg3615 are indicated with shaded circles. The secondary structure representation of Zg3615 is shown above the sequence alignment and the secondary structure representation of Zg3597 is shown below.

Figure 5. A. Stereo figure of the -1 active site of Zg3597 (gold) overlapped with ZgAhgB (Zg3615) (mauve) which is in complex with β-3,6-anhydro-L-galactose. Hydrogen bonds between ZgAhgB and the sugar are indicated. The adjacent ordered Ca^{2+} atom is also shown. B. An omit map was generated by omitting β-3,6-anhydro-L-galactose from the refinement. The resulting maps are maximum-likelihood/σ_A(55)-weighted $2F_{obs} - F_{calc}$ contoured at 2σ (blue) and $F_{obs} - F_{calc}$ contoured at 3σ (green) (0.72 and 0.19 electrons/Å^3, respectively).

Figure 6. A. Secondary structure overlap of Zg3597 and ZgAhgB (Zg3615) showing the Zg3597 loop flip. Zg3597 is coloured pink and ZgAhgB is coloured raspberry. B. Amino acids which contribute to the stability of the loop flip in Zg3597. C. The same loop in the opposing orientation in ZgAhgB.

Figure 7. Envelopes and resulting curves from small angle X-ray scattering. A. Envelope obtained by GASBOR (best fit χ^2 =1.6) using the scattering curve at 4.78 mg/ml. The marked angle of 100° (calculated as a dihedral angle) above the shape is a rough estimate of the relative orientation of the 2 catalytic pockets. B. Superimposition of the Zg3597 homodimer from the crystal structure onto the same SAXS envelope as in A (the orientation is roughly turned by 90°). C. Surface (transparent) and cartoon representation of the dimer of Zg3597 as determined by the crystal structure. The marked angle of 130° (calculated as a dihedral angle; shown below the shape) is a rough estimate of the relative orientation of the 2 catalytic pockets. In panels B and C the two-fold symmetry axis of the shapes lies in the plane (vertical), and out of the plane in panel A. D. Experimental scattering curve of Zg3597 in solution at 4.78 mg/ml (blue crosses) and best fit obtained by the *ab initio* shape calculation from GASBOR (red line). Inset: P(r) function of the experimental scattering curve at 4.78 mg/ml.

Figure 8. Unrooted phylogenetic tree of the GH117 homologues. The phylogenetic tree was generated using the Maximum Likelihood approach with the program Mega6. Bootstrap numbers are indicated. Accession numbers of the selected sequence homologues are available in supplemental table 1. The structures solved to date are marked with a diamond for the GH117 enzymes produced by *Z. galactanivorans* and by a circle for the others.

Figure 9. 3,6-anhydro-L-galactose structures. A. $^4C_1$ α-3,6-anhydro-L-galactose such as found in agarose. B. Neoagarobiose from the BpGH117 pdb 4AK7 with $B_{1,4}$ α-3,6-anhydro-L-galactose. Aspartate, the putative catalytic general base, is shown coordinating an ordered water just below the anomeric carbon. C. $^4C_1$ β-3,6-anhydro-L-galactose from the ZgAhgB (Zg3615) pdb 4U6D. Here the putative catalytic general base is shown hydrogen bonded to the β-hydroxyl group, in the same position as the ordered water from the BpGH117 structure.

Synopsis of the main findings of the article for inclusion in the Table of Contents:

Structural, functional and phylogenetic analyses reveal enzyme diversity within family 117 of the glycoside hydrolases which remove terminal alpha-1,3-linked 3,6-anhydro-L-galactose from neoagaro-oligosaccharides produced by β-agarases. A product complex with the unique β-3,6-anhydro-L-galactose provides strong crystallographic evidence for an inverting mechanism in this family of enzymes.

Table 1.

Sample [mg/ml]	Global Rg (Å)	Rg from Gunier approx (Å)	Rg from Porod (Å)	Rg from P(r) (Å)	Dmax (Å)	Calculated Mw (kDa)	Gasbor (χ^2)	Crysol (χ^2)
7.9	34.5±1.0	34.5	35.4	33.6	102	108	3.4	ND
4.8	35.1±1.2	35.3	36.2	33.8	100	108	1.61	3.4
4.1	34.1±1.0	34.3	34.4	33.6	100	113	2.07	ND
2.0	34.7±1.8	33.6	36.5	34.2	101	110	2.44	ND
Mean Rg	**34.6**				Mean MW	**109**		

Table 2.

Data collection statistics	Zg3615	Zg3597
Beamline	PROXIMA 1 SOLEIL	ID14-4 ESRF
Wavelength (Å)	0.980	0.979
Space group	$P2_12_12_1$	F222
Resolution (Å)	30.00-1.70 (1.80-1.70)	30.00-2.30 (2.42-2.30)
Cell dimension a, b,c (Å)	57.11, 226.13, 67.12	187.79, 223.52, 225.22
	90.00 90.00 90.00	90.00 90.0 90.00
α, β, γ (^{0})		
R_{merge} (%)	4.8 (38.2)	8.8 (43.7)
Completeness (%)	100.0 (100.0)	99.7 (99.8)
$<I/\sigma I>$	25.0 (5.0)	13.2 (3.9)
Redundancy	7.4 (7.2)	6.0 (6.1)
Total reflections	711242	624636
Unique reflections	96642	103865
Refinement statistics		
R (%)	18.7	17.8
R_{free} *(%)*	21.9	22.5
RMSD		
Bond lengths (Å)	0.014	0.017
Bond angles (^{o})	1.571	1.939
Average *B*-factors (Å^2)		
Protein Chain A	16.7	39.8
Protein Chain B	25.7	39.1
Protein Chain C	N/A	37.7
Protein Chain D	N/A	39.9
Water molecules	28.1	39.5
Sugar molecule	19.2	N/A
Metal ions	17.3	40.0
Number of atoms		
Protein Chain A	3108	2686
Protein Chain B	3095	2705
Protein Chain C	N/A	2693
Protein Chain D	N/A	2680
Water molecules	730	737
Sugar	11	N/A
Metal ions	4	9
Ramachandran statistics		
Most favored (%)	95.2	95.8
Additional allowed (%)	4.8	3.9
Disallowed (%)	0	0.3

Figure 1.

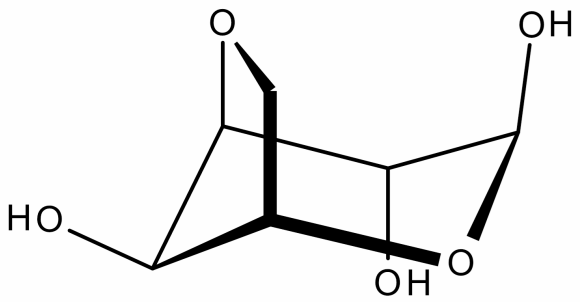


Figure 2.

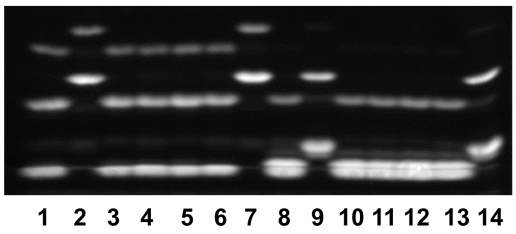


Figure 3.

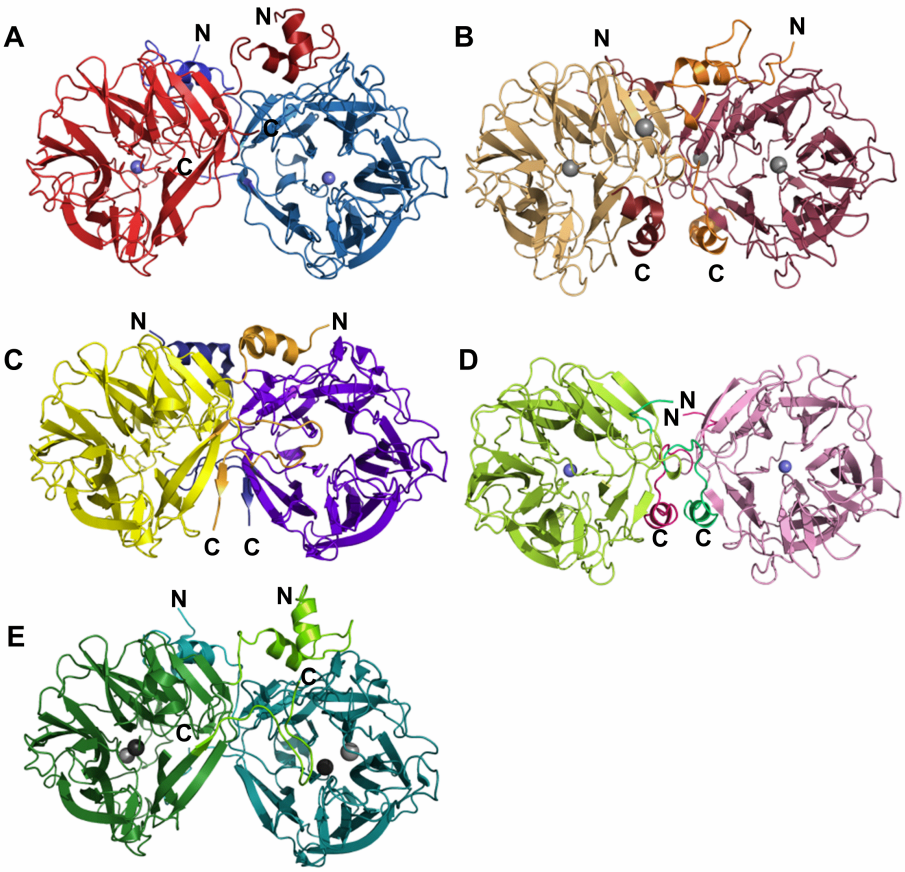


Figure 4.

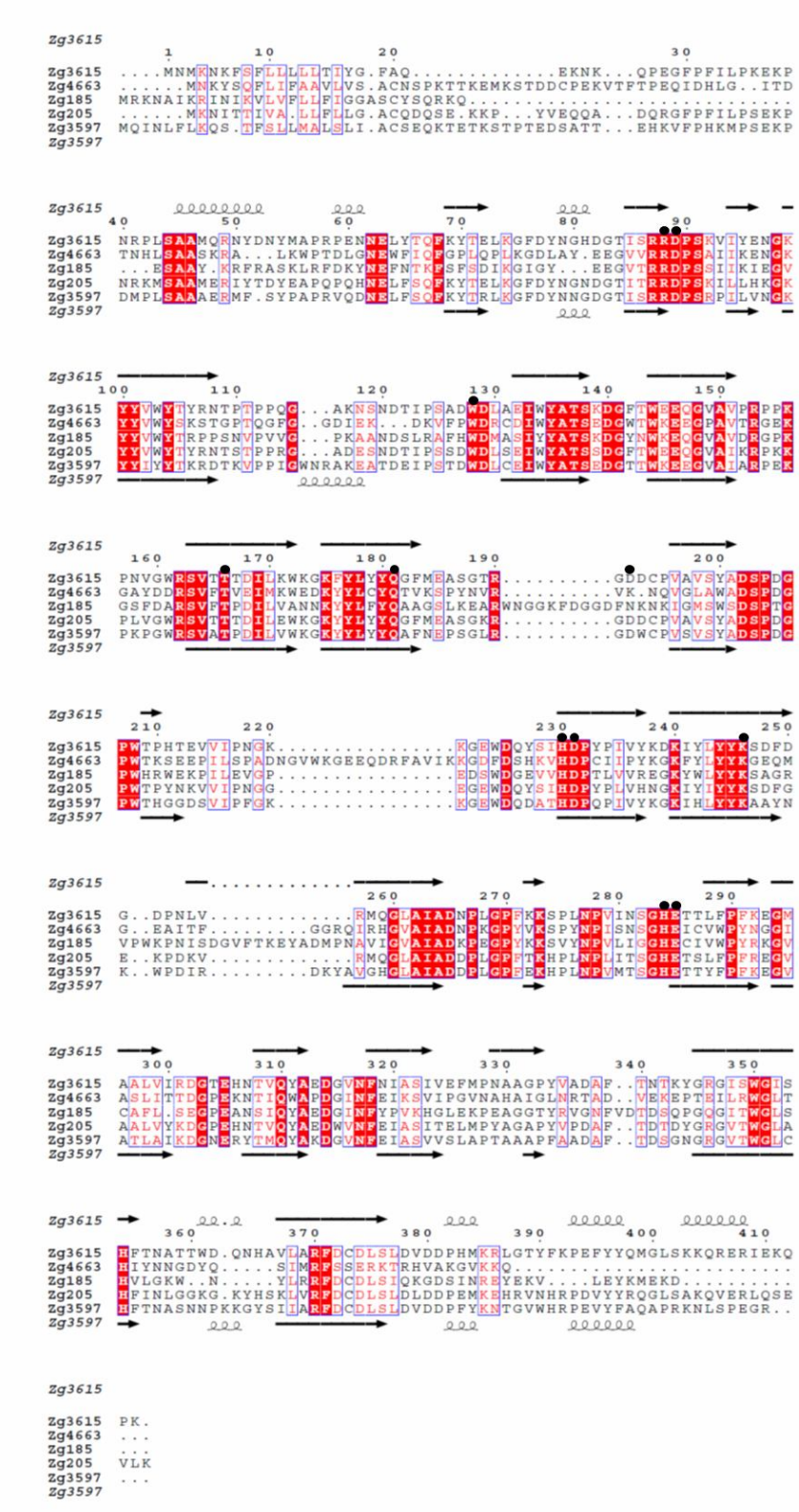

Zg3615
Zg4663
Zg185
Zg205
Zg3597

Figure 5.

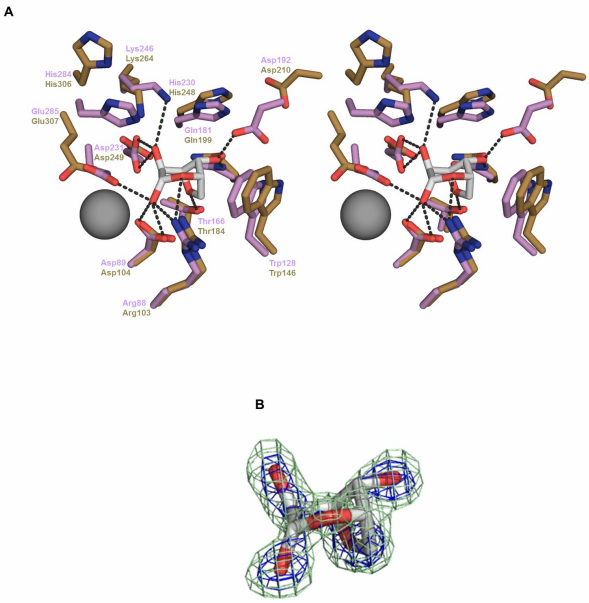


Figure 6.

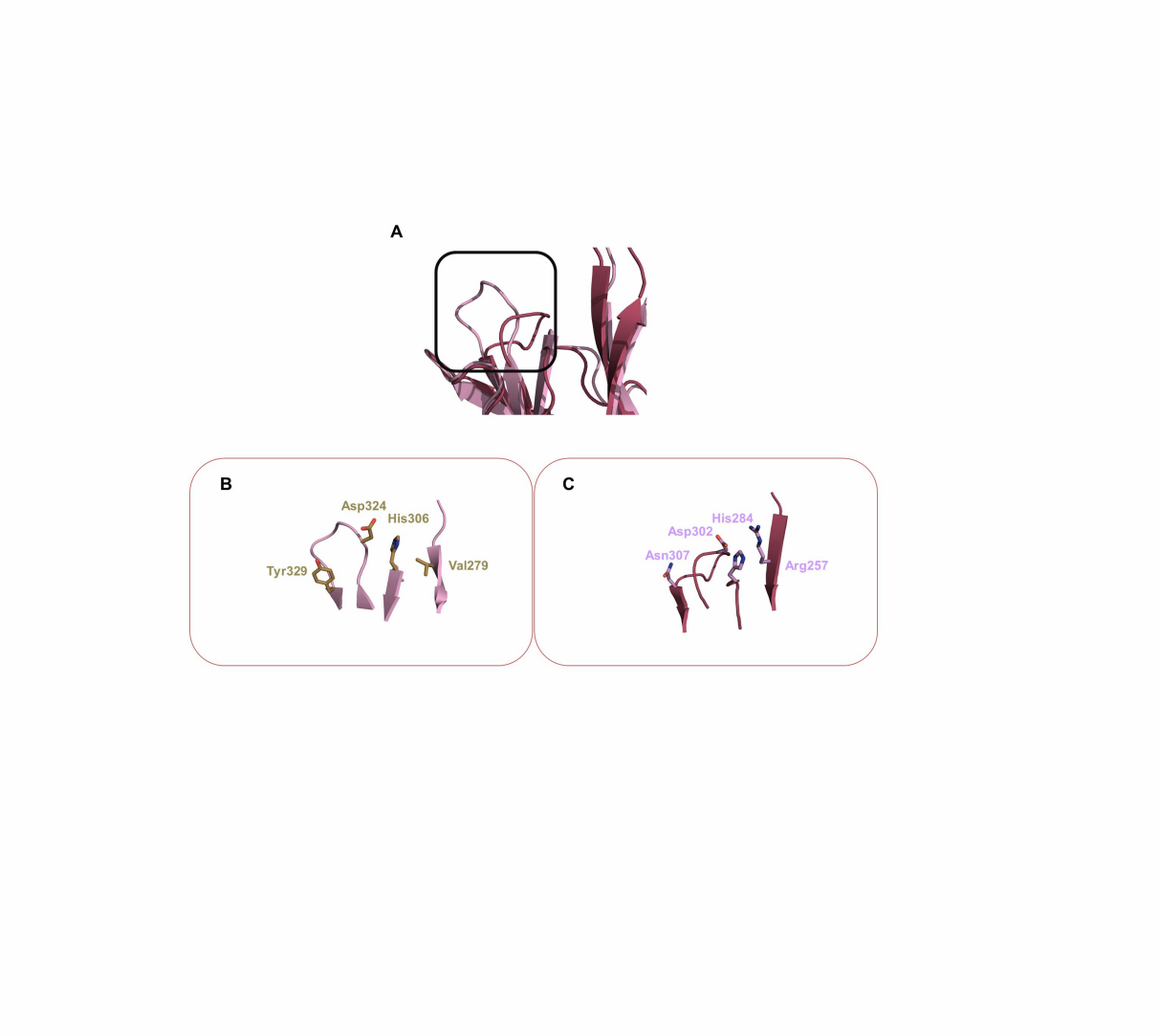

A
B
Asp324
His306
Tyr329
Val279
C
His284
Asp302
Asn307
Arg257

Figure 7.

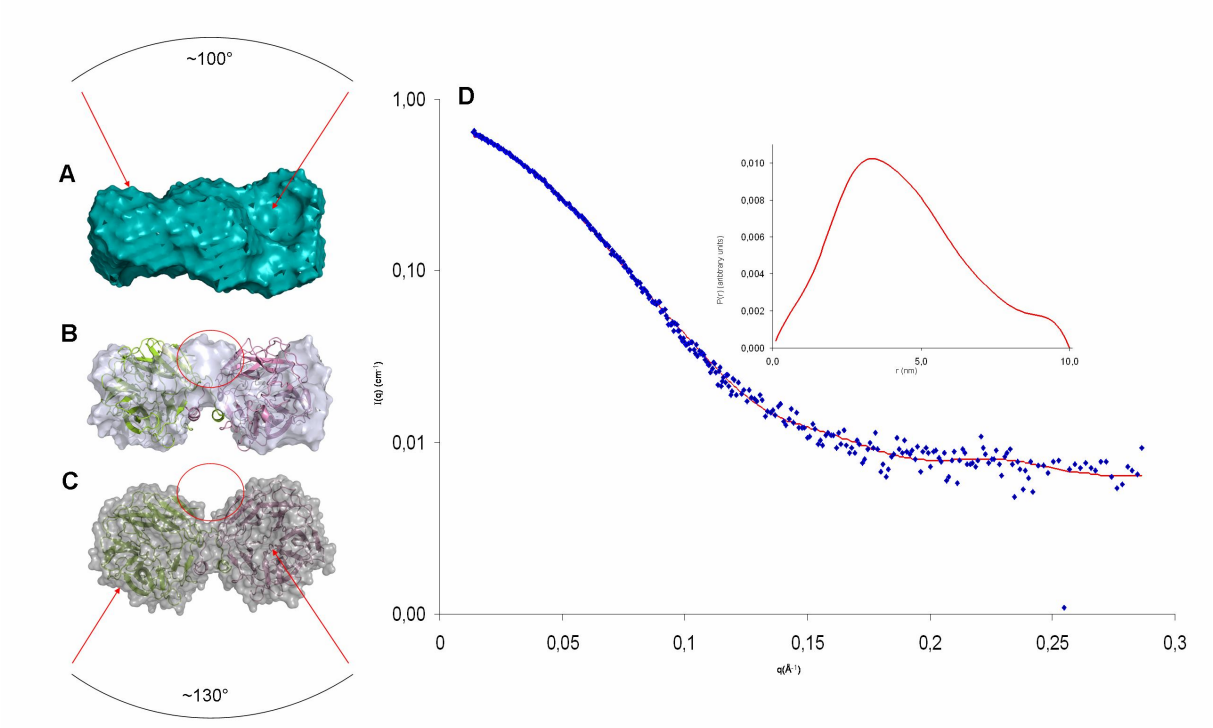
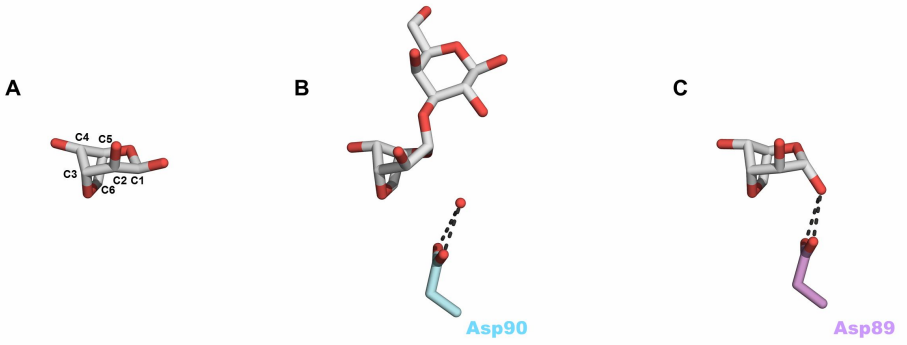

~100°
A
B
C
~130°
D
1,00
0,10
0,01
0,00
I(q) (cm⁻¹)
0
0,05
0,1
0,15
0,2
0,25
0,3
q(Å⁻¹)
0,010
0,008
0,006
0,004
0,002
0,000
P(r) (arbitrary units)
0,0
5,0
10,0
r (nm)

Figure 8.



$^4C_1$ α-3,6-anhydro-L-galactose $B_{1,4}$ 3,6-anhydro-L-galactose-α-1,3-D-galactose $^4C_1$ β-3,6-anhydro-L-galactose