Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

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,[a,b]* Delphine Duffieux,[a,b] Étienne Rebuffet,[a,b] Robert Larocque,[a,b] Agnes Groisillier,[a,b] Gurvan Michel[a,b]* and Mirjam Czjzek[a,b]*

[a]Sorbonne Universités, UPMC Université Paris 06, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, 29688 Roscoff CEDEX, France, and [b]CNRS, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, 29688 Roscoff CEDEX, France

Correspondence e-mail: efickoblean@sb-roscoff.fr, gurvan.michel@sb-roscoff.fr, czjzek@sb-roscoff.fr

Received 19 September 2014
Accepted 14 November 2014

PDB references: Zg3615, 4u6d; Zg3597, 4u6b

The family 117 glycoside hydrolase (GH117) enzymes have exo-α-1,3-(3,6-anhydro)-L-galactosidase activity, removing terminal nonreducing 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. The investigation of 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 the structure of this unique sugar produced by hydrolysis of the agarophyte red algal cell wall.

1. Introduction

Red seaweeds (Rhodophyta) are ancient photosynthetic eukaryotes and are considered to be the first organisms to have developed advanced multicellularity (Butterfield, 2000). These red macroalgae are important primary producers for 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 (Popper *et al.*, 2011). Red seaweeds contain either agars or carrageenans, both of which are sulfated galactans, and thus are referred to as agarophytes (*e.g.* *Gracillaria*, *Gelidium* and *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 (Kropf *et al.*, 1988). 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 referring 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 sulfation.

Both agarophyte and carrageenophyte algae synthesize an unusual carbohydrate moiety, the α-1,3-linked bicyclic 3,6-anhydrogalactose (LA monomer in agars and 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 (Rees, 1969). 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 (Lawson & Rees, 1970; Genicot-Joncour *et al.*, 2009). However, knowledge of the biosynthesis of sulfated galactans remains very limited and the first global view of the carbohydrate metabolism of red seaweeds has only recently been provided by the genome sequencing of the carrageenophyte *Chondrus crispus* (Collén *et al.*, 2013, 2014).

The regular structure of the agar backbone is often modified with methyl groups and variations in acidic side groups such as sulfate or pyruvate (Rees, 1969; Popper *et al.*, 2011). The main repeating disaccharides are agarobiose (G-β-1,4-LA) and 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 carbohydrate-active enzyme (CAZyme) families (Cantarel *et al.*, 2008) 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 (Flament *et al.*, 2007). 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 (for a review, see Martin *et al.*, 2014). Finally, the β-porphyranases, which cleave the β-1,4 bond between two neoporphyranobiose units (L6S-α-1,3-G), constitute new subfamilies within families GH16 (Hehemann *et al.*, 2010) and GH86 (Hehemann, Kelly *et al.*, 2012).

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 (Rebuffet *et al.*, 2011). ZgAhgA specifically removes 3,6-anhydro-L-galactose at the nonreducing 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 (Martin *et al.*, 2014). In particular, it can assimilate sulfated galactans from red seaweeds. *Z. galactanivorans* possesses one GH16 kappa-carrageenase (Barbeyron *et al.*, 1998) and three GH82 iota-carrageenases (Barbeyron *et al.*, 2000; Rebuffet *et al.*, 2010). This bacterium has also evolved a complex agarolytic system composed of four GH16 β-agarases and five GH16 β-porphyranases (Hehemann, Correc *et al.*, 2012). Several of these enzymes have been already biochemically and structurally characterized: the β-agarases ZgAgaA, ZgAgaB (Allouch *et al.*, 2003, 2004; Jam *et al.*, 2005) and ZgAgaD (Hehemann, Correc *et al.*, 2012) and the β-porphyranases ZgPorA and ZgPorB (Hehemann *et al.*, 2010; Hehemann, Correc *et al.*, 2012).

Z. galactanivorans possesses five GH117 enzymes, including the abovementioned ZgAhgA (Zg4663). The GH117 family has been divided into three phylogenetic clades (A, B and C), with ZgAhgA belonging to clade A (Rebuffet *et al.*, 2011). This enzyme adopts a five-bladed β-propeller fold and forms a dimer in solution by the 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* (Ha *et al.*, 2011) and BpGH117 from *Bacteroides plebeius* (Hehemann, Smyth *et al.*, 2012). These two enzymes are also α-1,3-(3,6-anhydro)-L-galactosidases and form dimers by swapping of the conserved N-terminal domain and also of a C-terminal extension that is absent in ZgAhgA. The structures of SdGH117 and BpGH117 were solved in complex with a D-galactose at subsite +1 (Ha *et al.*, 2011) and neoagarobiose at subsites +1 and −1 (Hehemann, Smyth *et al.*, 2012), respectively. A site-directed mutagenesis analysis identified the potential catalytic residues, and notably a conserved histidine was proposed to act as a general acid (Hehemann, Smyth *et al.*, 2012). To date, only clade A enzymes have been structurally and functionally characterized, and little is known for the two clade B enzymes. Based on sequence comparison within the active-site region of the clade C enzymes, which suggested the 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 (Hehemann, Smyth *et al.*, 2012).

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 potential functional diversity and/or synergy. New phyologenetic analysis reveals that within the previously defined *Z. galactanivorans* clade C orthologues (Rebuffet *et al.*, 2011) there are significant differences between enzymes; this divergence results in clade C enzymes which might not have the same specificities. With

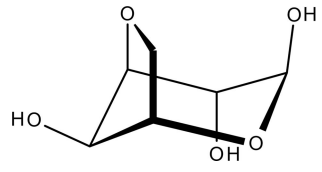


Figure 1
Chemical structure of $^4C_1$ β-3,6-anhydro-L-galactose.

the growth in the available genomic sequences, and the structural and biochemical characterization performed on the GH117 enzymes to date, we have new phylogenetic evidence that there are at least six clades within this highly unusual family.

In order to investigate the fine differences within the clades, we chose to study two of the (previously defined) GH117 clade C enzymes: Zg3615 (now clade E) and Zg3597 (now clade D). The structure of Zg3597 was determined in its native form and that of Zg3615 was determined in complex with its product β-3,6-anhydro-L-galactose (Fig. 1). This has important implications for 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.

2. Experimental procedures

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

2.1. Cloning

PCR was used to amplify gene fragments from *Z. galactanivorans* genomic DNA encoding the loci for Zg3615 and Zg3597 without their predicted signal peptides. *SignalP* 4.1 (Petersen *et al.*, 2011) 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 entry 3p2n; Rebuffet *et al.*, 2011) 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 (Groisillier *et al.*, 2010). The following primers and restriction-enzyme sites were used to clone Zg3615: forward primer 5′-TTTTTTGTCGACCAGCCCGAAGGGTTTCCGTTCA-3′ including *Sal*I (into the *Xho*I site of pFO4) and reverse primer 5′-AAAAAACTGCAGTTATTTAGGTTGTTTTTCGATG-CGTTC-3′ including a *Pst*I site (into the *Nsi*I site of pFO4). The following primers and restriction-enzyme sites were used to clone Zg3597: forward primer 5′-GGGGGGGGATCCT-GTTCCGAACAAAAAACGGAAACCA-3′ with a *Bam*HI site and reverse primer 5′-CCCCCCGAATTCTTATCGT-CCTTCGGGACTTAGGTTT-3′ with an *Eco*RI site.

2.2. Protein production and purification

Plasmids were transformed into competent *Escherichia coli* strain BL21 (DE3) and grown with shaking in 1 l LB medium containing 100 µg ml^{-1} ampicillin at 37°C to an optical density of ∼1.0, whereupon the temperature was lowered to 20°C for 1 h followed by overnight induction of protein production by the addition of IPTG to 0.5 m*M*. Cells were harvested the following morning by centrifugation at 3000 rev min^{-1} for 30 min. A chemical lysis procedure was used to lyse the cells. Briefly, the cells were resuspended in 20 ml 50 m*M* Tris pH 8.0 with 25% sucrose and 10 µg lysozyme and left stirring at 4°C for 20 min. Next, 40 ml 1% deoxycholate, 1% Triton X-100, 20 m*M* Tris pH 7.5, 100 m*M* NaCl was added and incubated with stirring at 4°C for 10 min. Finally, $MgCl_2$ to a concentration of 5 m*M* and 0.1 mg DNase were added and the lysis reaction was incubated at room temperature for 20 min until the viscosity decreased. The lysis solution was clarified at 13 000 rev min^{-1} over 45 min at 4°C. Supernatant containing the soluble proteins of interest was run on a 5 ml GE HisTrap HP column. Briefly, using an ÄKTA FPLC, the IMAC column was loaded with 0.45 µ*M* filtered supernatant in buffer *A* (20 m*M* Tris pH 8.0, 100 m*M* NaCl), washed with buffer *A* + 1% buffer *B* (20 m*M* Tris pH 8.0, 100 m*M* NaCl, 1 *M* imidazole) and eluted using a gradient of 1–100% buffer *B*. Fractions containing the protein of interest were pooled and concentrated using an Amicon stirred-cell concentrator (molecular-weight cutoff 3 kDa). This was followed by size-exclusion chromatography on a Superdex S200 column in 20 m*M* Tris pH 8.0, 100 m*M* NaCl. Again, fractions containing the protein of interest were pooled and concentrated using an Amicon stirred-cell concentrator (molecular-weight cutoff 3 kDa). Proteins were assessed as having >95% purity by SDS–PAGE. The concentration was calculated from the A_{280} and the extinction coefficient calculated using the *ProtParam* tool from ExPASy ($\varepsilon_{0.1\%}$ of 2.1 for Zg3615 and 2.0 for Zg3597). The final concentrations of Zg3615 and Zg3597 were 18.0 and 10.0 mg ml^{-1}, respectively.

2.3. Thin-layer chromatography (TLC)

Enzymatic reactions were performed in 0.5% oligosaccharide or polysaccharide, 100 m*M* Tris pH 7.5, 0.1 m*M* $ZnSO_4$, 100 m*M* NaCl, 25 m*M* $CaCl_2$ and 0.1 µg µl^{-1} enzyme. Neoagaro-oligosaccharides were purified as described previously (Jam *et al.*, 2005). 2 µl of the reaction volume was spotted onto Merck TLC Silica Gel 60 plates and developed in 3× butanol, 1× ethanol and 1× water. The plates were dried and then stained in 50 ml 10% H_2SO_4 in ethanol added to 25 ml of a 0.2% solution of naphthoresorcinol (1,3-dihydroxynaphthalene) in ethanol just before use. The plates were dried again and incubated at 60°C until spots appeared (approximately 45 min).

2.4. Fluorophore-assisted chromatography electrophoresis (FACE)

Enzymatic reactions were performed as above. 50 µg of oligosaccharide or polysaccharide reaction volume was dried down using a speed vacuum. Labelling was performed by adding 2 µl ANTS (8-aminonaphthalene-1,3,6-trisulfonic acid) in DMSO to the dried sample followed by the addition of 5 µl freshly prepared sodium cyanoborohydride in DMSO. The samples were mixed well, centrifuged and left at 37°C in the dark overnight. 20 µl 20% glycerol was then added as a gel-loading aid. 8 µl sample was loaded onto a 27% polyacrylamide gel. The running buffer used was 0.192 *M* glycine, 0.025 *M* Tris pH 8.5. Gels were migrated at 200 V for approximately 3 h and then visualized under UV light.

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

	R_g (Å)						χ^2	
Sample (mg ml^{-1})	Global	From Gunier approximation	From Porod	From $P(r)$	D_{max} (Å)	Calculated molecular mass (kDa)	*GASBOR*	*CRYSOL*
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	34.6					109		

2.5. Small-angle X-ray scattering (SAXS)

SAXS experiments were carried out on beamline X33 at the DORIS III storage ring at DESY (Deutsches Elektronen-Synchrotron), Hamburg, Germany. The wavelength 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 to 6 nm^{-1}. All experiments were performed at 16°C at a series of concentrations ranging from 2.0 to 7.9 mg ml^{-1} for the heterologously expressed and purified enzyme Zg3597. The scattering vector is defined as $q = 4\pi\sin\theta/\lambda$, where 2θ is the scattering angle. The detector used was a two-dimensional photon-counting PILATUS 1M-W pixel X-ray detector and ten successive frames of 10 s exposure time 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^{-1} 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 to eventually detect and eliminate concentration-dependent effects. Background scattering was quantified before and after each protein measurement by measuring the corresponding buffer, which was then subtracted from the protein patterns using the *PRIMUS* suite (Konarev *et al.*, 2003).

The radius of gyration R_g of each measurement was derived by the Guinier approximation $I(q) = I(0)\exp(-q^2R_g^2/3)$ up to $qR_g < 1.0$. The radii of gyration R_g calculated for the different protein concentrations displayed no evident concentration dependence. *GNOM* (Svergun, 1992) was used to compute the pair-distance distribution functions $P(r)$. This approach gives the maximum dimension of the macromolecule D_{max} and offers an alternative calculation of R_g 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* (Franke & Svergun, 2009; Svergun *et al.*, 2001). The scattering profiles were used up to q_{max} = 3 nm^{-1}. Nine low-resolution models obtained from different runs were compared using *DAMAVER* (Volkov & Svergun, 2003) 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 of the reconstructions filtered by *DAMFILT*. The atomic structures of the individual modules were then positioned into the low-resolution envelopes using *SUPCOMB* (Kozin & Svergun, 2001). For the Zg3597 dimer obtained from the crystal structure the theoretical SAXS profile, the R_g and the corresponding fit (represented by χ^2 values) to the experimental data were calculated using *CRYSOL* (Svergun *et al.*, 1995).

2.6. Crystallization

Initial hits were obtained by sitting-drop vapour diffusion with a Cartesian dispensing system (Genomic Solutions) using four different commercial screens: The JCSG+, PEGs, PEGs II and PACT Suites. 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 with a 1:1 ratio of protein solution (at 10 mg ml^{-1}) to mother liquor. Soaks were performed on the native protein crystals using 10 m*M* neoagarotetraose made up in mother liquor and mixed in a 1:1 ratio with the crystal drop. The optimized crystallization conditions for Zg3597 were 16% glycerol, 0.15 *M* ammonium sulfate, 20% PEG 4000 in a 1:1 ratio with protein solution at 5 and 2.5 mg ml^{-1} and 14% PEG 3350 and 75 m*M* sodium acetate with a 2:1 ratio of protein solution (at 7.5 mg ml^{-1}) to mother liquor. Crystals were flash-cooled in liquid nitrogen using a cryoprotectant of 20% glycerol or 20% ethylene glycol for Zg3597 and 30% ethylene glycol for Zg3615. Crystals of both proteins grew at 18°C.

2.7. X-ray crystallographic data collection and processing

Diffraction data were collected for Zg3615 on the PROXIMA1 beamline 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 were processed using *MOSFLM* (Powell, 1999) and scaled using *SCALA* (Evans, 2006). The structures were solved by molecular replacement using the one-to-one threading tool from *Phyre*2 (Kelley & Sternberg, 2009) to build a model (100% confidence) using monomer *A* from the GH117 Zg4663 (PDB entry 3p2n; Rebuffet *et al.*, 2011) and removing any nonconserved amino acids from the *MOLREP* (Vagin & Teplyakov, 2010) input. A library for the co-crystallized sugar was created using *PRODRG* from the *CCP*4 suite of programs (Winn *et al.*, 2011). Data were refined using *REFMAC*5 (Murshudov *et al.*, 2011) from the *CCP*4 suite of

programs. Waters were added using the FIND WATERS option in *Coot* (Emsley & Cowtan, 2004). Successive rounds of model building and refinement were performed using *Coot* and *REFMAC*5. All data-collection statistics are summarized in Table 2. Ramachandran statistics were calculated using *RAMPAGE* from the *CCP*4 suite of programs (Lovell *et al.*, 2003). The structure of Zg3597 has been deposited as PDB entry 4u6b and that of Zg3615 as PDB entry 4u6d.

2.8. Phylogeny

New GH117 sequences were identified using *BlastP* at the GenBank database. These new sequences have been added to the previous data set of GH117 sequences (Rebuffet *et al.*, 2011), and all sequences were aligned using *MAFFT* with the iterative refinement method and the BLOSUM62 scoring matrix (Katoh *et al.*, 2002). The *MAFFT* alignment of these sequences was manually refined using *BioEdit* on the basis of the superposition of the five available crystal structures of GH117 [ZgAhgA (Zg4663), PDB entry 3p2n; Zg3597, PDB entry 4u6b; Zg3615, PDB entry 4u6d; SdGH117, PDB entry 3r4y; BpGH117, PDB entry 4ak5]. This refined alignment allowed the calculation of model tests and maximum-likelihood trees with *MEGA* v.6.0.6 (Kumar *et al.*, 2004). Tree reliability was tested by bootstrap using 100 resamplings of the data set. The trees were displayed with *MEGA* v.6.0.6 (Kumar *et al.*, 2004). The accession numbers of the selected sequence homologues are available in Supplementary Table S1.[1]

3. Results

3.1. Activity assays

Two representatives (Zg3615 and Zg3597) from the recently discovered GH117 family of enzymes were chosen for investigation. The activities of Zg3615 and Zg3597 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; Rebuffet *et al.*, 2011), in case the activity of one enzyme is dependent on exposing the substrate to another. Exo-α-1,3-(3,6-anhydro)-L-galactosidase activity of Zg3615 was initially demonstrated using TLC, which showed the release of terminal 3,6-anhydro-L-galactose from neoagarotetraose, neoagarohexaose and neoagaro-octaose (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 were 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,the activity of Zg3615 is indistinguishable from that of Zg4663 by TLC. Therefore, FACE was employed to further investigate any potential differences in activity between the clades. Enzyme reactions using neoagarobiose/neoagarotetraose, neoagarotetraose/neoagarohexaose (Fig. 2) and agarose (not shown) as substrates were migrated on FACE gels. Both

Table 2
X-ray crystallographic data-collection statistics.

	Zg3615	Zg3597
Data-collection statistics		
Beamline	PROXIMA1, 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)
Unit-cell parameters		
a (Å)	57.11	187.79
b (Å)	226.13	223.52
c (Å)	67.12	225.22
$\alpha = \beta = \gamma$ (°)	90.00	90.00
R_{merge} (%)	4.8 (38.2)	8.8 (43.7)
Completeness (%)	100.0 (100.0)	99.7 (99.8)
$\langle I/\sigma(I)\rangle$	25.0 (5.0)	13.2 (3.9)
Multiplicity	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
R.m.s.d.		
Bond lengths (Å)	0.014	0.017
Bond angles (°)	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
No. 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 favoured (%)	95.2	95.8
Additional allowed (%)	4.8	3.9
Disallowed (%)	0	0.3

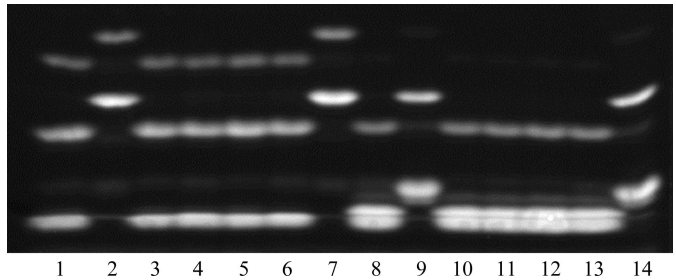


Figure 2
FACE of enzymatic digests of neoagarooligosaccharides. In lanes 1–7 digests were performed on a mixture of neoagarotetraose and neoagarohexaose. In lanes 8–14 digests were performed on a mixture of neoagarobiose and neoagarotetraose. Lane 1, Zg3615; lane 2, Zg3597; lane 3, Zg4663; lane 4, Zg3597, 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.

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: MN5079).

Zg3615 and Zg4663 were active on neoagaro-oligosaccharides; however, there was no discernible difference between the activities of Zg3615 and Zg4663 on these unsubstituted oligoagars. 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 (Fig. 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 (Fig. 2, lanes 8–14). The enzymatic digests of 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 although thorough screening of activity of Zg3597 on potential agarose substrates was undertaken, it was not possible to discern activity on any of the oligosaccharides or polysaccharides tested, suggesting a novel and as yet unidentified activity within the family.

3.2. Overall homodimer structural features

The X-ray crystal structure of Zg3615 was determined to 1.7 Å resolution and the structure of Zg3597 was determined

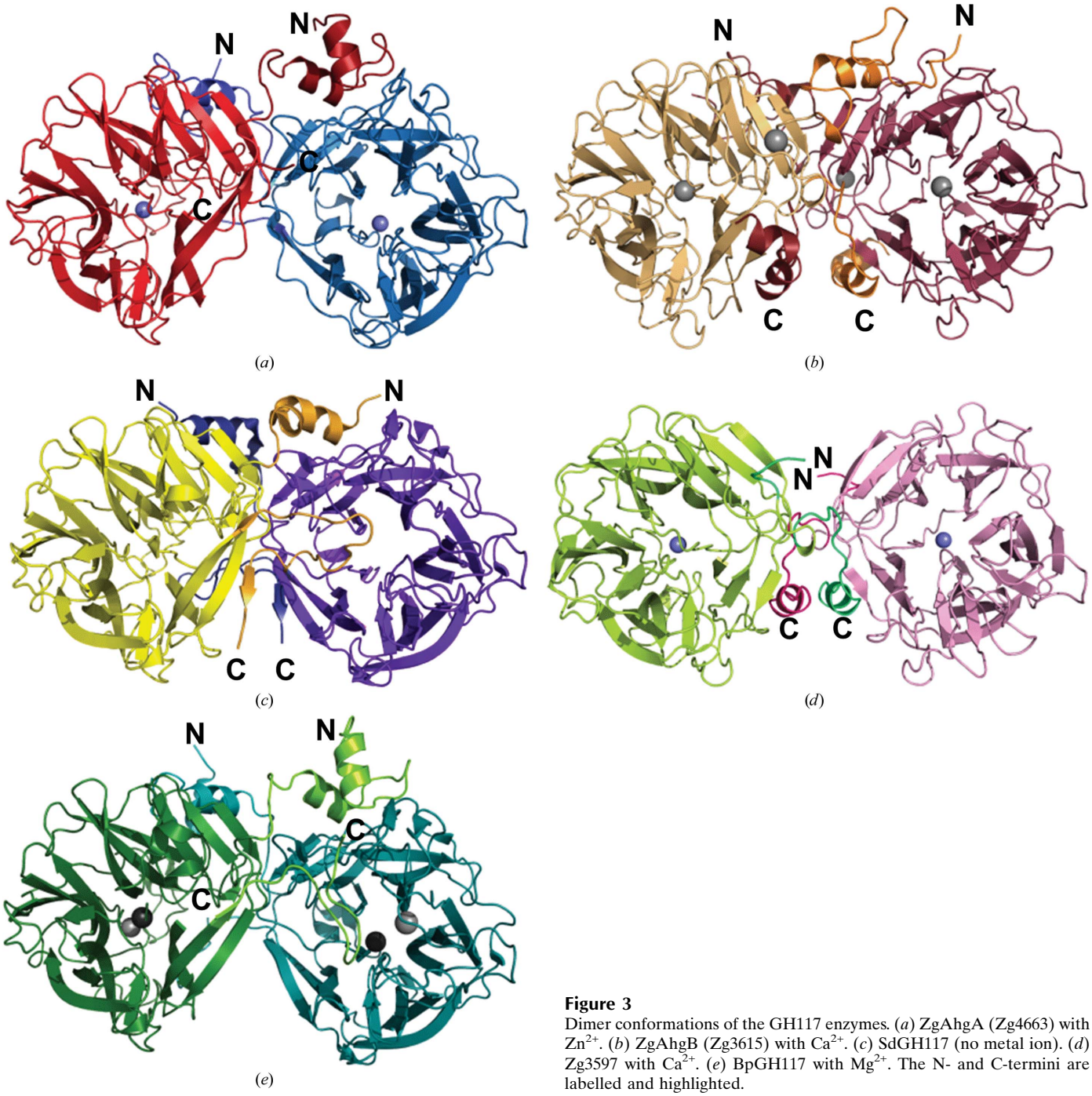
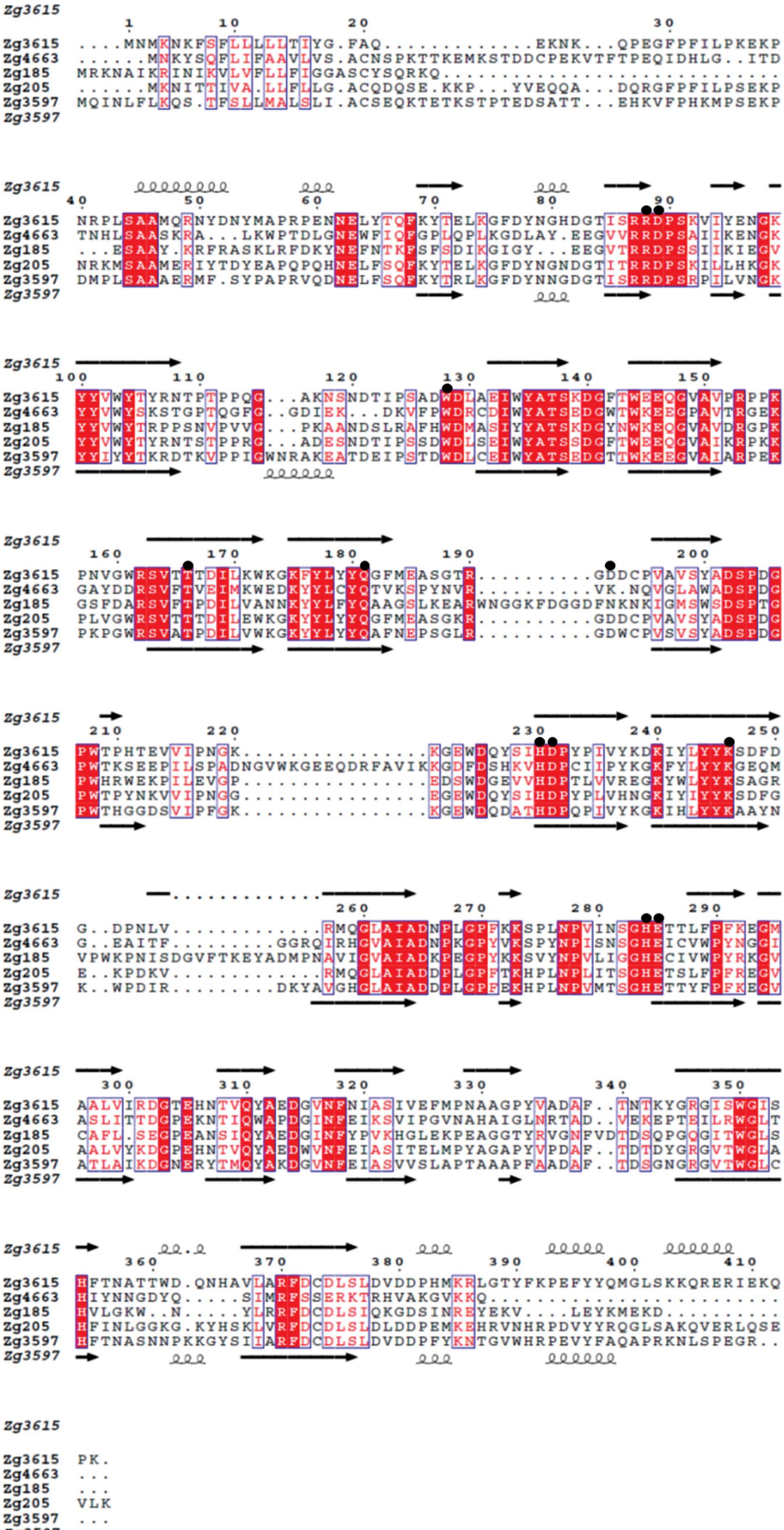


Figure 3
Dimer conformations of the GH117 enzymes. (*a*) ZgAhgA (Zg4663) with Zn^{2+}. (*b*) ZgAhgB (Zg3615) with Ca^{2+}. (*c*) SdGH117 (no metal ion). (*d*) Zg3597 with Ca^{2+}. (*e*) BpGH117 with Mg^{2+}. The N- and C-termini are labelled and highlighted.



to 2.3 Å resolution. The globular Zg3615 and Zg3597 GH117 monomers share the fivefold β-propeller fold with the other members of the family and each β-sheet has three or four antiparallel β-strands. The GH117 enzymes form a structural homodimer, previously shown to be mediated by both the N-terminal and the C-terminal domains (SdGH117 and BpGH117) or just the N-terminal domain (Zg4663) (Fig. 3). The structures of Zg3615 and Zg3597 reveal differences in their N-terminal and C-terminal domains. Zg3615 has a helix–random coil N-terminal motif (amino acids 28–61) which interacts with the other monomer through multiple surface hydrophobic residues. The N-terminus is in a structurally similar position to the N-terminal helix and is followed by a turn followed by a helix in 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 of only one helix motif (amino acids 408–

Figure 4
Sequence alignment between the five GH117 enzymes from *Z. galactanivorans* generated by *ESPript* 3.0 (Robert & Gouet, 2014). 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.

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 domain and this has resulted in a different dimer configuration for Zg3597 when compared with the other GH117 enzymes.

Using the *PDBePISA* (*Proteins, Interfaces, Structures and Assemblies*) web service (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 that it is essential for complex formation. To this, the N-terminal domain contributes approximately ∼1900 Å^2 of surface area and the C-terminal domain 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 that it is relevant to complex formation. Zg3597 is a tetramer in the asymmetric unit, although the surface interaction area (∼740 Å^2) that is not mediated by the N-terminal or C-terminal domains appears to be the result of crystal packing only (CSS = 0).

3.3. Active-site architecture

Both the Zg3615 and Zg3597 enzymes have a shallow cavity typical of exo-acting enzymes. The proposed catalytic residues are conserved (Fig. 4); in Zg3615 these are His284, which is proposed to be the catalytic acid, Asp89, which is proposed to be the general base, and Glu285, which is proposed to modulate the pK_a of the nearby general base (Hehemann, Smyth *et al.*, 2012). Although sequentially conserved (Fig. 4), the proposed catalytic acid (His306) in Zg3597 is shifted compared with Zg3615 (Fig. 5*a*) and with the structures of the other enzymes. This is a rather pronounced positional change, putting it well out of the hydrogen-bonding position to the glycosidic O atom observed in the BpGH117 complex (at a distance of approximately 6.9 Å compared with between 2.7 and 3.6 Å for the other enzymes).

Soaks with neoagarotetraose were performed on the native Zg3615 crystals. Examination of 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 fitted the density well in an unconstrained manner (Fig. 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 Arg88, Asp89 and Glu285 of Zg3615. Arg88 also hydrogen-bonds to the endocyclic O atom of the six-membered ring (Fig. 5*a*). The axial C2 hydroxyl interacts *via* hydrogen bonds with 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 cyclic O atom of the five-membered ring. Finally, a remarkable difference is observed when comparing the two monomers in the Zg3615 homodimer. The loop containing Asp192 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 owing to interactions with the carbohydrate moiety. Hydrophobic contributions to the product complex are provided by the imidazole face of His230, which interacts with the equatorial C4 arm of β-3,6-anhydro-L-galactose. Trp128 provides a hydrophobic platform for the C5 and C6 atoms of the sugar, which form part of the five-membered sugar ring, and the arm of Arg88 also provides hydrophobic contributions to this ring component.

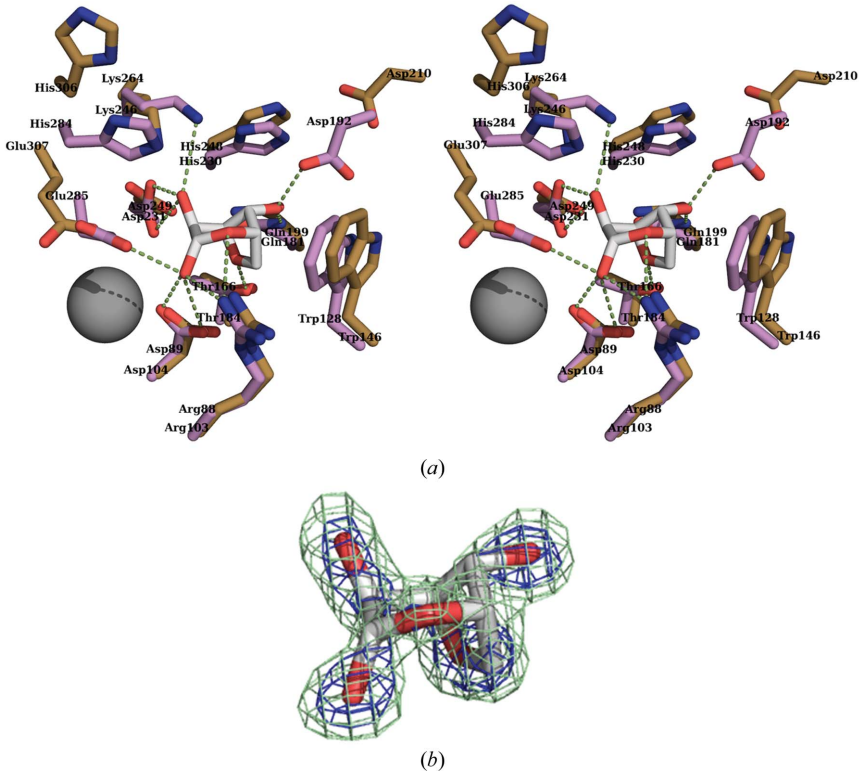


Figure 5
(*a*) Stereo figure of the −1 active site of Zg3597 (gold) overlaid with that of ZgAhgB (Zg3615; mauve), which is in complex with β-3,6-anhydro-L-galactose. Hydrogen bonds between ZgAhgB and the sugar are indicated. The adjacent ordered Ca^{2+} ion is also shown. (*b*) An OMIT map was generated by omitting β-3,6-anhydro-L-galactose from the refinement. The resulting maps are maximum-likelihood σ_A-weighted (Read, 1986) $2F_{obs} - F_{calc}$ maps contoured at 2σ (blue) and $F_{obs} - F_{calc}$ maps contoured at 3σ (green) (0.72 and 0.19 e Å^{-3}, respectively).

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 to two of the clade A enzymes, SdGH117 and BpGH117, the C-termini of which do participate in the active-site substructure and interact with the galactose component in the +1 subsite (His392 in BpGH117 and Ser358 and Tyr359 in SdGH117). The last residues of the C-terminal domain of the clade A Zg4663 are disordered in the native, noncomplexed crystal structure (His400–Gln408), although 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 the active-site substructure and the N-terminal dimerization. Firstly, Asp250 from Zg3615 forms a salt bridge with Arg58 from the N-terminus 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 side chain of Tyr227. Zg3597 replaces these two residues with a histidine residue (His48, which is not visible in the crystal structure) and Asp245, suggesting that this interaction is abrogated in Zg3597.

A loop comprising Ile322–Tyr329, which sits just beside the turn containing His306, is flipped in Zg3597 (Figs. 6*a* and 6*b*). One reason for this flip may be the hydrophobic Tyr329, which would otherwise be located on the surface. In the +1 subsite, Arg257 from Zg3615 is spatially, although 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 (Fig. 6*c*), which is the putative catalytic acid. Zg3597 replaces Arg257 with Val279, a much smaller nonpolar residue, leaving a pocket at the C4 site, possibly for accommodating a substituent on the galactose C4 hydroxyl group. His306, the putative catalytic acid in Zg3597, 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 His306 of Zg3597 lays sandwiched hydrophobically between Val279 and Asp324 (Fig. 6*b*). In the other structures, this aspartic acid residue is within hydrogen-bonding distance of His306. The change in the position of the His306 residue of Zg3597 represents a rather large deviation for a proposed catalytic acid. Although 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, although 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 BpGH117 Tyr186 residue is not spatially conserved in Zg3597 or Zg3615; however, the functionality of the Tyr

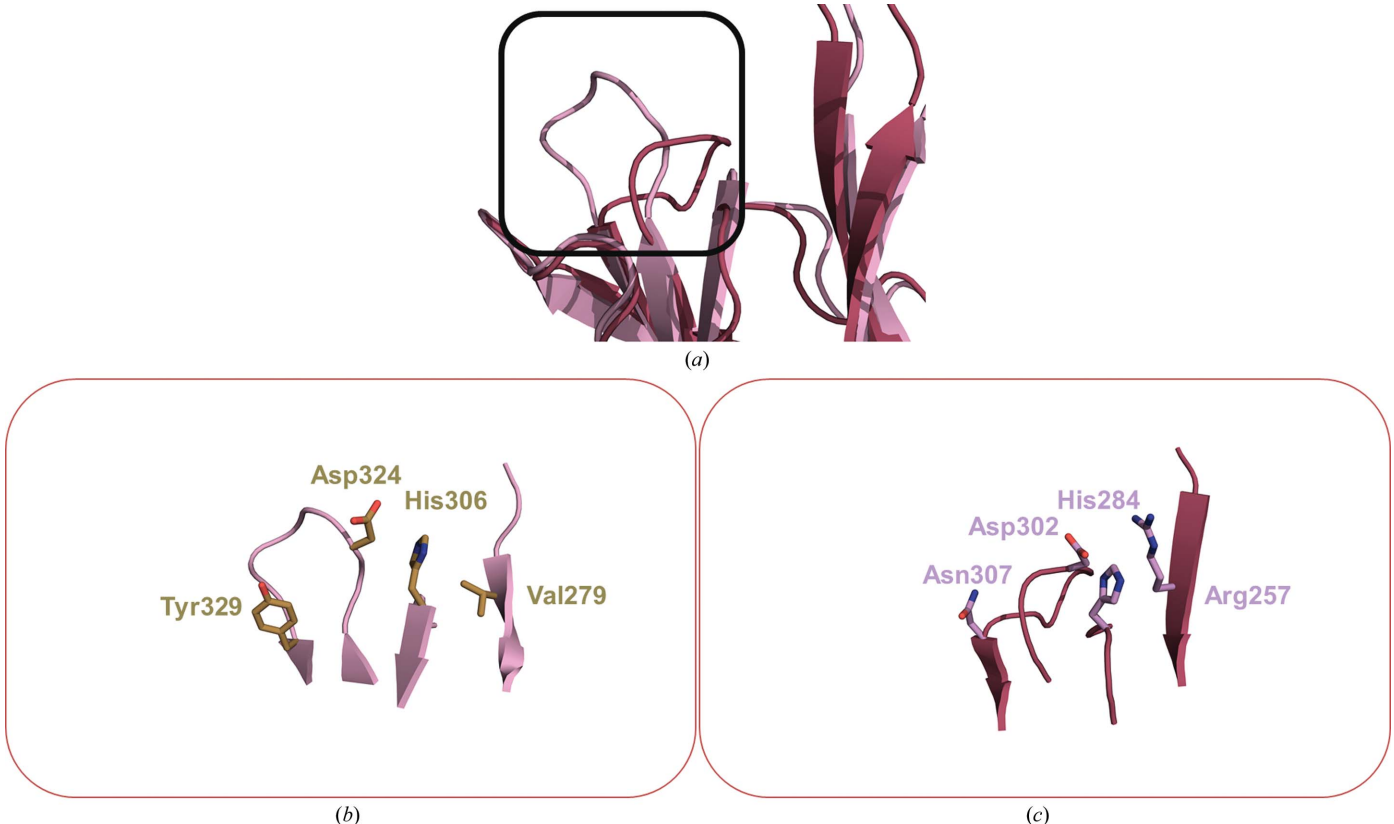


Figure 6
(*a*) Secondary-structure overlay of Zg3597 and ZgAhgB (Zg3615) showing the Zg3597 loop flip. Zg3597 is coloured pink and ZgAhgB is coloured raspberry. (*b*) Amino acids which contribute to the stability of the loop flip in Zg3597. (*c*) The same loop in the opposing orientation in ZgAhgB.

hydroxyl group is roughly conserved in 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 in Zg3615 and the polar Thr144 in the case of Zg3597. The C4 hydroxyl-interacting residue Asp192 in Zg3615 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 the Zg3615 and the Zg3597 enzymes have more open and accessible active-site entrances when compared with BpGH117 and SdGH117 (although not Zg4663), in particular Zg3597.

3.4. 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 was modelled. A Ca^{2+} molecule has been modelled for Zg3615 and Zg3597 in this same site, which is supported by the presence of seven organized water molecules in the solvation shell (Pérez *et al.*, 2012) and the similarity of the *B* factors between the metal atoms and their neighbouring water molecules. Furthermore, the distances between the water atoms and the Ca^{2+} molecule in Zg3615 (monomer *A*) range from 2.32 to 2.60 Å, while in Zg3597 (monomer *A*) they range from 2.18 to 2.58 Å; these values are consistent with the expected distribution range (Zheng *et al.*, 2008). 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 5σ (0.317 e^- Å^{-3}), consistent with the molecule being modelled incorrectly as a water. Furthermore, the distorted octahedral geometry of six ligands surrounding a central atom is incompatible with the central atom being a highly coordinated water molecule (Pérez *et al.*, 2012). Regardless, this metal-binding site appears to have some degree of plasticity and is conserved throughout the GH117 structures solved to date.

3.5. 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 (Fig. 7*d*). The slope and intensity at zero angle of the experimental scattering data allowed the R_g values to be calculated and the molecular weight in solution to be estimated (Table 1), which were coherent with the protein adopting a nonspherical, 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) were obtained when imposing twofold symmetry for a dimeric form of the enzyme. These data also allowed the determination of an R_g 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 (D_{max}) of 100 ± 2 Å (Fig. 7*d*). The molecular mass determined from the scattering intensity extrapolated to zero scattering angle indicated a molecular mass of approximately 109 kDa (Table 1), in close agreement with Zg3597 existing as a homodimer in solution, as is also observed by size-exclusion chromatography, DLS and in the crystal structure (data not shown). The molecular mass determined by SAXS also

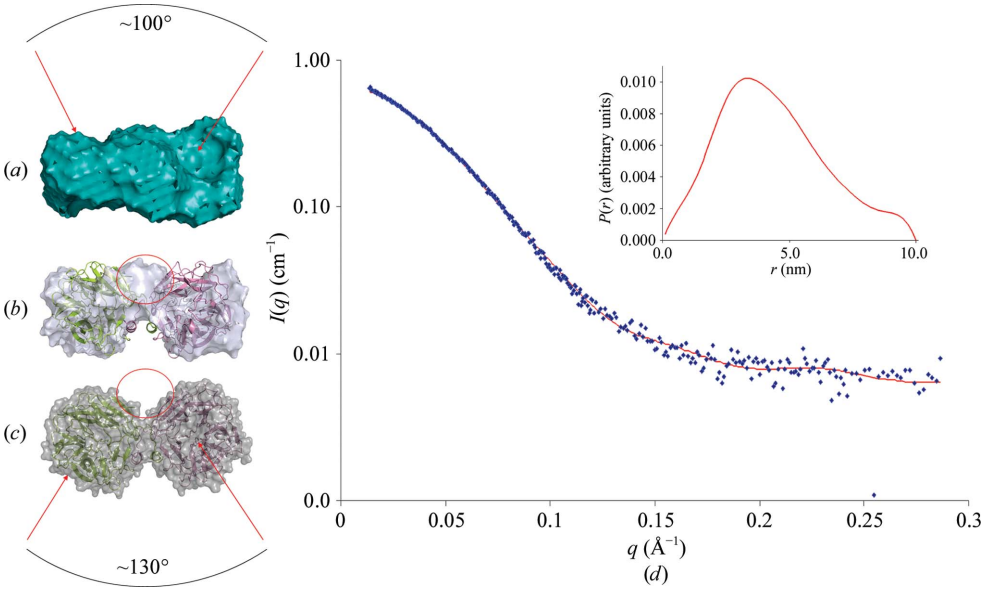


Figure 7
Envelopes and resulting curves from small-angle X-ray scattering. (*a*) Envelope obtained by *GASBOR* (best fit χ^2 = 1.6) using the scattering curve at 4.78 mg ml^{-1}. The marked angle of 100° (calculated as a dihedral angle) above the shape is a rough estimate of the relative orientation of the two catalytic pockets. (*b*) Superimposition of the Zg3597 homodimer from the crystal structure onto the same SAXS envelope as in (*a*) (the orientation is rotated by roughly 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 two catalytic pockets. In (*b*) and (*c*) the twofold symmetry axis of the shapes lies in the plane (vertical) and in (*a*) it is out of the plane. (*d*) Experimental scattering curve of Zg3597 in solution at 4.78 mg ml^{-1} (blue crosses) and the 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^{-1}.

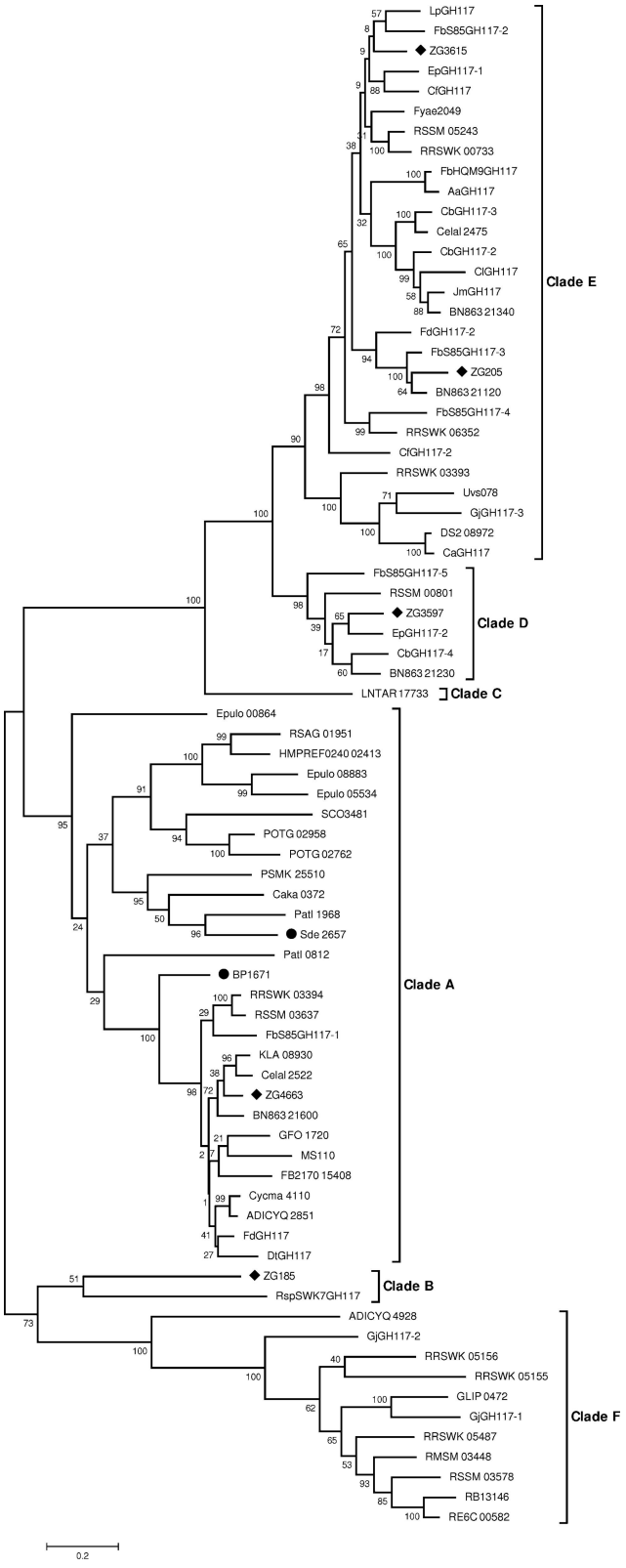


Figure 8
Unrooted phylogenetic tree of the GH117 homologues. The phylogenetic tree was generated using the maximum-likelihood approach with the program *MEGA* v.6.0.6. Bootstrap numbers are indicated. Accession numbers of the selected sequence homologues are available in Supplementary Table S1. The structures solved to date are marked with diamonds for the GH117 enzymes produced by *Z. galactanivorans* and by circles for the others.

compares well with the theoretical molecular mass 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 an R_g value (31.9 Å) of roughly the same order of magnitude as the experimental scattering curve, the χ^2 value of 3.4 describing the fit between the experimental and the 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 that is likely to correspond to the catalytic amino-terminal domains was present in the SAXS envelope (Figs. 7*a*, 7*b* and 7*c*). This may reflect conformational flexibility of the amino-terminal region with respect to the catalytic domains, with 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 (Fig. 7*c*). In addition, the SAXS envelope calculated by *GASBOR* (Fig. 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° (Fig. 7*a*), whereas this angle estimated from the crystal structure is about 130° (Fig. 7*c*). This would imply a rotational rearrangement of the two monomers, most probably owing to constraints in the crystal structure that might disfavour the establishment of interactions of the amino-terminal domains with the adjacent catalytic domains in the swapped geometry.

3.6. Phylogenetic analysis

In the initial phylogenetic analysis of the GH117 family, three clades were defined: clade A contained most of the sequences, while clades B and C comprised only few sequences and essentially paralogous proteins from *Z. galactanivorans* (clade B, Zg185; clade C, Zg205, Zg3597 and Zg3165; Rebuffet *et al.*, 2011). 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 (Fig. 8 and Supplementary Table S1) in order to revise the definition of the GH117 subfamilies. In the initial tree (Rebuffet *et al.*, 2011), clade B contained only 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 weak statistical support (51%), while RB13146 clusters together with ten 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): the first containing Zg3597 and five new sequences and the second encompassing 28 sequences including Zg205 and Zg3615. The difference in enzymatic activity between Zg3597 (which remains 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 different activity from those of Zg3597 and Zg3615. Altogether, we propose that the clade composed of LNTAR_17733 retains the name clade C, while the clades containing Zg3597 and Zg3615 are referred to as clade D and clade E, respectively.

4. Discussion

4.1. *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 (Jam *et al.*, 2005; Hehemann, Correc *et al.*, 2012), five β-porphyranases (Hehemann *et al.*, 2010) and one κ-carrageenase (Barbeyron *et al.*, 1998). 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 (Rebuffet *et al.*, 2011), 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 has thus been named ZgAhgA (Rebuffet *et al.*, 2011). 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 (Fig. 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 (Fig. 8). While the former clades B and C initially contained very few sequences (Rebuffet *et al.*, 2011), these subfamilies are largely expanded owing 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.

4.2. Implications for 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 mole-

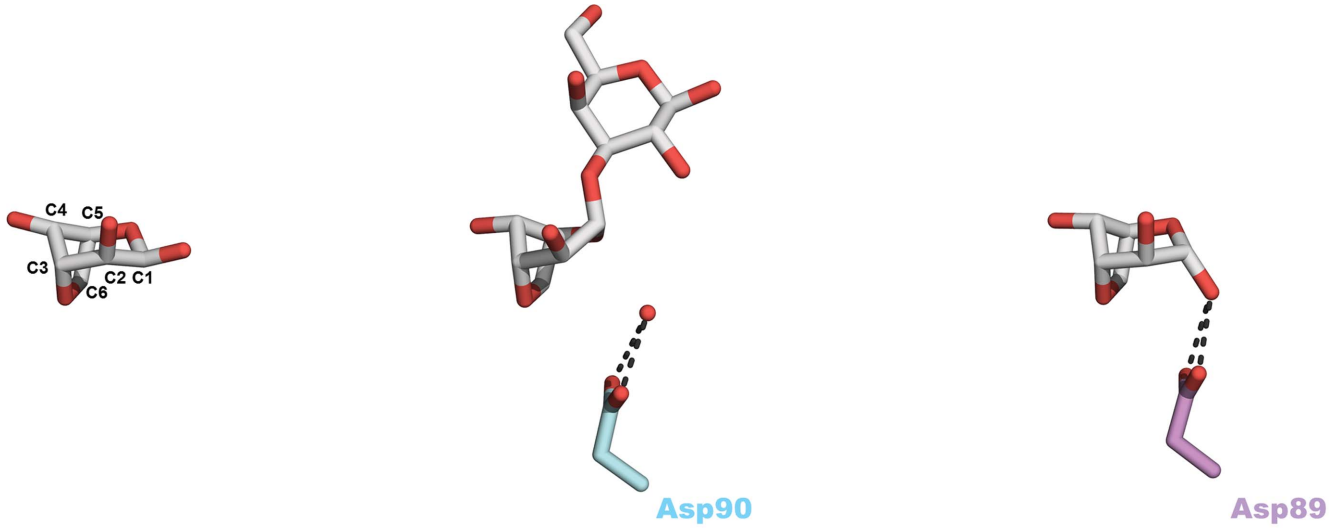


Figure 9
3,6-Anhydro-L-galactose structures. (*a*) $^4C_1$ α-3,6-anhydro-L-galactose as found in agarose. (*b*) Neoagarobiose from BpGH117 (PDB entry 4ak7) with $B_{1,4}$ α-3,6-anhydro-L-galactose. Aspartate, the putative catalytic general base, is shown coordinating an ordered water just below the anomeric C atom. (*c*) $^4C_1$ β-3,6-Anhydro-L-galactose from ZgAhgB (Zg3615; PDB entry 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.

cules located at the bottom of the active-site pocket (Pérez *et al.*, 2012). The divalent cation was proposed to play a role in the activation of the catalytic water molecule to provoke hydrolysis. This ion-binding site is conserved in the structures of all of the enzymes solved to date, although 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 (Fig. 9*b*) which aligns the anomeric centre for nucleophilic attack and demonstrates how the enzymes alter the sugar conformation to favour the oxocarbenium-ion transition state (Hehemann, Smyth *et al.*, 2012). Based on crystallographic and mutational evidence for BpGH117, Hehemann and coworkers proposed that Asp90 (Fig. 9*b*) acts as the general base with the adjacent Glu303 as a pK_a modulator (Hehemann, Smyth *et al.*, 2012). The positioning of the BpGH117 active-site residues suggests an inverting mechanism (Zechel & Withers, 2000) 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, although unusual glycoside hydrolase mechanisms are not unheard of (Jongkees & Withers, 2014). The three proposed catalytic residues, histidine, aspartic acid and glutamic acid, are conserved throughout the GH117 family, with the exception of two clade F enzymes (GjGH117-1 and RRSWK_05156; Fig. 8), 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 (Pluvinage *et al.*, 2013; Allouch *et al.*, 2004; Henshaw *et al.*, 2006), implying that this is not a high-energy intermediate (Figs. 9*a* and 9*c*). The unusual positioning of the C1 hydroxyl in a 'flattened' conformation is likely to be owing in part to the 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 an α- or β-anomeric form in solution (Rochas *et al.*, 1994; Ducatti *et al.*, 2011). 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 hydrogen-bonding distance of 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 of 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 (Hehemann, Smyth *et al.*, 2012). Therefore, the trapped β-3,6-anhydro-L-galactose provides substantial crystallographic evidence supporting a mechanism of stereochemical inversion at the anomeric centre, as has been previously suggested (Hehemann, Smyth *et al.*, 2012).

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 β-anomeric $^4C_1$ complex of ZgAhgB. 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 and Arg88) which provide hydrogen bonds to the β-anomeric hydroxyl group (Fig. 5*a*). The three residues which interact with the β-anomeric hydroxyl are conserved in all of 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 C atom (Fig. 9*b*). Asp90 of BpGH117 is positioned for in-line activation of the water molecule for direct attack on the $B_{1,4}$ anomeric C atom. The BpGH117 water molecule is coordinated by the same conserved residues as the C1 hydroxyl of the ZgAhgB $^4C_1$ β-anomer. The interactions at this key site are likely to 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 (Rye & Withers, 2000).

4.3. ZgAhgB and Zg3597 have novel modes of dimerization

Both ZgAhgB and Zg3597 have novel modes of dimerization compared with the clade A enzymes. Zg3597 is unusual among the GH117 enzymes as it lacks the 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 loop in ZgAhgB 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 domain, shifting its relative orientation compared with ZgAhgB and the clade A GH117 enzymes. ZgAhgB also dimerizes in a shifted state relative to the clade A enzymes, indicating that there is some plasticity in the dimerization conformations; however, the C-terminal domains of ZgAhgB and Zg3597 overlap in a similar structural position (Figs. 3*b* and 3*d*).

4.4. Zg3597 has a significantly modified active site

There are several possible scenarios that might arise from the divergent Zg3597 active-site substructure. Firstly, the substrate may be significantly different for this enzyme and thus His306 is shifted to accommodate the substrate. Secondly, a substrate-stabilized rearrangement may occur upon binding substrate, possibly implicating the N-terminus in dimerization as, for example, observed in the solution structure. Thirdly, it is possible that this protein is a noncatalytic protein as observed for the 'inactivated' GH18 chitin-binding proteins

(Hennig *et al.*, 1995; Bleau *et al.*, 1999). Finally, the crystal structure of Zg3597 may be in a non-native form induced by crystallization, although the SAXS analysis does reveal that the protein exists as a dimer in solution (Fig. 7*a*). In our actual working hypothesis we assume that the dimeric form of Zg3597 is the biologically active form, and the observed flexibility of the amino-terminal domain of Zg3597 is associated with its catalytic function, but as the natural substrate is not yet known we are not able to verify this assumption.

4.5. GH117 novel subsite structure

ZgAhgB, and Zg3597 in particular, have a more open entrance to their active sites when compared with their clade A counterparts. This may confer an advantage when presented with the diverse decorated sugars found in the agar polygalactan chain. Although it has been postulated that ZgAhgB has an extra −2 subsite (Hehemann, Smyth *et al.*, 2012), 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, Asp210 of Zg3597 is shifted considerably (Fig. 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 of the new clade F enzymes have a glycine in place of the tryptophan within their active sites and that 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 that 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 classical 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 (Hehemann, Smyth *et al.*, 2012), have trapped the 3,6-anhydro-L-galactose in a relaxed α-$^4C_1$ conformation (Pluvinage *et al.*, 2013; Allouch *et al.*, 2004; Henshaw *et al.*, 2006). This is the first time that the β-anomer $^4C_1$ conformation has been demonstrated experimentally, and it provides important insight into the mechanism of the GH117 enzymes as well as into the structure of these unique bicyclic sugars.

EFB was funded by a postdoctoral fellowship supported by the Région Bretagne through the program 'Algevol' with reference SAD_Obex_EMBRC 12010152. ER benefited from a PhD fellowship from the Région Bretagne through the program 211-B2-9/ARED 091539 'Iotase3D' and MC is grateful 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 regard to the IDEALG investment expenditure program with reference ANR-10-BTBR-04. We are indebted to the European Synchrotron Research Facility, Grenoble, France for regular access to X-ray beamlines and to all local contacts for their support during data collection on the MX beamlines ID23-1 and ID29. We also thank the German synchrotron at DESY, Hamburg for access to SAXS beamline X33 and especially Clement Blanchet (EMBL-Hamburg) for valuable help during SAXS data collection.

References

Allouch, J., Helbert, W., Henrissat, B. & Czjzek, M. (2004). *Structure*, **12**, 623–632.
Allouch, J., Jam, M., Helbert, W., Barbeyron, T., Kloareg, B., Henrissat, B. & Czjzek, M. (2003). *J. Biol. Chem.* **278**, 47171–47180.
Barbeyron, T., Gerard, A., Potin, P., Henrissat, B. & Kloareg, B. (1998). *Mol. Biol. Evol.* **15**, 528–537.
Barbeyron, T., Michel, G., Potin, P., Henrissat, B. & Kloareg, B. (2000). *J. Biol. Chem.* **275**, 35499–35505.
Bleau, G., Massicotte, F., Merlen, Y. & Boisvert, C. (1999). *EXS*, **87**, 211–221.
Butterfield, N. (2000). *Paleobiology*, **26**, 386–404.
Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2008). *Nucleic Acids Res.* **37**, D233–D238.
Collén, J. *et al.* (2013). *Proc. Natl Acad. Sci. USA*, **110**, 5247–5252.
Collén, J., Cornish, M. L., Craigie, J., Ficko-Blean, E., Hervé, C., Krueger-Hadfield, S. A., Leblanc, C., Michel, G., Potin, P., Tonon, T. & Boyen, C. (2014). *Adv. Bot. Res.* **71**, 53–89.
Ducatti, D. R. B., Colodi, F. G., Gonçalves, A. G., Duarte, M. E. R. & Noseda, M. D. (2011). *Rev. Bras. Farmacogn.* **21**, 296–304.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Flament, D., Barbeyron, T., Jam, M., Potin, P., Czjzek, M., Kloareg, B. & Michel, G. (2007). *Appl. Environ. Microbiol.* **73**, 4691–4694.
Franke, D. & Svergun, D. I. (2009). *J. Appl. Cryst.* **42**, 342–346.
Genicot-Joncour, S., Poinas, A., Richard, O., Potin, P., Rudolph, B., Kloareg, B. & Helbert, W. (2009). *Plant Physiol.* **151**, 1609–1616.
Groisillier, A., Hervé, 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. & Czjzek, M. (2010). *Microb. Cell Fact.* **9**, 45.
Ha, S. C., Lee, S., Lee, J., Kim, H. T., Ko, H.-J., Kim, K. H. & Choi, I.-G. (2011). *Biochem. Biophys. Res. Commun.* **412**, 238–244.
Hehemann, J. H., Correc, G., Barbeyron, T., Helbert, W., Czjzek, M. & Michel, G. (2010). *Nature (London)*, **464**, 908–912.
Hehemann, J. H., Correc, G., Thomas, F., Bernard, T., Barbeyron, T., Jam, M., Helbert, W., Michel, G. & Czjzek, M. (2012). *J. Biol. Chem.* **287**, 30571–30584.
Hehemann, J. H., Kelly, A. G., Pudlo, N. A., Martens, E. C. & Boraston, A. B. (2012). *Proc. Natl Acad. Sci. USA*, **109**, 19786–19791.
Hehemann, J. H., Smyth, L., Yadav, A., Vocadlo, D. J. & Boraston, A. B. (2012). *J. Biol. Chem.* **287**, 13985–13995.
Hennig, M., Jansonius, J. N., Terwisscha van Scheltinga, A. C., Dijkstra, B. W. & Schlesier, B. (1995). *J. Mol. Biol.* **254**, 237–246.
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. & Gilbert, H. J. (2006). *J. Biol. Chem.* **281**, 17099–17107.
Jam, M., Flament, D., Allouch, J., Potin, P., Thion, L., Kloareg, B., Czjzek, M., Helbert, W., Michel, G. & Barbeyron, T. (2005). *Biochem. J.* **385**, 703–713.
Jongkees, S. A. K. & Withers, S. G. (2014). *Acc. Chem. Res.* **47**, 226–235.
Katoh, K., Misawa, K., Kuma, K. & Miyata, T. (2002). *Nucleic Acids Res.* **30**, 3059–3066.
Kelley, L. A. & Sternberg, M. J. E. (2009). *Nature Protoc.* **4**, 363–371.

Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 1277–1282.
Kozin, M. B. & Svergun, D. I. (2001). *J. Appl. Cryst.* **34**, 33–41.
Kropf, D. L., Kloareg, B. & Quatrano, R. S. (1988). *Science*, **239**, 187–190.
Kumar, S., Tamura, K. & Nei, M. (2004). *Brief. Bioinform.* **5**, 150–163.
Lawson, C. J. & Rees, D. A. (1970). *Nature (London)*, **227**, 392–393.
Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins*, **50**, 437–450.
Martin, M., Portetelle, D., Michel, G. & Vandenbol, M. (2014). *Appl. Microbiol. Biotechnol.* **98**, 2917–2935.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Pérez, C., Muckle, M. T., Zaleski, D. P., Seifert, N. A., Temelso, B., Shields, G. C., Kisiel, Z. & Pate, B. H. (2012). *Science*, **336**, 897–901.
Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. (2011). *Nature Methods*, **8**, 785–786.
Pluvinage, B., Hehemann, J. H. & Boraston, A. B. (2013). *J. Biol. Chem.* **288**, 28078–28088.
Popper, Z. A., Michel, G., Hervé, C., Domozych, D. S., Willats, W. G., Tuohy, M. G., Kloareg, B. & Stengel, D. B. (2011). *Annu. Rev. Plant Biol.* **62**, 567–590.
Powell, H. R. (1999). *Acta Cryst.* D**55**, 1690–1695.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Rebuffet, E., Barbeyron, T., Jeudy, A., Jam, M., Czjzek, M. & Michel, G. (2010). *Biochemistry*, **49**, 7590–7599.
Rebuffet, E., Groisillier, A., Thompson, A., Jeudy, A., Barbeyron, T., Czjzek, M. & Michel, G. (2011). *Environ. Microbiol.* **13**, 1253–1270.
Rees, D. A. (1969). *Adv. Carbohydr. Chem. Biochem.* **24**, 267–332.
Robert, X. & Gouet, P. (2014). *Nucleic Acids Res.* **42**, W320–W324.
Rochas, C., Potin, P. & Kloareg, B. (1994). *Carbohydr. Res.* **253**, 69–77.
Rye, C. S. & Withers, S. G. (2000). *Curr. Opin. Chem. Biol.* **4**, 573–580.
Svergun, D. I. (1992). *J. Appl. Cryst.* **25**, 495–503.
Svergun, D., Barberato, C. & Koch, M. H. J. (1995). *J. Appl. Cryst.* **28**, 768–773.
Svergun, D. I., Petoukhov, M. V. & Koch, M. H. J. (2001). *Biophys. J.* **80**, 2946–2953.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Volkov, V. V. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 860–864.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Zechel, D. L. & Withers, S. G. (2000). *Acc. Chem. Res.* **33**, 11–18.
Zheng, H., Chruszcz, M., Lasota, P., Lebioda, L. & Minor, W. (2008). *J. Inorg. Biochem.* **102**, 1765–1776.