University of Dundee

Structural and functional analyses of glycoside hydrolase 138 enzymes targeting chain A galacturonic acid in the complex pectin rhamnogalacturonan II

Labourel, Aurore; Baslé, Arnaud; Munoz-Munoz, Jose; Ndeh, Didier; Booth, Simon; Nepogodiev, Sergey A.

Structural and functional analyses of glycoside hydrolase 138 enzymes targeting chain A galacturonic acid in the complex pectin rhamnogalacturonan II

Aurore Labourel[‡1], Arnaud Baslé[‡], Jose Munoz-Munoz[‡2], Didier Ndeh[‡], Simon Booth[‡], Sergey A. Nepogodiev[§], Robert A. Field[§], and Alan Cartmell[‡3]

From the ‡Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne NE2 4HH, United Kingdom and §Department of Biological Chemistry, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, United Kingdom

Edited by Joseph M. Jez

The metabolism of carbohydrate polymers drives microbial diversity in the human gut microbiome. The selection pressures in this environment have spurred the evolution of a complex reservoir of microbial genes encoding carbohydrate-active enzymes (CAZymes). Previously, we have shown that the human gut bacterium *Bacteroides thetaiotaomicron* (*Bt*) can depolymerize the most structurally complex glycan, the plant pectin rhamnogalacturonan II (RGII), commonly found in the human diet. Previous investigation of the RGII-degrading apparatus in *Bt* identified BT0997 as a new CAZyme family, classified as glycoside hydrolase 138 (GH138). The mechanism of substrate recognition by GH138, however, remains unclear. Here, using synthetic substrates and biochemical assays, we show that BT0997 targets the D-galacturonic acid-α-1,2-L-rhamnose linkage in chain A of RGII and that it absolutely requires the presence of a second D-galacturonic acid side chain (linked β-1,3 to L-rhamnose) for activity. NMR analysis revealed that BT0997 operates through a double displacement retaining mechanism. We also report the crystal structure of a BT0997 homolog, BPA0997 from *Bacteroides paurosaccharolyticus*, in complex with ligands at 1.6 Å resolution. The structure disclosed that the enzyme comprises four domains, including a catalytic TIM (α/β)₈ barrel. Characterization of several BT0997 variants identified Glu-294 and Glu-361 as the catalytic acid/base and nucleophile, respectively, and we observed a chloride ion close to the active site. The three-dimensional structure and bioinformatic analysis revealed that two arginines, Arg-332 and Arg-521, are key specificity determinants of BT0997 in targeting D-galacturonic acid residues. In summary, our study reports the first structural and mechanistic analyses of GH138 enzymes.

The human gut microbiota (HGM)[4] is a vast microbial community, of $>10^9$ bacterial cells, inhabiting the distal human colon (1). The HGM is essential for maintaining human health by preventing colonization of pathogens and providing secondary metabolites, which provide up to 10% of the host's calorific intake (2). When the HGM falls into a state of dysbiosis, potentially as a result of the host's diet being deficient in fiber, it has been linked to a number of disease states, such as colitis, colorectal cancer, and even Alzheimer's disease (3–5). Thus, a healthy HGM can be maintained by providing the community with a diet that is high in fiber (6). This fiber is in the form of plant glycans ingested by the host, which are then degraded by members of the HGM and utilized as a primary source of carbon. Competition within the HGM is intense, and bacteria of this community, such as those of the *Bacteroidetes* phyla, have evolved elaborate and complex mechanisms in the form of polysaccharide utilization loci (PUL) to acquire and degrade glycans. PULs are a set of colocalized and coregulated genes that are dedicated to degrade a target glycan (7). Understanding how the HGM degrades dietary fiber is therefore critical to devising prebiotic and probiotic strategies to improve human health. This reservoir of carbohydrate active enzymes is also an invaluable source for the discovery of enzymatic activities.

The pectic polysaccharide rhamnogalacturonan II (RGII) is highly conserved in terrestrial plants being found in the primary cell walls of fruits and vegetables (8). It is also highly enriched in luxury foods such as red wine and dark chocolate and as such is a common component of the human diet (9). RGII is the most complex glycan known in nature comprising 3 unique sugars and 23 unique glycosidic linkages. It was unknown if a single organism could degrade the polymer or if a consortium of bacteria would be required. Recently, our group showed that *Bacteroides thetaiotaomicron* (*Bt*), a member of the *Bacteroidetes* phyla and a model organism for studying glycan metabolism by the HGM, is able to cleave all but 1 of the 23 distinct glycosidic linkages present in RGII by deploying three

[4] The abbreviations used are: HGM, human gut microbiota; BPA, *Bacteroides paurosaccharolyticus*; *Bt*, *Bacteroides thetaiotaomicron*; CAZymes, carbohydrate active enzymes; CD4, C-terminal domain of BPA0997; D2, TIM barrel domain of BPA0997; D3, third domain of BPA0997; GalA, galacturonic acid pyranose form; GH, glycoside hydrolase; ND1, N-terminal domain of BPA0997; PUL, polysaccharide utilization loci; RGII, rhamnogalacturonan II; Rha, rhamnose; RMSD, root mean square deviation; SeMet, selenomethionine.

discrete PULs (10). The primary catalytic apparatus encoded by PULs is in the form glycoside hydrolases (GH). GHs are catalogued on the CAZy database by sequence homology into unique families (http://www.cazy.org)[5] (11, 35). Members within a family have the same tertiary fold, catalytic apparatus, and mechanism. As the complete degradative pathway utilized by *Bt* was dissected, seven new GH families were discovered and added to the CAZy database. One of these families, founded by the protein BT0997, was annotated as GH138 and was shown to be an α-D-galacturonidase targeting the α-1,2-D-galacturonic acid linked to rhamnose within RGII (10). In this paper, we describe the kinetic properties of BT0997 and its homologue WP_024993800 from *Bacteroides paurosaccharolyticus* (BPA0997). We also dissected the mechanism by which GH138 operates. Finally, the structural solution of BPA0997, in complex with ligands, allows the assignment of both the catalytic residues in the family and the amino acids that drive specificity.

Results

Biochemical properties of BT0997

In a previous study, we showed that BT0997 is a new family of α-1,2-D-galacturonidase specifically targeting the D-galacturonic acid (GalA)-α-1,2-L-rhamnose (Rha) contained within chain A of RGII (Fig. 1*A*). It was also demonstrated that BT0997 was inactive on intact RGII and was only active after chain A (10) was released from the backbone by an endo acting polysaccharide lyase 1 (PL1) and concomitant removal of both α-L-galactose by BT1010 and β-D-glucuronic acid by BT0996 from the top of chain A. After this initial degradation, BT0997 is then able to act and its action is required for further degradation of chain A. The mechanism of substrate recognition displayed by the enzyme, however, remained unclear. In the present study, we used synthetic substrates (Fig. 1, *B* and *C*) and showed that BT0997 recognizes the second β-linked GalA as an absolute specificity determinant. The enzyme was inactive on SN909 which only contains one of the two GalA residues (D-GalA-α-1,2-L-Rha-α-1,4-L-Fucose) whereas it displayed full activity against SN908, the doubly GalA decorated rhamnose substrate (D-GalA-α-1,2-L-Rha-β-1,3-D-GalA). A comparable activity was observed on G2RAX, a hexasaccharide produced by the growth of the deletion mutant $\Delta bt0997$ on RGII. The absence of BT0997 means that the α-1,2-D-GalA can no longer be cleaved, preventing the full digestion of chain A, leading to the accumulation of the hexasaccharide G2RAX in the growth media. These data suggest that BT0997 has three subsites, a -1 subsite where the GalA linked α-1,2 to Rha binds; a $+1$ subsite where Rha is situated; and finally a $+2NR$ subsite where the second GalA β-1,3 linked to Rha binds (Table 1) (subsite nomenclature of GHs is such that bond cleavage occurs between the -1 and $+1$ subsites with the catalytic apparatus being housed at the -1 subsite. Subsites then increase in number, $+2$, $+3$ etc. toward the reducing end of the glycan and decrease in number, -2, -3 etc. toward the nonreducing end of the glycan). As BT0997 was
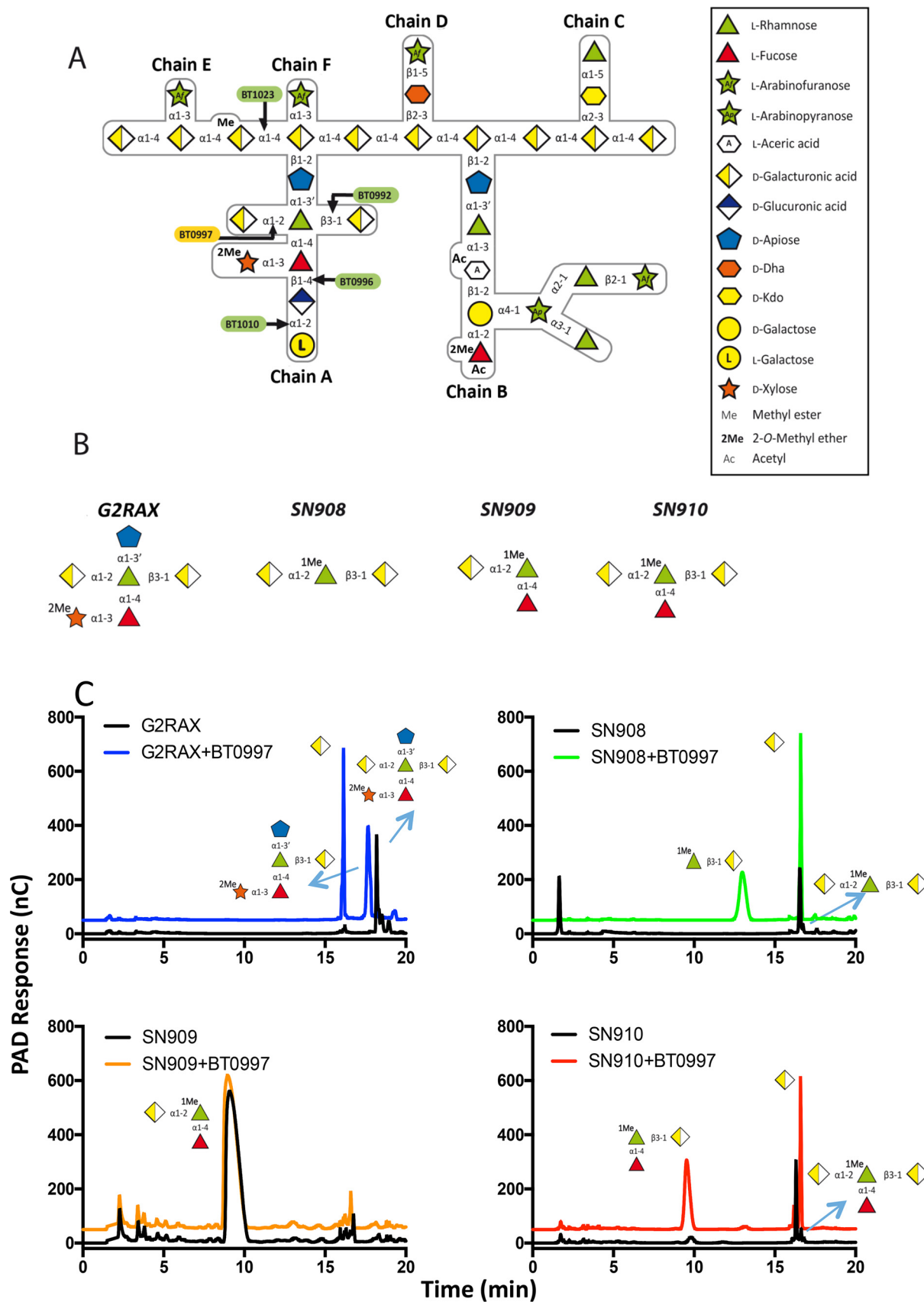
the founding member of a new GH family (GH138) we endeavored to reveal its catalytic mechanism. GHs typically operate through two catalytic mechanisms: an inverting mechanism, which is a single displacement reaction causing inversion of the anomeric configuration of the product compared with the substrate, or a retaining mechanism that is a double displacement reaction proceeding via a covalent intermediate, causing retention of the anomeric configuration in the products, with respect to the substrate (Fig. 2). The ^{1}H NMR spectra of the synthetic substrate SN910 were recorded prior to and at intervals after the addition of BT0997. This allowed the monitoring of the stereochemical course of the reaction and is shown in Fig. 2. After 1 min of incubation a doublet at about 5.22 ppm, assigned to H-1α of GalA (12), appeared and rapidly increased in intensity. Later in the incubation a small doublet at 4.5 ppm, because of H-1β of GalA, became noticeable and is from the mutarotation of the initially formed α-anomer. Therefore, the data clearly indicate that BT0997 catalyzes the hydrolysis of D-GalA-α1,2-L-Rha linkages at the nonreducing end of the oligosaccharide with retention of the anomeric configuration.

Crystal structure of BPA0997

Attempts to crystallize BT0997 proved unsuccessful, probably because of the heterogeneous oligomeric states and soluble aggregates formed by the protein, as evidenced by size-exclusion chromatography (Fig. S1). WP_024993800 from *B. paurosaccharolyticus* (BPA0997), a homologue that shares 70% identity with BT0997, displayed the same specificity as BT0997 and also had maximal activity on Rha decorated with both α-1,2– and β-1,3–linked GalA. BPA0997 produced two peaks when purified by size-exclusion chromatography, corresponding to a monomeric and dimeric state (Fig. S1). Protein corresponding to the monomeric state was carried forward into crystallographic trials. As BT0997 and BPA0997 are the founding members of the GH138 family, no corresponding structure was available in the Protein Data Bank (PDB). The structure of BPA0997 was solved by selenomethionine (SeMet), single wavelength anomalous X-ray scattering to a resolution of 2.00 Å (Table 2) (PDB code 6NZF). The SeMet protein crystallized in the primitive space group $P2_12_12_1$ and had two molecules in the asymmetric unit. BPA0997 is composed of four domains (Fig. 3). The N-terminal domain (ND1; residues 23–132) is made up of a five-stranded mixed β sheet with two parallel α helices beneath the sheet. The fifth β strand of the ND1 continues into the catalytic TIM barrel domain (D2; residues 133–468), which has a central barrel of eight β strands, 3–10, and is spiraled around by α helices 1–8. The third domain (D3; residues 469–724) comprises seven α helices, 11–17, which bundle together and run antiparallel to each other. Finally, the C-terminal domain (CD4; residues 725–893) presents a β-sandwich fold.

ND1 makes few interactions with the TIM barrel domain and displays higher average *B*-factors compared with the rest of the domains. Hydrophobic and polar contacts are made, however, between Phe-86, Tyr-107, and Glu-111 from ND1 and Trp-485 and Arg-183, respectively, from the TIM barrel domain. The interface area between ND1 and D2 is 827.5 Å^2. The D3 α helical bundle domain interacts with the TIM barrel through apo-

---

ASBMB

Figure 1. *A*, diagrammatic representation of rhamnogalacturnon II (RGII) highlighting the cleavage process and linkage cleaved by BT0997. *B*, the oligosaccharide substrates used in this study. *C*, HPLC chromatograms of 10 μM BT0997, incubated for 16 h, against the various oligosaccharides, at concentration of >1 mM, demonstrating only substrates with double galacturonic acid substitutions on rhamnose are targets for BT0997. All reactions were performed in 20 mM sodium phosphate, pH 7.0, and 150 mM sodium chloride.
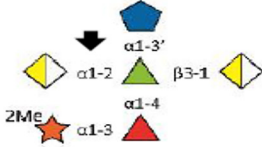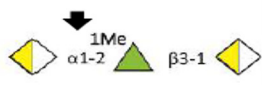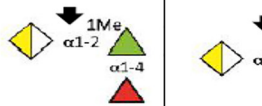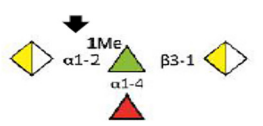
Table 1

Kinetic parameters of BT0997 and BPA0997 against RGII oligosaccharides

k_{cat}/K_m (min^{-1} M^{-1}) were determined in 20 mM sodium phosphate pH 7.0, 150 mM NaCl using 20–100 nM enzyme and 100 μM substrate. ΔCT refers to removal of the C-terminal domain of BPA0997. ND indicates that no significant activity was detected. n/a indicates activity was not tested.

	G2RAX	SN908	SN909	SN910
BT0997	(2.63 ± 0.1.8) × 10^4	(4.12 ± 0.26) × 10^4	ND	(6.09 ± 0.18) × 10^3
BPA0997	(1.49 ± 0.07) × 10^5	(3.63v± 0.13) × 10^5	ND	(1.04 ± 0.008) × 10^5
BPA0997ΔCT	(1.14 ± 0.031) × 10^3	n/a	n/a	n/a

lar interactions: Phe-485, Phe-482, and Trp-478 from helix 11 and Tyr-500 and Pro-508 from helix 12, with Trp-401, Pro-404, and Phe-406 from α helix 9 and Trp-454 and Tyr-463 from α helix 10 of the TIM barrel. These extensive interactions are reflected in the total surface interface of 2510.2 Å^2. The CD4 domain makes almost no interactions with the TIM barrel but sits on its top, burying a surface area of 504.8 Å^2. The CD4 domain appears to "lock" into the active site with an extended loop (Arg-814–Ser-820) that would prevent substrate binding. A glutamate, Glu-816, in the extended loop makes polar interactions with Arg-521 and His-542 forming a potentially "ionic lock" that secures the CD4 domain over the active site (Fig. 3, *B* and *C*). It is also possible that the location of the CD4 domain over the anterior surface of the TIM barrel is a consequence of the crystal packing. Indeed, the C-terminal linking loop makes close contacts with its symmetry-related copy resulting in the CD4 domain occluding the active site. Additionally, the CD4 domain makes contact with the helical domain of a symmetry-related mate. Thus, considering that BPA0997 displays activity in solution and that only few interactions occur between the catalytic domain and the CD4 domain, it is possible that out of the crystal context, the CD4 domain is mobile and could be involved in substrate recognition instead of preventing substrate binding.

Active site interactions of BPA0997 in complex with ligands

Despite the active site's being blocked by the CD4 domain both a serine, in the SeMet structure (which comes from the crystallization condition and is not biologically relevant so will not be discussed further), and a GalA, in the full-length E361S inactive mutant (cocrystallization with GalA) can be observed in the −1 active site, Fig. 3. These data suggest that the CD4 domain could close over a covalently bound GalA in the −1

subsite. As discussed earlier, the recognition of the β-1,3 Rha– linked GalA at the +2NR subsite is crucial for enzyme activity. To identify the key residues responsible for this feature, a BPA0997-G2RAX complex was necessary. However, the loop from the CD4 domain is likely to prevent the binding of G2RAX by blocking the active site in the crystal context. Thus, a C-terminally truncated version, eliminating residues 719–893 of BPA0997 (BPA0997ΔCT), was designed. This mutant enzyme displayed the same catalytic efficiency on G2RAX as the full-length enzyme (Table 1). Crystals of the inactive mutant BPA0997ΔCT$_{E361S}$ showed that the protein crystallized in the space group P22$_1$2$_1$ with one molecule in the asymmetric unit. Some crystals were thus soaked with a mixture of G2RAX and GRAX containing no free GalA (see "Materials and methods" for substrate generation and purification) and a 1.6 Å resolution structure was obtained. Although a complex was observed, only two GalA residues could be reliably built. The −1 GalA adopts both α and β anomers, whereas the +2NR GalA, which is partially disordered, could only be modeled as the β anomer. The O1 positions of the −1 α-GalA and the +2NR β-GalA are well-positioned to be connected to the O$_2$ and O3 of Rha, respectively (Fig. S2). There is some density suggesting a glycosidic linkage from the +2NR GalA to Rha but not for the Rha molecule itself. Also, there is no evidence of a connection between the −1 GalA and the Rha. This suggests that BPA0997ΔCT$_{E361S}$ had retained some residual catalytic activity against G2RAX and is likely bound to GalA and GRAX. The GRAX cannot be reliably built into the observed electron density, however, because of disorder.

Superimposition of the full-length native protein with BPA0997ΔCT$_{E361S}$, reveals that Glu-361 (sitting atop β strand 4) sits 3.5 Å from the anomeric carbon of the −1 GalA sugar and

A

C

B

C
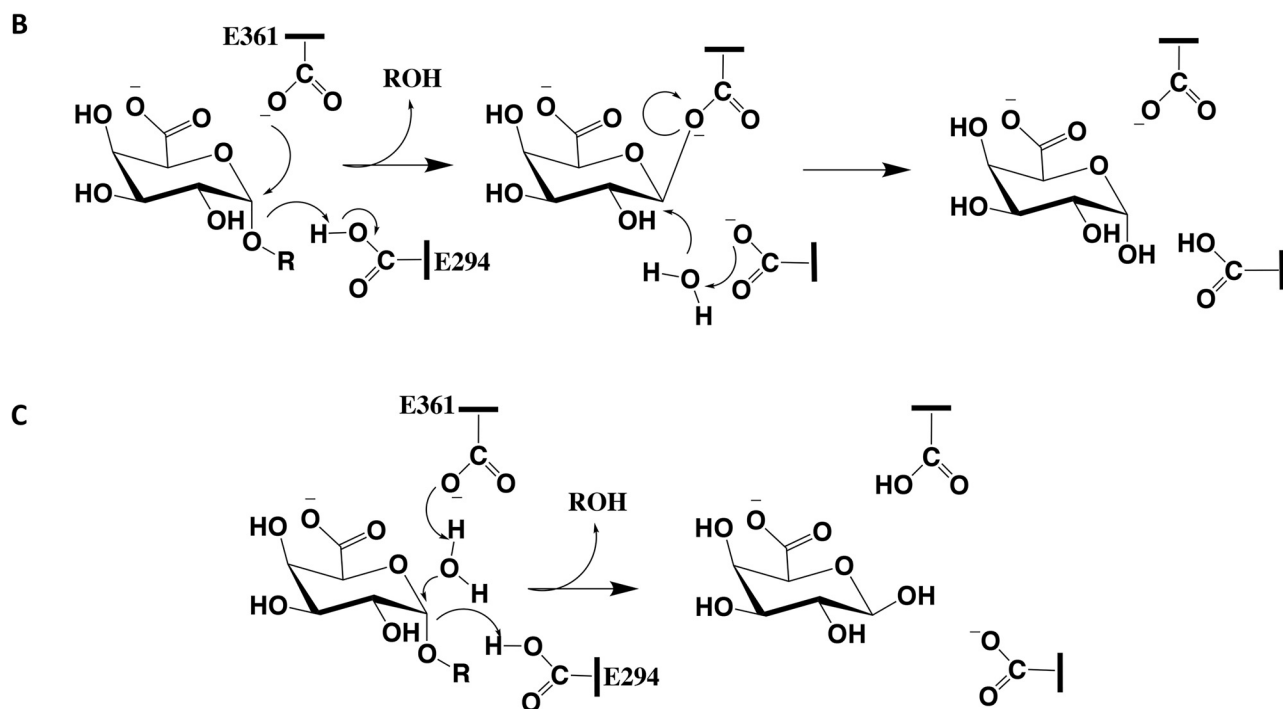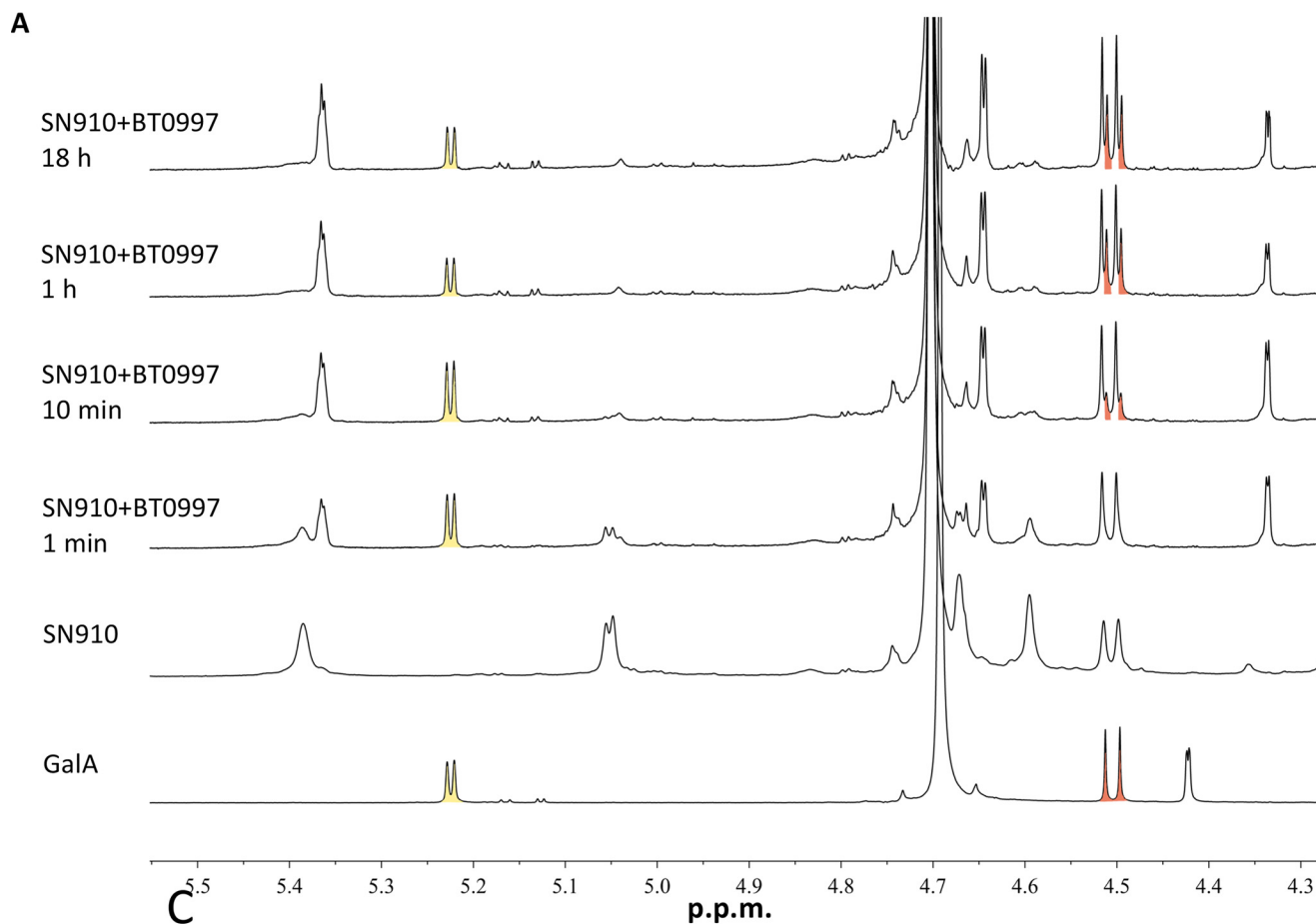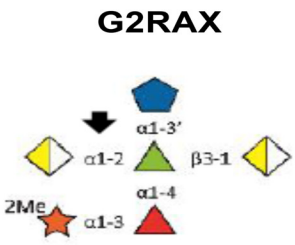
Figure 2. **¹H NMR spectra of SN910 treated with BT0997 in D₂O.** *A*, all reactions were carried out in 20 mM sodium phosphate buffer, pH 7.0, with 150 mM NaCl. The concentration of enzyme used was 145 μM against 12.5 mM SN910. *Yellow* and *red* highlight the signals corresponding to α-GalA and β-GalA, respectively. *B* and *C*, schematic representations of the retaining (double displacement mechanism) and the inverting mechanism (single displacement mechanism).

Table 2
Kinetic parameters of BPA0997 wildtype and mutants against G2RAX

k_{cat}/K_m (min^{-1} M^{-1}) were determined in 20 mM sodium phosphate pH 7.0, 150 mM NaCl using 20–100 nM enzyme and 100 μM substrate. ND indicates that activity could not be observed qualitatively (Fig. S2) and thus was not quantified. Errors are the S.E. from three independent experiments.

	G2RAX
BPA0997 (Wt)	$(1.5 \pm 0.1) \times 10^5$
E294S	ND
E294Q	ND
E361S	ND
E361Q	ND
R332A	ND
K357A	ND
H391A	ND
Y154A	$(1.15 \pm 0.05) \times 10^3$
W190A	$(8.80 \pm 0.5) \times 10^3$
Y241A	$(3.64 \pm 0.3) \times 10^2$
W432A	$(6.49 \pm 0.2) \times 10^3$
Y426A	$(2.17 \pm 0.03) \times 10^3$

is therefore the perfect candidate to be the catalytic nucleophile. Glu-294 (sitting atop β strand 6), on the other hand, is located at ~3 Å away from O1 of the −1 GalA and therefore should be fulfilling the role of the catalytic acid/base. Glu-361 and Glu-294 are ~5.7 Å apart which is well in line with the expected ~5.5 Å for catalytic residues in retaining enzymes. Their catalytic role is confirmed by the observation that the BPA0997 mutants E294S, E294Q, E361S, and E361Q are inactive when assayed biochemically (Table 2 and Fig. S3).

Interactions involving noncatalytic residues

The GalA at −1 makes several other interactions, mainly via its carboxylic acid group, forming a bidentate ionic interaction with Arg-332. Lys-357 and Tyr-241 also interact with the carboxylic acid with Lys-357 making a potential hydrogen bond to the endocyclic oxygen. Mutation of Arg-332 or Lys-357 to an alanine completely abolishes enzymatic activity, emphasizing the importance of these positively charged residues. The −1

GalA interacts with Oδ1 and NH$_2$ of Asn-520 via O2 and O3, respectively. The axial O4 of the −1 GalA, which is the unique chiral position of GalA, interacts with His-391 via Nϵ2. Mutation of this residue to an alanine completely knocks out catalytic capability of the enzyme. The effect of Tyr-241 is less detrimental, but still significant, suffering an ~72-fold loss in activity. There is a partially disordered GalA occupying what will be termed the +2NR subsite. Occupation of this subsite is critical for enzymatic activity and the interaction with the protein seems to be dominated by the carboxylic acid group forming a bidentate interaction with Arg-521. Additional interactions are potentially made by Thr-517 and Glu-518 with O3 and O4, respectively; however, the electron density for this part of the sugar is weak and may be reflective of the relative contribution of these interactions compared with Arg-521. Circular dichroism was performed on the inactive mutants with the data demonstrating they were properly folded (Fig. S4). Thus, inactivity was because of loss of the amino acid side chain in these mutants (Table 2).

Chloride-binding site

Also, in the vicinity of the active site is a chloride anion which is coordinated by NH1 and NH2 of Arg-332, Nδ2 of Asn-389, and a water molecule (Fig. 4). This is in agreement with the coordination sphere of the majority of the chloride anions observed in the protein crystal structures deposited in the PDB (13). Chloride anions have previously been observed to play an important role in the action of α-amylases. The anion binds to a common site close in the vicinity of the active site where it assists in the reaction mechanism (14). As Arg-332 and Asn-389 are two highly conserved residues (Figs. 4 and 5), we wanted to know whether the chloride anion present in BPA0997 is essential for catalytic activity. We attempted to remove NaCl from BPA0997 by gel filtration using 25 mM Hepes, pH 7.0, as the buffer, a method described previously (15). The reaction on G2RAX was subsequently performed post gel filtration in 25 mM Hepes, pH 7.0, with no chloride added. The enzyme was still active (data not shown) suggesting that the anion is not required for catalytic activity or that the anion is too tightly bound to be removed by the method described above.

Bioinformatic analysis

Sequence alignments of the top 500 results returned from Blastp, using BT0997 as the input sequence, reveal that all but 21 sequences show conservation of the catalytic residues Glu-294 and Glu-361, the acid/base and nucleophile, respectively. In addition the two arginine residues Arg-332 and Arg-521, which coordinate the carboxylic acid groups of the −1 and +2NR GalA residues and are required for activity, are also invariant in these 479 sequences (Fig. 5). These two arginine residues are likely the key specificity determinants for GH138 enzymes in targeting GalA residues in double substitutions. This demonstrates that BT0997 is very well-conserved in organisms where it is found, and therefore its function. The observation that three sequences retain both catalytic residues and Arg-332 but not Arg-521 may suggest that these protein sequences encode catalytically active enzymes but target residues where only one acidic sugar substitutes the +1 residue.
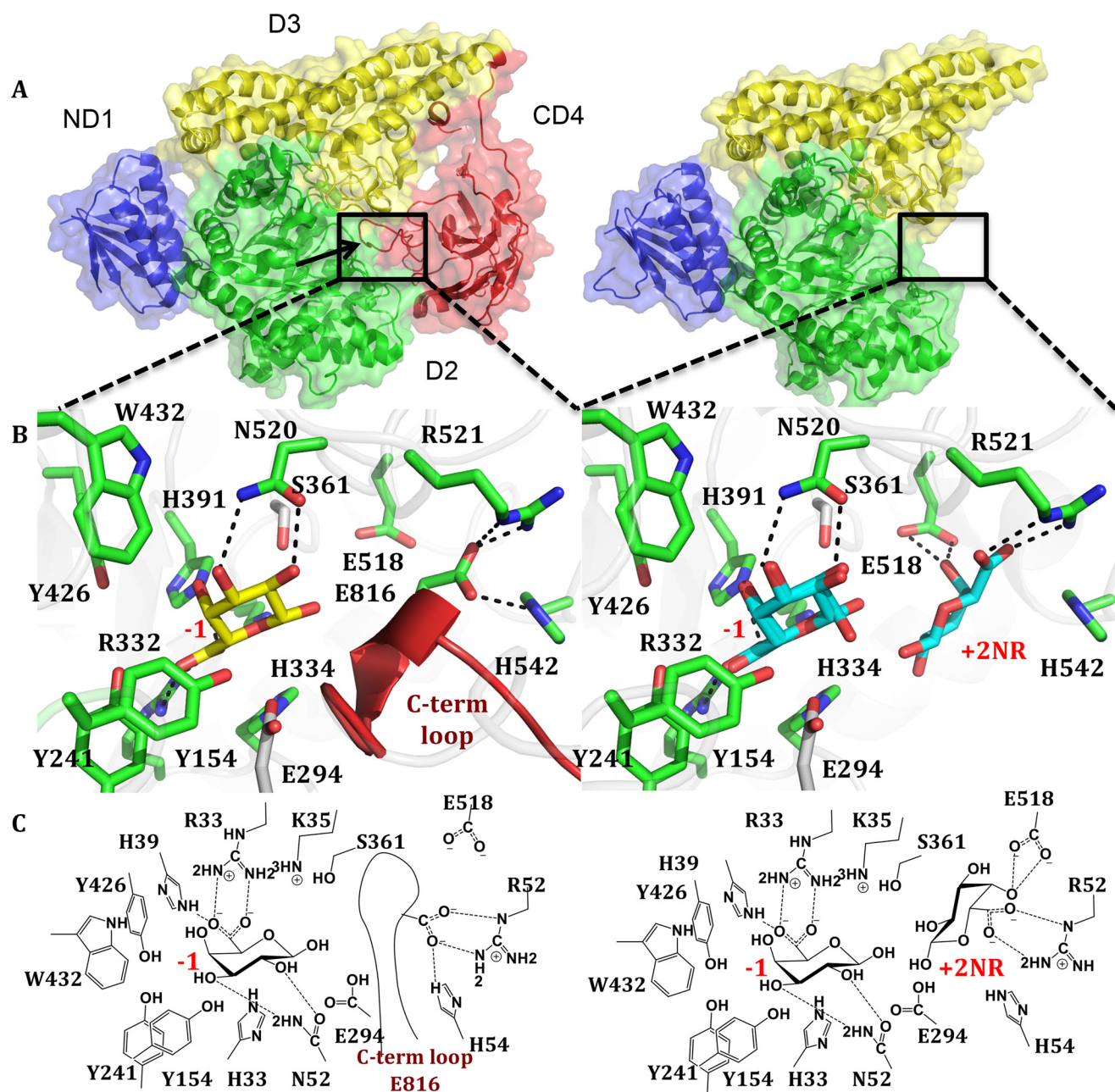
Figure 3. Structural details of BPA0997. *A*, cartoon representation of full-length BPA0997$_{E361S}$, *left*, and C-terminally truncated BPA0997ΔCT$_{E361S}$, *right*. The four domains, ND1 (*blue*), D2 (*green*), D3 (*yellow*), and CD4 (*red*) are shown (*blue* to *red* from its N to C terminus). The *black arrow* shows the loop of the CD4 domain locking into the active site of the D2 TIM barrel domain. The protein backbone is represented as cartoon. The surface of the protein is shown at 80% transparency. *B*, active site interactions of BPA0997, *left*, with GalA shown in *yellow*, and C-terminally truncated BPA0997ΔCT$_{E361S}$, *right*, with GalA shown in *cyan*. *C*, diagrammatic active site interactions of BPA0997$_{E361S}$, *left*, and C-terminally truncated BPA0997ΔCT$_{E361S}$, *right*.

The one sequence, which retains both catalytic amino acids, but neither of the basic arginines, may encode an enzyme targeting neutral sugars at the −1 subsite.

Discussion

BT0997 and BPA0997 are well-conserved in the PULs of RGII-degrading organisms, suggesting it is an important part of the degradative apparatus. Indeed, the genetic knockout of *bt0997* in *Bt* causes degradation of chain to cease at the point of the target linkage causing the undigested hexasaccharide, G2RAX from chain A, to be left over in the growth media (10). There are examples where multiple decorations of sugar resi-

dues cause a specific stereochemical blockage to cleavage of the glycosidic bond of the sugar to which they are appended also. This requires the evolution of enzymes that are specifically targeted to these residues. BT0997 and its homologues are part of a small number of glycoside hydrolases that have evolved to address this problem. The ability to target these residues usually places the requirement that both of the decorating sugars are required for activity and creates a more open active site cavity. This has resonance with the AXHd3, an arabinofuranosidase from *Humicola insolens* that specifically targets the α-1,3-L-arabinofuranose residues on doubly decorated xylose (16). The enzyme, analogous to BT0997, requires both the
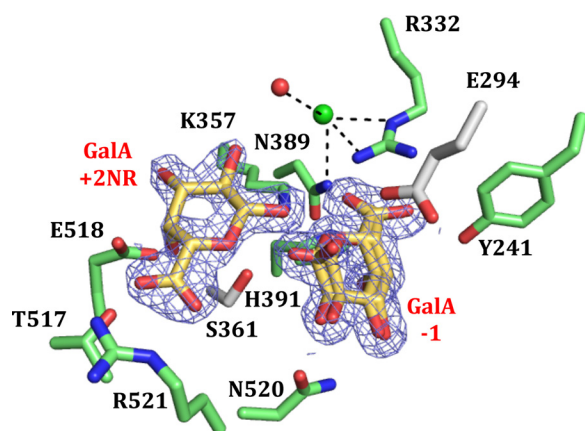
Figure 4. Active site of BPA0997ΔCT$_{E361S}$ in complex with two molecules of galacturonic acid and a chloride anion. The $2F_{obs} - F_c$ electron density map for the two bound galacturonic acid residues contoured to 1σ (0.35 e/A^3).

α-1,3-L-arabinofuranose and α-1,2 substitutions to be active but only cleaves the α-1,3 linkage.

A Dali search using the BPA0997ΔCT showed no significant structural homologs in the PDB database. The top three results were a GH67 (PDB code 1K9E), GH115 (PDB code 4C91) and a GH20 (PDB code 4C7D), with the best result (GH67) having only 11% identity, a root mean square deviation (RMSD) of 3.6 Å over 499 residues and a Z-score of only 18.7. These three enzymes also utilize different mechanisms to the GH138 with GH67 and GH115 utilizing an inverting mechanism whereas GH20 operates through a substrate-assisted mechanism. These data point to very little ancestry between the top hits from the Dali search and GH138 enzymes. GH67, however, also contains ND1 and a domain similar to D3 but lacks the CD4 domain. A Dali search was also run with ND1, D3, and CD4 individually. ND1 matches the N-terminal domains of several other GH67 members, with PDB code 1LN8 being the best match with a RMSD of 1.8 Å, an identity of 20% but a Z-score of only 13.6. The D3 domain best matched PDB code 2R03 which is involved in apoptosis and adopts a similar, extended, α helical fold. The Dali results, however, were poor, having an RMSD of 2.7 Å, an identity of 6%, and a Z-score of 9.4. The CD4 domain returns some evidence for similar folds in several other CAZyme families. Again the similarities were low with an RMSD of 3.1 Å, an identity of 8%, and a Z-score of 3.6. Further inspection of the overlays of these three structures with BPA0997 reveals that there are no gross similarities between the active sites; however, some commonalities can be observed with GH67. GH67 is predominantly xylan α-1,2-D-glucuronidases and Arg-318 is invariantly conserved when a Blastp is performed using PDB code 1K9E. This residue corresponds to Arg-332 in BPA0997. This demonstrates that GH67 and GH138 coordinate the carboxylic acids, glucuronic acid and GalA, respectively, in the same manner. The catalytic nucleophile Glu-361 also is spatially similar to the Asp-364 in GH67, but Asp-364 likely acts as a catalytic acid.

The structure of the BPA0997 revealed a TIM barrel fold, a common tertiary fold observed in many GH families and particularly in the clan A class of enzymes. BPA0997 constitutes a new family and is not part of clan A as the catalytic residues are

not found on β strands 4 and 7, but rather β strands 4 and 6, of the central barrel and the catalytic acid is not part of the NEP/HEP motif found in clan A enzymes (17). Presumably, if its evolutionary origins are distinct, and as other members are discovered in the future, it may form part of a new clan. The same may be said of the other three domains. These are not new folds but the poor values from the Dali search may imply very distant evolutionary relationships or again evolving from unique progenitor sequences. The role of the CD4 domain remains ambiguous as no effect on catalytic activity, and presumably stability, was observed with its removal. However, such ionic locks securing loop conformers in CAZymes have been observed before, notably in the recently discovered PL27 and *Cj*Man26C where removal of these features heavily impacted enzyme function (18, 19). It is plausible that the effects of removing the CD4 domain could be masked by measuring only the k_{cat}/K_m against G2RAX (because of limited substrate full kinetic parameters could not be derived). Although G2RAX and the CD4 domain cannot occupy the same space, the CD4 domain may close over the covalently bound −1 GalA and concomitantly drive departure of the leaving group helping to create a closed microenvironment for the second stage of catalysis. This would then require the CD4 domain to move out of the active site before the −1 GalA could depart and allow a new catalytic cycle. Thus, removal of the CD4 domain would lose this "microenvironment" but could allow the −1 GalA to dissociate faster at the end of the catalytic cycle. This could raise both the K_m and k_{cat} by commensurate levels and thus no effect could be observed measuring only k_{cat}/K_m. This is a limitation when substrate is bespoke and limited.

Conclusions

This study describes the first structural and mechanistic analysis of GH138, characterizing the proteins BT0997 and BPA0997. Both enzymes target the L-Rha-α-1,2-D-GalA linkage contained within chain A of RGII. The enzymes target, and require, rhamnose that is double substituted with GalA using two invariant arginines to drive specificity, performing catalysis through a double displacement, retaining, mechanism.

Materials and methods

Cloning, expression, and purification of BT0997 and of its homologue BPA0997

All recombinant forms of proteins used in this study were expressed in the cytoplasm of *Escherichia coli*. The gene encoding BT0997 (locus identifier: BT_0997, NCBI Reference Sequence: NC_004663.1) was described previously (10). DNA encoding *B. paurosaccharolyticus* GH138 (BPA0997, NCBI Reference Sequence: WP_024993800.1) was initially generated by GeneArt gene synthesis (Thermo Fisher Scientific). BPA0997 construct was cloned such that the encoded protein contains a C-terminal His$_6$ tag. BPA0997 constructs were cloned in the pET_SUMO vector displaying a modified MCS (provided by Dr. Patrick Moynihan from University of Birmingham) such that the encoded proteins contain a N-terminal SUMO tag that is subsequently removed by a specific protease following the instructions provided by the Champion™ pET SUMO Protein Expression System kit protocol. Site-directed
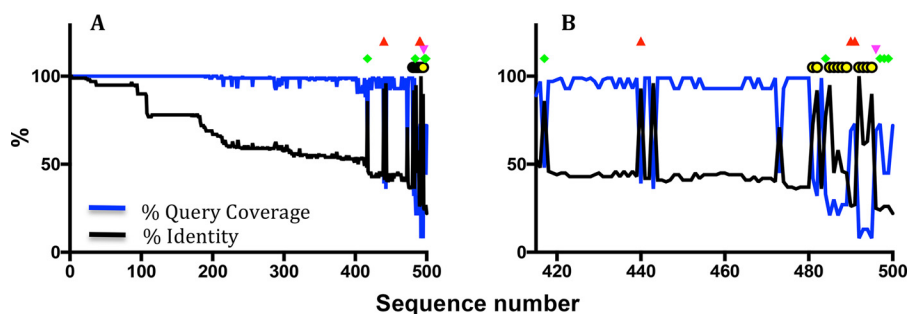
Figure 5. *A* and *B*, the same analysis of a BLASTP search using BT0997 as the subject. In *B*, however, only sequences 415 to 500 are shown for ease of visualization. The *black line* shows percentage identity of the sequence to BT0997; the *blue line* indicates the percentage of query coverage. 479 of the 500 sequences contain both the catalytic residues (Glu-294 and Glu-361) and the critical arginine specificity determinants (Arg-332 and Arg-521). The 21 sequences that do not contain all four residues are highlighted with *red triangles* indicating the absence of only Arg-521, upside down *magenta triangles* indicate the absence of both R332/R521, *green diamonds* indicate the absence of R332/R521/E361 and finally *yellow circles* indicate the absence of Arg-332/Arg-521/Glu-361/Glu-294.

mutagenesis was carried out using the PCR-based QuikChange method (Stratagene). To express the recombinant proteins, *E. coli* strain Tuner (DE3), harboring appropriate recombinant plasmids, was cultured to mid–exponential phase in Luria broth at 37 °C supplemented with 10 μg.ml^{-1} kanamycin. Isopropyl β-D-galactopyranoside to a final concentration of 0.2 mM was then added to induce recombinant gene expression, and the culture incubated for a further 18 h at 16 °C. Cells were harvested by centrifugation at 4424 $\times$ *g* for 10 min and resuspended in 20 mM Tris·HCl buffer, pH 8.0, containing 300 mM NaCl. Cells were lysed by sonication, and the cell-free extract was recovered by centrifugation at 27,216 $\times$ *g* for 30 min. The recombinant proteins were purified to 90% electrophoretic purity by immobilized metal ion affinity chromatography using Talon^{TM} (Clontech), cobalt-based matrix, and elution with 100 mM imidazole, as described previously (20) When preparing the selenomethionine derivatives of BPA0997 constructs for crystallography, the proteins were expressed in *E. coli* B834 (DE3), a methionine auxotroph, cultured in medium comprising 1 liter of SelenoMet^{TM} Medium Base, 50 ml of SelenoMet^{TM} nutrient mix (Molecular Dimensions), and 4 ml of a 10 mg.ml^{-1} solution of L-selenomethionine. Recombinant gene expression and protein purification were as described above. For crystallographic studies, BPA0997 constructs were further purified by gel filtration chromatography using HiLoad 16/600 Superdex 200 pg (GE Healthcare). The elution buffer was 20 mM Na-Hepes, pH 7.5, 150 mM NaCl.

Oligosaccharides production and purification—RGII-derived oligosaccharides

B. thetaiotaomicron strains containing specific gene deletions for *bt0997* (generates G2RAX) and *bt0992* (generates GRAX) were inoculated into minimal medium containing 1% RGII and grown in glass test tubes for 48 h at 37 °C, in an anaerobic cabinet (Whitley A35 Work station; Don Whitley, UK) to an $A_{600\,nm}$ of 2.0. Cells were harvested by centrifugation initially at 2400 $\times$ *g* for 10 min and later at 17,000 $\times$ *g* for another 10 min. The resulting supernatant was filtered through a 0.2 μm syringe filter (VWR^{TM}) and separated on two Bio-Gel P2 (Bio-Rad) size-exclusion columns (2.5 cm $\times$ 100 cm), ran in series, and eluted with 50 mM acetic acid at 0.2 ml·min^{-1}. Fractions (1 ml) were collected and analyzed by TLC using orcinol/sulfuric acid to

reveal the resolved sugars. Fractions containing oligosaccharides of interest were pooled and concentrated by freeze-drying using a CHRIST Gefriertrocknung ALPHA 1–2 freeze-dryer (Helmholtz-Zentrum Berlin) at −50 °C (10). The synthetic substrates SN908, SN909, and SN910 were prepared using protocols outlined previously (21, 22).

Enzyme assays

All enzyme assays unless otherwise stated were carried out in 20 mM sodium phosphate buffer, pH 7.0, containing 150 mM NaCl and performed in triplicate. Assays were carried out with 100 nM–1 μM enzyme against 100–300 μM substrate at 37 °C. Aliquots were taken over a 16-h time course, and samples and products were assessed by TLC and high pressure anion exchange chromatography (HPAEC) with pulsed amperometric detection. Sugars were separated on a Carbopac PA1 guard and analytical column in an isocratic program of 100 mM sodium hydroxide and then with a 40% linear gradient of sodium acetate over 60 min. Sugars were detected using the carbohydrate standard quad waveform for electrochemical detection at a gold working electrode with an Ag/AgCl pH reference electrode. Spectrophotometric, quantitative assays for BT0997 and BPA0997 were monitored by the formation of NADH, at $A_{340\,nm}$ using an extinction coefficient of 6230 M^{-1} cm^{-1}, with an appropriately linked enzyme assay system. The assays were adapted from purchased Megazyme International assay kit D-glucuronic acid/D-galacturonic acid assay kit (K-URONIC). A single substrate concentration was used to calculate catalytic efficiency (k_{cat}/K_m) and was checked to be $\ll K_m$ by halving and doubling the substrate concentration and observing an appropriate increase or decrease in rate. The equation $V_0 = (k_{cat}/K_m)[S][E]$ was used to calculate k_{cat}/K_m.

Circular dichroism

Concentrated proteins were diluted to 500 nM in 50 mM sodium phosphate pH 7.0, and analyzed with a JASCO J8–10 spectropolarimeter in a 1 mm pathlength quartz cuvette. Samples were scanned from 190 nm to 260 nm at 20 nm·min^{-1} with a bandwidth of 2 nm and a response time of 8 s. Nine circular dichroism scans were performed per sample at 37 °C and the runs averaged. Raw spectra were analyzed using the BestSel server (23).

Table 3
Crystallographic data and refinement statistics
Values in parentheses are for the highest resolution shell. R_{free} was calculated using a set (5%) of randomly selected reflections that were excluded from refinement.

	BPA0997 (SeMet)	BPA0997$_{E361S}$	BPA0997ΔCT$_{E361S}$
Data collection			
Date	17/12/16	04/03/17	09/12/17
Source	I03	I04	I24
Wavelength (Å)	0.978	0.979	0.979
Space group	P2$_1$	P2$_1$	P22$_1$2$_1$
Cell dimensions			
a, b, c (Å)	69.8 103.2 138.5	69.5 107.4 139.2	55.30 68.38 220.16
α, β, γ (°)	90.0 90.4 90.0	90.0 98.7 90.0	90.0 90.0 90.0
No. of measured reflections	1,721,396 (82,026)	210,454 (17,375)	514,191 (23,838)
No. of independent reflections	134,407 (6637)	55,515 (4532)	100,505 (4630)
Resolution (Å)	49.50–2.00 (2.03–2.00)	47.35–2.70 (2.78–2.70)	73.39–1.60 (1.63–1.60)
CC$_{1/2}$	0.999 (0.879)	0.992 (0.835)	0.991 (0.343)
$I/\sigma(I)$	21.0 (3.6)	10.5 (2.1)	5.9 (1.7)
Completeness (%)	100.0 (99.9)	99.8 (99.7)	90.5 (84.7)
Redundancy	12.8 (1.4)	3.8 (3.8)	5.1 (5.1)
Anomalous completeness (%)	99.6 (98.9)		
Anomalous multiplicity	6.4 (6.1)		
Refinement			
R_{work}/R_{free}	17.3/20.8	21.3/28.2	15.4/20.5
No. atoms			
Protein	13,898	13,926	5630
Ligand/ions	19	36	41
Water	686		769
B-factors			
Protein	41.8	48.2	20.4
Ligand/ions	32.6	40.8	25.7
Water	42.0		34.3
PDB code	**6HZF**	**6HZE**	**6HZG**

Crystallization, data collection, structure solution, and refinement

Crystallization screening was undertaken with the robotic NanoDrop dispensing systems (mosquito LCP; TTP Labtech) using commercial screens. SeMet-derivative crystals of BPA0997 were initially obtained in Morpheus H10 (Molecular Dimensions). SeMet BPA0997 was manually optimized and single crystals were obtained as follows: 1 μl of enzyme at 12 mg.ml^{-1} was mixed with 1 μl of a 1 ml reservoir solution containing 100 mM Tris(base)/bicine buffer, pH 8.5, 30 mM each amino acid (glutamate, alanine, glycine, lysine, serine), 8% v/v ethylene glycol, and 16% w/v PEG 8000 in hanging drops at 20 °C. Prior to flash-cooling in liquid nitrogen, single crystals were cryo-protected with 20% w/v PEG 400. As no appropriate molecular replacement model was available for BPA0997 in the Protein Data Bank, the structure was solved with experimental phasing by SeMet-SAD (single wavelength anomalous diffraction). A fluorescence energy scan was performed around the selenium K atomic absorption edge of 12658 eV, confirming the presence of selenomethionine in the sample, and a SAD data set was collected at the optimal peak wavelength of 0.979 Å. Diffraction data were collected at the Diamond Light Source, Didcot, United Kingdom, on beamlines I03, I04, and I24 at a temperature of 100 K. Data were indexed and integrated with XDS (24) and scaled using Aimless (25). Space groups were confirmed using Pointless (25). Phases were determined experimentally using the anomalous scattering from the selenium atoms with the SHELXC/D/E (26) suite using HKL2MAP. The structure was built automatically in Buccaneer (27) and ARP_wARP (28) consecutively. Native crystals of the mutant BPA0997$_{E361S}$ were obtained in presence of 100 mM D-galacturonic acid in 30 mM sodium fluoride, 30 mM sodium bromide, 100 mM imidazole/MES, pH 6.5, 25% 2-methyl-2,4-pentanediol. BPA0997ΔCT$_{E361S}$ crystals were obtained in 9% butanol, 100 mM

Tris, pH 8.4, and 16% PEG 8000. The data were integrated and indexed with XDS (24) or Xia2 (29) and scaled with Aimless (25). The phase problem for both mutants was solved by molecular replacement using the SeMet model with Molrep (30) or Phaser (31). For all data, 5% of the observations were randomly selected for the R_{free} set. The models underwent recursive cycles of model building in Coot (32) and refinement cycles in Refmac5 (33). The models were validated using Coot (32) and MolProbity. Structural figures were made using PyMOL (The PyMOL Molecular graphics system, Version 2.0 Schrodinger, LLC) and all other programs used were from the CCP4 suite (34). The data processing and refinement statistics are reported in Table 3.

NMR spectroscopy

All samples were diluted in 20 mM sodium phosphate buffer, pH 7.0, containing 150 mM NaCl, freeze-dried and resuspended three times in D$_2$O before the experiment. Initial spectra were recorded with 800 μl of 12.5 mM S910 in reaction buffer before initiating the reaction by the addition of 200 μl of BT0997 (final concentration, 30 μM). ^{1}H NMR spectra were recorded in D$_2$O on a Bruker Avance III HD 500 MHz NMR spectrometer operating at 500.15 MHz at regular intervals. Each spectrum was acquired with nine scans. Spectra of D-galacturonic acid (20 mM in reaction buffer) were also recorded.

Author contributions—A. L., A. B., J. M.-M., D. N., and A. C. investigation; A. L., A. B., S. B., S. A. N., and A. C. methodology; A. L., A. B., R. A. F., and A. C. writing-review and editing; A. L. and A. C. conceptualization; A. C. formal analysis; A. C. writing-original draft.

References

1. Porter, N. T., and Martens, E. C. (2017) The critical roles of polysaccharides in gut microbial ecology and physiology. *Annu. Rev. Microbiol.* **71,** 349–369 CrossRef Medline

2. McNeil, N. I. (1984) The contribution of the large intestine to energy supplies in man. *Am. J. Clin. Nutr.* **39,** 338–342 CrossRef Medline

3. Hickey, C. A., Kuhn, K. A., Donermeyer, D. L., Porter, N. T., Jin, C., Cameron, E. A., Jung, H., Kaiko, G. E., Wegorzewska, M., Malvin, N. P., Glowacki, R. W., Hansson, G. C., Allen, P. M., Martens, E. C., and Stappenbeck, T. S. (2015) Colitogenic *Bacteroides thetaiotaomicron* antigens access host immune cells in a sulfatase-dependent manner via outer membrane vesicles. *Cell Host Microbe* **17,** 672–680 CrossRef Medline

4. Zitvogel, L., Galluzzi, L., Viaud, S., Vetizou, M., Daillere, R., Merad, M., and Kroemer, G. (2015) Cancer and the gut microbiota: An unexpected link. *Sci. Transl. Med.* **7,** 271ps1 CrossRef Medline

5. Jiang, C., Li, G., Huang, P., Liu, Z., and Zhao, B. (2017) The gut microbiota and Alzheimer's disease. *J. Alzheimers Dis.* **58,** 1–15 CrossRef Medline

6. Desai, M. S., Seekatz, A. M., Koropatkin, N. M., Kamada, N., Hickey, C. A., Wolter, M., Pudlo, N. A., Kitamoto, S., Terrapon, N., Muller, A., Young, V. B., Henrissat, B., Wilmes, P., Stappenbeck, T. S., Núñez, G., and Martens, E. C. (2016) A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. *Cell* **167,** 1339–1353e1321 CrossRef Medline

7. Martens, E. C., Lowe, E. C., Chiang, H., Pudlo, N. A., Wu, M., McNulty, N. P., Abbott, D. W., Henrissat, B., Gilbert, H. J., Bolam, D. N., and Gordon, J. I. (2011) Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol.* **9,** e1001221 CrossRef Medline

8. O'Neill, M. A., Warrenfeltz, D., Kates, K., Pellerin, P., Doco, T., Darvill, A. G., and Albersheim, P. (1996) Rhamnogalacturonan-II, a pectic polysaccharide in the walls of growing plant cell, forms a dimer that is covalently cross-linked by a borate ester. *In vitro* conditions for the formation and hydrolysis of the dimer. *J. Biol. Chem.* **271,** 22923–22930 CrossRef Medline

9. Apolinar-Valiente, R., Williams, P., Romero-Cascales, I., Gomez-Plaza, E., Lopez-Roca, J. M., Ros-Garcia, J. M., and Doco, T. (2013) Polysaccharide composition of Monastrell red wines from four different Spanish terroirs: Effect of wine-making techniques. *J. Agric. Food Chem.* **61,** 2538–2547 CrossRef Medline

10. Ndeh, D., Rogowski, A., Cartmell, A., Luis, A. S., Baslé, A., Gray, J., Venditto, I., Briggs, J., Zhang, X., Labourel, A., Terrapon, N., Buffetto, F., Nepogodiev, S., Xiao, Y., Field, R. A., *et al.* (2017) Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature* **544,** 65–70 CrossRef Medline

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

12. Pitson, S. M., Mutter, M., van den Broek, L. A., Voragen, A. G., and Beldman, G. (1998) Stereochemical course of hydrolysis catalysed by α-L-rhamnosyl and α-D-galacturonosyl hydrolases from *Aspergillus aculeatus*. *Biochem. Biophys. Res. Commun.* **242,** 552–559 CrossRef Medline

13. Carugo, O. (2014) Buried chloride stereochemistry in the Protein Data Bank. *BMC Struct. Biol.* **14,** 19 CrossRef Medline

14. Aghajari, N., Feller, G., Gerday, C., and Haser, R. (2002) Structural basis of α-amylase activation by chloride. *Protein Sci.* **11,** 1435–1441 CrossRef Medline

15. Feller, G., Bussy, O., Houssier, C., and Gerday, C. (1996) Structural and functional aspects of chloride binding to *Alteromonas haloplanctis* α-amylase. *J. Biol. Chem.* **271,** 23836–23841 CrossRef Medline

16. McKee, L. S., Peña, M. J., Rogowski, A., Jackson, A., Lewis, R. J., York, W. S., Krogh, K. B., Viksø-Nielsen, A., Skjøt, M., Gilbert, H. J., and Marles-Wright, J. (2012) Introducing endo-xylanase activity into an exo-acting arabinofuranosidase that targets side chains. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 6537–6542 CrossRef Medline

17. Zhang, Y., Ju, J., Peng, H., Gao, F., Zhou, C., Zeng, Y., Xue, Y., Li, Y., Henrissat, B., Gao, G. F., and Ma, Y. (2008) Biochemical and structural characterization of the intracellular mannanase AaManA of *Alicyclobacillus acidocaldarius* reveals a novel glycoside hydrolase family belonging to clan GH-A. *J. Biol. Chem.* **283,** 31551–31558 CrossRef Medline

18. Cartmell, A., Topakas, E., Ducros, V. M., Suits, M. D., Davies, G. J., and Gilbert, H. J. (2008) The *Cellvibrio japonicus* mannanase *Cj*Man26C displays a unique *exo*-mode of action that is conferred by subtle changes to the distal region of the active site. *J. Biol. Chem.* **283,** 34403–34413 CrossRef Medline

19. Munoz-Munoz, J., Cartmell, A., Terrapon, N., Baslé, A., Henrissat, B., and Gilbert, H. J. (2017) An evolutionarily distinct family of polysaccharide lyases removes rhamnose capping of complex arabinogalactan proteins. *J. Biol. Chem.* **292,** 13271–13283 CrossRef Medline

20. Charnock, S. J., Brown, I. E., Turkenburg, J. P., Black, G. W., and Davies, G. J. (2002) Convergent evolution sheds light on the anti-β-elimination mechanism common to family 1 and 10 polysaccharide lyases. *Proc. Natl. Acad. Sci. U.S.A.* **99,** 12067–12072 CrossRef Medline

21. Chauvin, A. L., Nepogodiev, S. A., and Field, R. A. (2005) Synthesis of a 2,3,4-triglycosylated rhamnoside fragment of rhamnogalacturonan-II side chain A using a late stage oxidation approach. *J. Org. Chem.* **70,** 960–966 CrossRef Medline

22. Nepogodiev, S. A., Fais, M., Hughes, D. L., and Field, R. A. (2011) Synthesis of apiose-containing oligosaccharide fragments of the plant cell wall: Fragments of rhamnogalacturonan-II side chains A and B, and apiogalacturonan. *Org. Biomol. Chem.* **9,** 6670–6684 CrossRef Medline

23. Micsonai, A., Wien, F., Bulyáki, É., Kun, J., Moussong, É., Lee, Y.-H., Goto, Y., Réfrégiers, M., and Kardos, J. (2018) BeStSel: A web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic Acids Res.* **46,** W315–W322 CrossRef Medline

24. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 125–132 CrossRef Medline

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

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

27. Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62,** 1002–1011 CrossRef Medline

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

29. Winter, G., Lobley, C. M., and Prince, S. M. (2013) Decision making in xia2. *Acta Crystallogr. D Biol. Crystallogr.* **69,** 1260–1273 CrossRef Medline

30. Vagin, A., and Teplyakov, A. (2010) Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 22–25 CrossRef Medline

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

32. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 486–501 CrossRef Medline

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

34. Collaborative Computational Project. (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50,** 760–763 CrossRef Medline

35. Terrapon, N., Lombard, V., Drula, É., Lapébie, P., Al-Masaudi, S., Gilbert, H. J., and Henrissat, B. (2017) PULDB: The expanded database of polysaccharide utilization loci. *Nucleic Acids Res.* **46,** D677–D683 CrossRef Medline

≋ASBMB

J. Biol. Chem. (2019) 294(19) 7711–7721 **7721**