Growth of *Chitinophaga pinensis* on Plant Cell Wall Glycans and Characterisation of a Glycoside Hydrolase Family 27 β-ʟ-Arabinopyranosidase Implicated in Arabinogalactan Utilisation

Lauren S. McKee[1,2], Harry Brumer[1,2,3]*

1 Division of Glycoscience, School of Biotechnology, Royal Institute of Technology (KTH), AlbaNova University Centre, 106 91, Stockholm, Sweden, **2** Wallenberg Wood Science Centre, Teknikringen 56–56, 100 44, Stockholm, Sweden, **3** Michael Smith Laboratories and Department of Chemistry, University of British Columbia, 2185 East Mall, Vancouver, V6T 1Z4, BC, Canada

* brumer@msl.ubc.ca

Abstract

The genome of the soil bacterium *Chitinophaga pinensis* encodes a diverse array of carbohydrate active enzymes, including nearly 200 representatives from over 50 glycoside hydrolase (GH) families, the enzymology of which is essentially unexplored. In light of this genetic potential, we reveal that *C. pinensis* has a broader saprophytic capacity to thrive on plant cell wall polysaccharides than previously reported, and specifically that secretion of β-ʟ-arabinopyranosidase activity is induced during growth on arabinogalactan. We subsequently correlated this activity with the product of the Cpin_5740 gene, which encodes the sole member of glycoside hydrolase family 27 (GH27) in *C. pinensis*, *Cp*Arap27. Historically, GH27 is most commonly associated with α-ᴅ-galactopyranosidase and α-ᴅ-*N*-acetylgalactosaminidase activity. A new phylogenetic analysis of GH27 highlighted the likely importance of several conserved secondary structural features in determining substrate specificity and provides a predictive framework for identifying enzymes with the less common β-ʟ-arabinopyranosidase activity.

Introduction

Microorganisms with the capacity to selectively and efficiently degrade plant-derived carbohydrates are of great interest to research and industry as a source of new tools for the characterisation and degradation of plant biomass [1–7]. *Chitinophaga pinensis* is a motile, gram-negative bacterium that was originally characterised by its ability to utilise the eponymous insect and fungal polysaccharide chitin [8, 9]. However, *C. pinensis* was in fact isolated from pine forest leaf litter, an environment which would be expected to provide a rich source of plant cell wall glycans. More recently, the complete genome sequence of *C. pinensis*, generated as part of the

Joint Genome Institute's Genomic Encyclopedia of Bacteria and Archaea (GEBA) project [10], revealed a large variety of catabolic carbohydrate-active enzymes (CAZymes), including 193 glycoside hydrolases from 56 of the 130 known glycoside hydrolase (GH) families (see http://www.cazy.org/b1017.html) [11, 12]. The predicted diversity of these enzymes expands well beyond the handful of GH families implicated in chitin degradation. Indeed, only 10 of the GH enzymes encoded by this genome are predicted to act on chitin (chitinases and *N*-acetylglucosaminidases) [12]. Thus, it seems likely that *C. pinensis* has a wider ability than previously appreciated to grow on complex plant biomass polysaccharides in its native environment [8]. At the same time, the limited ability of this bacterium to grow on cellulose and starch suggests that *C. pinensis* may preferentially degrade the amorphous matrix glycans that are ubiquitous and abundant in plant cell walls. As such, the genome of *C. pinensis* constitutes a rich resource for the discovery of new CAZymes. Indeed, the CAZyme complement of *C. pinensis* ranks highly among other well-endowed Bacteroidetes from gut microbiota, which likewise must address a diversity of complex plant glycans [13].

In the present study, we explored this catalytic potential by surveying the growth of *C. pinensis* on a panel of purified polysaccharide substrates to reveal that this bacterium is in fact a prolific degrader of plant -derived glycans. Due to our continuing interest in the evolution of glycosidase diversity [5, 14–17], and in particular specificity and mechanism in glycoside hydrolase family 27 (GH27) [18, 19], we fully characterised the product of the sole GH27 gene in *C. pinensis*, Cpin_5740, secretion of which is induced during growth on arabinogalactan. Biochemical analysis of the recombinant wild-type enzyme, henceforth referred to as *Cp*Arap27 [20], and site-directed mutants showed that this enzyme is exquisitely specific for β-L-arabinopyranosides, *vis-a-vis* the well-known specificity of GH27 members for *galacto*-configured substrates. Building on previous work [21, 22], we incorporate these new data into an updated phylogenetic analysis of GH27 that includes the more recent discovery of β-L-arabinopyranosidase activity in the family, and assess the reliability of conserved secondary structural elements as predictors of enzyme activity.

Experimental Procedures

Carbohydrates

The 4-nitrophenyl glycoside substrates used in this study (Gal-α-PNP and L-Ara*p*-β-PNP) were purchased from Sigma-Aldrich. The polysaccharides beech wood xylan and Gum Arabic (acacia) were also purchased from Sigma-Aldrich. Xyloglucan (tamarind seed), sugar beet arabinan, barley β-glucan, wheat arabinoxylan, larch arabinogalactan, konjac glucomannan, carob galactomannan and guar bean galactomannan were purchased from Megazyme.

Strain growth

All reagents were purchased from Sigma-Aldrich, unless otherwise stated, and were of microbiological grade. A lyophilised pellet of cultured *Chitinophaga pinensis* strain UQM 2034^T was purchased from DSMZ (designated DSM–2588), and propagated on LB media plates supplemented with kanamycin at 50 μg mL^{-1}, to which the bacterium has innate immunity [8].

Samples (100 μl) from starter cultures (10 mL) grown in LB were initially inoculated into M9 minimal media (15 mL) containing glucose (0.5%) or an alternative carbohydrate as a carbon source (0.5%) to study growth on a range of polysaccharides [23]. Samples were taken at regular intervals and OD readings (A$_{600}$) used as an indirect measure of cell density according to a protocol employed for other filamentous Bacteroidetes [24]. Samples of some cultures were also retained for later analysis of carbohydrate structures by HPAEC-PAD (see below). Minimal medium not supplemented with carbohydrate served as a control, and supported no

growth. Cultures were incubated at 30°C with rotary shaking at 180 rpm. For subsequent analyses of proteins produced during growth, cultures of 50 mL were inoculated (three biological replicates).

Preparation of protein fractions from cultures of *C. pinensis*

The total secretome of a culture was collected by centrifugation at 6,000 *g* for 15 minutes at 4°C to pellet cells. Based on initial growth curve analyses, secretomes were harvested approximately at late exponential phase; specifically, this was day 7–14, depending on the growth rate for each carbon source. For assays of secreted activity, secretomes (50 mL) were filtered (0.25 μm) (Nalgene, USA), then concentrated around 10 times, desalted and washed several times into dH_2O using 5 kDa cut-off Amicon Ultra centrifugal filters (Millipore).

Periplasmic proteins were collected using an osmotic shock method outlined in Larsbink *et al.*, 2011 [25]. Briefly, cells were washed with 10 mL of 50 mM Tris-HCl (pH 7.7) and collected by centrifugation at 4400 *g* for 10 min at 4°C; the media contained secreted proteins. The pellet was resuspended in 50 mL of 30 mM Tris-HCl, 20% (w/v) sucrose and 1 mM EDTA (pH 8.0), and the cells were incubated at room temperature for 10 min. The cells were then collected by centrifugation at 4400 *g* for 15 min at 4°C. Ice-cold 5 mM $MgSO_4$ (50 mL) was added and the cells were incubated on ice for 10 min. The cells were again collected by centrifugation at 14000 *g* for 10 min at 4°C and the supernatant also retained.

The cell pellet was resuspended in 50 mM sodium phosphate buffer (pH 7.4) and sonicated to lyse cells. The lysate was centrifuged at 5000 *g* for 10 min at 4°C to remove debris. Using an ultracentrifuge, the supernatant fluid was centrifuged at 100000 *g* for 1 h at 4°C. The supernatant liquid from this round of centrifugation contained soluble proteins and was retained. The pellet, containing membrane-bound and membrane-associated proteins, was resuspended in 100 mM sodium carbonate buffer (pH 9) to remove trapped soluble proteins and/or weakly membrane-associated proteins and centrifuged again at 100000 *g* for 1 h at 4°C to obtain integral membrane proteins. The supernatant fluid from this step again contained soluble proteins and/or weakly membrane-associated proteins. The final pellet, containing membrane proteins, was resuspended in 1 mL of 50 mM sodium phosphate buffer (pH 7.4).

Production of recombinant proteins

Cloning. The Cpin_5740 gene was amplified from genomic DNA by PCR and inserted into the vector pNIC28-Bsa4 by ligation independent cloning. The resulting expression construct contained a hexahistidine tag and a TEV-protease cleavage site (MHHHHHHSSGVD LGTENLYFQS) at the N-terminus. Correct in-frame insertion was confirmed by plasmid sequencing. Cloning was performed at the Karolinska Institutet/SciLifeLab Protein Science Facility (http://psf.ki.se).

Site-directed mutagenesis. Point mutations were introduced into the plasmid harbouring the Cpin_5740 gene by PCR amplification using the Pfx enzyme and buffer system (Life Technologies/Thermo Fisher Scientific). Table A in S1 File describes the primers used to generate each mutant plasmid. The PCR program utilised in each case was as follows: 94°C 5 minutes, 22 cycles of [94°C 30 seconds, 55°C 1 minute, 68°C 6 minutes], 68°C 15 minutes. Following Dpn1 treatment to degrade methylated parental plasmid DNA and subsequent PCR Clean-Up (Qiagen), mutated plasmids were transformed into OneShot Top10 cells (Life Technologies) by heat shock at 42°C for 30 seconds. Plasmid sequences were confirmed to contain the desired mutation by sequencing, performed by Eurofins Genomics, Germany.

Expression and purification. Plasmids containing the wild-type and mutant Cpin_5740 genes were transformed into *E. coli* BL21 (DE3) (Life Technologies) cells by heat shock at 42°C

for 30 seconds. The cells were grown at 37°C with shaking in LB medium containing kanamycin (50 μg mL^{-1}), to an OD$_{600}$ of 0.4–0.6, at which point protein expression was induced by addition of 0.2 mM IPTG (isopropyl-D-galactopyranoside) and the temperature was lowered to 25°C. Protein expression continued for 2 days, after which the cells were collected by centrifugation at 4000 *g* for 10 minutes. The cells were resuspended in Buffer A (20 mM sodium phosphate pH 7.4, 500 mM sodium chloride, 20 mM imidazole) and lysed by sonication, followed by centrifugation at 17000 *g* for 30 minutes. The supernatant liquid was loaded onto 5 mL HiTrap IMAC FF columns (GE Healthcare) using an ÄKTA FPLC system (GE Healthcare Life Sciences) and washed thoroughly with Buffer A. Each protein variant was purified on a separate, unused column. His-tagged proteins were eluted using a linear gradient of 0–100% Buffer B (20 mM sodium phosphate pH 7.4, 500 mM sodium chloride, 500 mM imidazole) over typically 4 column volumes. Eluted proteins were concentrated and exchanged into 50mM sodium phosphate pH 7.4 using Amicon Ultra centrifugal filters (Millipore). Liquid chromatography electrospray ionisation MS was used to verify the correct molecular weight of purified proteins [26]. Each mutated variant of the protein was purified separately on virgin resin to avoid any cross-contamination [27].

Size exclusion chromatography (SEC). An ÄKTA FPLC system was used to assess the apparent molecular mass of the Cpin_5740 gene product in solution by SEC on a Sephacryl S–300 HR (750 ml) column (GE Healthcare Life Sciences). Protein was loaded onto the column at 2 g L^{-1}, with a volume of 2 mL, and eluted with 50 mM Tris–HCl buffer pH 7.0, 100 mM NaCl with a flow rate of 0.4 mL min^{-1}. The void volume of the column was determined to be 102 mL using blue dextran. The molecular mass of the protein was assessed by comparing the elution volume with that of a series of standard proteins of known molecular weight in the range of 6.5 kDa to 66 kDa (Sigma Aldrich product code MWGF70).

Enzyme activity assays

PNP-glycoside assays. Assays in which the PNP-glycosides D-Gal-α-PNP and L-Ara*p*-β-PNP were used as substrates were monitored for the release of 4-nitrophenolate at A_{410}, using a Cary 50 spectrophotometer. For an initial activity screen of *C. pinensis* secretomes, stopped assays were performed: substrate at 2 mM was incubated with 100μl of concentrated secretome for 2 hours at 30°C in 50 mM sodium phosphate buffer, pH 7. Following incubation, an equal volume of 200 mM Na$_2$CO$_3$ was added to terminate the reactions by raising the pH to 11. An extinction coefficient of 18500 M^{-1} cm^{-1} was used to calculate product concentration from absorbance values [28].

A stopped assay was also used to determine the optimum pH and temperature conditions for the enzyme. Substrate (2 mM) was incubated with enzyme (125 nM) in a total reaction volume of 200 0 mine the optimum pH and temperature conditions for the enzyme. Substrate (2 mM) was incubated with formate, sodium acetate, sodium succinate, HEPES, sodium phosphate, glycyl glycine, and glycine, over a pH range from 2.5 to 10.0. The optimum pH for the wild-type enzyme was determined to be pH 5.0 (50 mM sodium citrate buffer) using the pNP-β-L-Ara*p* substrate (*vide infra*, Results and Discussion). This buffer was used to perform the same reaction at a range of temperatures, and the optimum was found to be 30°C (*vide infra*, Results and Discussion). These conditions of pH and temperature were used for all subsequent kinetic analyses of all enzyme variants acting on pNP-β-L-Ara*p*, pNP-α-D-Gal*p* and larch arabinogalactan.

For kinetic analyses of hydrolysis of PNP-glycosides by pure enzyme, a continuous assay was used in a Cary 300 spectrophotometer. A standard curve for pNP in 50 mM sodium citrate buffer, pH 5.0 gave the extinction coefficient 1415 M^{-1}cm^{-1}, which was used to calculate

product concentration from absorbance values. The range of substrate concentrations utilised in kinetic analysis was 0–25 mM for pNP-β-ʟ-Ara*p*, and 0–40 mM for pNP-α-ᴅ-Gal*p*. A control experiment without enzyme was performed for each rate analysis, to account for spontaneous substrate hydrolysis. Kinetic experiments were performed in 50 mM sodium citrate buffer at pH 5.0 and 30°C. All quantitative assays of enzyme activity were performed in triplicate.

Assay for the specific detection of arabinose release. A linked galactose dehydrogenase/galactose mutarotase assay kit (Megazyme product code E-GALMUT) was used to quantify the release of arabinose from arabinogalactan [29]. The release of arabinose led to the stoichiometric reduction of NAD$^+$ to NADH, giving an increase in A_{340} (ε 6230 M^{-1} cm^{-1} at pH 7 [30]), which was read continuously using a Cary 300 spectrophotometer. Kinetic experiments were performed in 50 mM sodium citrate buffer at pH 5.0 and 30°C. The range of substrate concentrations used in kinetic analysis was 0–160 mg mL^{-1}. A control experiment without enzyme was performed for each rate analysis, to account for spontaneous substrate hydrolysis. All quantitative assays of enzyme activity were performed in triplicate.

Enzyme product profiles. Reaction products were analysed on a Dionex ICS-3000 HPLC system operated by Chromeleon software version 6.80 (Dionex) using a Dionex Carbopac PA1 column. Solvent A was water, solvent B was 1 M sodium hydroxide and solvent C was 200 mM NaOH with 170 mM Na acetate. The following programme was employed: pre-wash and column calibration -14–-7 min 60% B, 40% C (1 mL min^{-1}); -6–0 min 100% A (1 mL min^{-1}); sample injection 0–5 min 100% A (0.5 mL min^{-1}); gradient elution 5–20 min 0–30% B (0.5 mL min^{-1}).

To analyse the products of polysaccharide degradation by secretomes, 1 mL assays were prepared using substrate at 1 mg mL^{-1}, which was incubated with ~400 μg mL^{-1} protein for up to 4 days at 30°C in 50 mM sodium citrate buffer, pH7, prior to analysis by HPAEC-PAD. Samples (200 μl) were also taken from cultures during growth on a range of carbohydrates. These samples were boiled to stop all enzyme reactions, concentrated by lyophilisation, resuspended in 50 μl of water, and analysed by HPAEC-PAD to observe oligosaccharide production and degradation during growth. Finally, pure protein (100 nM) was incubated with polysaccharide at 1 mg mL^{-1}, for 16 hours at 30°C in 50 mM sodium citrate buffer, pH 5.0, prior to analysis by HPAEC-PAD.

Identification of the Cpin_5740 gene product in the native secretome

Generation and purification of antibodies specific for the Cpin_5740 gene product. For eventual use in a Western blot intended to probe for the presence of the protein, antibodies were raised in rabbits against the recombinant Cpin_5740 gene product (AgriSera AB, Vännäs, Sweden). Pre-sera from rabbits to be used in antibody generation were screened by Western blot for natural antibodies to the protein of interest, and were determined not to be reactive to the Cpin_5740 gene product. The immunisation procedure was as follows; Immunisation 1: 200 μg of antigen (1.25 mg mL^{-1}) and FCA (Freund's complete adjuvant); Immunisations 2, 3 and 4 (respectively 1, 2 and 3 months later): 100 μg of antigen (1.25 mg mL^{-1}) and FCI (Freund's incomplete adjuvant). The final bleed was performed 10 days after the final immunisation.

Polyclonal antibodies were purified from the final serum by affinity purification at Agrisera. Briefly, the recombinant Cpin_5740 gene product protein was first coupled to a 1 mL HiTrap NHS-activated HP Column (GE Healthcare) according to the manufacturer's instructions. The column was washed using an ÄKTA Prime system (GE Healthcare) with several column volumes of PBS pH 7.4. Two mL of 10×PBS were added to 20 mL of antiserum and this solution was applied to the recombinant protein-coupled HiTrap column. The column was washed with several column volumes of PBS. Antibodies bound to the column were eluted with 200

mM glycine in 1.0 mL fractions, into tubes containing 50 μl of 1 M Tris to neutralise the eluent. Fractions with A_{280}>0.1 were pooled and precipitated with saturated ammonium sulphate overnight. The solution was then centrifuged at 5000 *g* for 15 minutes and the pellet was dissolved in 1 x PBS pH 7.4. Traces of ammonium sulphate were removed using PD10 columns from GE Healthcare. The concentration of the purified antibody was 2.7 g L^{-1}.

Western blot analysis of the *C. pinensis* secretome. A Western blot analysis was performed to probe for the Cpin_5740 gene product in native *C. pinensis* secretomes. As described above, total secretomes were collected from 50 mL liquid cultures of *C. pinensis* by centrifugation to pellet cells. Secretomes were concentrated approximately 50 times and utilised in Western blots. A total of 100 μg of each secretome was loaded and run on SDS-PAGE. TBST buffer (tris-buffered saline (50 mM Tris-HCl, 150 mM NaCl, pH 7.4) with 0.1% Tween–20) was used for washing and dilutions throughout the Western blot protocol. The membrane was washed 5–6 times (~5 minutes each) between each step in the following procedure. After blotting proteins from an SDS-PAGE gel, the membrane was first blocked with a solution of TBST buffer + 3% BSA for 1 hour at room temperature to reduce non-specific binding. The purified antibodies were used as the primary antibody (1:1000 dilution, 1 hour incubation at room temperature with 1% BSA), with anti-rabbit IgG coupled to HRP (Sigma Aldrich) as the secondary antibody (1:10,000, 1 hour incubation at room temperature with 1% BSA). Final visualisation was achieved by chemiluminescence using the luminol-based Amersham ECL Western Blotting Detection Reagent (GE Healthcare) and a Fujifilm Intelligent Dark Box with LAS–1000 camera and software.

Bioinformatics

A sequence alignment of the GH27 catalytic domains of 71 protein sequences (49 characterised and 22 uncharacterised proteins, identified by BLAST searching of GH27 enzymes with known activity and/or structure) was performed using the online Clustal Omega server [31]. Table B in S1 File provides details of the uncharacterised proteins included in the analysis. Full sequences of the catalytic domains of these proteins were used for an initial sequence alignment, which revealed the presence in *Cp*Arap27 of a number of loop insertions in conserved positions, comprising between 6 and 34 amino acids. For subsequent sequence alignments which were used to generate a phylogenetic tree, these loops were removed from the sequences, as in previous phylogenetic analyses of the family [21]. The software PhyML was used to produce a phylogenetic tree of the alignment results using the Blosum62 model of amino acid substitution, with bootstrapping of results (100 replicates) [32]. The software MEGA6 was used to view the tree [33] and Adobe Illustrator CS5 was used to produce the final figure.

Results and Discussion

Chitinophaga pinensis is capable of growth on complex polysaccharides

Chitinophaga pinensis was screened for the ability to grow on a wide range of soluble carbohydrates of diverse structural complexity (Fig 1). In common with previous observations, growth on starch was very poor [8, 11]. As shown in Fig 1, stronger growth was sometimes observed on complex polysaccharides than on the equivalent monosaccharides. This observation has also been made for other *Bacteroidetes* species [24, 34]. Those polysaccharides supporting the greatest growth were konjac glucomannan and arabinan. Less effective growth substrates were xylans (wheat arabinoxylan and beech wood glucuronoxylan), galactomannans (from carob and guar seeds), xyloglucan and arabinogalactan. Very low levels of growth were also observed on barley β-glucan and gum arabic, but this was too weak to be accurately measured. However, it should also be noted that the liquid culture conditions may not fully reflect the native growth conditions in solid forest litter.
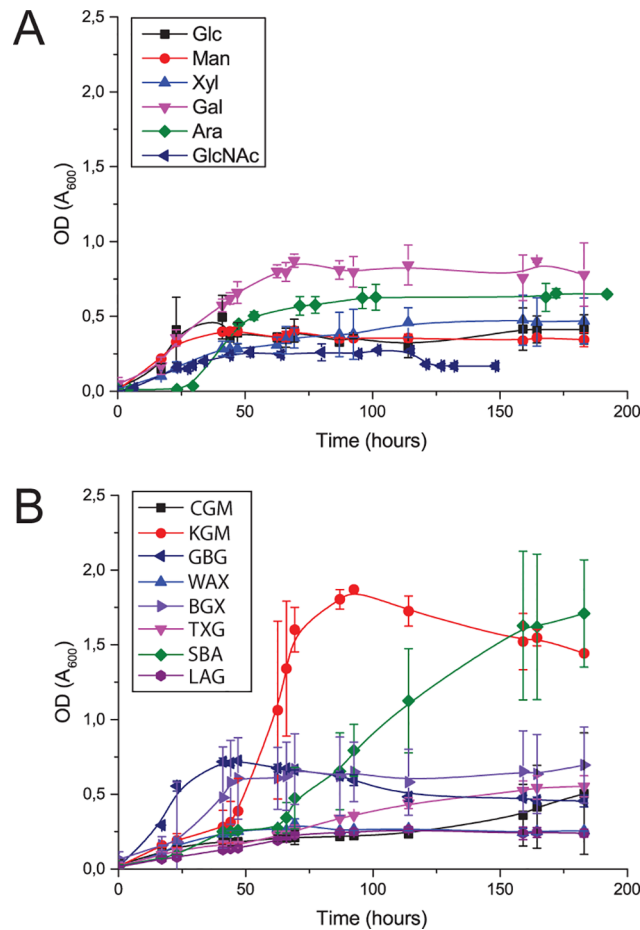
Fig 1. Growth curves for *C. pinensis* on a range of soluble carbohydrates. Growth was determined by measuring OD_{600} of samples taken at regular intervals from triplicate 15 mL cultures of *C. pinensis* in M9 media supplemented with various carbohydrates [24]. Growth curves for several monosaccharides (A) and polysaccharides (B) are shown. Large errors on some data points are ascribed to turbidity effects resulting from the filamentous growth habit of the bacterium; however, no visible clumps were observed during the measurement of OD_{600} values. Abbreviations for the monosaccharides are as follows: Glc, glucose; Man, mannose; Xyl, xylose; Gal, galactose; Ara, arabinose; GlcNAc, *N*-acetylglucosamine. Abbreviations for the polysaccharides are as follows: CGM: carob galactomannan. KGM: konjac glucomannan. GBG: guar bean galactomannan. WAX: wheat arabinoxylan. BGX: beechwood glucuronoxylan. TXG: tamarind xyloglucan. SBA: sugar beet arabinan. LAG: larchwood arabinogalactan.

doi:10.1371/journal.pone.0139932.g001

β-L-arabinopyranosidase activity is secreted during growth on arabinogalactan

Sparked by an interest in galactomannan utilisation, we initially identified locus Cpin_5740, the sole GH27-encoding gene in *C. pinensis*, as an enzyme of interest due to the predominance of α-galactosidases in this GH family. However, recombinant expression of Cpin_5740 in *E. coli* subsequently revealed this enzyme to be a strict β-L-arabinopyranosidase (*vide infra*). We therefore expanded our initial screen to focus on carbohydrates containing the β-L-Ara*p* structure. There is evidence that β-linked L-Ara*p* residues are found in Type I and Type II arabinogalactans, although they seem to be relatively rare compared to L-Ara*f* residues [35–37] (Fig 2). Other polysaccharides which may contain small amounts of β-L-Ara*p* residues are gum arabic and sugar beet arabinan [38–42]. Therefore, the ability of *C. pinensis* to grow on the branched arabinose-containing polysaccharides larch arabinogalactan [35–37, 43], sugar beet arabinan
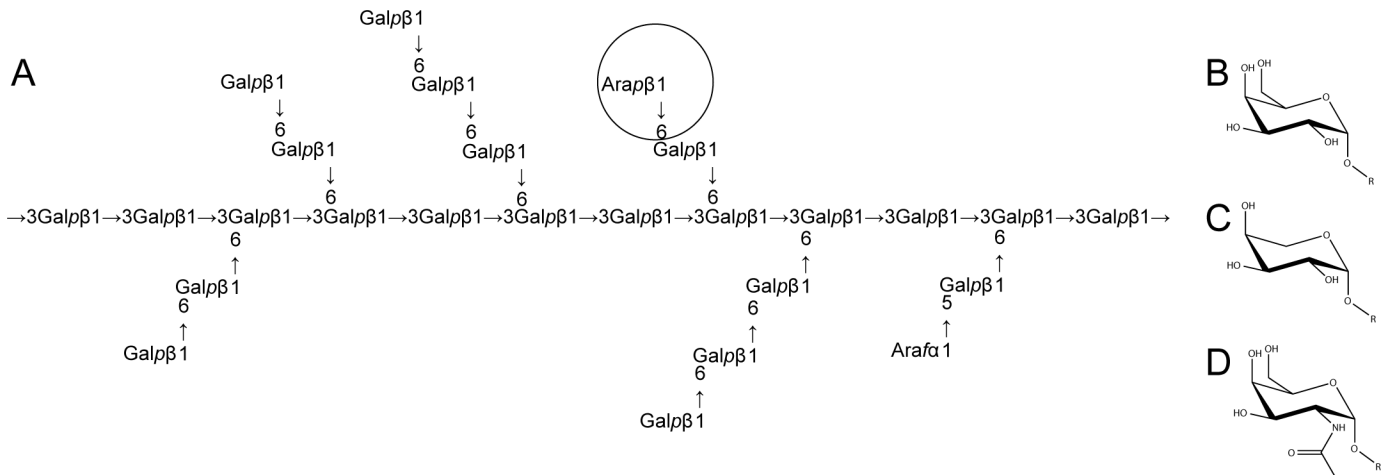
Fig 2. Carbohydrates known to be substrates for GH27 enzymes. (A) The complex structure of larch wood arabinogalactan comprises a backbone of β-1,3-Gal*p* residues, with β-1,6-Gal*p* decorations and side chains, as well as some terminal Ara*f* and β-L-Ara*p* (circled) substitutions. (B-D) The monosaccharide substrates of characterised GH27 enzymes are, respectively, α-D-Gal*p*, β-L-Ara*p*, and *N*-acetylgalactosamine. The complexities of carbohydrate nomenclature obfuscate the fact that β-L-Ara*p* and α-D-Gal*p* differ only in the absence or presence, respectively, of a hydroxymethyl group on the C5 position of the sugar ring [48, 49].

doi:10.1371/journal.pone.0139932.g002

[40, 44, 45], and gum arabic [46, 47], as well as their component monosaccharides, was analysed in more detail. Growth on the simple substrate glucose was also analysed, as a control experiment. Despite a comparable doubling rate during the exponential phase, growth on arabinogalactan was poor compared to growth on the constituent monosaccharides arabinose and galactose, with a longer lag phase and lower final OD (Fig 1). The bacterium was able to grow only very slightly on the structurally similar gum arabic, which may reflect a paucity of hydrolytic enzymes for this polysaccharide, including the apparent inability of the Cpin_5740 gene product to release monosaccharides from it (*vide infra*). Growth on sugar beet arabinan was strong and conformed more closely to the aforementioned pattern observed for other *Bacteroidetes* species where growth on polysaccharide is stronger than growth on simpler carbohydrates [24, 34]. During growth on arabinogalactan, the medium was sampled regularly for analysis by HPAEC-PAD, which showed no *endo*-hydrolysis of the polysaccharide.

To further probe the behaviour of the bacterium under these conditions, proteins produced during growth on galactose-containing polysaccharides were tested for hydrolysis of the chromogenic *exo*-glycosidase substrates pNP-β-L-Ara*p* and pNP-α-D-Gal*p*, which revealed a significant induction of β-L-arabinopyranosidase activity during growth on arabinogalactan (Fig 3A). Moreover, this activity was predominantly localised to the secretome, versus the periplasm, cytosol, and cellular membranes (Fig 3B). Incubation of cell-free secretomes from glucose-grown control cultures and arabinogalactan-grown cultures with arabinogalactan as an assay substrate revealed that only the latter was able to release a very small amount of arabinose from the polysaccharide, in which the β-L-Ara*p* structure is quite rare [35–37].

Locus Cpin_5740 encodes a secreted β-L-arabinopyranosidase induced by arabinogalactan

GH27 is a family of retaining enzymes notably containing α-D-galactosidases and some α-*N*-acetylgalactosaminidases [12], which recently has seen an expansion of its catalytic repertoire to include β-L-arabinopyranosidases [41, 42, 50, 51]. Locus Cpin_5740 encodes the sole member of GH27 in the *C. pinensis* genome [11], which we predicted to be an extracellular protein

due to the presence of an SpI signal peptide [52]. We therefore identified this gene as a likely candidate to encode the extracellular β-ʟ-arabinopyranosidase activity identified in the secretome analysis. Cpin_5740 is located on the chromosome adjacent to a gene (Cpin_5739) encoding a GH51 member, a predicted α-ʟ-arabinofuranosidase that may also target arabinogalactan (Fig 2). Indeed, independent assays also detected hydrolysis of pNP-α- ʟ-Ara*f* in the arabinogalactan-induced secretome (data not shown). No other predicted CAZyme-encoding genes are located nearby, but the *C. pinensis* genome does encode other enzymes likely to be involved in degradation of arabinogalactan, including potential β-galactosidases (members of families GH1, 2, 16, 35, and 43), α-ʟ-arabinofuranosidases (families GH2, 43 and 51), and galactanases (families GH5, 16, 30, 35 and 53) [11, 12]. In this context, it is interesting to note that *C. pinensis*, a member of the phylum Bacteroidetes, does not appear to co-locate carbohydrate-active enzymes and carbohydrate-binding proteins into Polysaccharide Utilisation Loci common in *Bacteroides* species and some other Bacteroidetes [53–55].

The cloned Cpin_5740 gene expressed well in *E. coli*, typically yielding 10–30 mg protein from a 1 L culture. The hexahistidine-tagged, recombinant protein (henceforth referred to as *Cp*Arap27) was readily purified using immobilised metal affinity chromatography (IMAC) (Fig A in S1 File). Analysis by size-exclusion chromatography (SEC) indicated that *Cp*Arap27 is monomeric in solution (Fig B in S1 File). Western blot analysis using rabbit antibodies raised against this recombinant protein confirmed that secretion of *Cp*Arap27 is indeed induced during growth on arabinogalactan but not glucose (Fig 4).

The purified recombinant enzyme was subjected to an activity screen on artificial, chromogenic pNP substrates (see full list in Experimental Procedures) and was found to be strictly specific for pNP-β-ʟ-Ara*p* (Table 1, full v_o vs. [S] plots are given in Fig D, S1 File). Using pNP-β-ʟ-Ara*p*, optimum conditions of pH and temperature for the enzyme were determined to be pH 5.0 and 30°C (Fig D, S1 File). The observation of strict pNP-β-ʟ-Ara*p* activity contrasts with other characterised β-ʟ-arabinopyranosidases, which have shown low activity against the structurally similar pNP-α-ᴅ-Gal*p* (Table 2) [41, 42, 51, 56, 57]. HPAEC-PAD analysis
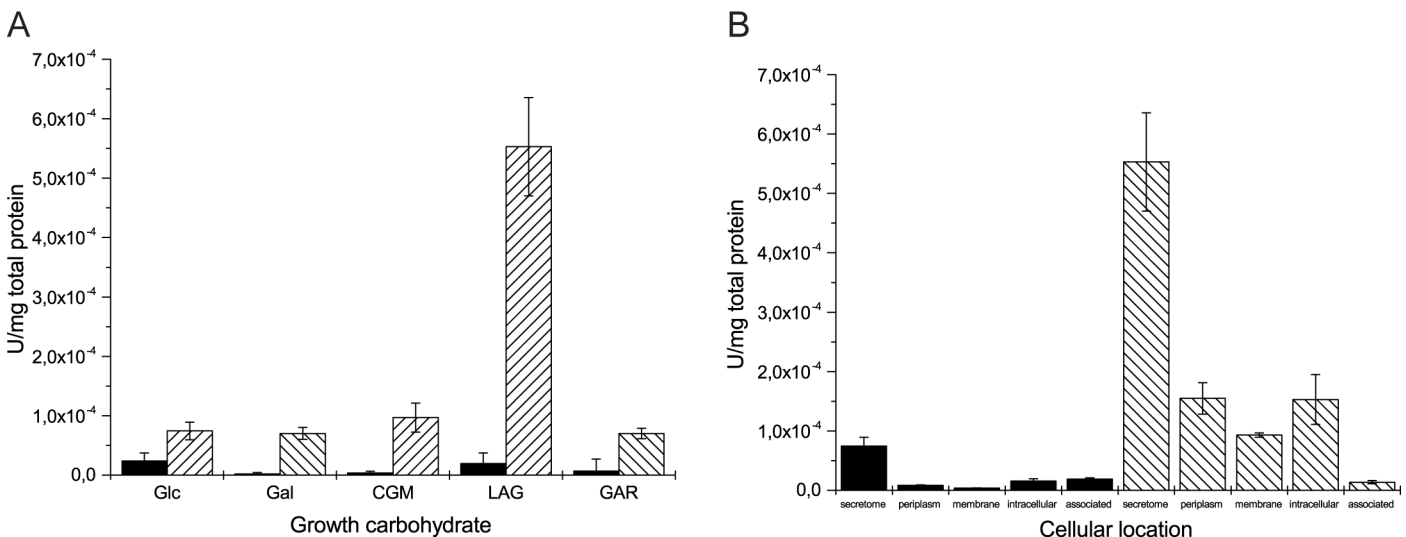
Fig 3. Growth on arabinogalactan induces the secretion of pNP-β-ʟ-Arabinopyranosidase activity. (A) Hydrolysis of pNP-α-ᴅ-Gal*p* (solid) and pNP-β-ʟ-Ara*p* (striped) by secretomes induced by different growth carbohydrates. (B) pNP-β-ʟ-arabinopyranosidase activity is more strongly enriched in the secreted protein fraction than other cellular locations, and is much more significant in cultures induced by arabinogalactan (striped) compared to cultures induced by glucose (solid). The term 'associated' denotes proteins which are weakly associated with cell membranes. Abbreviations are as follows: Glc, glucose; Gal, galactose; CGM, carob galactomannan; LAG, larch wood arabinogalactan; GAR, gum arabic.

doi:10.1371/journal.pone.0139932.g003
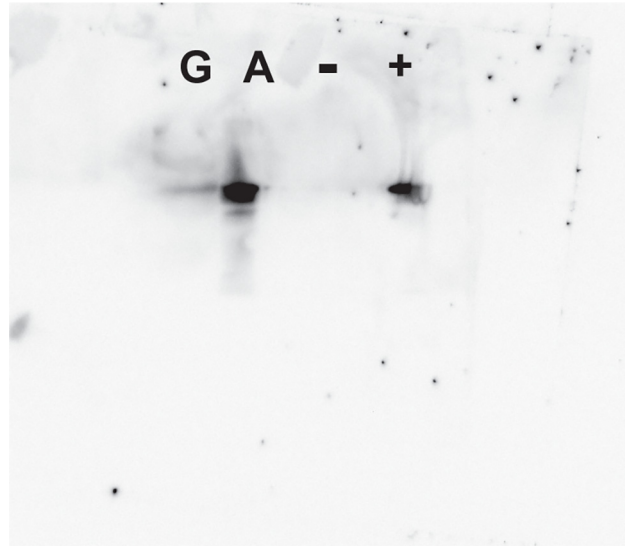
Fig 4. Western blot evidence that *Cp*Arap27 is secreted during growth on arabinogalactan. Western blot analysis of *C. pinensis* secretomes induced by growth on glucose and arabinogalactan. A band is visible which corresponds to the ~48 kDa *Cp*Arap27 protein in the arabinogalactan secretome. Protein samples on the blot are labelled as follows; G, glucose-induced secretome; A, arabinogalactan-induced secretome; -, BSA (negative control); +, pure recombinant His$_6$-tagged *Cp*Arap27 (positive control). The PageRuler Plus protein ladder (Life technologies) was used to assess that the visible band was of the correct size. A total of 100 μg of protein was loaded for each of the secretomes.

doi:10.1371/journal.pone.0139932.g004

demonstrated that *Cp*Arap27 is able to release arabinose from larch arabinogalactan as the sole reaction product (Fig C, S1 File). Using a linked assay to measure arabinose release, kinetic analysis of this reaction was performed for wild-type *Cp*Arap27 (Table 1, and Fig D, S1 File). The high K_m for this reaction, which is derived from the polysaccharide concentrations in the assays, likely reflects the low abundance of the target structure in arabinogalactan [35, 58, 59] (Fig 2). No arabinose release was detected when *Cp*Arap27 was assayed against the arabinose-containing polysaccharides sugar beet arabinan and gum arabic, either because the β-ʟ-Ara*p* structure was of too low abundance or sterically inaccessible to *Cp*Arap27 in these substrates.

Structural determinants of specificity in family GH27

A key question regarding *Cp*Arap27 in the context of GH27 is which active site features of the enzyme determine specificity for the β-ʟ-Ara*p* substrate over the similar α-ᴅ-Gal*p* substrate. To explore this, we performed sequence alignments with previously characterised GH27 enzymes with differing substrate specificities, and subsequently produced a homology model of *Cp*Arap27 (*vide infra*). As the sequence alignment in Fig 5 shows, likely candidates for the two catalytic Asp residues of *Cp*Arap27 were identified as Asp187 and Asp242. Individual site-directed alanine mutants of these residues had drastically reduced activity against pNP substrates compared to the wild-type enzyme (Table 1, and Fig D, S1 File) [27].

Geobacillus stearothermophilus Abp, another GH27 enzyme, was recently described with the ability to hydrolyse pNP-β-ʟ-Ara*p* and remove arabinose from arabinogalactan, with limited activities against pNP-α-ᴅ-Gal*p* and pNP-α-ʟ-Ara*f* [42]. The importance of the residue Ile67 in this enzyme has recently been demonstrated by structural determination and site-directed mutation; a crystal structure of *Gs*Abp in complex with ʟ-Ara suggests that a steric clash would occur between Ile67 and galactopyranosides in the active site pocket [51]. Sequence analysis (Fig 5) shows that *Cp*Arap27 possesses a homologous isoleucine (Ile56), a

Table 1. Kinetic parameters for wild-type and mutant forms of *CpArap27* against artificial and natural substrates.

Enzyme	pNP-β-L-arabinopyranoside			pNP-α-D-galactopyranoside			Larch wood arabinogalactan		
	k_{cat} (min^{-1})	K_m (mM)	k_{cat} / K_m	k_{cat} (min^{-1})	K_m (mM)	k_{cat} / K_m	k_{cat} (min^{-1})	K_m (mg mL^{-1})	k_{cat} / K_m
Wild-type	1940 ± 215	12.9 ± 3.27	150	0.0124 ± 2.52 × 10^{-3}	NA	3.51 × 10^{-4}	1580 ± 135	18.0 ± 4.46	87.5
I56D	0.431 ± 0.0465	8.63 ± 2.14	0.0499	6.31 ± 3.11	35.3 ± 11.9	0.177	160.0 ± 20.6	36.0 ± 12.9	4.44
I56E	31.7 ± 6.88	6.56 ± 3.39	4.83	ND[a]	35.6 ± 15.2	2.45 ± 0.2	7.95 ± 1.11	21.6 ± 8.56	0.368
I56C	24.2 ± 7.14	21.5 ± 10.4	1.13	ND	ND	0.539 ± 0.0176		NA[b]	
I56A	ND	ND	0.372 ± 0.0561		ND			NA	
D187A	ND	ND	2.35 × 10^{-3} ± 5.18 × 10^{-4}		NA			NA	
D242A	ND	ND	5.44 × 10^{-3} ± 3.78 × 10^{-4}		NA			NA	

Kinetic parameters are provided for wild-type and variant forms of *CpArap27* acting on three different substrates. Throughout, values are given to 3 significant figures, with standard errors of means also provided.

[a] ND indicates a kinetic parameter which could not be determined. For reactions for which individual k_{cat} and K_m values are not given, values for k_{cat} / K_m were determined from linear curve fitting of initial rate data, at low substrate concentrations (Fig D, S1 File). Otherwise, values for k_{cat} and K_m were determined using Origin 9.1.

[b] NA denotes that no activity against a particular substrate was detected.

doi:10.1371/journal.pone.0139932.t001

Table 2. Summary of specificities of several characterised GH27 enzymes, and motivation for the mutations of *Cp*Arap27.

*Cp*Arap27 variant form		Equivalent previously characterised enzyme		
Enzyme	Ara:Gal ratio [a]	Enzyme	k_{cat} / K_m	Ara:Gal ratio [a]
Wild-type	Ara only	*Gs*Abp	Ara:495 [42]	Ara (trace pNP-α-D-Gal*p*, pNP-α-L-Ara*f*)
I56E	Ara only	*Sa*Arap27A	Ara:88.1 Gal:0.5 [41]	196:1
I56D	142:1	*Sca*Gal	Gal:63.6 [21]	Gal only
I56D	142:1	*Sa*Arap27A E99D	Ara:1.5 Gal:6.7 [41]	1:4.4
I56C	1:2.2	*Fo*Ap1	Ara:9.2 Gal: 9.9 [50]	1.1:1
I56C	1:2.2	*Fo*Ap2	Ara:8.1 Gal:0.6 [50]	13.6:1
I56A	1:1.4	*No examples*	-	-

Kinetic parameters for several GH27 enzymes, with different substrate specificities and different amino acids in a key active site position, are compared. Kinetic parameters for all forms of *Cp*Arap27 are given in Table 1. The Ara:Gal ratios were determined by dividing k_{cat} / K_m values and normalising the smaller value to 1.

[a] The Ara:Gal ratio of each enzyme or enzyme variant compares k_{cat} / K_m for hydrolysis of pNP-β-L-Ara*p* and pNP-α-D-Gal*p* respectively.

doi:10.1371/journal.pone.0139932.t002

series of loop insertions, and an overall 53% sequence identity vis-à-vis *Gs*Abp. A homology model of the *Cp*Arap27 structure was generated by the Phyre2 server, using the *Gs*Abp crystal structure as a template [51, 62]. In light of the high primary structural similarity between the template and the modelled protein, a high level of tertiary structural similarity was correspondingly observed.

Superimposing the *Cp*Arap27 model structure with L-Ara and D-Gal from ligand-bound structures of *Sa*Arap27A (the first GH27 shown to possess β-L-arabinopyranosidase activity) suggests that the side chain of Ile56 in *Cp*Arap27 would bias specificity towards arabinopyranose in the same manner as Ile67 in *Gs*Abp does (Fig 6D and 6E). The model also indicates that the loop insertions in *Cp*Arap27 and *Gs*Abp, identified in sequence alignments (Fig 5) and not present in *Sa*Arap27A, may be significant for activity, as they contribute to the architecture of the active site pocket.

In characterised GH27 enzymes, the amino acid position corresponding to Ile56 contains an Asp in α-D-galactopyranosidases [21], a Glu or Ile in β-L-arabinopyranosidases [41, 42], and a Cys in two catalytically flexible fungal enzymes, *Fo*Ap1 and *Fo*Ap2 [50]. Due to the apparent correlation between this residue and enzyme specificity (Table 2), we were intrigued by the possibility of engineering *Cp*Arap27 to reflect the other specificities displayed by members of family GH27. Ile56 was mutated to each of these alternate amino acids and the specificity and kinetic parameters of these variants were explored (Table 1, and Fig D, S1 File)

As predicted, mimicking the active site of typical GH27 α-D-galactosidases by generating the I56D variant form of *Cp*Arap27 did introduce hydrolytic activity toward pNP-α-D-Gal*p* (Table 1), as was previously demonstrated for *Gs*Abp [51] (Table 2). Similarly, the I56E variant showed some catalytic promiscuity (Table 1), as has been observed for *Sa*Arap27A [41] (Table 2). Both of these variants were catalytically feeble and in neither case was hydrolysis of the galactopyranoside substrate the most significant activity. For *Cp*Arap27 I56D, which has the Asp typical of GH27 α-D-galactosidases, and *Cp*Arap27 I56E, comparison of the k_{cat}/K_m values for hydrolysis of pNP-β-L-Ara*p* and pNP-α-D-Gal*p* indiciates Ara*p*:Gal*p* preferences of 12:1 and 39:1 (Ara*p*:Gal*p* for the wild-type enzyme = 1:0), respectively. The *Cp*Arap27 I56E mutant is a mimic of the wild-type *Sa*Arap27A (Ara:Gal ratio 67:1) and has similar levels of preference for the L-arabinopyranosyl substrate, but is catalytically much weaker [41]. Similarly, the I67D mutation of *Gs*Abp induced a 3-fold increase in hydrolysis of pNP-α-D-Gal*p*, as well as a 2.7-fold decrease in hydrolysis of pNP-β-L-Ara*p* [51]. Further, although *Sa*Arap27A
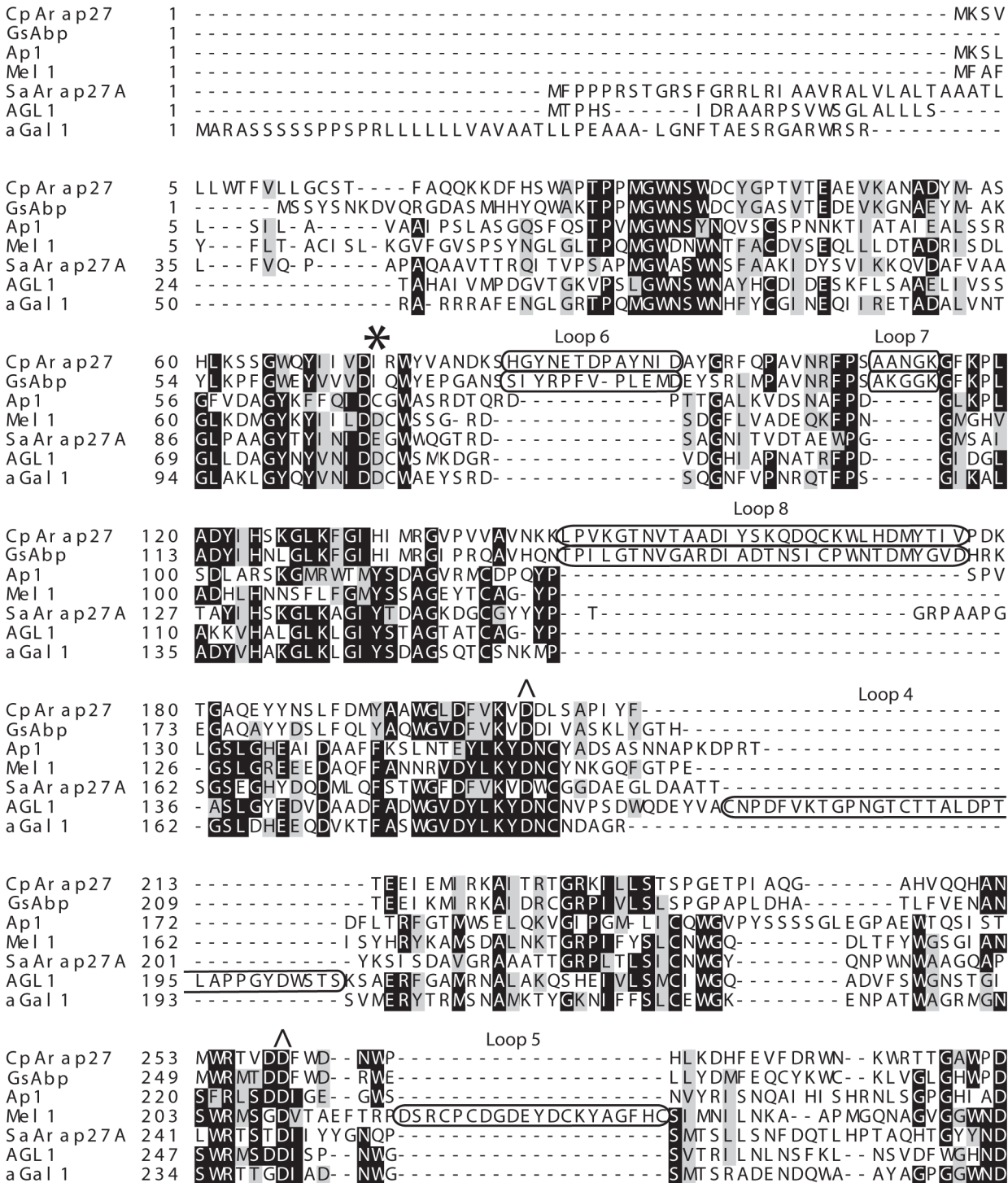
Fig 5. Sequence alignment of key variable GH27 regions. Sequence alignment of *Cp*Arap27 (*C. pinensis*), *Gs*Abp (*G. stearothermophilus*) [51], Ap1 (*F. oxysporium*) [50], Mel 1 (*S. cerevisiae*) [21], *Sa*Arap27A (*S. avermitilis*) [41], AGL1 (*T. reesei*) [60], and αGal1 (*O. sativa*) [61]. This alignment highlights variation of amino acid in a key position for specificity (indicated with an asterisk), and shows significant loop insertions in different GH27 enzymes, highlighted in boxes. Catalytic amino acids are indicated by a caret.

doi:10.1371/journal.pone.0139932.g005
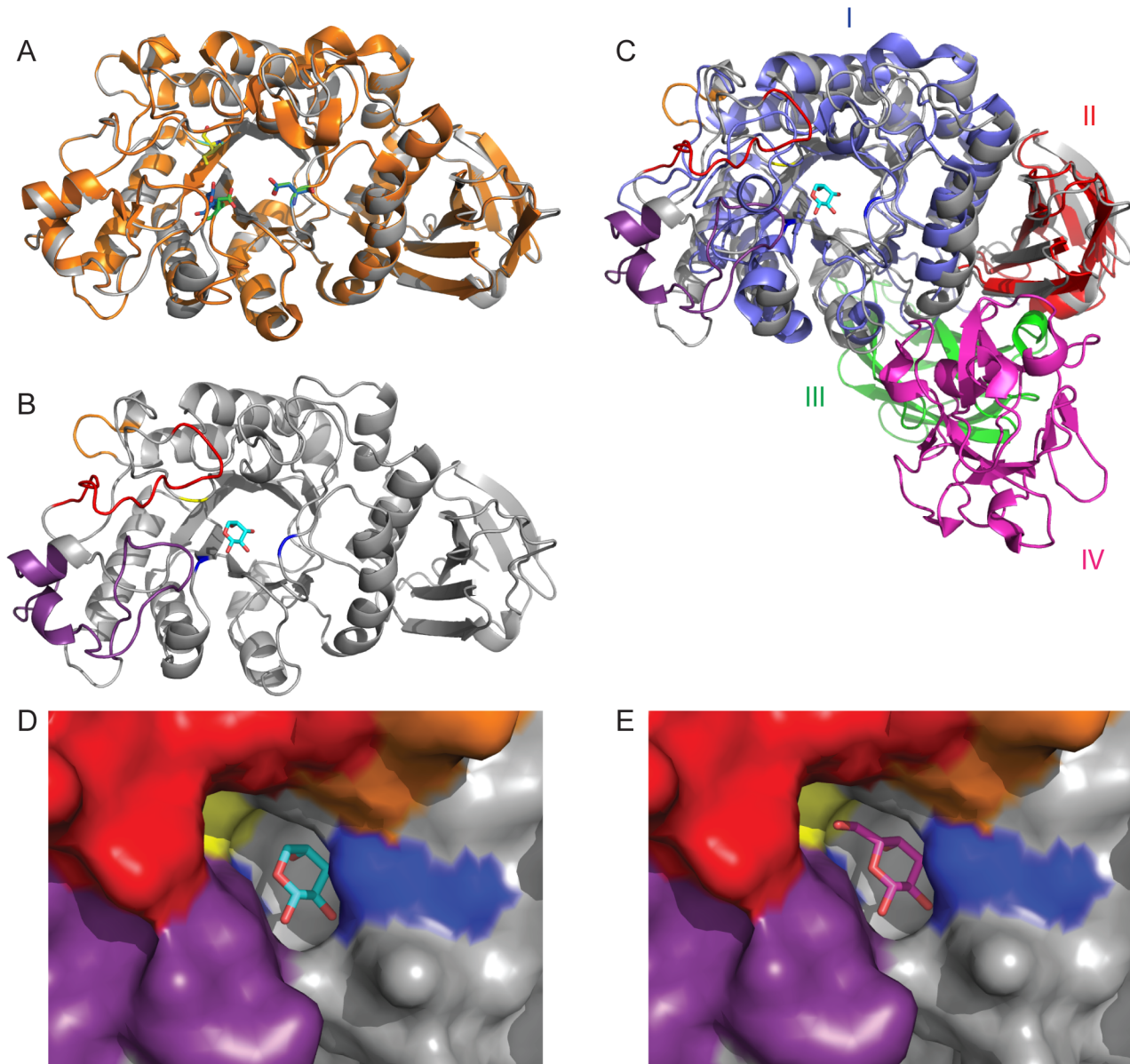
Fig 6. Structural comparison of *Cp*Arap27 with *Gs*Abp and *Sa*Arap27A. (A) Cartoon representation of the *Cp*Arap27 homology model (grey) and the crystal structure of *Gs*Abp (orange). Catalytic amino acids for the enzymes are picked out respectively in blue/green, while the key Isoleucine specificity determinant is yellow/cyan. (B) Secondary structural representation of *Cp*Arap27, with loop insertions highlighted as follows: L6 is red, L7 is orange, and L8 is purple. Ile56 is shown in stick form and highlighted in yellow. Catalytic amino acids are highlighted in blue. (C) Comparison of the modular structures of *Sa*Arap27A (Domains I-IV) and the homology model of *Cp*Arap27 (Domains I and II). *Sa*Arap27A is shown in 4 colours to highlight the 4 domains of this protein. *Cp*Arap27 is coloured as in panel B. A molecule of L-Ara is shown in the active site, taken from the structure of *Sa*Arap27A. (D) and (E) Surface representations of the active site of the *Cp*Arap27 homology model overlaid with, respectively, L-Ara and D-Gal from the *Sa*Arap27A structures. Catalytic amino acids are shown in blue, Ile56 is shown in yellow, and the loop insertions are again shown in purple, red and orange. The Isoleucine shown here is predicted to hydrophobically clash with a hypothetical D-Gal substrate, as shown for *Gs*Abp [51].

doi:10.1371/journal.pone.0139932.g006

was previously modified to prefer the galactoside by mutation of the key Glu to an Asp (Table 2), the resulting E99D mutant (Ara:Gal ratio 1:9) was also catalytically enfeebled in the wild-type activity, similar to the results obtained here for the *Cp*Arap27 I56D variant (Tables 1 & 2, and Fig D, S1 File). Notably, the *Cp*Arap27 I56A and I56C variants are able to hydrolyse

both pNP-α-D-Gal*p* and pNP-β-L-Ara*p* with roughly equal efficiency, but are nonetheless also poor catalysts (Table 1).

In all cases, despite alterations in the activities toward the artificial chromogenic glycosides, no gain-of-function for hydrolysis of arabinogalactan was observed for any of the *Cp*Arap27 variants. For the I56D and I56E variants, arabinose could be identified as the sole hydrolysis product of larch arabinogalactan, as was observed for the wild-type enzyme (Fig C, S1 File). Despite prolonged incubation with high enzyme concentration, no reaction products were detected by HPAEC-PAD for *Cp*Arap27 I56A, I56C or the catalytically inactive D187A and D242A. Neither the wild-type nor any of the mutant enzymes showed any activity on sugar beet arabinan, gum arabic, linear galactan or galactomannan polysaccharides.

Although all of the variants examined do have some ability to bind and hydrolyse the chromogenic galactopyranoside, it is clear that key enzyme-substrate interactions that enable efficient hydrolysis in natural GH27 α-D-galactopyranosidases still have not been fully accounted. As mentioned above, *Cp*Arap27 and *Gs*Abp also share other important structural features, including several inserted loop regions which are not present in *Sa*Arap27A or in α-D-galactosidases of the family. Loop insertions into the general (α/β)₈ barrel fold are known to significantly affect the substrate specificity and oligomerisation of GH27 enzymes [21, 22], and indeed all TIM-barrel containing GHs [63, 64]. Furthermore, past attempts to engineer the specificity of non-glycosidase TIM-barrel containing enzymes via loop exchange have been more successful than simple mutagenic alterations [65, 66]. To provide a broader view of structural variations within GH27 and their impact on substrate specificity, we performed a detailed phylogenetic analysis including all functionally and structurally characterised members of the family.

Phylogenetic analysis of GH27

A phylogenetic analysis of GH27 presented by Fernández-Leiro *et al* in 2010 showed the importance of loop insertions in controlling protein oligomerisation and enzyme specificity, including the preference of α-galactosidases for the terminal or inner galactosyl side-chains of a polysaccharide [21]. A significant increase in knowledge of the specificities of GH27 enzymes, in particular the recent revelation of β-L-arabinopyranosidase activity in the family, warranted an update to this phylogeny, which is presented in Fig 7. Our analysis reveals several new clades which share distinct patterns of loop insertions and specificity-determining amino acid residues, and shows again that there is a strong correlation between specificity and the presence of specific loops. Specifically, loop insertions possessed by members of the group to which *Cp*Arap27 belongs were not identified by previous phylogenetic analyses [21, 22]. From this new tree it is clear that certain subsets of GH27 enzymes are very well studied, while other groups still require investigation in order to better understand the full complexity of this enzyme family. The groups identified by this analysis, described individually below, may have application in predicting the activity of GH27 enzymes yet to be characterised.

The first apparent group, Group I, comprises only two proteins, which are closely related enzymes from *Bacteroides fragilis*. Neither of these has been biochemically characterised so no predictive conclusions can be drawn from this group, although crystal structures are available for both (unpublished). These proteins have an Asp at the critical position described above, and two loop insertions (L1 & L2) of currently unknown influence on specificity, but which appear not to contribute directly to the active sites.

Group II comprises mostly mammalian enzymes and includes examples of both α-galactosidases [22] and α-*N*-acetylgalactosaminidases [68]. These enzymes have very highly conserved active sites with an Asp in the specificity-influencing position (Table 2), and as was previously noted [21, 22], all include a loop (L3) which is a major specificity determinant. This loop has

previously been referred to as the "2 position recognition loop", and the presence of a very short insertion in this region causes a structural rearrangement that allows the active site to preferentially accommodate the bulkier GalNAc residue over a Gal residue [22, 69, 70]. These enzymes also include a conserved loop with amino acids important in dimerisation (dL). The reader is referred to the insightful work of Garman and Garboczi for a detailed discussion of enzyme structure-function relationships in this clade [22].

Groups III and IV comprise α-Galactosidases (with one important exception in Group IV), but differ in the specificity of the enzymes for targeting the branching galactosyl residues on substrates such as galactomannan. All enzymes have an Asp in the key position mentioned above. The loop insertion L4 in most members of Group IV may contribute to the specificity of these enzymes for polysaccharides with galactosyl branches along their length [60], while members of Group III, which lack this loop, mostly hydrolyse galactose branches on terminal backbone residues of polysaccharides, although some are flexible in this specificity. All of the enzymes in these groups appear to be monomeric, lacking the loop insertions necessary for oligomerisation, as exemplified by the structure of the catalytically flexible Group III rice (*Oryza sativa*) α-galactosidase [70].

It should be noted here that Group IV also includes the first characterised β-ʟ-arabinopyranosidase, *Sa*Arap27A [41]. This enzyme is a significant outlier in this group, lacking the very highly conserved L4 insertion, and possessing a Glu rather than an Asp at the critical position. Further, *Sa*Arap27A possesses two additional domains (domains III and IV shown in Fig 6) which are not present in the other members of Group IV, or in other β-ʟ-arabinopyranosidases. Domain III, which has a β-jellyroll conformation, makes contact directly with the enzyme active site in domain I, while the C-terminal domain IV of *Sa*Arap27A is a family 13 carbohydrate-binding module (CBM13) that may mediate association to large polysaccharide substrates such as arabinogalactan [41]. These extensive structural modifications, plus the presence of the Glu in the conserved active site–adjacent position, may explain the different specificity of *Sa*Arap27A compared to the other members of Group IV, although the manner in which these significant differences arose is currently unclear in the absence of a larger number of characterised examples. A BLAST search against the non-redundant protein database indicates that the domain architecture of this enzyme is common to predicted α-galactosidases from *Streptomyces* species, many of which contain the active site Glu, indicating that they may also be β-ʟ-arabinopyranosidases.

Group V includes several characterised α-galactosidases. Enzymes in this group have an Asp in the critical position, and contain L5, a loop insertion identified previously as having involvement in protein oligomerisation and in the creation of binding sites [21]. This loop is the structural determinant which restricts the access of branching galactosyl residues to the active site, causing an enzyme to be specific for terminal galactosyl residues [21].

With respect to previous phylogenetic analyses [21, 22], Group VI is a newly apparent clade that includes two structural representatives (*Gs*Abp and Bh1870) and two characterised β-ʟ-arabinopyranosidases (*Cp*Arap27 and *Gs*Abp). The presence in these enzymes of an Ile at the key position, which engenders specificity for ʟ-arabinosyl substrates over ᴅ-galactosyl structures ([42, 51] and the present work), distinguishes the members of this group from other GH27s. Further, all members of this group possess the inserted loops L6, L7 & L8, with L6 and L7 contributing directly to the active site architecture. L8, in particular, is found at the dimerisation interface of *Gs*Abp [51]. Interestingly, whereas *Gs*Abp has been shown by SEC to be a tetramer in solution, and by crystallography to comprise a 'dimer of dimers', our analysis indicates that *Cp*Arap27 is monomeric in solution (Fig B in S1 File). The amino acid sequence of L8 similar, but not identical between these two proteins (Fig 5), and analysis of the structure of
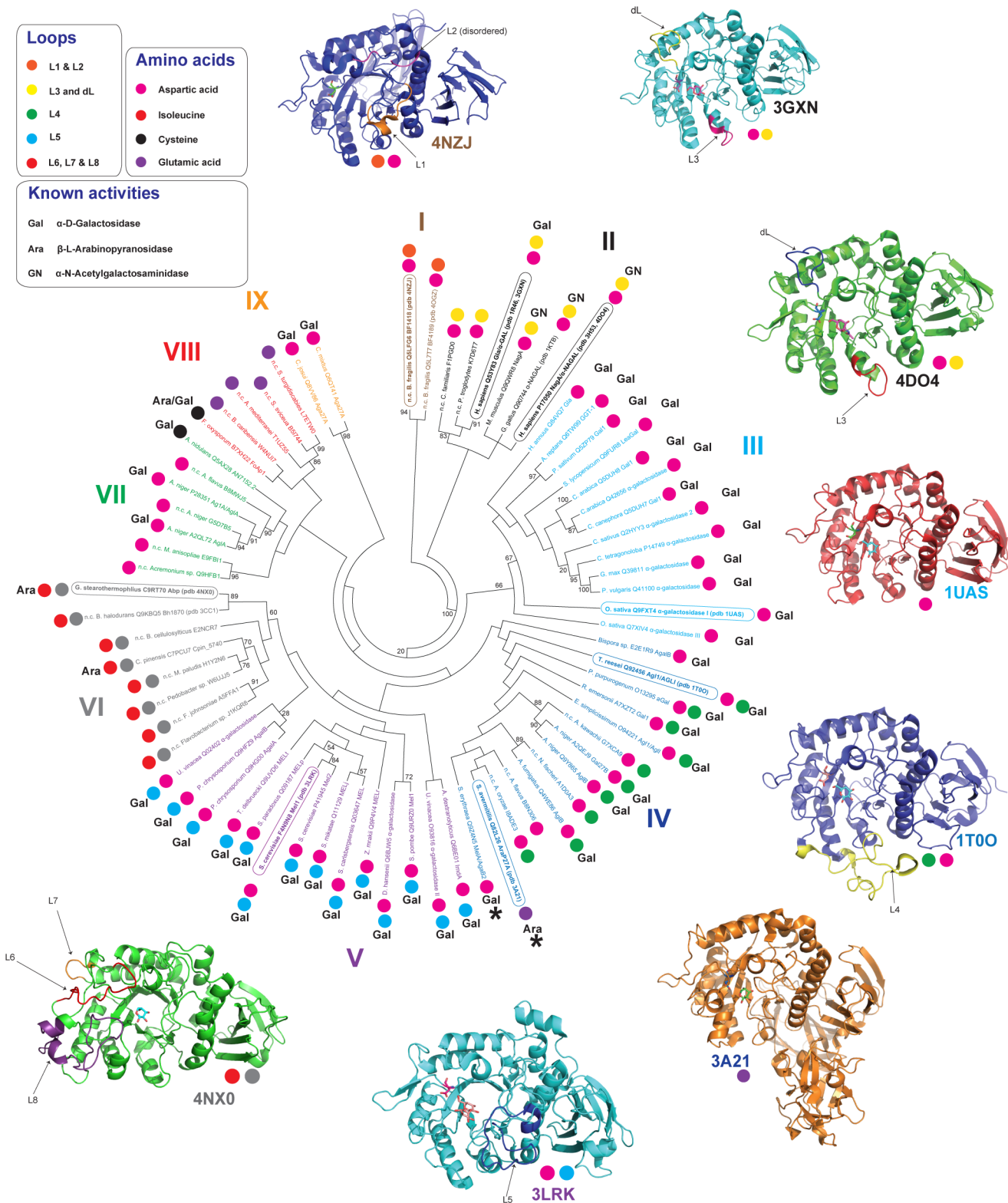
Fig 7. Phylogenetic analysis of family GH27. Proteins are labelled with organism name and *UNIPROT* code. The phylogenetic tree includes all characterised GH27 enzymes identified as such on the CAZy database at the time of writing, plus several as yet uncharacterised gene products, identified by sequence homology to characterised enzymes using pBLAST analysis. Proteins for which a 3D structure is available are shown in bold and pdb codes are provided. Examples are shown in cartoon form around the tree to highlight conserved structural elements within a clade, and the name of an illustrated

protein is circled on the tree. Each clade is labelled in a specific colour. Characterised enzymes are labelled with Ara or Gal to indicate their preferred substrate. Currently uncharacterised proteins are indicated with n.c. Finally, proteins not conforming to the pattern of loop insertions or active site amino acids otherwise well conserved for their clade are highlighted with an asterisk. The phylogenetic tree was produced using a Clustal Omega alignment [31] and the PhyML software [32], and the tree was visualised using MEGA5 [67].

the *Gs*Abp tetramer suggest that specific amino acids in this loop mediate dimerisation, which rationalises this apparent discrepancy in oligomerisation behaviour [51].

The remaining three groups have relatively few members, and possess no distinguishing loop insertions into the overall protein fold. Groups VII and VIII are broadly distinguished by the presence of Asp (Group VII) or Glu (Group VII) in the key position, with certain exceptions. An *Aspergillus nidulans* [71] galactosidase of Group VII has a Cys residue in this position. Likewise, the only characterised member of Group VIII is an enzyme annotated as a bifunctional α-ᴅ-galactosidase/β-ʟ-arabinopyranosidase (*Fo*Ap1) which also has a Cys in this position; this enzyme is able to cleave both substrates [50] but has a preference for galactose. In light of the variation between Asp and Cys in members of these groups, prediction of enzyme specificity by extrapolation from these examples is limited. Likewise, the implications of a Cys in an important, specificity-determining position in members of Groups VII and VIII, which appear to be otherwise highly similar in the key structural elements discussed here, is currently unclear. Finally, Group IX is a budding clade of α-galactosidases which possess an Asp in the key position. The two enzymes represented in this group are both active on galactomannan [61, 72].

In summary, our phylogenetic analysis highlights how the $(\alpha/\beta)_8$ barrel fold of GH27 enzymes has been modified in nature to incorporate specific active site residues and loop insertions which affect specificity and oligomerisation. The presence or absence of many of these features may help guide functional prediction and provides a framework for the characterisation of novel GH27 members.

Conclusions

C. pinensis is a free-living, saprophytic bacterium with a significant capacity to secrete diverse glycoside hydrolases for the utilisation of complex polysaccharides for growth. Among these is the GH27 enzyme *Cp*Arap27, which is highly specific for the hydrolysis of β-ʟ-arabinopyranoside substrates. Sequence alignment and phylogenetic analysis have demonstrated that a key amino acid position in the family influences specificity in GH27 enzymes, with individual members having absolute specificity for β-ʟ-Ara*p*, increasing levels of activity towards α-ᴅ-galactopyranose, or absolute specificity for α-ᴅ-Gal*p* (Fig 7, Table 2). However, we and others have shown that manipulation of the specificity of these enzymes by site-directed mutagenesis of this amino acid is inevitably accompanied by a significant penalty to catalysis. A strictly reductionist approach to specificity engineering is therefore clearly limited in the GH27 system, which indicates that other structural features make important contributions to specificity. Indeed, the β-ʟ-arabinopyranosidases of this family are distinguished not just by the presence of an important Isoleucine in the active site, but also by major loop insertions, which are unique to the phylogenetic clade in which they are found. Nonetheless, the phylogeny presented here serves as a useful guide for predicting further β-ʟ-arabinopyranosidases in GH27, thereby informing future bioinformatics and enzyme structure-function studies.

Supporting Information

S1 File. Supporting information. Tables A-B and Figures A-D.
(PDF)

Acknowledgments

Author Contributions

Conceived and designed the experiments: HB LSM. Performed the experiments: LSM. Analyzed the data: LSM HB. Wrote the paper: LSM HB.

References

1. Bouws H, Wattenberg A, Zorn H. Fungal secretomes—nature's toolbox for white biotechnology. Appl Microbiol Biotechnol. 2008; 80(3):381–8. Epub 2008/07/19. doi: 10.1007/s00253-008-1572-5 PMID: 18636256.

2. Dashtban M, Schraft H, Qin W. Fungal bioconversion of lignocellulosic residues; opportunities & perspectives. International journal of biological sciences. 2009; 5(6):578–95. Epub 2009/09/24. PMID: 19774110; PubMed Central PMCID: PMCPmc2748470.

3. Hasunuma T, Okazaki F, Okai N, Hara KY, Ishii J, Kondo A. A review of enzymes and microbes for lignocellulosic biorefinery and the possibility of their application to consolidated bioprocessing technology. Bioresource technology. 2013; 135:513–22. Epub 2012/12/01. doi: 10.1016/j.biortech.2012.10.047 PMID: 23195654.

4. McCartney L, Gilbert HJ, Bolam DN, Boraston AB, Knox JP. Glycoside hydrolase carbohydrate-binding modules as molecular probes for the analysis of plant cell wall polymers. Analytical biochemistry. 2004; 326(1):49–54. Epub 2004/02/11. doi: 10.1016/j.ab.2003.11.011 PMID: 14769335.

5. Gilbert HJ, Stalbrand H, Brumer H. How the walls come crumbling down: recent structural biochemistry of plant polysaccharide degradation. Curr Opin Plant Biol. 2008; 11(3):338–48. Epub 2008/04/24. doi: 10.1016/j.pbi.2008.03.004 PMID: 18430603.

6. Schubert C. Can biofuels finally take center stage? Nat Biotech. 2006; 24(7):777–84.

7. Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, et al. The Path Forward for Biofuels and Biomaterials. Science. 2006; 311(5760):484–9. doi: 10.1126/science.1114736 PMID: 16439654

8. Sangkhobol V, Skerman VBD. Chitinophaga, a New Genus of Chitinolytic Myxobacteria. International Journal of Systematic Bacteriology. 1981; 31(3):285–93. doi: 10.1099/00207713-31-3-285

9. Sly LI, Taghavi M, Fegan M. Phylogenetic position of Chitinophaga pinensis in the Flexibacter-Bacteroides-Cytophaga phylum. Int J Syst Bacteriol. 1999; 49 Pt 2:479–81. Epub 1999/05/13. PMID: 10319467.

10. Joint Genome Institute (United States Department of Energy) J. Phylogenetic Diversity [Accessed on: 28-7-2015]. Available from: http://jgi.doe.gov/our-science/science-programs/microbial-genomics/phylogenetic-diversity/.

11. Glavina Del Rio T, Abt B, Spring S, Lapidus A, Nolan M, Tice H, et al. Complete genome sequence of Chitinophaga pinensis type strain (UQM 2034). Stand Genomic Sci. 2010; 2(1):87–95. Epub 2011/02/10. doi: 10.4056/sigs.661199 PMID: 21304681.

12. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014; 42(1):D490–5. Epub 2013/11/26. doi: 10.1093/nar/gkt1178 PMID: 24270786.

13. El Kaoutari A, Armougom F, Gordon JI, Raoult D, Henrissat B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. Nature reviews Microbiology. 2013; 11(7):497–504. Epub 2013/06/12. doi: 10.1038/nrmicro3050 PMID: 23748339.

14. Eklöf JM, Shojania S, Okon M, McIntosh LP, Brumer H. Structure-Function Analysis of a Broad Specificity Populus trichocarpa Endo-β-glucanase Reveals an Evolutionary Link between Bacterial Licheninases and Plant XTH Gene Products. The Journal of Biological Chemistry. 2013; 288(22):15786–99. doi: 10.1074/jbc.M113.462887 PMC3668736. PMID: 23572521

15. Larsbrink J, Rogers TE, Hemsworth GR, McKee LS, Tauzin AS, Spadiut O, et al. A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. Nature. 2014; 506

(7489):498–502. Epub 2014/01/28. doi: 10.1038/nature12907 PMID: 24463512; PubMed Central PMCID: PMCPmc4282169.

16. Larsbrink J, Thompson AJ, Lundqvist M, Gardner JG, Davies GJ, Brumer H. A complex gene locus enables xyloglucan utilization in the model saprophyte Cellvibrio japonicus. Molecular Microbiology. 2014; 94(2):418–33. doi: 10.1111/mmi.12776 PMID: 25171165

17. Aspeborg H, Coutinho PM, Wang Y, Brumer H 3rd, Henrissat B. Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). BMC Evol Biol. 2012; 12:186. Epub 2012/09/21. doi: 10.1186/1471-2148-12-186 PMID: 22992189; PubMed Central PMCID: PMCPmc3526467.

18. Guce AI, Clark NE, Salgado EN, Ivanen DR, Kulminskaya AA, Brumer H, et al. Catalytic Mechanism of Human α-Galactosidase. The Journal of Biological Chemistry. 2010; 285(6):3625–32. doi: 10.1074/jbc.M109.060145 PMC2823503. PMID: 19940122

19. Hart DO, He S, Chany CJ 2nd, Withers SG, Sims PF, Sinnott ML, et al. Identification of Asp–130 as the catalytic nucleophile in the main alpha-galactosidase from Phanerochaete chrysosporium, a family 27 glycosyl hydrolase. Biochemistry. 2000; 39(32):9826–36. Epub 2000/08/10. PMID: 10933800.

20. Henrissat B, Teeri TT, Warren RAJ. A scheme for designating enzymes that hydrolyse the polysaccharides in the cell walls of plants. FEBS Letters. 1998; 425(2):352–4. http://dx.doi.org/10.1016/S0014-5793(98)00265-8. PMID: 9559678

21. Fernandez-Leiro R, Pereira-Rodriguez A, Cerdan ME, Becerra M, Sanz-Aparicio J. Structural analysis of Saccharomyces cerevisiae alpha-galactosidase and its complexes with natural substrates reveals new insights into substrate specificity of GH27 glycosidases. J Biol Chem. 2010; 285(36):28020–33. Epub 2010/07/02. doi: 10.1074/jbc.M110.144584 PMID: 20592022; PubMed Central PMCID: PMCPmc2934667.

22. Garman SC, Garboczi DN. The molecular defect leading to Fabry disease: structure of human alpha-galactosidase. J Mol Biol. 2004; 337(2):319–35. Epub 2004/03/09. doi: 10.1016/j.jmb.2004.01.035 PMID: 15003450.

23. Miller JH. Experiments in Molecular Genetics. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 1975.

24. McBride MJ, Xie G, Martens EC, Lapidus A, Henrissat B, Rhodes RG, et al. Novel features of the polysaccharide-digesting gliding bacterium *Flavobacterium johnsoniae* as revealed by genome sequence analysis. Appl Environ Microbiol. 2009; 75(21):6864–75. Epub 2009/09/01. doi: 10.1128/AEM.01495-09 PMID: 19717629.

25. Larsbrink J, Izumi A, Ibatullin FM, Nakhai A, Gilbert HJ, Davies GJ, et al. Structural and enzymatic characterization of a glycoside hydrolase family 31 alpha-xylosidase from Cellvibrio japonicus involved in xyloglucan saccharification. Biochem J. 2011; 436(3):567–80. Epub 2011/03/24. doi: 10.1042/bj20110299 PMID: 21426303.

26. Sundqvist G, Stenvall M, Berglund H, Ottosson J, Brumer H. A general, robust method for the quality control of intact proteins using LC-ESI-MS. J Chromatogr B Analyt Technol Biomed Life Sci. 2007; 852 (1–2):188–94. Epub 2007/02/03. doi: 10.1016/j.jchromb.2007.01.011 PMID: 17267305.

27. Ly HD, Withers SG. Mutagenesis of glycosidases. Annu Rev Biochem. 1999; 68:487–522. Epub 2000/06/29. doi: 10.1146/annurev.biochem.68.1.487 PMID: 10872458.

28. Brumer H, Sims PFG, Sinnott ML. Lignocellulose degradation by Phanerochaete chrysosporium: purification and characterization of the main alpha-galactosidase. Biochemical Journal. 1999; 339:43–53. PMID: 10085226

29. McCleary BV, Charnock S. Assay for determination of free D-galactose and/or L-arabinose. US Patent US7785771. 2004.

30. Melrose J, Sturgeon RJ. An enzymic assay of l-arabinose, using β-d-galactose dehydrogenase: Its application in the assay of α-l-arabinofuranosidase. Carbohydrate research. 1983; 118:247–53. http://dx.doi.org/10.1016/0008-6215(83)88052-5.

31. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology. 2011; 7:539. Epub 2011/10/13. doi: 10.1038/msb.2011.75 PMID: 21988835; PubMed Central PMCID: PMCPmc3261699.

32. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic biology. 2010; 59(3):307–21. Epub 2010/06/09. doi: 10.1093/sysbio/syq010 PMID: 20525638.

33. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution. 2013; 30(12):2725–9. doi: 10.1093/molbev/mst197 PMID: 24132122

34. Koropatkin NM, Martens EC, Gordon JI, Smith TJ. Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. Structure. 2008; 16(7):1105–15. Epub 2008/07/10. doi: 10.1016/j.str.2008.03.017 PMID: 18611383; PubMed Central PMCID: PMCPmc2563962.

35. Huisman MMH, Brüll LP, Thomas-Oates JE, Haverkamp J, Schols HA, Voragen AGJ. The occurrence of internal (1→5)-linked arabinofuranose and arabinopyranose residues in arabinogalactan side chains from soybean pectic substances. Carbohydrate Research. 2001; 330(1):103–14. http://dx.doi.org/10.1016/S0008-6215(00)00269-X. PMID: 11217953

36. Jones JKN, Reid PE. Structural studies on the water-soluble arabinogalactans of mountain and European larch. Journal of Polymer Science Part C: Polymer Symposia. 1963; 2(1):63–71. doi: 10.1002/polc.5070020108

37. Haq S, Adams GA. STRUCTURE OF AN ARABINOGALACTAN FROM TAMARACK (LARIX LARICINA). Canadian Journal of Chemistry. 1961; 39(8):1563–73. doi: 10.1139/v61-199

38. Cardoso SM, Silva AM, Coimbra MA. Structural characterisation of the olive pomace pectic polysaccharide arabinan side chains. Carbohydr Res. 2002; 337(10):917–24. Epub 2002/05/15. PMID: 12007474.

39. Shofiqur Rahman AKM, Kato K, Kawai S, Takamizawa K. Substrate specificity of the α-l-arabinofuranosidase from Rhizomucor pusillus HHT–1. Carbohydrate Research. 2003; 338(14):1469–76. http://dx.doi.org/10.1016/S0008-6215(03)00203-9. PMID: 12829392

40. Ishii T, Konishi T, Ito Y, Ono H, Ohnishi-Kameyama M, Maeda I. A β-(1→3)-arabinopyranosyltransferase that transfers a single arabinopyranose onto arabino-oligosaccharides in mung bean (Vigna radiate) hypocotyls. Phytochemistry. 2005; 66(20):2418–25. http://dx.doi.org/10.1016/j.phytochem.2005.08.007. PMID: 16171834

41. Ichinose H, Fujimoto Z, Honda M, Harazono K, Nishimoto Y, Uzura A, et al. A beta-l-Arabinopyranosidase from Streptomyces avermitilis is a novel member of glycoside hydrolase family 27. J Biol Chem. 2009; 284(37):25097–106. Epub 2009/07/18. doi: 10.1074/jbc.M109.022723 PMID: 19608743; PubMed Central PMCID: PMCPMC2757213.

42. Salama R, Alalouf O, Tabachnikov O, Zolotnitsky G, Shoham G, Shoham Y. The abp gene in Geobacillus stearothermophilus T–6 encodes a GH27 beta-L-arabinopyranosidase. FEBS Lett. 2012; 586 (16):2436–42. Epub 2012/06/13. doi: 10.1016/j.febslet.2012.05.062 PMID: 22687242.

43. Odonmažig P, Ebringerová A, Machová E, Alföldi J. Structural and molecular properties of the arabinogalactan isolated from Mongolian larchwood (Larix dahurica L.). Carbohydrate Research. 1994; 252 (0):317–24. http://dx.doi.org/10.1016/0008-6215(94)90028-0.

44. Westphal Y, Kuhnel S, de Waard P, Hinz SW, Schols HA, Voragen AG, et al. Branched arabino-oligosaccharides isolated from sugar beet arabinan. Carbohydr Res. 2010; 345(9):1180–9. Epub 2010/05/11. doi: 10.1016/j.carres.2010.03.042 PMID: 20452576.

45. Beldman G, Schols HA, Pitson SM, Searle-van Leuwen MJF, Voragen AG. Arabinans and Arabinan Degrading Enzymes. Advances in Macromolecular Carbohydrate Research (R J Sturgeon). 1: JAI Press Inc; 1997.

46. Goodrum LJ, Patel A, Leykam JF, Kieliszewski MJ. Gum arabic glycoprotein contains glycomodules of both extensin and arabinogalactan-glycoproteins. Phytochemistry. 2000; 54(1):99–106. Epub 2000/06/10. PMID: 10846754.

47. Nie S-P, Wang C, Cui SW, Wang Q, Xie M-Y, Phillips GO. A further amendment to the classical core structure of gum arabic (Acacia senegal). Food Hydrocolloids. 2013; 31:42–8.

48. Stick RV, Williams S. Carbohydrates: The Essential Molecules of Life, Second Edition Elsevier; 2008.

49. Withers SG. Anomeric centre (alpha and beta). in *CAZypedia*, URL http://wwwcazypediaorg/indexphp/Anomeric_centre_%28alpha_and_beta%29. Accessed 20-4-2015.

50. Sakamoto T, Tsujitani Y, Fukamachi K, Taniguchi Y, Ihara H. Identification of two GH27 bifunctional proteins with beta-L-arabinopyranosidase/alpha-D-galactopyranosidase activities from Fusarium oxysporum. Appl Microbiol Biotechnol. 2010; 86(4):1115–24. Epub 2009/11/26. doi: 10.1007/s00253-009-2344-6 PMID: 19937437.

51. Lansky S, Salama R, Solomon HV, Feinberg H, Belrhali H, Shoham Y, et al. Structure-specificity relationships in Abp, a GH27 beta-L-arabinopyranosidase from Geobacillus stearothermophilus T6. Acta Crystallogr D Biol Crystallogr. 2014; 70(Pt 11):2994–3012. Epub 2014/11/06. doi: 10.1107/s139900471401863x PMID: 25372689.

52. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Meth. 2011; 8(10):785–6.

53. Martens EC, Kelly AG, Tauzin AS, Brumer H. The Devil Lies in the Details: How Variations in Polysaccharide Fine-Structure Impact the Physiology and Evolution of Gut Microbes. J Mol Biol. 2014; 426 (23):3851–65. doi: 10.1016/j.jmb.2014.06.022 PMID: WOS:000345880200004.

54. Martens EC, Koropatkin NM, Smith TJ, Gordon JI. Complex Glycan Catabolism by the Human Gut Microbiota: The Bacteroidetes Sus-like Paradigm. J Biol Chem. 2009; 284(37):24673–7. doi: 10.1074/jbc.R109.022848 PMID: ISI:000269734000001.

55. Hemsworth GR, Dejean G, Davies GJ, Brumer H. Learning from Microbial Strategies for Polysaccharide Degradation. Biochemical Society Transactions. 2015;in press.

56. Dey PM. β-l-arabinosidase from Cajanus inducus: A new enzyme. Biochimica et Biophysica Acta (BBA)—Enzymology. 1973; 302(2):393–8. http://dx.doi.org/10.1016/0005-2744(73)90167-8.

57. Dey PM. Further characterization of β-L-arabinosidase from Cajanus indicus. Biochimica et Biophysica Acta (BBA)—Enzymology. 1983; 746:8–13.

58. Willför S, Sjöholm R, Laine C, Holmbom B. Structural features of water-soluble arabinogalactans from Norway spruce and Scots pine heartwood. Wood Science and Technology. 2002; 36(2):101–10. doi: 10.1007/s00226-001-0137-x

59. Ponder GR, Richards GN. Arabinogalactan from Western larch, Part III: alkaline degradation revisited, with novel conclusions on molecular structure. Carbohydrate Polymers. 1997; 34(4):251–61. http://dx.doi.org/10.1016/S0144-8617(97)00099-4.

60. Golubev AM, Nagem RA, Brandao Neto JR, Neustroev KN, Eneyskaya EV, Kulminskaya AA, et al. Crystal structure of alpha-galactosidase from Trichoderma reesei and its complex with galactose: implications for catalytic mechanism. Journal of molecular biology. 2004; 339(2):413–22. Epub 2004/05/12. doi: 10.1016/j.jmb.2004.03.062 PMID: 15136043.

61. Jindou S, Karita S, Fujino E, Fujino T, Hayashi H, Kimura T, et al. alpha-Galactosidase Aga27A, an enzymatic component of the Clostridium josui cellulosome. J Bacteriol. 2002; 184(2):600–4. Epub 2001/12/26. PMID: 11751843; PubMed Central PMCID: PMCPmc139563.

62. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protocols. 2015; 10(6):845–58. doi: 10.1038/nprot.2015.053 PMID: 25950237

63. Rigden DJ, Jedrzejas MJ, de Mello LV. Identification and analysis of catalytic TIM barrel domains in seven further glycoside hydrolase families. FEBS Lett. 2003; 544(1–3):103–11. Epub 2003/06/05. PMID: 12782298.

64. Wierenga RK. The TIM-barrel fold: a versatile framework for efficient enzymes. FEBS Letters. 2001; 492(3):193–8. http://dx.doi.org/10.1016/S0014-5793(01)02236-0. PMID: 11257493

65. Ochoa-Leyva A, Barona-Gómez F, Saab-Rincón G, Verdel-Aranda K, Sánchez F, Soberón X. Exploring the Structure–Function Loop Adaptability of a (β/α)8-Barrel Enzyme through Loop Swapping and Hinge Variability. J Mol Biol. 2011; 411(1):143–57. http://dx.doi.org/10.1016/j.jmb.2011.05.027. doi: 10.1016/j.jmb.2011.05.027 PMID: 21635898

66. Ochoa-Leyva A, Soberón X, Sánchez F, Argüello M, Montero-Morán G, Saab-Rincón G. Protein Design through Systematic Catalytic Loop Exchange in the (β/α)8 Fold. J Mol Biol. 2009; 387(4):949–64. http://dx.doi.org/10.1016/j.jmb.2009.02.022. doi: 10.1016/j.jmb.2009.02.022 PMID: 19233201

67. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011; 28(10):2731–9. Epub 2011/05/07. doi: 10.1093/molbev/msr121 PMID: 21546353; PubMed Central PMCID: PMCPmc3203626.

68. Clark NE, Garman SC. The 1.9 a structure of human alpha-N-acetylgalactosaminidase: The molecular basis of Schindler and Kanzaki diseases. Journal of molecular biology. 2009; 393(2):435–47. Epub 2009/08/18. doi: 10.1016/j.jmb.2009.08.021 PMID: 19683538; PubMed Central PMCID: PMCPmc2771859.

69. Garman SC, Hannick L, Zhu A, Garboczi DN. The 1.9 A structure of alpha-N-acetylgalactosaminidase: molecular basis of glycosidase deficiency diseases. Structure. 2002; 10(3):425–34. Epub 2002/05/15. PMID: 12005440.

70. Fujimoto Z, Kaneko S, Momma M, Kobayashi H, Mizuno H. Crystal structure of rice alpha-galactosidase complexed with D-galactose. J Biol Chem. 2003; 278(22):20313–8. Epub 2003/03/27. doi: 10.1074/jbc.M302292200 PMID: 12657636.

71. Bauer S, Vasu P, Persson S, Mort AJ, Somerville CR. Development and application of a suite of polysaccharide-degrading enzymes for analyzing plant cell walls. Proceedings of the National Academy of Sciences of the United States of America. 2006; 103(30):11417–22. Epub 2006/07/18. doi: 10.1073/pnas.0604632103 PMID: 16844780; PubMed Central PMCID: PMCPmc1544100.

72. Dias FM, Vincent F, Pell G, Prates JA, Centeno MS, Tailford LE, et al. Insights into the molecular determinants of substrate specificity in glycoside hydrolase family 5 revealed by the crystal structure and kinetics of Cellvibrio mixtus mannosidase 5A. J Biol Chem. 2004; 279(24):25517–26. Epub 2004/03/12. doi: 10.1074/jbc.M401647200 PMID: 15014076.