The structure and function of an arabinan-specific α-1,2-arabinofuranosidase identified from screening the activities of bacterial GH43 glycoside hydrolases

Alan Cartmell[1,2]*, Lauren McKee[1,2]*, Maria J. Peña[2], Johan Larsbrink[3], Harry Brumer[3], Satoshi Kaneko[4], Hitomi Ichinose[4], Richard J. Lewis[1], Anders Viksø-Nielsen[5], Harry J. Gilbert[1,2]¶ and Jon Marles-Wright[1]

[1]Institute for Cell and Molecular Biosciences, Newcastle University, The Medical School, Newcastle upon Tyne NE2 4HH, U.K; [2]The Complex Carbohydrate Research Center, The University of Georgia, 315 Riverbend Road, Athens, GA 30602, USA; [3]School of Biotechnology, Royal Institute of Technology, AlbaNova University Centre, 10691 Stockholm, Sweden; [4] Food Biotechnology Division, National Food Research Institute, 2-1-12 Kannondai, Tsukuba, Ibaraki 305-8642, Japan, [5]Novozymes A/S, Krogshoejvej 36, 2880 Bagsvaerd, Denmark
*These authors contributed equally
¶Corresponding author: Address: Institute for Cell and Molecular Biosciences, Newcastle University, The Medical School, Framlington Place, Newcastle upon Tyne NE2 4HH, U.K. Phone: +44 191 2228800. Fax: +44 2227424.
E-mail: hgilbert@ncl.ac.uk
Running title: Arabinan specific arabinofuranosidase

Reflecting the diverse chemistry of plant cell walls, microorganisms that degrade these composite structures synthesize an array of glycoside hydrolases. These enzymes are organized into sequence-, mechanism- and structure-based families. Genomic data have shown that several organisms that degrade the plant cell wall contain a large number of genes encoding family 43 (GH43) glycoside hydrolases. Here we report the biochemical properties of the GH43 enzymes of a saprophytic soil bacterium, *Cellvibrio japonicus*, and a human colonic symbiont, *Bacteroides thetaiotaomicron*. The data show that *C. japonicus* uses predominantly exo-acting enzymes to degrade arabinan into arabinose, while *B. thetaiotaomicron* deploys a combination of endo- and side chain-cleaving glycoside hydrolases. Both organisms, however, utilize an arabinan-specific α-1,2-arabinofuranosidase in the degradative process, an activity that has not previously been reported. The enzyme can cleave α-1,2-arabinofuranose decorations in single or double substitutions, the latter being recalcitrant to the action of other arabinofuranosidases. The crystal structure of the *C. japonicus* arabinan-specific α-1,2-arabinofuranosidase, *Cj*Abf43A, displays a 5-bladed β-propeller fold. The specificity of the enzyme for arabinan is conferred by a surface cleft that is complementary to the helical backbone of the polysaccharide. The specificity of *Cj*Abf43A for α-1,2-L-arabinofuranose side chains is conferred by a polar residue that orientates the arabinan backbone such that O2 arabinose decorations are directed into the active site pocket. A shelf-like structure adjacent to the active site pocket accommodates O3 arabinose side chains, explaining how the enzyme can target O2 linkages that are components of single or double substitutions.

Plant cell walls comprise a diverse repertoire of chemically complex polysaccharides (1). The microbial degradation of these composite structures, which is mediated by an extensive repertoire of hydrolytic enzymes, is of considerable biological and industrial significance. The release of soluble sugars, principally monosaccharides, from the

1

plant cell wall is not only required to maintain microbial ecosystems, but the volatile fatty acids generated by these microbiota are essential nutrients for higher order organisms such as mammalian herbivores (2). Within an industrial context the microbial enzymes that catalyze this process are integral to the exploitation of lignocellulose as an environmentally sustainable substrate for the biofuel and bioprocessing industries (3,4). An important limitation in the industrial exploitation of lignocellulose is the cost of the pre-treatment process, reflecting the low activity displayed by the degradative enzymes against cell wall structures. This reflects, in part, the chemical complexity of the plant cell wall, particularly within the pectin group of polysaccharides (5). There is, therefore, an urgent need to discover glycoside hydrolases that exhibit novel activities against these complex structural polysaccharides. Such enzymes will not only have the potential to contribute to the degradative process, but will also be useful reagents with which to probe the molecular architecture of cell walls.

Within the plant cell wall, cellulose fibres are cross-linked with hemicellulosic and pectic polysaccharides (1). The backbones of these matrix polysaccharides are often decorated with arabinofuranose-containing side chains. For example, rhamnogalacturonan I (RGI) is a pectic polysaccharide found in primary cell walls, which plays an important role in plant cell growth and tissue development. RGI contains a backbone of the repeating disaccharide 4-α-D-galacturonic acid-1,2-α-L-rhamnose-1, which is extensively decorated with a variety of oligosaccharides of which arabinan is the most abundant (5). Arabinan consists of an α-1,5-L-arabinofuranose backbone that is heavily substituted with α-1,3-L-arabinofuranose and/or α-1,2-L-arabinofuranose side chains (5). The β-1,4-xylose backbone of the major

hemicellulosic polysaccharide in cereals, arabinoxylan, is also extensively decorated at O2 and/or O3 with single arabinofuranose units (6).

Within the CAZy database, glycoside hydrolases are grouped into families based on sequence, structural and mechanistic similarities (7). Enzymes that hydrolyse arabinofuranose-linked glycosidic bonds are predominantly located in glycoside hydrolase families (GHs) 43, 51, 54 and 62 (7). Of particular note is GH43 which, in addition to containing arabinofuranosidases and arabinanases (8-10), also contains β-xylosidases (11), exo-β-1,3-galactanases (12) and xylanases (13). An intriguing feature of several organisms, which utilize the plant cell wall as an important nutrient, is that their genomes encode a large number of GH43 enzymes. For example the soil saprophyte *Cellvibrio japonicus*, which degrades all the major components of the cell wall, contains 14 genes encoding GH43 enzymes (14), while *Bacteroides thetaiotaomicron*, a human colonic symbiont that targets pectins and mammalian glycans, has the genetic capacity to synthesise 33 different GH43 hydrolases (15). Superficially, GH43 appears to contain relatively few enzyme activities and thus the biological rationale for the expansion of this family in certain organisms is unclear. It is likely, however, that GH43 arabinofuranosidases may target different hemicellulosic and pectic polysaccharides.

Here we have explored the range of activities displayed by the *C. japonicus* and *B. thetaiotaomicron* GH43 enzymes against arabinose-containing polysaccharides. The data underlines the different strategies used by the two microorganisms to degrade arabinan. A common feature of the two systems is an arabinan-specific α-1,2-arabinofuranosidase, an activity that has not previously been reported. It is likely

2

that the primary role of the enzyme is to target O2-linked arabinose residues that are components of double substitutions. The crystal structure of the *C. japonicus* arabinan-specific α-1,2-arabinofuranosidase, *Cj*Abf43A, reveals a curved substrate binding cleft that accommodates the helical structure of the arabinan backbone. The active site, which comprises a deep pocket in the middle of the cleft, displays considerable conservation with other GH43 exo-acting enzymes. Asn165, located in the +1 subsite, in combination with aromatic residues, appear to orientate the polysaccharide chain such that the O2 arabinose decoration is housed in the active site. Thus, the shape of the cleft in combination with specific interactions with the arabinan backbone, confer the specificity displayed by this novel enzyme.

MATERIALS AND METHODS

Gene Cloning: The genes encoding the mature forms (lacking their signal peptides) of the *C. japonicus* and *B. thetaiotaomicron* GH43 enzymes were amplified by PCR from genomic DNA with the thermostable DNA polymerase Kod, and the primers listed in supplemental information, **Tables S1** and **S2**. The PCR products were cloned into the expression vectors pRSET, pET22b, pET28b and pET32b to generate the plasmids listed in **Tables S1** and **S2.** All of the recombinant enzymes contained either a His$_6$-tag or were fused to glutathione S-transferase (GST).

Protein Expression and Purification: *Escherichia coli* BL21 or Tuner cells, transformed with the relevant plasmid, were cultured in Luria-Bertani broth at 37 °C to mid-exponential phase (A_{600}=0.6). Recombinant protein expression was induced by the addition of 1 mM isopropyl-1-thio-β-D-galactopyranoside and grown for a further 12 h at 16 °C. Cell-free extracts were prepared as described previously (16) and proteins with a His$_6$ tag were purified by immobilised metal ion affinity chromatography (IMAC) using Talon^{TM} resin and eluted with 50 mM Tris/HCl buffer, pH 8.0, containing 300 mM NaCl and 100 mM imidazole. Proteins fused to GST were purified using a glutathione resin and eluted with 10mM Tris/HCl buffer, pH 8.0, containing 10mM reduced glutathione. Enzymes used in biochemical assays were dialysed against 50 mM sodium phosphate buffer, pH 7.0, at 4°C. Proteins destined for crystallographic analysis were subjected to further purification on a 26/60 S75 Superdex gel filtration column equilibrated with 10 mM Tris/HCl pH 8.0 containing 150 mM NaCl. Fractions containing the purified protein were pooled and dialysed against ultrapure water. To produce a selenomethionine derivative of *Cj*Abf43A, the *E. coli* methionine auxotroph B834 (DE3) containing pAC12 (encodes *Cj*Abf43A) was cultured as described previously (16) and the arabinofuranosidase was purified by IMAC and gel filtration as described above, except that the buffers were supplemented with 5 mM β-mercaptoethanol or 10 mM DL-dithiothreitol, respectively. The purified protein was dialysed extensively against 10 mM DL- dithiothreitol.

Site-directed mutagenesis: Site-directed mutagenesis was conducted using the PCR-based QuikChange site-directed mutagenesis kit (Stratagene) according to the manufacturer's instruction, using pAC12 as the template and the primer pairs listed in **Table S3**.

Enzyme Assays: The substrates 4-nitrophenyl-α-L-arabinofuranoside (4NPA) and sugar beet arabinan were purchased from Sigma and Megazyme, respectively. Assays in which 4NPA was the substrate were monitored by the release of 4-nitrophenolate at a wavelength of 400 nm. The product was quantified using a molar

extinction coefficient of 10500 M^{-1} cm^{-1}. When using polysaccharides such as arabinan as the substrate, arabinose release was monitored using galactose dehydrogenase (Megazyme), an enzyme which catalyses the oxidation of galactose/arabinose with concomitant reduction of NAD^+ to NADH (17). The concentration of NADH generated was quantified by measuring ΔA_{340nm} using a molar extinction coefficient of 6230 M^{-1} cm^{-1}. All reactions were carried out in 50 mM sodium phosphate buffer pH 7.0, at 25 °C for the *C. japonicus* enzymes and at 37 °C for the *B. thetaiotaomicron* enzymes. For pH optima assays a range of different buffers were used. Assays were monitored discontinuously and time points were terminated by the addition of an equal volume of 500 mM Na_2CO_3. The addition of Na_2CO_3 raises the pH to 11, where 4-nitrophenolate has an extinction coefficient of 24150 M^{-1} cm^{-1}. Reducing sugar assays were carried out using 3,5-dinitrosalicylic acid (DNSA) according to (18). At various time-points, a 500 µl aliquot of the reaction was taken and stopped by the addition of an equal volume of DNSA reagent. After boiling for 20 min, the A_{575nm} values were determined and compared to a standard curve of arabinose. The enzymes used in these assays contained their respective His_6- or GST-tags. All enzyme assays were carried out in triplicate and the standard error of the mean is reported for the individual kinetic parameters.

HPAEC of monosaccharide and oligosaccharide reaction products: Oligosaccharides and monosaccharides, derived from polysaccharide hydrolysis by either enzymatic or acid treatment, were analysed using a CARBOPAC™ PA-100 anion exchange column (Dionex) equipped with a CARBOPAC™ PA-100 guard column. The fully automated system (LC25 Chromatography Oven, GP40 Gradient Pump, ED40 Electrochemical Detector and AS40 Autosampler) had a loop size of 200 µl, flow rate of 1.0 ml/min, with sugars detected by pulsed amperometric detection (PAD). The PAD settings were E_1 = +0.05, E_2 = +0.6 and E_3 = -0.6. Typical elution conditions were 0-5 min 66 mM NaOH, 5-30 min 66 mM NaOH with a 0-75 mM sodium acetate linear gradient, 30-40 min 500 mM sodium acetate in 66 mM NaOH, 40-50 min 500 mM sodium hydroxide and then from 50-60 min 66 mM NaOH. Appropriate oligosaccharides were used as standards at a concentration of 50 µM. Samples to be analysed were centrifuged at 13000 x g for 10 min. Data were collected and manipulated using the Chromeleon™ Chromatography Management System V.6.8 (Dionex) *via* a Chromeleon™ Server (Dionex). Final graphs were drawn with Prism 4.02 (GraphPad).

Thin Layer Chromatography: TLC plates (Silicagel 60, 20 × 20, Merck) spotted with samples (2 x 1 µl drops) were developed in 1-butanol/acetic acid/water in a ratio of 2:1:1 (v/v). Visualisation of migrated sugars was achieved by immersing the TLC plates in developer (sulphuric acid/ethanol/water 3:70:20 v/v, orcinol 0.1 %) for 1-2 s. Plates were again dried and heated between 80 and 100 °C.

Nuclear Magnetic Resonance: Arabinooligosaccharides were prepared for NMR analysis by the controlled acid hydrolysis of sugar beet arabinan. HCl was added to 10 ml of a 30 mg/ml solution of sugar beet arabinan to a final concentration of 50 mM and the solution was boiled for 30 min. The reaction was then put on ice and neutralised with 500 µl of 1 M NaOH. The large polymers in the mixture were precipitated in 80 % (v/v) ethanol and the arabinooligosaccharide-containing supernatant was vacuum dried. The arabinan oligosaccharides were dissolved in 5 mM sodium phosphate, pH 7.0, at 5 mg/ml, and incubated for 60 min in the presence and absence of 10 µM *Cj*Abf43A at 25 °C. After the incubation, the whole

reaction without separation of the reaction products or the enzyme was lyophilized. The dried material was dissolved in D_2O (99.9%; Cambridge Isotope Laboratories) and ^1H and ^{13}C NMR two dimensional spectra were recorded with a Varian Inova NMR spectrometer operating at 600 MHz and with a sample temperature of 298 K. Gradient version of COSY (Correlation spectroscopy), HSQC (heteronuclear single quantum coherence), and HMBC (heteronuclear multiple quantum coherence) experiments were recorded using standard Varian pulse programs. Heteronuclear spectra were recorded with 512×512 complex points. These data were processed typically with zero filling to obtain a 1024×1024 matrix. Chemical shifts were measured relative to internal acetone (δ_H 2.225, δ_C 30.89). Data were processed using MestRe-C software (Universidad de Santiago de Compostela, Spain).

Crystallisation, Data Collection, Solution and Refinement of *Cj*Abf43A:
A selenomethionine derivative of *Cj*Abf43A in ultrapure water was crystallised at 30 mg/ml in 1.5 M ammonium sulphate and 100 mM Tris-HCl buffer, pH 7.5. Crystals, which grew overnight at 20 °C, were transferred to a cryoprotectant solution of the crystallisation condition supplemented with 25 % (v/v) glycerol. Native protein in water was crystallised at 10 mg/ml in 13 % PEG 3350 and 150 mM sodium acetate at 20 °C, with crystals reaching maximum size in four weeks. The purified catalytic base mutant was dialysed against 50 mM sodium chloride and crystallised at 20 mg/ml in the presence of 20 mM arabinotetraose in 20 % PEG 3350 and 200 mM ammonium tartrate at 20 °C, with crystals growing over a two month period. These two crystal forms were cryoprotected with the addition of 25 % v/v ethylene glycol to the respective crystallisation conditions. Crystals were flash cooled in liquid nitrogen and

diffraction data, for the native arabinofuranosidase and the enzyme-ligand complex, were collected at Diamond Light Source on beamline I02 and at the Advanced Photon Source on beamline 22-BM, respectively.

As no appropriate molecular replacement model was available for *Cj*Abf43A in the protein databank, the structure was solved by single wavelength anomalous dispersion (SAD) using the selenomethionine derivative. A fluorescence energy scan was performed around the selenium K atomic absorption edge of 12.658 keV, confirming the presence of selenomethionine in the sample, and a SAD dataset was collected at the optimal peak wavelength of 0.979 Å. The data was integrated using iMosflm (19) and scaled and merged in Scala (20). The SAD dataset had an overall anomalous correlation of 0.31. The scaled data were input into the SHELX C/D/E pipeline (21), for experimental phasing, where the anomalous signal, as indicated by the <d"/sig> statistics, was 3.3 in the resolution shell up to 8 Å, and was 0.98 in the highest resolution shell, between 3.2 and 3.0 Å. In total, 30 selenium atom positions were identified and phasing and density modification unambiguously determined the hand of the data to generate an interpretable initial electron density map. The initial phases were improved through the use of solvent flattening and two-fold non-crystallographic symmetry averaging, with operators determined from the selenium positions, with DM (22). An initial model was automatically built with ARP/wARP (23) and was extended and refined using Buccaneer (24), this model had an R_{cryst} of 0.200 and R_{free} of 0.236 and was used as a molecular replacement model for the native high-resolution dataset.

The native and the D41A:ligand-enzyme complex data were integrated using iMosflm (19) and scaled and merged with

Scala (20). The scaled data for the two crystal forms were input into Phaser (25) and the structures solved by molecular replacement using the selenomethionine structure as the search model. Refinement of all structures was performed using Refmac5 (26) and phenix.refine (27) interspersed with manual model correction and building in Coot (28) until convergence. Waters were added using ARP/waters (23) and checked manually. Data collection and refinement statistics are given in **Table 1**. It should be noted that the D41A mutation introduces a dual conformation of H267. One of these conformations, with a χ_1 angle of -65 °, points towards the bound oligosaccharide and is stabilised in this position by the presence of a solute network that co-incides spatially with the other rotamer position of H267 ($\chi_1 = -161$ °), which is oriented away from the oligosaccharide. In the apo structure, H267 adopts a single conformation. Interestingly, a related GH43 enzyme, PDBid 3C7G, contains a bound sodium ion in a similar location, but our diffraction data do not permit the exact identification of the bound solutes for rigorous alternative chemical occupancy refinement, and the deposited model co-ordinates do not include this partially occupied solute constellation.

RESULTS
Screening of the activities displayed by *B. thetaiotaomicron* and *C. japonicus* GH43 enzymes: Previous studies have shown that the GH43 family contains a large number of enzymes that contribute to the degradation of polysaccharides containing arabinofuranose residues (8-10). To explore the relationship between the genetic potential of *C. japonicus* and *B. thetaiotaomicron* to synthesize GH43 enzymes, and the capacity of the organisms to metabolise arabinose-containing complex carbohydrates, the genes encoding the GH43 proteins were expressed in *E. coli*. In total 23 of the 33 *B. thetaiotaomicron*, and 11 out of the 14

C. japonicus GH43 proteins were produced in soluble form and purified to electrophoretic homogeneity (data not shown). The *orf* nomenclature used herein corresponds to that adopted in the annotation of the two genomes (14,15).

The catalytic activity of the GH43 proteins was assessed using 4-nitrophenyl-α-L-arabinofuranose (4NPA) and a range of arabinose-containing polysaccharides (and oligosaccharides) including wheat and rye arabinoxylan, sugar beet, linear arabinan and debranched rhamnogalacturonan I. The data, summarized in **Tables S4** and **S5**, show that six enzymes, three from both *B. thetaiotaomicron* and *C. japonicus* displayed significant 4NPAase activity, two *B. thetaiotaomicron* enzymes displayed endo-arabinanase activity, while two *C. japonicus* GH43 proteins were shown to function as exo-α-1,5-arabinofuranosidases (and also hydrolyzed 4NPA). A single β-xylosidase, CJA_3070 from *C. japonicus*, was identified. Finally, an enzyme, common to *B. thetaiotaomicron* and *C. japonicus*, Bt0369 and CJA_3018 (designated hereafter as *Cj*Abf43A), respectively, hydrolyzed 4NPA and released arabinose from sugar beet arabinan, but was not significantly active against linear arabinan or any of the other polysaccharides tested, **Tables 2, S4** and **S5**.

A surprising feature of the biochemical studies described above is that nine of the GH43 proteins displayed no significant activity against polysaccharides, while 17 of the proteins displayed trace endoxylanase activity and one enzyme, CJA_3601, appeared to function as an extremely weak xylan-specific arabinofuranosidase. Although these activities could be detected by HPLC, after reactions were incubated for ~48 h, the concentrations of the products were too low to be quantified by reducing sugar assays. It is evident, therefore, that this xylanase activity is too low to be of

biological significance, exemplified by the observation that *B. thetaiotaomicron* is unable to grow on xylan (29). These data may have some resonance with a previous report suggesting (through zymogram analysis) that a GH43 protein functions as an endoxylanase (13), although true GH43 xylanases have recently been identified (30).

Characterization of arabinan degrading enzymes: *C. japonicus:* Biochemical analysis of the two exo-α-1,5-arabinofuranosidases, CJA_3012 and CJA_0806, **Table 2**, showed that both enzymes displayed highest catalytic activity against linear arabinan and arabinooligosaccharides, releasing exclusively arabinose, likely from the non-reducing end (all exo-acting GH43 enzymes attack the non-reducing end of polysaccharides (8,11,31,32)). The two enzymes released only very small amounts of arabinose from sugar beet, reflecting the limited concentration of terminal undecorated α-1,5-arabinofuranose units in this highly branched polysaccharide (data not shown). The catalytic efficiency of the two enzymes against arabinooligosaccharides with different degrees of polymerization (d.p.) was assessed. The data show that CJA_3012 is ~300-fold more active against arabinobiose than 4NPA, **Table 2**, suggesting that the interaction of arabinofuranose with the +1 subsite plays a critical role in substrate binding and catalysis (the -1 subsite houses the sugar at the non-reducing end of the substrate and the scissile glycosidic bond is between sugars bound at -1 and +1 (33)). The enzyme did not display elevated activity against substrates with a d.p. >2, indicating that the arabinofuranosidase contains only two kinetically significant subsites. Indeed the longer arabinoligosaccharides appeared to be hydrolyzed more slowly than the disaccharide. The similar activity displayed by CJA_0806 against 4NPA and arabinobiose indicates that the +1 subsite

exhibits limited specificity for arabinose, although the enzyme appears to be optimized to hydrolyse arabinotriose as the trisaccharide is a better substrate than either di- or tetra-saccharides. It is possible that the +2 subsite in CJA_0806, and the +1 subsite in CJA_3012, preferentially bind to arabinose at the reducing end of substrates by targeting its pyranose conformation, or the β anomer of the furanose sugar.

An intriguing feature of the two exo-α-1,5-arabinanases is that although they display similar catalytic efficiencies against 4NPA, the K_m and k_{cat} of CJA_0806 was much higher than CJA_3012. The sequence identity between these two enzymes is 49 %. It is possible that the +1 site of CJA_3012 binds much more tightly to 4-nitrophenyl than CJA_0806, which may result in the low K_m value. This tight binding, however, may limit aglycone departure from the active site, which would impact upon the catalytic rate.

Kinetic analysis of *Cj*Abf43A showed that the enzyme displayed similar catalytic efficiency against 4NPA and sugar beet arabinan, releasing monomeric arabinose, **Figure 1**. The enzyme did not show activity against any other plant cell wall polysaccharides evaluated such as linear arabinan, xylans (arabinoxylans derived from oat, wheat, rye and birchwood), pectin backbone structures (homopolygalacturonic acid and rhamnogalacturonan) and β-glucans (cellulose, β-1,4,1,3 mixed linkage glucans and laminarin), **Table S4**. *Cj*Abf43A was ~10^4-fold less active against arabinohexaose (an α-1,5 linked arabinofuranose hexasaccharide) than sugar beet arabinan, **Table 2**. Sugar beet arabinan consists of an α-1,5 linked L-arabinofuranose backbone that is ~60 % monosubstituted with α-1,3-L-arabinofuranose side chains and, less frequently, with single α-1,2 L-arabinofuranose (5). Some backbone

residues are doubly substituted with both α-1,2 and α-1,3 side chains (5). The high activity of the enzyme against sugar beet arabinan, and the low activity against α-1,5 araibinooligosaccharides, suggests *Cj*Abf43A hydrolyses the α-1,2 and/or α-1,3 side chains present in sugar beet arabinan. To further explore *Cj*Abf43A specificity, arabinan oligosaccharides were subjected to two-dimensional NMR spectroscopy analysis after incubation without (control) or with the enzyme. Resonances in the 2D spectra were assigned to specific arabinosyl residues on the basis of their scalar coupling patterns and chemical shifts, and by comparison to the published data (34) and **Table S6**. The HSQC NMR spectrum of the control contained two distinct cross peaks at δ 4.170, 87.6 and δ 4.314, 85.6 that disappeared after the treatment with the enzyme, **Figure 2**. The resonance at δ 87.6 is diagnostic of α-1,2 side chains since the chemical shift of the C2 of the α-1,5 linked L-arabinofuranose residues shifted markedly downfield when the residues are substituted at the O2 position. The resonance at δ 85.6 corresponded to arabinosyl residues substituted at O2 and O3 position. Other resonances within the spin systems of the single and double substituted arabinosyl residues also disappeared after the treatment with the enzyme. These NMR data are consistent with the view that *Cj*Abf43A hydrolyses α-1,2-L-arabinofuranose glycosidic bonds within the context of double or single substitutions. The specificity of the enzyme for the O2 linkage to backbone arabinose residues that carry two substitutions (at O2 and O3) was further examined through synergy experiments with *Hi*AXd3, an arabinofuranosidase that targets exclusively O3 linkages in double substitutions of arabinan and arabinoxylan ((35) and LM unpublished data). Arabinan, which had been pre-treated with *Cj*Abf43A, was not hydrolyzed by *Hi*AXd3. *Cj*Abf43A, however, was able to release arabinose residues from arabinan

pre-treated with *Hi*AXd3. These results are consistent with the NMR data in showing that *Cj*Abf43A targets O2 linkages in backbone arabinose residues that are singly or doubly substituted. *Cj*Abf43A releases 33 % more arabinose than *Hi*AXd3, indicating a ratio of O2 single and O2+O3 double substitutions of 1:3

B. thetaiotaomicron: While both Bt0360 and Bt0367 display endo-arabinanase activity they differ in specificity for linear and branched arabinan. Bt0360 shows a preference for branched arabinan, while Bt0367 exhibits higher activity against linear arabinan, **Table 2**. Bt0369, an arabinan-specific arabinofuranosidase, did not release arabinose from arabinan pre-treated with *Cj*Abf43A, while the *Cellvibrio* arabinofuranosidase was inactive against the polysaccharide that had been incubated with the *Bacteroides* enzyme. NMR of the reaction products showed that Bt0369 removed O2 linked arabinose side chains in the context of both single and double substitutions. These data show that Bt0369 and *Cj*Abf43A display the same substrate specificity, consistent with their structural similarity (the primary sequence of the two enzymes are 65 % identical). The functional significance of the substrate specificities of the *B. thetaiotaomicron* and *C. japonicus* enzymes, within the wider context of arabinan degradation by the two bacteria, is discussed below.

Phylogenetic analysis of *B. thetaiotaomicron* and *C. japonicus* GH43s: The *B. thetaiotaomicron* and *C. japonicus* GH43 proteins were subjected to phylogenetic analysis. The analysis incorporated selected enzymes from other organisms with known activities. The data, presented in **Figure 3**, show that enzymes that displayed arabinanase and α-1,5-exo-arabinanase activity are in clades with glycoside hydrolases that exhibit similar activities. By contrast, and consistent with their novel activity, *Cj*Abf43A

8

(CJA_3018) and Bt0369 formed a clade that contains no other enzymes with known catalytic properties. The proteins displaying "trace" xylanase activity are not clustered into a specific region of the family tree, suggesting that this minor activity may be a generic feature of GH43. Furthermore, one might speculate that an ancestral GH43 enzyme displayed significant xylanase activity, however, as new specificities were introduced into the GH43 fold, the endogenous capacity to hydrolyze xylan was lost. The requirement for a GH43 xylan degrading enzyme may also have been greatly reduced as cell wall degrading organisms acquired more powerful GH10 and GH11 xylanases (see (36,37) for reviews). It should be noted that nine out of the 34 recombinant proteins from the two bacteria, although soluble, displayed no biologically significant enzyme activity. It is possible that, through redundancy, there was no requirement for the bacteria to retain functional forms of these enzymes. Alignment of eight of these enzymes of unknown function show that they lack one or more of the catalytic residues, providing support for the view that, indeed, they are not catalytically active. This view is further illustrated by the structure of Bt2959 (PDB 3NQH), which shows that the protein lacks the canonical catalytic base. It is also possible that for those, "apparently inactive", enzymes that contain the expected catalytic residues, the appropriate substrate was not evaluated in this study.

Three dimensional structure of CjAbf43A: The crystal structure of *Cj*Abf43A was solved as a selenomethionine derivative, a native form and a native catalytic base mutant in complex with ligand to resolutions of 3.0, 1.6 and 1.8 Å, respectively. The selenomethionine and ligand bound crystals of *Cj*Abf43A were in the tetragonal space group $P4_12_12$, while the native crystals were in the monoclinic space group $P2_1$. The structures each consist of two molecules of *Cj*Abf43A in the asymmetric unit, equating to residues 29-325 in the full length enzyme. The two molecules in each asymmetric unit were not restrained to one another during the course of model refinement, but nonetheless can be considered identical, superimposing with r.m.s.d.s of ~0.2 Å on 295 matched C_α atoms. *Cj*Abf43A displays a five bladed β-propeller fold, **Figure 4**, typical of GH43 enzymes (8,10,11,32,38). It has a cylindrical shape with a diameter and height of 35 Å. The propeller is based upon a 5-fold repeat of blades formed by β-sheets that adopt the classical "W" topology of four antiparallel β-strands. The blades are arranged radially from the centre of the propeller. In some proteins that display a five-bladed β-propeller fold, the fifth blade consists of strands from the N- and C- terminus. Such closure of β-propeller proteins is colloquially termed "molecular velcro", and appears to provide considerable stabilization to the fold (39). In *Cj*Abf43A the fifth strand comprises only β-strands from the C-terminal sequence of the domain, typical of other GH43 enzymes, and thus the propeller fold is not "closed" by the fifth blade. The N- and C-terminal blades, however, are in close association and stability is provided by strong hydrogen bonds between D78 in blade 1 and several residues in blade 5. *Cj*Abf43A lacks a C-terminal β-jelly roll domain seen in several exo-acting GH43 enzymes (8,10,11,31).

Structural comparison of *Cj*Abf43A with GH43 enzymes: Structural comparison of *Cj*Abf43A by secondary structure matching of the global structure, revealed that the *Bacillus subtilis* xylan-specific GH43 arabinofuranosidase (10), *Bs*AXHm43 (PDB 3C7G), was the closest structural homologue with an rmsd of 1.4 Å over 271 aligned Cα atoms and a sequence identity of 37 %. The other most meaningful structural homologue was the

α-L-arabinanase (38) from *Cellvibrio japonicus* (1GYD), which shares 24 % sequence identity with *Cj*Abf43A and can be superimposed on 237 matched Cα atoms yielding an r.m.s.d. of 1.9 Å. Comparisons of the active sites of *Cj*Abf43A with *Bs*AXHm43 and other GH43 enzymes are discussed in brief below.

The interaction of *Cj*Abf43A with susbstrate: *Cj*Abf43A was also crystallised in the presence of arabinotetraose (Megazyme). To trap the complex between the protein and the ligand, we utilised the D41A mutant, which is enzymatically inactive (Table 3, see below). We expected to observe a linear oligosaccharide defining the active site, but the solved structure revealed an α-1,5 arabinotriose molecule with an α-1,3 linked arabinofuranose side chain on the middle arabinose (3^2-Ara-Ara$_3$) in one of the two *Cj*Abf43A molecules in the asymmetric unit, chain B. The second *Cj*Abf43A molecule (chain A) has the same oligosaccharide bound, but the electron density is too poor to define with certainty the position of the arabinofuranose branch, and so instead an α-1,5 arabinotriose has been modelled in this molecule.

The surface representation of *Cj*Abf43A, **Figure 4**, reveals a deep pocket in the centre of a highly curved cleft. Structural conservation and mutagenesis studies (see below) support the view that the catalytic apparatus is housed in the pocket, which therefore comprises the active site of the enzyme. The rim of the pocket abuts onto a shelf-like structure that accommodates the O3-linked arabinose branch. The α-1,5-trisaccharide is located in the central region of the curved surface confirming its identity as the substrate binding cleft that accommodates the arabinan backbone. The pseudosymmetry of arabinofuranosides makes determining their orientation difficult because oligosaccharide chains

built in either direction fit equally into the electron density. The orientation of the built oligosaccharide was guided by the knowledge that N185, a residue that plays a key role in arabinan recognition (see below), only makes a key hydrogen bond with the substrate in the modelled orientation. Furthermore, O2 of the central backbone arabinose, which participates in the glycosidic bond hydrolyzed by the enzyme, points directly into the active site pocket confirming that the orientation of the arabinan backbone has been built correctly. The O3-linked arabinose in the 3^2-Ara-Ara$_3$ ligand is located on the shelf adjacent to the active site pocket, demonstrating how the enzyme has plasticity for α-1,2-L-arabinofuranose side chains in single and double substitutions.

The subsite topology of *Cj*Abf43A is defined as follows: the scissile bond is between the arabinose decoration at -1 (the active site) and the backbone arabinose at +1. Subsites extended towards the reducing end of the arabinan backbone (from the +1 subsite) are defined as +2R, +3R etc, while subsites extending to the non-reducing end of the polymer are designated +2NR, +3NR and so forth. The subsite accommodating the arabinose linked O3 to the +1 subsite (and is therefore not part of the backbone) is defined as +2NR* subsite. Thus, the bound ligand occupies subsites +2NR, +2NR*, +1 and +2R. Based on the topology of the enzyme, the substrate binding cleft is unlikely to extend distal to the +2R subsite, although the enzyme may contain at least one additional non-reducing subsite (+3NR).

Subsite Interactions: *Positive Subsites:* The arabinose at the +2R subsite is sandwiched between the aromatic residues, F66 and W164, The high activity displayed by F66A suggests that the phenylalanine contributes little to substrate binding. By contrast W164A is 10-30 fold less active than the wild type enzyme,

10

suggesting that the tryptophan does make a contribution to arabinan recognition, **Table 3**.

At the +1 subsite the Nδ2 of N185 interacts with the endocylic ring oxygen and the glycosidic oxygen that links sugars at the +1 and +2NR subsites. Oϵ2 of E215, the catalytic acid (see below), interacts with O2 of the +1 arabinose, which comprises the scissile glycosidic bond. Consistent with its catalytic role, mutation of E215 completely inactivates the enzyme, **Table 3**. The N185A mutation actually causes a modest increase in activity against 4NPA, reflecting a slight lowering of K_M. This implies that N185 is not able to interact with the aryl-glycoside as it does not offer any hydrogen bond donors or acceptors. Replacing the polar side chain of asparagine with alanine increases the volume and the hydrophobicity of the +1 subsite, which is thus more able to accommodate the bulky phenyl ring of 4NPA. The amino acid substitution, however, results in a ~3000-fold decrease in activity against arabinan, **Table 3**. The substantial effect of the N185A mutation on arabinan hydrolysis is consistent with the hydrogen bonds made by the Asn with the +1 arabinose and the glycosidic bond between substrate bound at +1 and +2NR. Indeed, it is likely that the interactions made by N185 play a role in defining the specificity of *Cj*Abf165A for the O2 rather than the O3 linkage. The only polar interaction between the substrate and *Cj*Abf43A that is unique, when O2 is orientated into the -1 pocket, is the hydrogen bond between N185 and the glycosidic oxygen linking the sugars at the +1 and +2NR subsites, **Figure 5**. As stated above, if the oligosaccharide was built such that O3 is orientated into the active site, N185 is unable to interact with the glycosidic oxygen at the +1/+2NR interface. Attempts at using NMR to show that N185A was able to hydrolyze both O2 and O3 linkages was compromised by the high abundance of O3 over O2 linkages,

and the very low activity of the N185A mutant, preventing unequivocal quantitative comparisons. Furthermore, it is possible that hydrophobic residues, such as F234, which make weak apolar contacts with C3-C5 of the +1 arabinose, also contribute to arabinan orientation, and thus randomization of O2 and O3 hydrolysis may require multiple amino acid substitutions, which would likely render the enzyme completely inactive. A BLAST search of *Cj*Abf43A against the UNIPROT database showed that of the 50 top hits, which displayed identities ranging from 44-82 % (data not shown), 46 contained an Asn in the motif WGN. Significantly, Bt0369, which is 66 % identical to *Cj*Abf43A, and also displays arabinan-specific α-1,2-arabinofuranosidase activity, contains the WGN motif. These data indicate that N185 in *Cj*Abf43A, and N186 in *Bt*0369, confer specificity for the O2 linkage.

At the +2NR subsite Nϵ of W164 and Oδ1 of N185 make hydrogen bonds with O2 of the bound arabinose, while F234 makes extensive hydrophobic contacts with the substrate. Of particular note are the hydrophobic contacts made with C5 of the +2NR arabinose, which, as discussed above, may assist in orientating the arabinan backbone. The importance of these hydrophobic interactions is illustrated by the reduction in activity when F234 is substituted with alanine, **Table 3**.

At the +2NR* subsite, which houses the α-1,3 linked arabinofuranose side chain in double substitutions, the sugar makes apolar contacts with F234, while the O2 and O3 atoms of the sugar appear to be in hydrogen bonding distance with Y293 and Q292, respectively. The potential polar contacts with Y293 and Q292, however, are unlikely to be functionally significant as the activity of the mutants Y293A and Q292A are similar to the wild type enzyme, **Table 3**. To explore the

11

significance of the +2NR* subsite, the double substitutions in arabinan were converted to O2 monosubstituted decorations by treatment with *Hi*AXHd3 (the enzyme removes O3 linked arabinose residues from double substitutions (35)). The activity of the arabinan-specific α-1,2-arabinofuranosidase was similar against arabinan and the *Hi*AXHd3-treated polysaccharide. Thus, it is unlikely that the putative interactions with the arabinose at the +2NR subsite are functionally significant, and therefore F234 likely only contributes to substrate binding at the +2NR subsite.

As described above, *Cj*Abf43A is likely to contain a +3NR subsite that may simply accommodate or, potentially, make productive interactions with the substrate. Although T186 and T214 could comprise a functional +3NR subsite, substituting these residues with alanine had no significant effect on catalysis, suggesting that the subsite does not bind to the substrate, **Table 3**. Replacing the two threonines with bulky residues (T186W, T214W and T186W/T214W), however, caused a substantial reduction in activity against arabinan, but had a very limited effect on 4NPA hydrolysis, **Table 3**. The introduction of the tryptophan residues is predicted to occlude the +3NR subsite, and likely generates steric clashes with the arabinose at the +2NR subsite, explaining why the mutations reduced activity against arabinan. Indeed, it is possible that changing the binding cleft into a blind canyon type topology may have restricted the specificity of the enzyme to only O2 arabinose decorations at the non-reducing terminus of the arabinan chains, which could explain the substantial increase in K_M.

The -1 subsite: The pocket housing the -1 subsite has no arabinose present but a molecule of ethylene glycol was observed. The ethylene glycol 'stacks' against W103 and binds to R295 via Nη1 and Nη2,

Figure 6. Furthermore, comparing the structure of *Cj*Abf43A with other exo-acting GH43 enzymes reveals several conserved features that likely contribute to substrate binding and transition state stabilization, **Figure 6**. Surprisingly, the residues that interact with the arabinose in the *Streptomyces* exo-α-1,5-arabinofuranosidase Ara*f*43A (8) are conserved not only in other arabinofuranosidases, such as *Cj*Abf43A, but also in GH43 β-xylosidases exemplified by XynB3. It is difficult, therefore, to understand how these enzymes distinguish between an Ara*f* and Xyl*p* at the critical -1 subsite since arabinofuranosidases that hydrolyze 4NP-Ara*f* are not active on 4NP-Xyl*p*. Thus D41, E215 and D168, which function as the catalytic base, catalytic acid and the modulator of the pKa of the catalytic acid, respectively, are conserved in GH43 enzymes (8,10,11,31,32). Consistent with their catalytic function, alanine substitution of these three residues inactivates the arabinofuranosidase, **Table 3**. W103 is also invariant in the other GH43 structures. The structural equivalent to this residue, W101 in Ara*f*43A and W76 in *Bs*AXHm43, provides a hydrophobic platform by stacking against the respective sugar rings, and thus W103 is likely to play an equivalent role in *Cj*Abf43A. The importance of the tryptophan is illustrated by the substantial reduction in catalytic activity (against arabinan the enzyme is completely inactive) when W103 is substituted for Ala, **Table 3**. I167 and F67 are also highly conserved in other GH43 enzymes where the equivalent residues make hydrophobic contacts with substrate, **Figures 6**. I167 sits at the top of the active site pocket and is 3.6 Å from the catalytic acid, E215, and thus may also create an apolar environment, which contributes to the elevated pKa of the acidic residue. The observation that the amino acid substitutions, F67A and I167A, cause a >100-fold reduction in activity suggests that I167 and F67 also make significant

contributions to substrate binding in the active site.

In addition to the catalytic residues, R295 and H267 are also likely to make polar contacts with arabinose at the active site. These residues are highly conserved in other GH43 enzymes, and the equivalent amino acids make direct polar contacts with arabinose and xylose residues at the -1 subsite. The R295A mutant did not express in *E. coli* and thus could not be studied. While mutation of H267 did not affect catalytic efficiency against 4NPA, the K_m was substantially increased (too high to measure), suggesting that there was a commensurate increase in k_{cat}. It is possible that the reduction in substrate binding, due to the loss of the imidazole side chain, leads to an increase in the rate of product departure, which results in the elevation in k_{cat}. Against arabinan the H267A mutation reduced activity by ~30-fold, which suggests that the decrease in substrate affinity at the -1 subsite does not influence product departure as the main chain of the polysaccharide, primarily the arabinose at the +1 subsite, makes additional interactions with the enzyme.

DISCUSSION

The mechanism of arabinan degradation: This study reports on the capacity of the large number of GH43 enzymes, produced by *B. thetaiotaomicron*, and *C. japonicus*, to hydrolyze arabinose-containing polysaccharides. The data presented here, in concert with previous biochemical and transcriptomic data, has led to the following models for arabinan degradation in the two bacteria. In *B. thetaiotaomicron*, genes encoding enzymes of related function are clustered into polysaccharide utilization loci (PUL) (40). PUL7, which is upregulated by sugar beet arabinan and orchestrates the metabolism of this polysaccharide, contains three GH43 and two GH51 enzymes (Eric C. Martens

personal communication). The GH43 endo-arabinanases Bt0360 and Bt0367 display a preference for decorated and linear arabinan, respectively. These enzymes act in consort with the arabinan-specific α-1,2-arabinofuranosidase Bt0369, which is also encoded by PUL7. Bt0369 represents a 'pre-treatment' stage of arabinan processing, converting all the double linked arabinose into single O3 decorations, as well as removing all single O2 arabinose side chains. Thus, Bt0360, Bt0367 and Bt0369 degrade arabinan to arabinooligosaccharides that are either linear or contain single O3-linked arabinose decorations. These oligosaccharides, which are transported into the periplasm by two SusC/ SusD-like complexes (41), are then further metabolized by two GH51 arabinofuranosidases. At least one of the two periplasmic GH51 enzymes is likely to remove the single O3-arabinose decorations to generate linear oligomers, a feature typical of enzymes in this family (42). It is possible that the other GH51 arabinofuranosidase converts the linear oligosaccharides into arabinose, although no GH51 enzyme has been reported to display α-1,5-exo-arabinanase activity.

In *C. japonicus Cj*Abf43A converts the double substitutions into single O3 linked decorations, which are then removed by the extracellular, membrane associated, GH51 arabinofuranosidase (43). The linear arabinan generated is then hydrolyzed, exclusively, to arabinotriose by an endo-processive arabinanase (9). The trisaccharide is metabolized, likely in the periplasm, by the GH43 α-1,5-exo-arabinanase CJA_0806 that displays a moderate preference for arabinotriose. The arabinobiose generated is transported into the cytoplasm where it is hydrolyzed by the second GH43 α-1,5-exo-arabinanase, CJA_3012, which appears to display a preference for the disaccharide.

13

A key feature of arabinan metabolism is the hydrolysis of one of the linkages in double substitutions, which makes the substrate accessible to other exo- and endo-acting enzymes. It is interesting that while *C. japonicus* and *B. thetaiotaomicron* have adopted different strategies to hydrolyze the arabinan backbone, the two bacteria utilize the same enzyme activity to cleave double substitutions, a critical component of the degradative hierarchy. There is a paucity of information on enzymes that hydrolyze double substitutions. Indeed, the only other enzyme known to hydrolyze such structures is the GH43 arabinofuranosidase AXHd3 from the colonic bacterium *Bifidobacterium adolescentis* (44). AXHd3, however, displays broader specificity than Bt3069 or *Cj*Abf43A as it hydrolyzes double substitutions in both arabinoxylan and arabinan. Such an activity is not required by *B. thetaiotaomicron* as the bacterium does not metabolize xylan (29). By contrast *C. japonicus* has an extensive xylan degrading apparatus (36,45), and thus likely contains an additional arabinofuranosidase that targets xylose residues decorated at O2 and O3.

Specificity of CjAbf43A for Sugar Beet Arabinan: The specificity of *Cj*Abf43A for sugar beet arabinan is due to the surface topography of the enzyme. F66, W164 and F234 contribute to the formation of a curved surface cleft around the -1 pocket. This topology is complementary to the extended helical structure of the α-1,5-L-arabinofuranose backbone of sugar beet arabinan, **Figure 4**, explaining why the enzyme targets this polysaccharide. The enzyme is unable to hydrolyse arabinoxylan, which also contains α-1,2-L-arabinofuranose side chains. The helical pitch of arabinoxylans (3-fold screw axis) is shorter than that of arabinans (6-fold), and this constrained structure is arabinoxylan is likely to make steric clashes with the curved surface of the substrate binding cleft. Conversely, the GH43 xylan-specific arabinofuranosidase, *Bs*AXHm2,3 (10), is unable to attack arabinan side chains; despite structural conservation with *Cj*Abf43A at the -1 subsite, the arabinan backbone could not be accommodated by the linear substrate binding cleft that houses the xylan backbone. Thus, it is the architecture of the surface substrate binding cleft, curved in *Cj*Abf43A and linear in *Bs*AXH-m2,3, that dictates the specificity of these enzymes. Apart from F234, residues that line the substrate binding cleft distal to the +1 subsite (which plays a key role in providing the binding energy required for substrate distortion in the active site) bind weakly to the arabinan backbone. This is consistent with the requirement for the polysaccharide backbone to dissociate prior to the release of arabinose from the active site pocket.

Conclusion: The expansion of GH43 enzymes in microorganisms from varied habitats points to a complex array of specificities within this family. Through the analysis of GH43s from *C. japonicus* and *B. thetaiotaomicron* some definable activities have been described, but a function could not be assigned for many of the proteins. This may partly reflect the loss in activity in these proteins due to functional redundancy. It is also likely, however, that the function of active enzymes could not be assigned as the substrates for these biocatalysts were not used, or the activities of these GH43 are only apparent when they are acting in synergy with other degradative enzymes. Despite the problems in assigning functions to all the GH43 proteins, the data presented here indicate that the main chain of arabinan, a target substrate for numerous GH43 enzymes, is degraded by an endo- and exo-mechanism in the gut symbiont and soil saprophyte, respectively. Although distinct, the mechanism of arabinan degradation in the two bacteria display an element of

convergence; both *C. japonicus* and *B. thetaiotaomicron* remove O2-linked arabinose decorations, in the context of single or double substitutions, through the action of an arabinan-specific α-1,2-arabinofuranosidase, an activity that has not previously been reported. The biological rationale for such an activity likely reflects the capacity of the enzyme to convert double substitutions into single decorations, which will then be accessible to the GH51 arabinofuranosidases expressed by these bacteria. The crystal structure of *Cj*Abf43A reveals several topological features, such as a curved substrate binding cleft, a shelf-like structure adjacent to the active site, and the targeting of the only asymmetric oxygen in arabinan, which confer the specificity displayed by the enzyme. The arabinan-specific α-1,2-arabinofuranosidase activity identified here will add to the toolbox of biocatalysts required to degrade and to understand the molecular architecture of plant cell walls.

REFERENCES
1. Burton, R. A., Gidley, M. J., and Fincher, G. B. *Nat Chem Biol* **6**, 724-732
2. Hooper, L. V., Midtvedt, T., and Gordon, J. I. (2002) *Annu Rev Nutr* **22**, 283-307
3. Himmel, M. E., and Bayer, E. A. (2009) *Curr Opin Biotechnol* **20**, 316-317
4. Himmel, M. E., Ding, S. Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W., and Foust, T. D. (2007) *Science* **315**, 804-807
5. Caffall, K. H., and Mohnen, D. (2009) *Carbohydr Res* **344**, 1879-1900
6. Gruppen, H., Kormelink, F. J. M., and Voragen, A. G. J. (1993) *J Cereal Sci* **18**, 111-128
7. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) *Nucl Acids Res* **37**, D233-238
8. Fujimoto, Z., Ichinose, H., Maehara, T., Honda, M., Kitaoka, M., and Kaneko, S. (2010) *J Biol Chem* **285**, 34134-34143
9. McKie, V. A., Black, G. W., Millward-Sadler, S. J., Hazlewood, G. P., Laurie, J. I., and Gilbert, H. J. (1997) *Biochem J* **323**, 547-555
10. Vandermarliere, E., Bourgois, T. M., Winn, M. D., van Campenhout, S., Volckaert, G., Delcour, J. A., Strelkov, S. V., Rabijns, A., and Courtin, C. M. (2009) *Biochem J* **418**, 39-47
11. Brux, C., Ben-David, A., Shallom-Shezifi, D., Leon, M., Niefind, K., Shoham, G., Shoham, Y., and Schomburg, D. (2006) *J Mol Biol* **359**, 97-109
12. Ichinose, H., Yoshida, M., Kotake, T., Kuno, A., Igarashi, K., Tsumuraya, Y., Samejima, M., Hirabayashi, J., Kobayashi, H., and Kaneko, S. (2005) *J Biol Chem* **280**, 25820-25829
13. Gosalbes, M. J., Perez-Gonzalez, J. A., Gonzalez, R., and Navarro, A. (1991) *J Bacteriol* **173**, 7705-7710

14. DeBoy, R. T., Mongodin, E. F., Fouts, D. E., Tailford, L. E., Khouri, H., Emerson, J. B., Mohamoud, Y., Watkins, K., Henrissat, B., Gilbert, H. J., and Nelson, K. E. (2008) *J Bacteriol* **190**, 5455-5463

15. Xu, J., Bjursell, M. K., Himrod, J., Deng, S., Carmichael, L. K., Chiang, H. C., Hooper, L. V., and Gordon, J. I. (2003) *Science* **299**, 2074-2076

16. Charnock, S. J., Bolam, D. N., Turkenburg, J. P., Gilbert, H. J., Ferreira, L. M., Davies, G. J., and Fontes, C. M. (2000) *Biochemistry* **39**, 5013-5021

17. Melrose, J., and Sturgeon, R. J. (1983) *Carbohydr Res* **118**, 247-253

18. Miller, G. L. (1959) *Anal Chem* **31**, 426-428

19. Leslie, A. G. W. (1992) *Joint CCP4 and ESF-EACBM Newsletters on Protein Crystallography* **26**

20. Evans, P. (2006) *Acta Crystallogr* **D62**, 72-82

21. Sheldrick, G. M. (2008) *Acta Crystallogr* **A64**, 112-122

22. Collaborative Computational Project Number 4. (1994) *Acta Crystallogr* **D50**, 760-763

23. Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008) *Nature Protocols* **3**, 1171-1179

24. Cowtan, K. D., and Main, P. (1993) *Acta Crystallogr* **D49**, 148-157

25. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) *J Appl Crystallogr* **40**, 658-674

26. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr* **D53**, 240-255

27. Zwart, P. H., Afonine, P. V., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., McKee, E., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., Storoni, L. C., Terwilliger, T. C., and Adams, P. D. (2008) *Methods Mol Biol*, 419-435

28. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr* **D60**, 2126-2132

29. Cooper, S. W., Pfeiffer, D. G., and Tally, F. P. (1985) *J Clin Microbiol* **22**, 125-126

30. Zhao, S., Wang, J., Bu, D., Liu, K., Zhu, Y., Dong, Z., and Yu, Z. (2010) *Appl Environ Microbiol* **76**, 6701-6705

31. Brunzelle, J. S., Jordan, D. B., McCaslin, D. R., Olczak, A., and Wawrzak, Z. (2008) *Arch Biochem Biophys* **474**, 157-166

32. Proctor, M. R., Taylor, E. J., Nurizzo, D., Turkenburg, J. P., Lloyd, R. M., Vardakou, M., Davies, G. J., and Gilbert, H. J. (2005) *Proc Natl Acad Sci U S A* **102**, 2697-2702

33. Davies, G. J., Wilson, K. S., and Henrissat, B. (1997) *Biochem J* **321** 557-559

34. Swamy, N. R., and Salimath, P. V. (1991) *Phytochemistry* **30**, 263-265

35. Sorensen, H. R., Jorgensen, C. T., Hansen, C. H., Jorgensen, C. I., Pedersen, S., and Meyer, A. S. (2006) *Appl Microbiol Biotechnol* **73**, 850-861

36. Gilbert, H. J. *Plant Physiol* **153**, 444-455

37. Gilbert, H. J., Stalbrand, H., and Brumer, H. (2008) *Curr Opin Plant Biol* **11**, 338-348

38. Nurizzo, D., Turkenburg, J. P., Charnock, S. J., Roberts, S. M., Dodson, E. J., McKie, V. A., Taylor, E. J., Gilbert, H. J., and Davies, G. J. (2002) *Nat Struct Biol* **9**, 665-668

39. Neer, E. J., and Smith, T. F. (1996) *Cell* **84**, 175-178

40. Bjursell, M. K., Martens, E. C., and Gordon, J. I. (2006) *J Biol Chem* **281**, 36269-36279

41. Martens, E. C., Koropatkin, N. M., Smith, T. J., and Gordon, J. I. (2009) *J Biol Chem* **284**, 24673-24677

42. Beylot, M. H., McKie, V. A., Voragen, A. G., Doeswijk-Voragen, C. H., and Gilbert, H. J. (2001) *Biochem J* **358**, 607-614

43.   Beylot, M. H., Emami, K., McKie, V. A., Gilbert, H. J., and Pell, G. (2001) *Biochem J* **358**, 599-605

44.   van den Broek, L. A., Lloyd, R. M., Beldman, G., Verdoes, J. C., McCleary, B. V., and Voragen, A. G. (2005) *Appl Microbiol Biotechnol* **67**, 641-647

45.   Pell, G., Szabo, L., Charnock, S. J., Xie, H., Gloster, T. M., Davies, G. J., and Gilbert, H. J. (2004) *J Biol Chem* **279**, 11777-11788

46.   Edgar, R. C. (2004) *BMC Bioinformatics* **5**, 113

47.   Guindon, S., and Gascuel, O. (2003) *Syst Biol* **52**, 696-704

48.   Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007) *Mol Biol Evol* **24**, 1596-1599

FIGURE LEGENDS

Figure 1   Catalytic activity of Bt3069 and *Cj*Abf43A
The enzymes at 50 nM were incubated with linear arabinan and the reaction products
generated were analyzed by HPAEC.

**Figure 1   Heteronuclear Single Quantum Coherence (HSQC) 1-bond 13C–1H
                correlation spectroscopy.**
Reactions were performed in 5 mM sodium phosphate buffer pH 7.0. *Cj*Abf43A (10 µM) was
incubated with acid hydrolysed sugarbeet arabinan for 1 hour. Chemical shifts are shown in
Table S6. Blue circle indicates the interaction of H2C2 of α-1,2 linked arabinose in single
and double substitutions. Red circle indicates H1C1 interaction of α-1,2 linked arabinose in
single and double substitutions

Figure 3   Phylogenetic tree of *C. japonicus* and *B thetaiotaomicron* GH43 enzymes
The phylogenetic tree was derived from the protein sequences of the GH43 members of *C.
japonicus* and *B. thetaiotaomicron* together with a number of enzymes with known activities
as markers, which are boxed. The two enzymes shown to display arabinan specific α-1,2-
arabinofuranosidase activity are boxed in red with a salmon background. Proteins whose
biochemical properties were not be explored, as they could not be expressed in recombinant
form, are listed as inactive on the tree. Bt3516 contains two GH43 modules which are defined
as Bt3516A (N-terminal GH43 module) and Bt3516B (C-terminal GH43 module). The
predicted GH43 superfamily domain(s) of the genes were aligned using MUSCLE (46). The
tree also contained a selection of enzymes that displayed known activities (boxed). Proteins
that lacked one or more of the catalytic residues contain an asterisk. The alignment used to
construct the phylogenetic tree, shown in **Figure S1**, deployed the maximum likelihood
method in the program PhyML (47). The reliability of the tree was analysed by bootstrap
analysis of 100 resamplings of the dataset. The tree was displayed using MEGA4 (48).

Figure 4 Crystal structure of *Cj*Abf43A
Panel A displays a protein cartoon of *Cj*Abf43A colour ramped from N-terminus (blue) to C-
terminus (red). *Panel B* shows the solvent accessible surface of *Cj*Abf43A in complex with
3^2-Ara-Ara$_3$. The bound α-1,5-L-arabinotriose backbone is shown in yellow (carbon), while
the O3-linked arabinofuranose decoration is in cyan (carbon). The residues that contribute to
the curved shape of the cleft and make hydrophobic and polar contacts with the arabinan
backbone, are shown in stick format. This figure, and other structure figures, was drawn with
PyMol (DeLano Scientific; http://pymol.sourceforge.net/).

Figure 5 The active site of *Cj*Abf43A
Panel A shows the 3D position of amino acids of *Cj*Abf43A that contribute directly to
substrate recognition through polar interactions. The carbon atoms of the amino acids are in
green (carbons in green). In the bound substrate, 3^2-Ara-Ara$_3$ the α-1,5-L-arabinotriose
backbone is shown in yellow (carbon), while the O3-linked arabinofuranose decoration is in
cyan (carbon). The maximum likelihood/σ_a-weighted $2F_{obs} - F_{calc}$ electron density map for 3^2-
Ara-Ara$_3$ is shown in *grey mesh* and contoured at 1σ (0.30 e$^-$/Å^3). The structure is shown in
divergent (wall-eyed) stereo. *Panel B* is a schematic of the residues that make polar contacts
with the substrate.

18

Figure 6 Overlay of the active site of *Cj*Abf43A with *Bs*AXH-m2,3 and Ara*f*43A
The figure shows an overlap of the three enzymes at the -1 active site with *Cj*Abf43A in green, *Bs*AXH-m2,3 (PDB 3C7G) in cyan and Ara*f*43A (PDB 3AKH) in yellow. The silver-coloured arabinose is derived from the Ara*f*43A structure and ethylene glycol (magenta) is from the *Cj*Abf43A structure.

Table 1 Data collection and refinement statistics.

	Selenomethionine	Native	D41A-Ligand
Date collection			
Data collection wavelength (Å)	0.98	0.97	1.00
Space group	$P4_12_12$	$P2_1$	$P4_12_12$
Unit cell parameters (Å, °)	$a = b = 192.65$ $c = 132.5$	$a = 47.29$, $b = 139.27$ $c = 50.46$, $\beta = 116.0$	$a = b = 85.64$ $c = 195.65$
Resolution (Å)	72.17 – 2.99 (3.15 – 2.99)	36.29 – 1.64 (1.73 – 1.64)	44.73 – 1.79 (1.89 – 1.79)
No. of observations	728565	260148	1008728
No of unique reflections	50829	71488	69576
Mean $I/\sigma I$	14.9 (5.2)	11.5 (4.1)	16 (5.4)
R_{merge}	0.22 (0.57)	0.10 (0.40)	0.11 (0.50)
R_{anom}	0.07 (0.29)	-	-
Completeness (%)	100 (100)	99.8 (100)	100 (100)
Anomalous completeness (%)	100 (100)	-	-
Multiplicity	14.3 (14.7)	3.6 (3.7)	14.5 (14.5)
Refinement			
R_{work}/R_{free}	0.200/0.236	0.153/0.184	0169/0.198
RMSD bond lengths (Å)	0.011	0.010	0.007
RMSD bond angles (°)	1.35	1.226	1.142
Ramachandran plot*			
Favoured (%)	94.7	96.5	96.3
Allowed (%)	99.9	99.7	100
Mean B-factor			
Wilson B ($Å^2$)	51.5	11.20	26.50
Main chain ($Å^2$)	21.03	9.60	24.77
Side chain ($Å^2$)	21.82	10.18	28.25
Ligand/water ($Å^2$)	-/-	-/22.00	35.33/36.38
PDB codes	3QED	3QEE	3QEF

Values in parentheses correspond to the highest resolution shell
* Ramachandran statistics generated using molprobity.

Table 2 Kinetic properties of *C. japonicus* and *B. thetaiotaomicron*. GH43 enzymes

Enzyme	Substrate	K_M (mM)	k_{cat} (min⁻¹)	k_{cat}/K_M (min⁻¹ M⁻¹)
CJA_3012	4NPA	0.1 ± 0.02	0.6 ± 0.1	7.0×10^3
	Arabinobiose	0.3 ± 0.03	755 ± 100	2.2×10^6
	Arabinotetraose	1.1 ± 0.85	708 ± 19	6.6×10^5
	Arabinohexaose	1.1 ± 0.08	693 ± 11	6.3×10^5
	Sugar beet arabinan	-	-	$2.0 \times 10^4 \pm 1.1 \times 10^{3*}$
CJA_0806	4NPA	-	-	$2.6 \times 10^4 \pm 2.0 \times 10^3$
	Arabinobiose	7.6 ± 1.27	220 ± 22.1	2.9×10^4
	Arabinotriose	3.4 ± 0.80	357 ± 18.8	1.1×10^5
	Arabinotetraose	4.0 ± 0.2	268 ± 11.0	6.8×10^4
	Arabinohexaose	4.3 ± 1.3	321 ± 41.7	7.4×10^4
	Sugar beet arabinan	-	-	$1.0 \times 10^3 \pm 2.0 \times 10^{1*}$
*Cj*Abf43A (CJA_3018)	4NPA	3.1 ± 0.5	1896 ± 319	6.2×10^5
	Arabinohexaose	-	-	$1.1 \times 10^2 \pm 7.0 \times 10^{-3}$
	Sugar beet arabinan	$0.3 \pm 0.02^*$	308 ± 8	1.1×10^6
Bt0369	4NPA	8.0 ± 2.4	$4.0 (\pm0.7) \times 10^4$	5.0×10^6
	Arabinohexaose	-	-	$3.6 \times 10^3 \pm 3.2 \times 10^2$
	Sugar beet arabinan	$2.1(\pm0.1) \times 10^{-5*}$	2681 ± 246	1.3×10^7
Bt0360	Sugar beet arabinan	-	-	$508.5 \pm 56.9^{**}$
	Linear arabinan	-	-	$57.8 \pm 10.2^{**}$
Bt0367	Sugar beet arabinan	-	-	$52.4 \pm 3.2^{**}$
	Linear arabinan			$323.9 \pm 21.8^{**}$
Bt2852	4NPA	0.4 ± 0.02	155.5 ± 19.2	44.3×10^5
Bt3094	4NPA	2.3 ± 0.2	14.6 ± 6.1	6.3×10^3
Bt3655	4NPA	1.0 ± 0.1	101.3 ± 3.4	1.0×10^5
*Cj*Xyl43A (CJA_3070)	4NPX	-	-	$2.03 \times 10^3 \pm 2.8 \times 10^2$
	Xylotriose	-	-	$4.09 \times 10^3 \pm 2.54 \times 10^2$
	Xylotetraose	-	-	$6.93 \times 10^3 \pm 7.46 \times 10^2$
	Xylohexaose	-	-	$7.15 \times 10^3 \pm 1.37 \times 10^3$

The *C. japonicus* and *B. thetaiotaomicron* enzymes were assayed in 50 mM sodium phosphate buffer pH 7.0 at 25 ºC and 37 ºC, respectively
*Sugarbeet arabinan at 1 mg/ml was enzymatically hydrolyzed to completion to calculate effective substrate concentration. The data were used to determine K_M or k_{cat}/K_M
**Substrate concentration was mg/ml and thus units for k_{cat}/K_M are min⁻¹ mg⁻¹ ml
– indicates parameter could not be determined.

Table 3 Kinetic constants of wild type and variants of *Cj*Abf43A

*Cj*Abf43A	4-NPA			Sugar beet arabinan		
	K_M (mM)	k_{cat} (min^{-1})	k_{cat}/K_M (min^{-1} M^{-1})	K_M (mM)	k_{cat} (min^{-1})	k_{cat}/K_M (min^{-1} M^{-1})
Wild type	3.1 ± 0.51	1896 ± 320	6.2×10^5	0.3 ± 0.02	308 ± 8	1.1×10^6
D41A	-	-	-	-	-	-
D168A	-	-	-	-	-	-
E215A	-	-	-	-	-	-
F66A	-	-	$2.2 \times 10^5 \pm 2.99 \times 10^4$	1.4 ± 0.1	427 ± 11	3.2×10^5
F67A	-	-	$1.6 \times 10^3 \pm 2.58 \times 10^2$	1.6 ± 0.5	14 ± 2	9.0×10^3
D101A	2.5 ± 0.2	524 ± 43	2.1×10^5	0.2 ± 0.02	266 ± 27	1.5×10^6
W103A	-	-	$2.5 \times 10^2 \pm 5.3 \times 10^1$	-	-	-
R121A	4.7 ± 1.6	134 ± 46	2.9×10^4	0.7 ± 0.08	47.27 ± 2	6.7×10^4
F129A	3.4 ± 1.2	906 ± 245	2.6×10^5	0.5 ± 0.09	293 ± 18	5.8×10^5
W164A	-	-	$6.4 \times 10^4 \pm 1.1 \times 10^4$	-	-	$3.6 \times 10^4 \pm 4.0 \times 10^3$
I167A	-	-	1.9×10^3	-	-	$2.5 \times 10^3 \pm 3.9 \times 10^2$
N185A	0.9 ± 0.05	1676 ± 147	1.8×10^6	-	-	$3.0 \times 10^2 \pm 3.3 \times 10^1$
T186A	4.9 ± 0.2	727 ± 98	1.5×10^5	0.4 ± 0.03	260 ± 7	3.4×10^5
T186W	11.2 ± 0.4	586 ± 1	5.3×10^4	-	-	33 ± 1.5
T214A	3.5 ± 0.1	717 ± 57	2.1×10^5	0.8 ± 0.08	154 ± 8	2.0×10^5
T214W	4.3 ± 0.5	2315 ± 403	5.4×10^5	4.1 ± 0.4	27 ± 2	6.6×10^3
F234A	9.4 ± 0.8	354 ± 17	3.8×10^4	-	-	$2.5 \times 10^3 \pm 3.9 \times 10^3$
H267A	-	-	$8.8 \times 10^5 \pm 1.9 \times 10^4$	0.8 ± 0.2	30 ± 0.3	3.7×10^4
Q292A	2.9 ± 0.5	797 ± 210	2.8×10^5	0.5 ± 0.04	329 ± 1	6.5×10^5
Y293A	-	-	$1.7 \times 10^4 \pm 2.2 \times 10^2$	1.4 ± 0.2	415 ± 69	3.1×10^5
T186W/T21W	3.9 ± 1.0	606 ± 252	1.6×10^5	-	-	$8.8 \times 10^2 \pm 1.8 \times 10^2$

Enzyme concentrations varied from 5 nM–10 μM. Reactions were carried out at 25 °C in 50 mM sodium phosphate buffer pH 7.0. Kinetic parameters were determined by non-linear regression analysis.
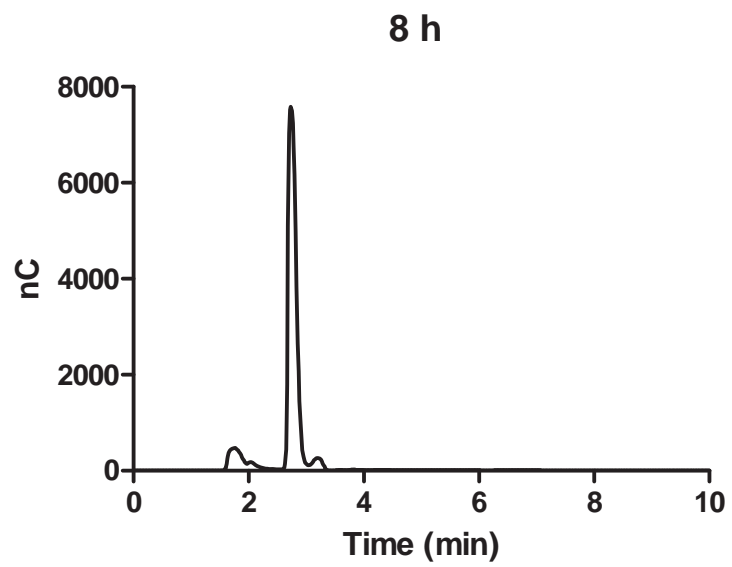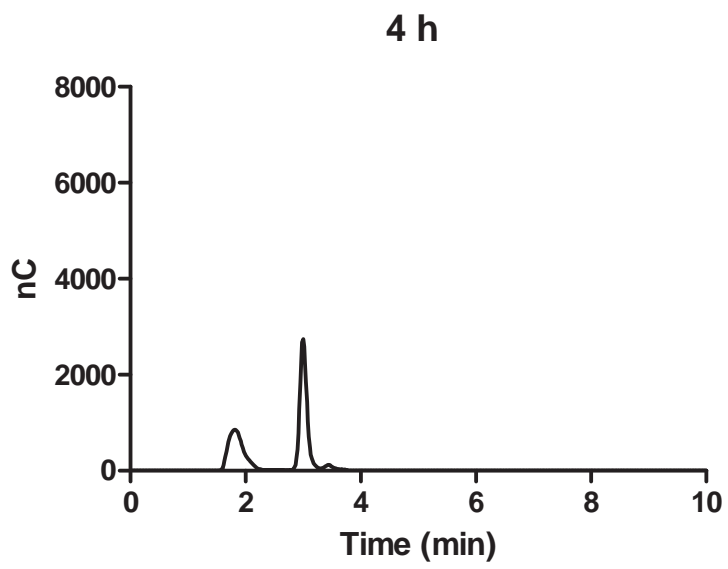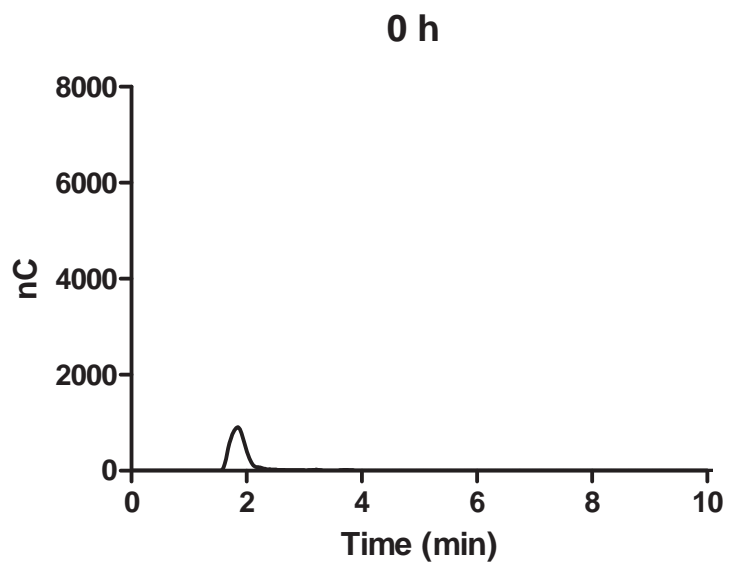– indicates parameter could not be determined.
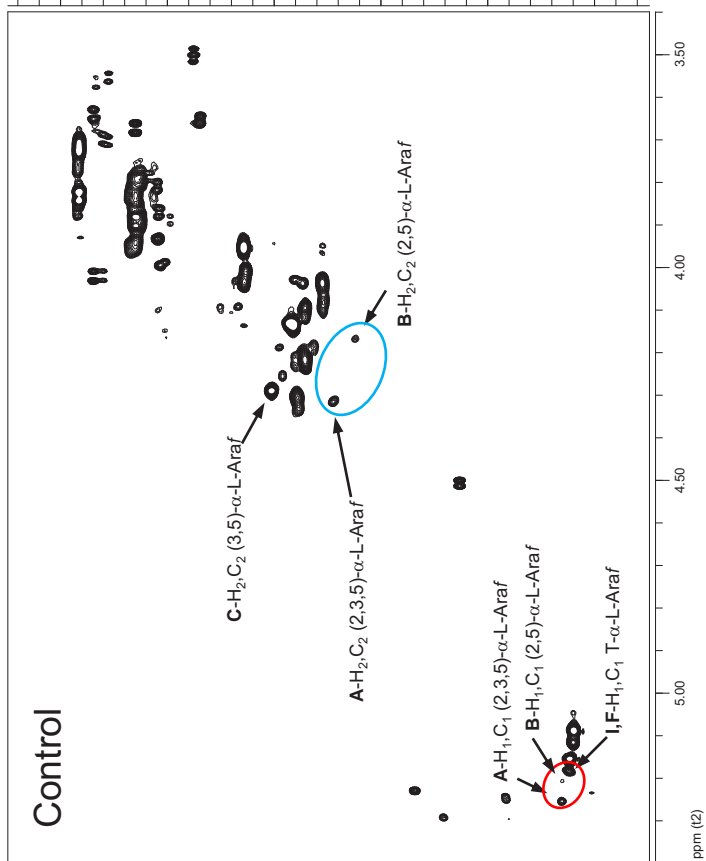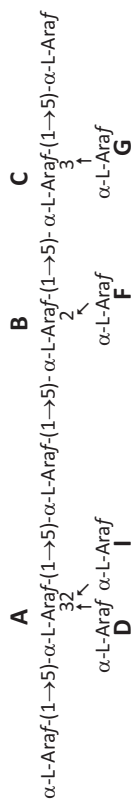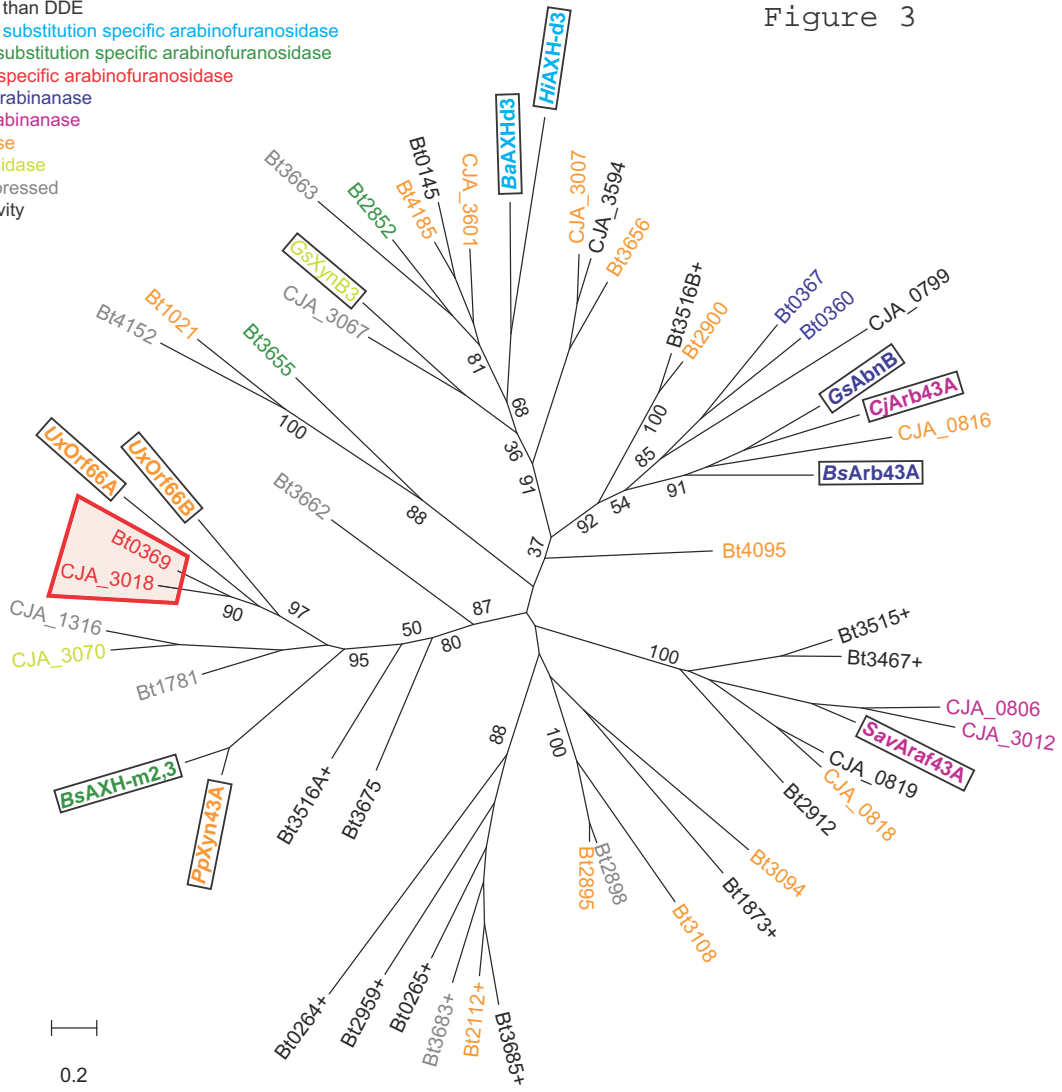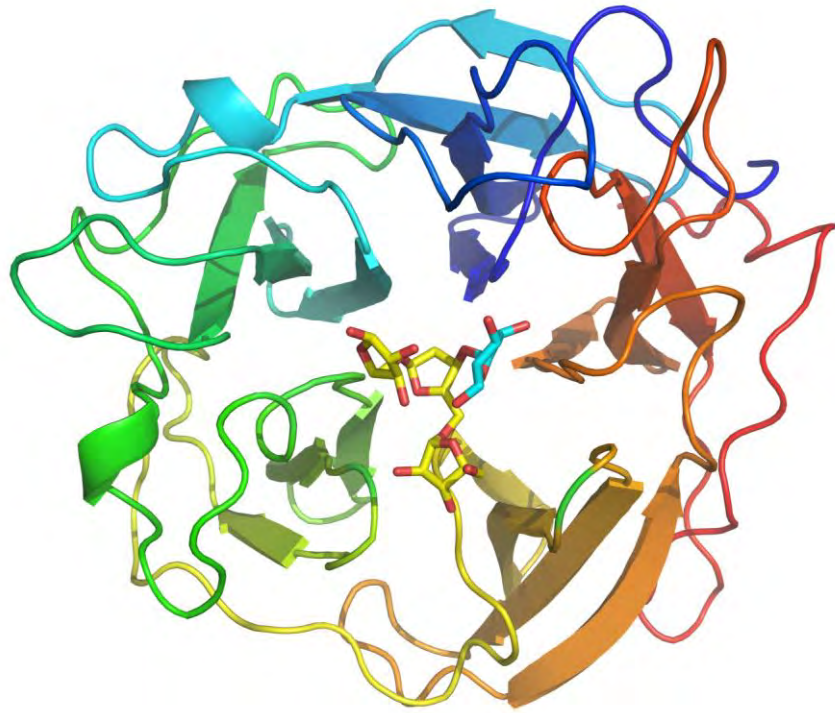
22
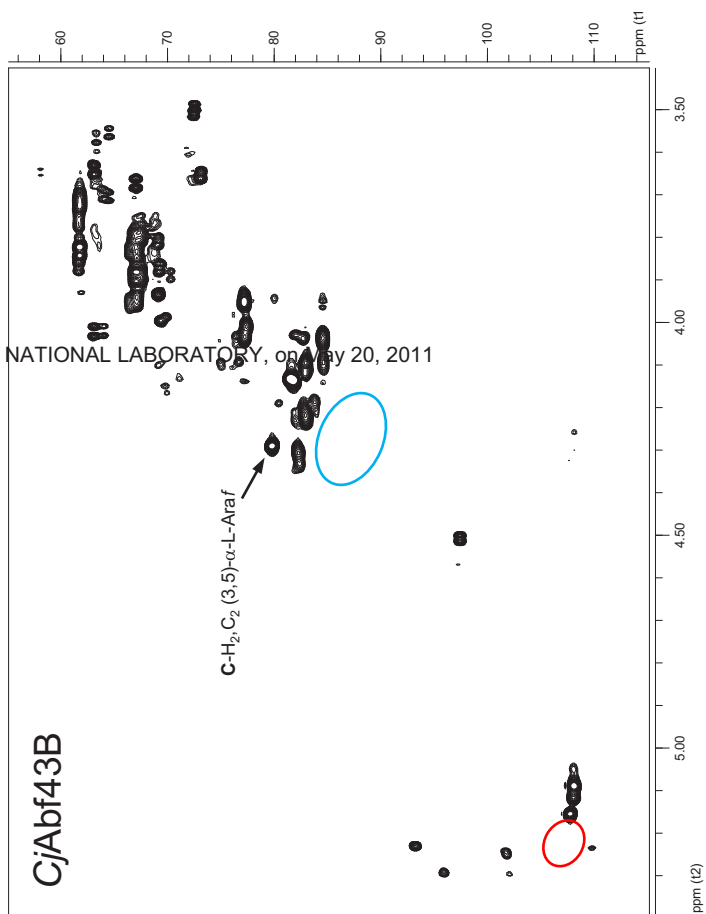
Figure 1

Figure 2

Figure 3

+ lacks catalytic residues or displays a different
  motif than DDE
double substitution specific arabinofuranosidase
single substitution specific arabinofuranosidase
α(1,2)-specific arabinofuranosidase
endo-arabinanase
exo-arabinanase
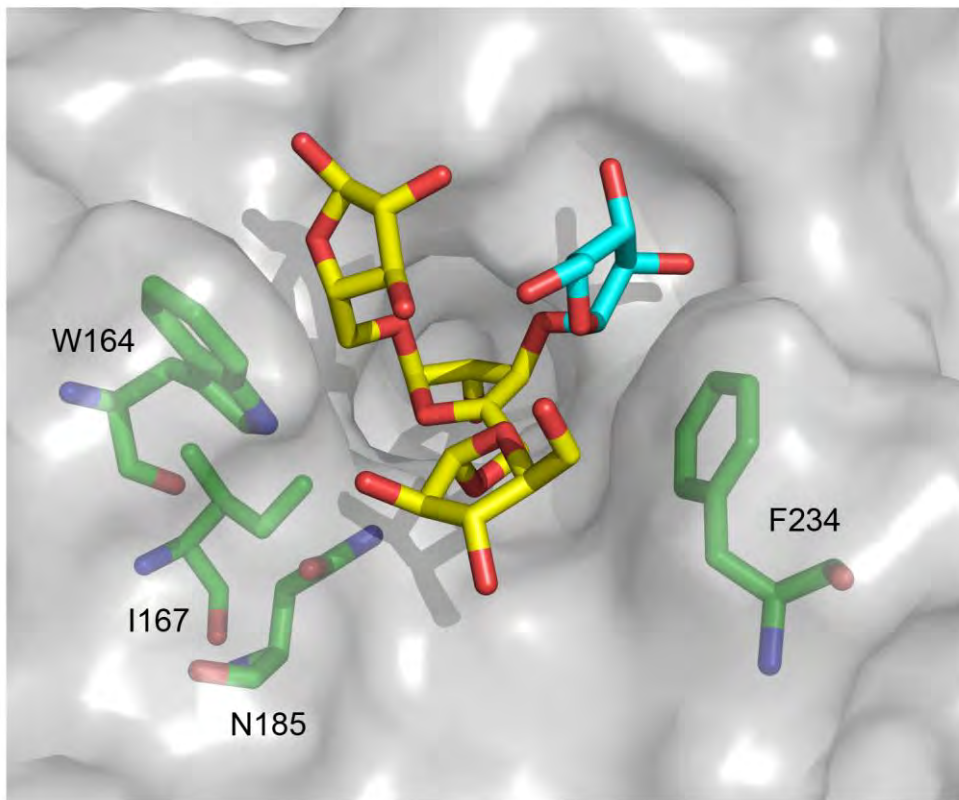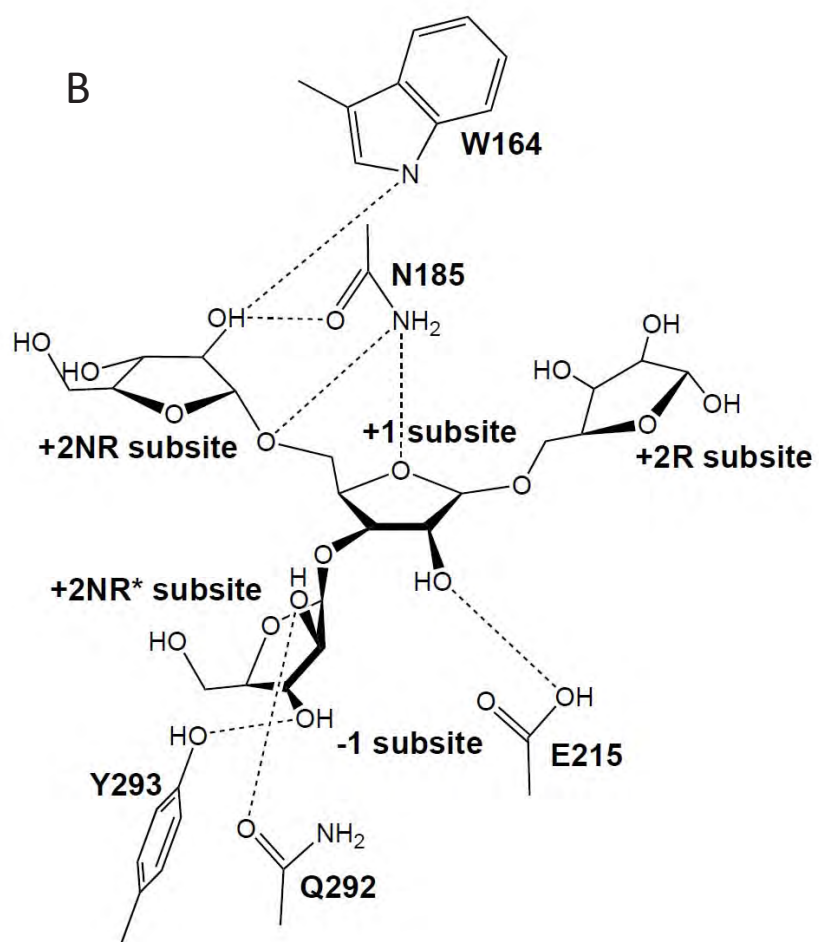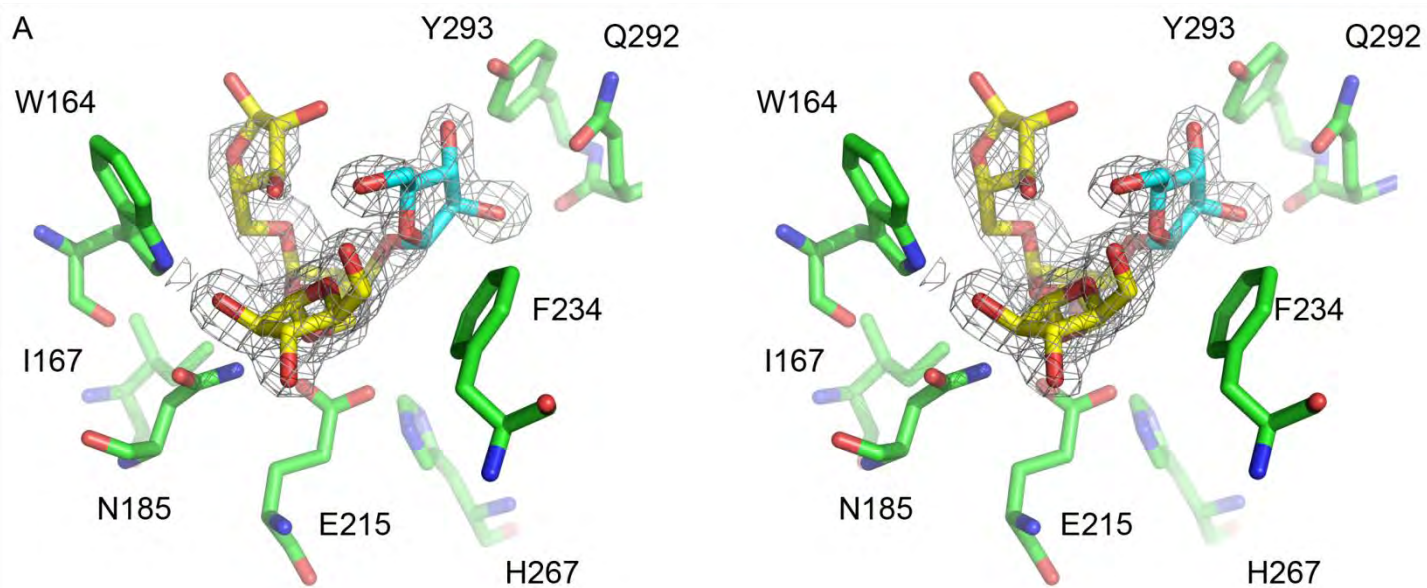xylanase
β-xylosidase
not expressed
no activity

A

B

Figure 4

Figure 5
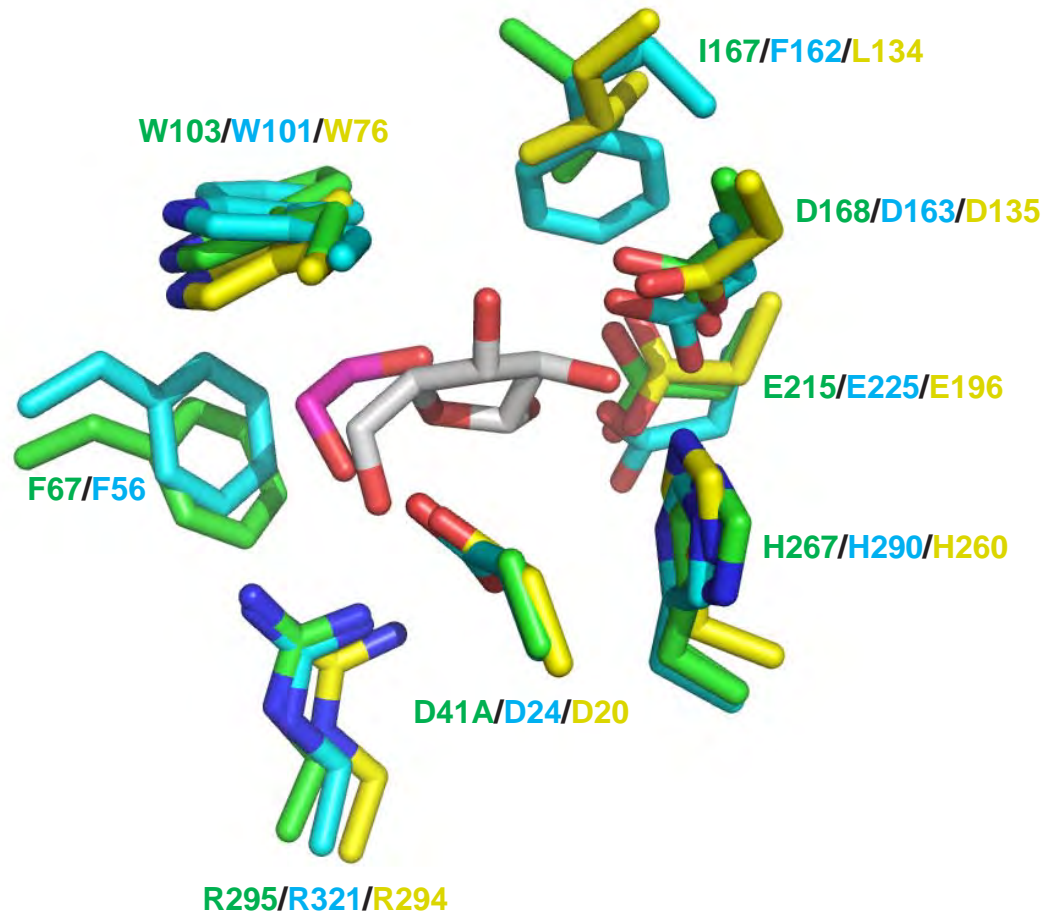
I167/F162/L134

W103/W101/W76

D168/D163/D135

E215/E225/E196

F67/F56

H267/H290/H260

D41A/D24/D20

R295/R321/R294

Figure 6