The Structure and Function of an Arabinan-specific α-1,2-Arabinofuranosidase Identified from Screening the Activities of Bacterial GH43 Glycoside Hydrolases**

Received for publication, December 23, 2010, and in revised form, February 16, 2011. Published, JBC Papers in Press, February 21, 2011, DOI 10.1074/jbc.M110.215962

Alan Cartmell†1, Lauren S. McKee†1, Maria J. Peña§1, Johan Larsbrink‡, Harry Brumer‡, Satoshi Kaneko‡, Hitomi Ichinose§, Richard J. Lewis1, Anders Vikso-Nielsen§**, Harry J. Gilbert†**‡, and Jon Marles-Wright§‡

From the §Institute for Cell and Molecular Biosciences, Newcastle University, The Medical School, Newcastle upon Tyne NE2 4HH, United Kingdom, the †Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602, the ♦School of Biotechnology, Royal Institute of Technology, AlbaNova University Centre, 10691 Stockholm, Sweden, the ‡Food Biotechnology Division, National Food Research Institute, 2-1-12 Kannondai, Tsukuba, Ibaraki 305-8642, Japan, and **Novozymes A/S, Krogshoejvej 36, 2880 Bagsvaerd, Denmark

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, whereas B. thetaiotaomicron deploys a combination of endo- and side-chain-cleaving glycosidase 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, CjAbf43A, displays a five-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 CjAbf43A for α-1,2-L-arabinofuranose side chains is conferred by a polar residue that orients the arabinan backbone such that O2 arabinose decorations are directed into the active site pocket. A shelflike 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.

The plant cell wall comprises 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. Not only is the release of soluble sugars, principally monosaccharides, from the plant cell wall required to maintain microbial ecosystems, but the volatile fatty acids generated by these microorganisms 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 pretreatment 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 fibers 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 is a pectic polysaccharide found in primary cell walls, which plays an important role in plant cell growth and tissue development. Rhamnogalacturonan I contains a backbone of the repeating disaccharide 4-α-L-rhamnopyranosyl acid-1,2-α-L-rhamnopyranose-1, which is extensively decorated with a variety of oligosaccharides, of which arabinan is the most abun-
Arabinan-specific Arabinofuranosidase

dant (5). Arabinan consists of an α-1,5-1-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-xyllose 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 data base, glycoside hydrolases are grouped into families based on sequence, structural, and mechanistic similarities (7). Enzymes that hydrolyze arabinofuranose-linked glycosidic bonds are predominantly located in glycoside hydrolase families GH43, GH51, GH54, and GH62 (7). Of particular note is GH43, which, in addition to containing arabinofuranosidases and arabinanases (8–10), also contains β-xyllosidases (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 of the major components of the cell wall, contains 14 genes encoding GH43 enzymes (14), whereas Bacteroides thetaiotaomicron, a human colonic symbiont that targets pectins and mammalian glycans, has the genetic capacity to synthesize 33 different GH43 hydrolases (15). Superficially, GH43 appears to differ from mammalian glycans, has the genetic capacity to synthesize 33 different strategies used by the two microorganisms to degrade arabinose-containing polysaccharides. The data underline the similarity of the two systems is an arabinan-specific α-1,2-arabinofuranosidase, an enzyme that has not previously been reported. It is likely 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, CjAbf43A, 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, confers 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 Tables S1 and S2. The PCR products were cloned into the expression vectors pSET, pET21d, pET22b, pET28b, and pET32b to generate the plasmids listed in supplemental Tables S1 and S2. All of the recombinant enzymes contained either a His6 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 midexponential phase (A600 = 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 His6 tag were purified by immobilized metal ion affinity chromatography (IMAC) using TalonTM 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 10 mM Tris/HCl buffer, pH 8.0, containing 10 mM reduced glutathione. Enzymes used in biochemical assays were dialyzed 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 dialyzed against ultrapure water. To produce a selenomethionine derivative of CjAbf43A, the E. coli methionine auxotroph B834 (DE3) containing pAC12 (encodes CjAbf43A) 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 dialyzed extensively against 10 mM DL-dithiothreitol.

Site-directed Mutagenesis.—Site-directed mutagenesis was conducted using the PCR-based QuickChange site-directed mutagenesis kit (Stratagene) according to the manufacturer’s instructions, using pAC12 as the template and the primer pairs listed in supplemental Table S3.

Enzyme Assays.—The substrates 4-nitrophenyl-α-L-arabinofuranoside (4NPA)3 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 10,500 M cm–1. When using polysaccharides such as arabinan as the substrate, arabinose release was monitored using galactose dehydrogenase (Megazyme), an enzyme that catalyzes the oxidation of galactose/arabinose with concomitant reduction of NAD+ to NADH (17). The concentration of NADH generated was quantified by measuring ΔA340 nm using a molar extinction coefficient of 6230 M 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 Na2CO3. The addition of Na2CO3 raises the pH to 11, where 4-nitrophenolate has an extinction coefficient of 24,150 M cm–1. Reducing sugar assays were carried out using 3,5-dinitrosalicylic acid

3 The abbreviations used are: 4NPA, 4-nitrophenyl-α-L-arabinofuranoside; HPAEC, high performance anion exchange chromatography.
Arabinan-specific Arabinofuranosidase

according to Ref. 18. At various time points, a 500-µl aliquot of the reaction was taken and stopped by the addition of an equal volume of 3,5-dinitrosalicylic acid reagent. After boiling for 20 min, the A$_{275}$ nm values were determined and compared with 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 S.E. 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 analyzed using a CARBOPAC$^TM$ PA-100 anion exchange column (Dionex) equipped with a CARBOPAC$^TM$ 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 0.2 µl and flow rate of 1.0 ml/min, with sugars detected by pulsed amperometric detection. The pulsed amperometric detection settings were as follows: E$_1$ = +0.05, E$_2$ = +0.6, and E$_3$ = −0.6. Typical elution conditions were as follows: 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 50–60 min, 66 mM NaOH. Appropriate oligosaccharides were used as standards at a concentration of 50 µM. Samples to be analyzed were centrifuged at 13,000 × g for 10 min. Data were collected and manipulated using the Chromeleon$^TM$ chromatography management system version 6.8 (Dionex) via a Chromeleon$^TM$ Server (Dionex). Final graphs were drawn with Prism 4.02 (GraphPad).

Thin Layer Chromatography—TLC plates (silica gel 60, 20 × 20, Merck) spotted with samples (two 1-µl drops) were run in 1-butanol/acetate acid/water in a ratio of 2:1:1 (v/v/v). Visualization of migrated sugars was achieved by immersing the TLC plates in developer (sulfuric acid/ethanol/water (3:70:20, v/v/v), 20% 1-butanol/acetic acid/water in a ratio of 2:1:1 (v/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 neutralized 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 CjAbf43A 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$_2$O (99.9%; Cambridge Isotope Laboratories), and $^1$H 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 versions of COSY, heteronuclear single quantum coherence, and 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$_t$ = 2.225, δ$_C$ = 30.89). Data were processed using MestRe-C software (Universidad de Santiago de Compostela, Spain).

Crystalization, Data Collection, Solution, and Refinement of CjAbf43A—A selenomethionine derivative of CjAbf43A in ultrapure water was crystallized at 30 mg/ml in 1.5 M ammonium sulfate and 100 mM Tris/HCl buffer, pH 7.5. Crystals, which grew overnight at 20 °C, were transferred to a cryo-protectant solution of the crystallization condition supplemented with 25% (v/v) glycerol. Native protein in water was crystallized at 10 mg/ml in 13% PEG 3350 and 150 mM sodium acetate at 20 °C, with crystals reaching maximum size in 4 weeks. The purified catalytic base mutant was dialyzed against 50 mM sodium chloride and crystallized 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 2-month period. These two crystal forms were cryoprotected with the addition of 25% (v/v) ethylene glycol to the respective crystallization 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 102 and at the Advanced Photon Source on beamline 22-BM, respectively.

Because no appropriate molecular replacement model was available for CjAbf43A in the Protein Data Bank, the structure was solved by single wavelength anomalous dispersion 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 single wavelength anomalous dispersion data set was collected at the optimal peak wavelength of 0.979 Å. The data were integrated using Mosfim (19) and scaled and merged in Scala (20). The single wavelength anomalous dispersion data set 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$^2$/sig), 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 2-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$_{crys}$ of 0.200 and R$_{free}$ of 0.236 and was used as a molecular replacement model for the native high resolution data set.

The native and the D41A ligand-enzyme complex data were integrated using Mosflm (19) and scaled and merged with Scala (20). The scaled data for the two crystal forms were input into Phaser (25), and the structures were solved by molecular replacement using the selenomethionine structure as the search model. Refinement of all structures was performed using Refmac5 (26) and phenix.reduce (27) interspersed with manual
Arabinan-specific Arabinofuranosidase

| TABLE 1 | Data collection and refinement statistics |
|---------|-----------------------------------------|
| Selenomethionine | Native | D41A-Ligand |
| Data collection | | | |
| Data collection wavelength (Å) | 0.98 | 0.97 | 1.00 |
| Space group | P4₁,2,2 | P2₁ | P4₁,2,2 |
| Unit cell parameters (Å) | a = b = 192.65, c = 132.5 | a = 47.29, b = 139.27, c = 50.46 | a = b = 85.64, c = 195.65 |
| Unit cell parameters (degrees) | | | |
| 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 | 728,565 | 260,148 | 1,008,728 |
| No. of unique reflections | 50,829 | 71,488 | 69,576 |
| Mean I/σI | 14.9 (5.2) | 11.5 (4.1) | 16 (5.4) |
| Rmerge | 0.22 (0.57) | 0.10 (0.40) | 0.11 (0.50) |
| Rfree | 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 | | | |
| Work/Rfree | 0.200/0.236 | 0.153/0.184 | 0.169/0.198 |
| Root mean square deviations | | | |
| Bond lengths (Å) | 0.011 | 0.010 | 0.007 |
| Bond angles (degrees) | 1.35 | 1.226 | 1.142 |
| Ramachandran plot | | | |
| Allowed (%) | 94.7 | 96.5 | 96.3 |
| Favored (%) | 99.9 | 99.7 | 100 |
| Mean B-factor | 51.5 | 11.20 | 26.50 |
| Wilson B (Å²) | 21.03 | 9.60 | 24.77 |
| Main chain (Å²) | 21.82 | 10.18 | 28.25 |
| Side chain (Å²) | 22.00 | -/22.00 | 35.33/36.38 |
| Ligand/water (Å²) | 3QED | 3QEE | 3QEF |
| Protein Data Bank codes | | | |

a Values in parentheses correspond to the highest resolution shell.

b Ramachandran statistics generated using molprobity.

c 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 His267. One of these conformations, with a χ₁ angle of −65°, points toward the bound oligosaccharide and is stabilized in this position by the presence of a solute network that coincides spatially with the other rotamer position of His267 (χ₁ = −161°), which is oriented away from the oligosaccharide. In the apo structure, His267 adopts a single conformation, and the deposited model coordinates 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 metabolize 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 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 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 supplemental Tables S4 and S5, show that six enzymes, three from both B. thetaiotaomicron and C. japonicus, displayed significant 4NPase activity, two B. thetaiotaomicron enzymes displayed endo-arabinanase activity, and 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, respectively (designated hereafter as CJAAbf43A), hydrolyzed 4NPA and released arabinose from sugar beet arabian but was not significantly active against linear arabinan or any of the other polysaccharides tested (Table 2 and supplemental Tables 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, whereas 17 of the proteins displayed trace endo-xylanase 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 endo-xylanase.
Characterization of Arabinan-degrading Enzymes

C. japonicus—Biochemical analysis of the two \exo-\alpha-1,5-arabinofuranosidases, CJA_3012 and CJA_0806 (Table 2), showed that both enzymes displayed the highest catalytic activity against linear arabinan and arabinooligosaccharides, releasing exclusively arabinose, probably 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 \alpha-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 was assessed. The data show that CJA_3012 is \sim 300-fold more active against arabinobiase 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 degree of polymerization of >2, indicating that the arabinofuranosidase contains only two kinetically significant subsites. Indeed, the longer arabinooligosaccharides appeared to be hydrolyzed more slowly than the disaccharide. The similar activity displayed by CJA_0806 against 4NPA and arabinobiase indicates that the +1 subsite exhibits limited specificity for arabinose, although the enzyme appears to be optimized to hydrolyze arabinotriose because the trisaccharide is a better substrate than either di- or tetrasaccharides. 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 \beta-anomer of the furanose sugar.

An intriguing feature of the two \exo-\alpha-1,5-arabinanases is that although they display similar catalytic efficiencies against 4NPA, the \Km and \kc of CJA_0806 were much higher than for CJA_3012. The sequence identity between these two enzymes is 49%. It is possible that the +1 site of CJA_3012 binds much tighter to the aglycone departure from the active site, which would impact upon the catalytic rate.

Kinetic analysis of CJAAbf43A showed that the enzyme displayed similar catalytic efficiency against 4NPA and sugar beet arabinan, releasing monomeric arabinose (Fig. 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 \beta-glucans (cellulose, \beta-1,4,1,3 mixed linkage glucans, and laminarin) (supplemental Table S4). CJAAbf43A was \sim 10-fold less active against arabinohexaose (an \alpha-1,5-linked arabinofuranose hexasaccharide) than sugar beet arabinan (Table 2). Sugar beet arabinan consists of an \alpha-1,5-linked \alpha-arabinofuranose backbone that is 60% monosubstituted with \alpha-1,3-\alpha-arabinofuranose side chains and, less frequently, with single \alpha-1,2 \alpha-arabinofuranose (5). Some backbone residues are doubly substituted with both \alpha-1,2 and

### Table 2

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

The C. japonicus and B. thetaiotaomicron enzymes were assayed in 50 mM sodium phosphate buffer, pH 7.0, at 25 and 37 °C, respectively.

| Enzyme           | Substrate | \( K_m \) (mg/ml) | \( k_{cat} \) (min\(^{-1}\)) | \( k_{cat}/K_m \) (min\(^{-1}\) mg\(^{-1}\) ml) |
|------------------|-----------|------------------|-----------------|------------------|
| CJA_3012         | 4NPA      | 0.1 ± 0.02       | 0.6 ± 0.1       | 7.0 \times 10^3  |
|                  | Arabinobiase | 0.3 ± 0.03     | 755 ± 100       | 2.2 \times 10^4  |
|                  | Arabinotetraose | 1.1 ± 0.08   | 708 ± 19        | 6.6 \times 10^4  |
|                  | Arabinohexaose | 1.1 ± 0.08    | 693 ± 11        | 6.3 \times 10^4  |
| CJA_0806         | 4NPA      | —                | —               | —               |
|                  | Arabinobiase | 7.6 ± 1.27     | 220 ± 22.1      | 9.2 \times 10^4  |
|                  | Arabinotriose | 3.4 ± 0.80     | 357 ± 18.8      | 1.1 \times 10^5  |
|                  | Arabinotetraose | 4.0 ± 0.2     | 268 ± 11.0      | 6.8 \times 10^4  |
|                  | Arabinohexaose | 4.3 ± 1.3      | 321 ± 41.7      | 7.4 \times 10^4  |
| CJAAbf43A (CJA_3018) | 4NPA     | 3.1 ± 0.5       | 1896 ± 319      | 6.2 \times 10^3  |
|                  | Arabinohexaose | —                | —               | —               |
|                  | Sugar beet arabinan | 0.3 ± 0.02    | 308 ± 8         | 1.1 \times 10^5  |
| Bt0369           | 4NPA      | 8.0 ± 2.4       | 4.0 ± 0.7 × 10^4 | 5.0 \times 10^4  |
|                  | Arabinohexaose | —                | —               | —               |
|                  | Sugar beet arabinan | 2.1 ± 0.1 × 10^-5 | 2681 ± 246 | 1.3 \times 10^5  |
| Bt0360           | Sugar beet arabinan | —                | —               | —               |
|                  | Linear arabinan | —                | —               | 5085 ± 66^9      |
|                  | Sugar beet arabinan | —                | —               | 57.8 ± 1.0^2     |
| Bt0367           | Sugar beet arabinan | —                | —               | 524 ± 3.2^9      |
|                  | Linear arabinan | —                | —               | 323.9 ± 21.8^6   |
| Bt2852           | 4NPA      | 0.4 ± 0.02      | 1555 ± 19.2     | 44.3 \times 10^4 |
| Bt3094           | 4NPA      | 2.3 ± 0.2       | 14.6 ± 6.1      | 6.3 \times 10^4  |
| Bt3655           | 4NPA      | 1.0 ± 0.1       | 1013 ± 3.4      | 1.0 \times 10^5  |
| CJAAbf43A (CJA_3070) | 4NPA     | —                | —               | —               |
|                  | Xylohexaose | —                | —               | —               |

\( ^* \) parameter could not be determined.

\( ^{\dagger} \) Sugar beet arabinan at 1 mg/ml was enzymatically hydrolyzed to completion to calculate effective substrate concentration. The data were used to determine \( K_m \) or \( k_{cat}/K_m \).

\( ^{\ddagger} \) Substrate concentration was mg/ml; thus, activity units for \( k_{cat} \) are \text{min}^{-1} \text{mg}^{-1} \text{ml}.

(13), although true GH43 xylanases have recently been identified (30).
The high activity of the enzyme against sugar beet arabinan and the low activity against \( -1,5 \) arabinoo- 
ligosaccharides suggest that \( \text{CjAbf43A} \) hydrolyzes \( -1,2 \) and/or \( -1,3 \) side chains present in sugar beet arabinan. To further explore \( \text{CjAbf43A} \) specificity, arabinan oligosaccharides were subjected to two-dimensional NMR spectroscopy analysis after incubation without (control) or with the enzyme. Resonances in the two-dimensional spectra were assigned to specific arabinosyl residues on the basis of their scalar coupling patterns and chemical shifts and by comparison with the published data (34) (supplemental Table S6). The heteronuclear single quantum coherence NMR spectrum of the control contained two distinct cross-peaks at \( \delta 4.170, 87.6 \) and \( \delta 4.314, 85.6 \) that disappeared after the treatment with the enzyme (Fig. 2). The resonance at \( \delta 87.6 \) is diagnostic of \( -1,2 \) side chains because the chemical shift of the C2 of the \( -1,5 \)-linked L-arabinofuranose residues shifted markedly downfield when the residues were substituted at the O2 position. The resonance at \( \delta 85.6 \) corresponded to arabinosyl residues substituted at the O2 and O3 positions. 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 \( \text{CjAbf43A} \) hydrolyzes \( -1,2 \)-, \( -1,3 \)-, \( -1,5 \)-linked \( \alpha \)-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 \( \text{HiAXHd3} \), an arabinofuranosidase that targets exclusively O3 linkages in double substitutions of arabinan and arabinoxylan (35).4 Arabinan that had been pretreated with \( \text{CjAbf43A} \) was not hydrolyzed by \( \text{HiAXHd3} \). \( \text{CjAbf43A} \), however, was able to release arabinose residues from arabinan pretreated with \( \text{HiAXHd3} \). These results are consistent with the NMR data in showing that \( \text{CjAbf43A} \) targets O2 linkages in backbone arabinose residues that are singly or doubly substituted. \( \text{CjAbf43A} \) releases 33% more arabinose than \( \text{HiAXHd3} \), indicating a ratio of O2 single and O2 + O3 double substitutions of 1:3.

**B. thetaiotaomicron**—Although both Bt0360 and Bt0367 display *endo*-arabinanase activity, they differ in specificity for linear and branched arabinan. Bt0360 shows a preference for branched arabinan, whereas Bt0367 exhibits higher activity against linear arabinan (Table 2). Bt0369, an arabinan-specific arabinofuranosidase, did not release arabinose from arabinan pretreated with \( \text{CjAbf43A} \), whereas 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 \( \text{CjAbf43A} \) display the same substrate specificity, consistent with their structural similarity (the primary sequences 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 Fig. 3, show that enzymes that displayed arabinanase and \( -1,5 \)-exo-arabinanase activity are in clades with glycoside hydrolases that exhibit similar activities. By contrast, and consistent with their novel activity, \( \text{CjAbf43A} \) (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 xyla-

---

4 L. McKee, unpublished data.

---
Arabinan-specific Arabinofuranosidase

The crystal structure of CjAbf43A 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 CjAbf43A were in the tetragonal space group P41212, whereas the native crystals were in the monoclinic space group P21. The structures each consist of two molecules of CjAbf43A 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 root mean square deviations of ~0.2 Å on 295 matched Ca atoms. CjAbf43A displays a five-bladed β-propeller fold (Fig. 4), typical of GH43 enzymes (8, 10, 11, 31). 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 center 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 CjAbf43A, 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 Asp78 in the N-terminal blade 1 and several residues in blade 5. CjAbf43A lacks a C-terminal β-jelly roll domain seen in several exo-acting GH43 enzymes (8, 10, 11, 31).

Structural Comparison of CjAbf43A with GH43 Enzymes

Structural comparison of CjAbf43A by secondary structure matching of the global structure revealed that the Bacillus subtilis xylan-specific GH43 arabinofuranosidase (10), BsAXH-m2,3 (Protein Data Bank entry 3C7G), was the closest structural homologue, with a root mean square deviation of 1.4 Å over 271 aligned Ca atoms and a sequence identity of 37%. The
other most meaningful structural homologue was the α-L-arabinanase (38) from *C. japonicus* (Protein Data Bank entry 1GYD), which shares 24% sequence identity with CjAbf43A and can be superimposed on 237 matched Ca atoms, yielding a root mean square deviation of 1.9 Å. Comparisons of the active sites of CjAbf43A with BsAXH-m2,3 and other GH43 enzymes are discussed in brief below.

**The Interaction of CjAbf43A with Substrate**

CjAbf43A was also crystallized in the presence of arabinofuranose (Megazyme). To trap the complex between the protein and the ligand, we utilized 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°-Ara-Ara₂) in one of the two CjAbf43A molecules in the asymmetric unit, chain B. The second CjAbf43A 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, so instead an α-1,5-arabinotriose has been modeled in this molecule.

The surface representation of CjAbf43A (Fig. 4) reveals a deep pocket in the center 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 shelllike 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 Asn₁₈⁵, a residue that plays a key role in arabinan recognition (see below), only makes a key hydrogen bond with the substrate in the modeled orientation. Furthermore,

![Phylogenetic tree of C. japonicus and B. thetaiotaomicron GH43 enzymes](image-url)

**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. 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 a +. The alignment used to construct the phylogenetic tree, shown in supplemental Fig. S1, deployed the maximum likelihood method in the program PhyML (47). The reliability of the tree was analyzed by bootstrap analysis of 100 resamplings of the data set. The tree was displayed using MEGA4 (48).
Thus, the bound ligand occupies subsites dating the arabinose-linked O3 to the designated subsites extending to the non-reducing end of the polymer (from the active site) and the backbone arabinose at \textsuperscript{1}H11001. How the enzyme has plasticity for presentation of the arabinan backbone has been built correctly. The O3-linked arabinose in the 32-Ara-Ara3 ligand is located directly into the active site pocket, confirming that the orientation of the arabinan backbone is shown in stick format. This figure and other structure figures were drawn with PyMOL (DeLano Scientific, LLC).

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 32-Ara-Ara3 ligand is located on the shelf adjacent to the active site pocket, demonstrating how the enzyme has plasticity for \( \alpha-1,2,\text{-arabinofuranose side chains in single and double substitutions.} \)

The subsite topology of \textit{CjAbf}43A 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 toward the reducing end of the arabinan backbone (from the \(+1\) subsite) are defined as \(+2R, +3R\) etc., whereas 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 the \(+2NR\) 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 arabinosyl at the \(+2R\) subsite is sandwiched between the aromatic residues, Phe\textsuperscript{66} and Trp\textsuperscript{164}. 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, suggesting that the tryptophan does make a contribution to arabinan recognition (Table 3).

At the \(+1\) subsite, the N82 of Asn\textsuperscript{185} interacts with the endocyclic ring oxygen and the glycosidic oxygen that links sugars at the \(+1\) and \(+2NR\) subsites. Oe2 of Glu\textsuperscript{215}, 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 Glu\textsuperscript{215} 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 Asn\textsuperscript{185} is not able to interact with the aryl-glycoside because 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 Asn\textsuperscript{185} play a role in defining the specificity of \textit{CjAbf}165A for the \(O2\) rather than the \(O3\) linkage. The only polar interaction between the substrate and \textit{CjAbf}43A that is unique, when O2 is orientated into the \(-1\) pocket, is the hydrogen bond between Asn\textsuperscript{185} and the glycosidic oxygen linking the sugars at the \(+1\) and \(+2NR\) subsites (Fig. 5). As stated above, if the oligosaccharide was built such that O3 is orientated into the active site, Asn\textsuperscript{185} 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 were 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 Phe\textsuperscript{234}, 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 probably render the enzyme completely inactive. A BLAST search of \textit{CjAbf}43A against the UNIPROT data base showed that of the 50 top hits, which displayed identities ranging from 44 to 82\% (data not shown), 46 contained an Asn in the motif WGN. Significantly, Bt0369, which is 66\% identical to \textit{CjAbf}43A and also displays arabinan-specific \(\alpha-1,2\)-arabinofuranosidase activity, contains the WGN motif. These data indicate that Asn\textsuperscript{185} in \textit{CjAbf}43A and Asn\textsuperscript{186} in Bt0369 confer specificity for the O2 linkage.
Arabinan-specific Arabinofuranosidase

### Table 3

**Kinetic constants of wild type and variants of CjAbf43A**

Enzyme concentrations varied from 5 nM to 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.

| CjAbf43A   | 4NPA                  | Sugar beet arabinan                        |
|------------|-----------------------|--------------------------------------------|
|            | $K_{m}$ ($\mu$M) | $k_{cat}$ ($\text{min}^{-1}$) | $k_{cat}/K_{m}$ ($\text{min}^{-1} \mu$M$^{-1}$) | $K_{m}$ ($\mu$M) | $k_{cat}$ ($\text{min}^{-1}$) | $k_{cat}/K_{m}$ ($\text{min}^{-1} \mu$M$^{-1}$) |
| Wild type  | 3.1 ± 0.51          | 1896 ± 320      | 6.2 × 10^2                                    | 0.3 ± 0.02      | 308 ± 8            | 1.1 × 10^6                                    |
| D41A       | —                    | —               | —                                           | —               | —                 | —                                            |
| D168A      | —                    | —               | —                                           | —               | —                 | —                                            |
| E21A       | —                    | —               | —                                           | —               | —                 | —                                            |
| F66A       | —                    | —               | —                                           | —               | —                 | —                                            |
| F67A       | —                    | —               | —                                           | —               | —                 | —                                            |
| D101A      | 2.5 ± 0.2            | 524 ± 43        | 2.1 × 10^2                                    | 1.4 ± 0.1      | 427 ± 11          | 3.2 × 10^3                                    |
| W103A      | —                    | —               | —                                           | 1.6 ± 0.5      | 14 ± 2            | 9.0 × 10^3                                    |
| R121A      | 4.7 ± 1.6            | 134 ± 46        | 2.9 × 10^2                                    | 0.2 ± 0.02      | 266 ± 27          | 1.5 × 10^4                                    |
| F219A      | 3.4 ± 1.2            | 906 ± 245       | 2.6 × 10^2                                    | 0.3 ± 0.02      | 293 ± 18          | 3.8 × 10^3                                    |
| W164A      | —                    | —               | —                                           | —               | —                 | —                                            |
| A67A       | —                    | —               | —                                           | —               | —                 | —                                            |
| N185A      | 0.9 ± 0.05           | 1676 ± 147      | 1.8 × 10^6                                    | 0.5 ± 0.02      | 47.27 ± 2         | 6.7 × 10^7                                    |
| T186A      | 4.9 ± 0.2            | 727 ± 98        | 1.5 × 10^7                                    | 0.2 ± 0.02      | 293 ± 18          | 3.8 × 10^3                                    |
| T186W      | 11.2 ± 0.4           | 586 ± 1         | 5.3 × 10^8                                    | 0.2 ± 0.02      | 30 ± 0.3           | 2.5 × 10^5                                    |
| T214A      | 3.5 ± 0.1            | 717 ± 57        | 2.1 × 10^6                                    | 0.8 ± 0.02      | 154 ± 8           | 2.0 × 10^6                                    |
| T214W      | 4.3 ± 0.5            | 2315 ± 403      | 5.4 × 10^6                                    | 4.4 ± 0.4       | 27 ± 2            | 6.6 × 10^6                                    |
| F234A      | 9.6 ± 0.8            | 358 ± 17        | 3.8 × 10^9                                    | 2.9 ± 0.2       | 30 ± 0.3           | 3.7 × 10^4                                    |
| H267A      | —                    | —               | —                                           | —               | —                 | —                                            |
| Q292A      | 2.9 ± 0.5            | 797 ± 210       | 2.8 × 10^6                                    | 0.5 ± 0.04      | 329 ± 1           | 6.5 × 10^5                                    |
| Y293A      | 4.7 ± 0.5            | 1706 ± 220      | 2.7 × 10^6                                    | 1.4 ± 0.2       | 415 ± 6           | 3.1 × 10^5                                    |
| T186W/T21W | 3.9 ± 1.0            | 606 ± 252       | 1.6 × 10^7                                    | —               | —                 | —                                            |

* —, parameter could not be determined.

At the +2NR subsite, Ne of Trp$^{164}$ and O$\delta$1 of Asn$^{185}$ make hydrogen bonds with O2 of the bound arabinose, whereas Phe$^{234}$ 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 Phe$^{234}$ is substituted with alanine (Table 3).

The +2NR subsite, which houses the α-1,3-linked arabinofuranose side chain in double substitutions, the sugar makes apolar contacts with Phe$^{234}$, whereas the O2 and O3 atoms of the sugar appear to be within hydrogen bonding distance of Tyr$^{293}$ and Gln$^{292}$, respectively. The potential polar contacts with Tyr$^{293}$ and Gln$^{292}$, however, are unlikely to be functionally significant because the activities of the mutants Y293A and Q292A are similar to that of the wild type enzyme (Table 3). To explore the significance of the +2NR subsite, the double substitutions in arabinan were converted to O2-mono-substituted decorations by treatment with HiAXHd3 (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 HiAXHd3-treated polysaccharide. Thus, it is unlikely that the putative interactions with the arabinose at the +2NR subsite are functionally significant, and therefore Phe$^{234}$ probably only contributes to substrate binding at the +2NR subsite.

As described above, CjAbf43A is likely to contain a +3NR subsite that may simply accommodate or, potentially, make productive interactions with the substrate. Although Thr$^{186}$ and Thr$^{214}$ 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 probably 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 Trp$^{101}$ and binds to Arg$^{295}$ via N$\gamma$1 and N$\gamma$2 (Fig. 6). Furthermore, comparing the structure of CjAbf43A with other exo-acting GH43 enzymes reveals several conserved features that probably contribute to substrate binding and transition state stabilization (Fig. 6). Surprisingly, the residues that interact with the arabinose in the Streptomyces exo-α-1,5-arabinofuranosidase Arf43A (8) are conserved not only in other arabinofuranosidases, such as CjAbf43A, but also in GH43 β-xyllosidases, exemplified by XynB3. It is difficult, therefore, to understand how these enzymes distinguish between an Araf and Xylp at the critical −1 subsite because arabinofuranosidases that hydrolyze 4NP-Araf are not active on 4NP-Xylp. Thus, Asp$^{121}$, Glu$^{213}$, and Asp$^{168}$, which function as the catalytic base, catalytic acid, and modulator of the pK$_{a}$ 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). Trp$^{103}$ is also invariant in the other GH43 structures. The structural equivalent to this residue, Trp$^{101}$ in Arf43A and Trp$^{76}$ in BsAXH-m2,3, provides a hydrophobic platform by stacking
against the respective sugar rings, and thus Trp103 is likely to play an equivalent role in CjAbf43A. The importance of the tryptophan is illustrated by the substantial reduction in catalytic activity (against arabinan, the enzyme is completely inactive) when Trp103 is substituted for Ala (Table 3). Ile167 and Phe67 are also highly conserved in other GH43 enzymes where the equivalent residues make hydrophobic contacts with substrate (Fig. 6). Ile167 sits at the top of the active site pocket and is 3.6 Å from the catalytic acid, Glu215 and thus may also create an apolar environment, which contributes to the elevated pK_a of the acidic residue. The observation that the amino acid substitutions, F67A and I167A, cause a >100-fold reduction in activity suggests that Ile167 and Phe67 also make significant contributions to substrate binding in the active site.

In addition to the catalytic residues, Arg295 and His267 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. Although mutation of His267 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 because 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...
Arabinan-specific Arabinofuranosidase

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 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

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 (PULs) (40). PUL7, which is up-regulated by sugar beet arabinan and orchestrates the metabolism of this polysaccharide, contains three GH43 and two GH51 enzymes. The GH43 exo-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 "pretreatment" stage of arabinan processing, converting all of the double linked arabinose into single O3-linked decorations, which are then removed by two GH51 enzymes. Apart from Phe234, residues that line the substrate binding cleft distal to 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 CjAbf43A and linear in BsAXH-m2,3, that dictates the specificity of these enzymes. The trisaccharide is metabolized, probably in the periplasm, by the GH43 α-1,5-exo-arabinanase CjA_0806 that displays a moderate preference for arabinotriose. The arabinobiase 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.

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 whereas 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 CjAbf43A because 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 probably contains an additional arabinofuranosidase that targets xylose residues decorated at O2 and O3.

Specificity of CjAbf43A for Sugar Beet Arabinan—The specificity of CjAbf43A for sugar beet arabinan is due to the surface topography of the enzyme. Phe<sup>66</sup>, Trp<sup>164</sup>, and Phe<sup>234</sup> 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-arabinofuranosyl backbone of sugar beet arabinan (Fig. 4), explaining why the enzyme targets this polysaccharide. The enzyme is unable to hydrolyze 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, BsAXH-m2,3 (10), is unable to attack arabinan side chains; despite structural conservation with CjAbf43A 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 CjAbf43A and linear in BsAXH-m2,3, that dictates the specificity of these enzymes. Apart from Phe<sup>234</sup>, 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 because 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 of 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. thetaiatotaomicron remove O2-linked arabinose decorations, in the context of single or double substitutions, through the action of an arabinanspecific α,1,2-arabinofuranosidase, an activity that has not previously been reported.

The biological rationale for such an activity probably 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 CjAbf43A reveals several topological features, such as a curved substrate binding cleft, a shell-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.

Acknowledgments—We thank the staff at the beamlines used for data collection at Diamond Light Source and the Advanced Photon Source.