



# Discovery and characterization of family 39 glycoside hydrolases from rumen anaerobic fungi with polyspecific activity on rare arabinosyl substrates

Received for publication, March 30, 2017, and in revised form, May 30, 2017. Published, Papers in Press, June 6, 2017, DOI 10.1074/jbc.M117.789008

Darryl R. Jones<sup>†1</sup>, Muhammed Salah Uddin<sup>†§1</sup>, Robert J. Gruninger<sup>‡</sup>, Thi Thanh My Pham<sup>¶</sup>, Dallas Thomas<sup>‡</sup>, Alisdair B. Boraston<sup>||</sup>, Jonathan Briggs<sup>\*\*</sup>, Benjamin Pluvinage<sup>||</sup>, Tim A. McAllister<sup>‡</sup>, Robert J. Forster<sup>‡</sup>, Adrian Tsang<sup>¶</sup>, L. Brent Selinger<sup>§</sup>, and D. Wade Abbott<sup>†§2</sup>

From the <sup>†</sup>Agriculture and Agri-Food Canada, Lethbridge Research and Development Centre, Lethbridge, Alberta T1J 4B1, Canada, the <sup>§</sup>Department of Biological Sciences, University of Lethbridge, Lethbridge, Alberta T1K 6T5, Canada, the <sup>¶</sup>Centre for Structural and Functional Genomics, Concordia University, Montreal, Quebec H4B 1R6, Canada, the <sup>||</sup>Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia V8W 3P6, Canada, and the <sup>\*\*</sup>School of Biology, Ridley Building 2, Newcastle University, Claremont Road, Newcastle upon Tyne NE1 7RU, United Kingdom

Edited by Ruma Banerjee

Enzyme activities that improve digestion of recalcitrant plant cell wall polysaccharides may offer solutions for sustainable industries. To this end, anaerobic fungi in the rumen have been identified as a promising source of novel carbohydrate active enzymes (CAZymes) that modify plant cell wall polysaccharides and other complex glycans. Many CAZymes share insufficient sequence identity to characterized proteins from other microbial ecosystems to infer their function; thus presenting challenges to their identification. In this study, four rumen fungal genes (*nf2152*, *nf2215*, *nf2523*, and *pr2455*) were identified that encode family 39 glycoside hydrolases (GH39s), and have conserved structural features with GH51s. Two recombinant proteins, NF2152 and NF2523, were characterized using a variety of biochemical and structural techniques, and were determined to have distinct catalytic activities. NF2152 releases a single product,  $\beta$ 1,2-arabinobiose (Ara<sub>2</sub>) from sugar beet arabinan (SBA), and  $\beta$ 1,2-Ara<sub>2</sub> and  $\alpha$ -1,2-galactoarabinose (Gal-Ara) from rye arabinosyl xylan (RAX). NF2523 exclusively releases  $\alpha$ -1,2-Gal-Ara from RAX, which represents the first description of a galactose ( $\alpha$ -1,2)-arabinosidase. Both  $\beta$ -1,2-Ara<sub>2</sub> and  $\alpha$ -1,2-Gal-Ara are disaccharides not previously described within SBA and RAX. In this regard, the enzymes studied here may represent valuable new biocatalytic tools for investigating the structures of rare arabinosyl-containing glycans, and potentially for facilitating their modification in industrial applications.

This work was supported by Agriculture and Agri-Food Canada (AAFC) through Agriculture Innovation Program Grant AIP-P022 and Elanco Animal Health, and Natural Sciences and Engineering Research Council of Canada Discovery Grant FRN 04355 (to A. B. B.). The intellectual property surrounding the field of use for these enzymes has been protected with a United States provisional patent (No. 62/380,741 under AAFC's patent docket, PAT-13830). D. R. J., M. S. U., T. A. M., A. T., and D. W. A. are co-inventors on this patent.

The atomic coordinates and structure factors (code 5U22) have been deposited in the Protein Data Bank (<http://www.pdb.org/>).

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) MF326649, MF326650, MF326651, and MF326652.

<sup>1</sup> Both authors contributed equally to the results of this work.

<sup>2</sup> To whom correspondence should be addressed. E-mail: wade.abbott@agr.gc.ca.

The rumen microbiome is recognized as one of the most efficient microbial ecosystems in the degradation of plant biomass (1, 2). It contains a diverse microbial community with large numbers of bacteria, anaerobic fungi, ciliate protozoa, and bacteriophages. Of these, rumen bacteria and fungi are considered to be indispensable for plant fiber digestion. It has been estimated that at least 85% of the microbes inhabiting the rumen have not been cultured using traditional approaches (1). In recent years, our knowledge of structure and function of the rumen microbial diversity has drastically increased due to improvements in sequencing technologies (3–5).

The plant cell wall comprises cellulose, hemicellulose, and pectin (6). These polysaccharides are often interconnected, contain many diverse sugar chemistries, and display a variety of decorations and branching. Cellulose is the simplest of these three classes of polysaccharides and is composed of linear polymers of  $\beta$ -1,4-linked D-glucose (7). Hemicellulose is a group of branched polysaccharides that are classified according to the primary sugar within the backbone of the polymer (e.g. xylan is composed of  $\beta$ -1,4-linked D-xylose). Hemicelluloses contain extensive variations in their repeating structure. This can be seen in arabinosyl xylans, such as rye arabinosyl xylan (RAX).<sup>3</sup> RAX consists of a polymer chain of  $\beta$ -1,4-linked D-xylose units, many of which are 2- or 3-monosubstituted or 2,3-disubstituted by  $\alpha$ -L-arabinose (Ara) (8). Pectin is found within the primary cell wall and the middle lamella, which punctuates the junctions between primary walls of neighboring cells, and participates in intercellular connections (9). It is a structurally complex polysaccharide that is enriched in D-galacturonic acid and divided into three distinct classes of pectic polysaccharides: homogalacturonan, rhamnogalacturonan-I (RG-I), and rhamnogalac-

<sup>3</sup> The abbreviations used are: RAX, rye arabinosyl xylan; RG-I, rhamnogalacturonan-I; SBA, sugar beet arabinan; AGP, arabinogalactan protein; GH, glycoside hydrolase; CBM, carbohydrate-binding module; ABF,  $\alpha$ -L-arabinofuranosidase; DP, degree of polymerization; HPAEC-PAD, high performance anion-exchange chromatography with pulsed amperometric detection; FACE, fluorophore-assisted carbohydrate electrophoresis; ANTS, 8-aminonaphthalene-1,3,6-trisulfonic acid; bis-tris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol; PDB, Protein Data Bank; ABN, arabinanases; r.m.s., root mean square.

turonan-II that display substantial variability in their structure (10, 11). For example, the side chains of RG-I can be heavily decorated with arabinans (12), such as sugar beet arabinan (SBA), which consists of an  $\alpha$ -1,5-L-arabinofuranosyl backbone decorated with  $\alpha$ -1,2- and  $\alpha$ -1,3-L-arabinofuranosyl side chains.

Also present in the plant cell wall are structurally diverse cell-surface glycoproteins that are collectively referred to as arabinogalactan proteins (AGPs). This structurally diverse protein family is enriched in the amino acids: hydroxy-Pro/Pro, Ala, and Ser/Thr and heavily glycosylated (90–98% w/w). AGPs are thought to play important roles in various aspects of plant growth and development including reproduction, cell signaling, and microbial interactions; and may serve to anchor the pectic and hemicellulosic polysaccharide networks (13).

Enzymes that modify carbohydrates are referred to as carbohydrate active enzymes (CAZymes). CAZymes are categorized into “classes” based upon their catalytic function and “families” based upon sequence relatedness (14). Family relatedness does not necessarily equate to functional relatedness, however, as many CAZyme families have now been described to contain members with variations in their enzyme specificities (see Fig. 1) (15, 16), a property that is referred to as “polyspecificity.”

The hydrolysis of a glycosidic linkage between two or more carbohydrates, or a carbohydrate and non-carbohydrate adduct, leading to the generation of a new reducing end is catalyzed by enzymes known as glycoside hydrolases (GHs) (17). Generally, hydrolysis is performed by two primary amino acid residues within the enzyme active site. Inverting GHs, which generate products that have an inverted anomeric configuration at the nascent reducing end, possess a general acid (*i.e.* proton donor) and a catalytic base (*i.e.* proton acceptor). Alternatively, retaining GHs, which generate products that retain anomeric configuration at the nascent reducing end, possess an acid/base and a nucleophile and catalyze a double displacement reaction. Often the activity of GHs is potentiated by carbohydrate-binding modules (CBMs) through “targeting” or “concentrating” effects (18). Polyspecificity is also a common feature in the binding specificities of CBM families, such as CBM13 (19). Using CBMs to identify associations with uncharacterized genes, therefore, may assist in the discovery of new enzyme activities.

Comparative analysis of fungal genomes and metatranscriptomes has revealed that rumen fungi exhibit tremendous diversity in the number and types of their CAZymes (20). Harnessing this genetic reservoir holds vast potential for biotechnological applications. For example, *in vitro* studies have demonstrated that rumen fluid supplemented with select combinations of CAZymes noticeably boosts the release of cellobiose, glucose, and xylose from plant cell wall structural polysaccharides (21). Anaerobic fungi isolated from rumen or herbivore dung, such as phylum *Neocallimastigomycota*, are known to participate in the deconstruction of plant cell wall substrates by invasive rhizoidal growth, which physically disrupts recalcitrant tissues; and secrete a diverse arsenal of CAZymes (22). Many of these CAZymes share little sequence identity to characterized proteins from other microbial ecosystems (20, 23).

In this study, we identified four genes from anaerobic fungi that are predicted to encode proteins with an N-terminal GH39 followed by a CBM13 (19): *nf2152*, *nf2523*, and *nf2215* from *Neocallimastix frontalis*; and *pr2455* from *Piromyces rhizinflatus*. These “fungal GH39s” comprise a new subfamily within GH39. To characterize their functions, we synthesized the genes, expressed them in *Escherichia coli*, and purified their gene products for biochemical characterization. The enzymes were screened against a library of plant cell wall substrates, and activity was observed on rare substrates present within SBA and RAX. Analyses of the product profiles of NF2152 and NF2523 revealed that two distinct products were released, underpinning that this subfamily is polyspecific with activities that are unique from what was reported for a related sequence, *bgxg1*, from *Orpinomyces* sp. strain C1A (24). Structural analysis of NF2152 provides insights into the molecular basis of recognition by this enzyme family.

## Results

### Phylogenetic analysis of fungal proteins

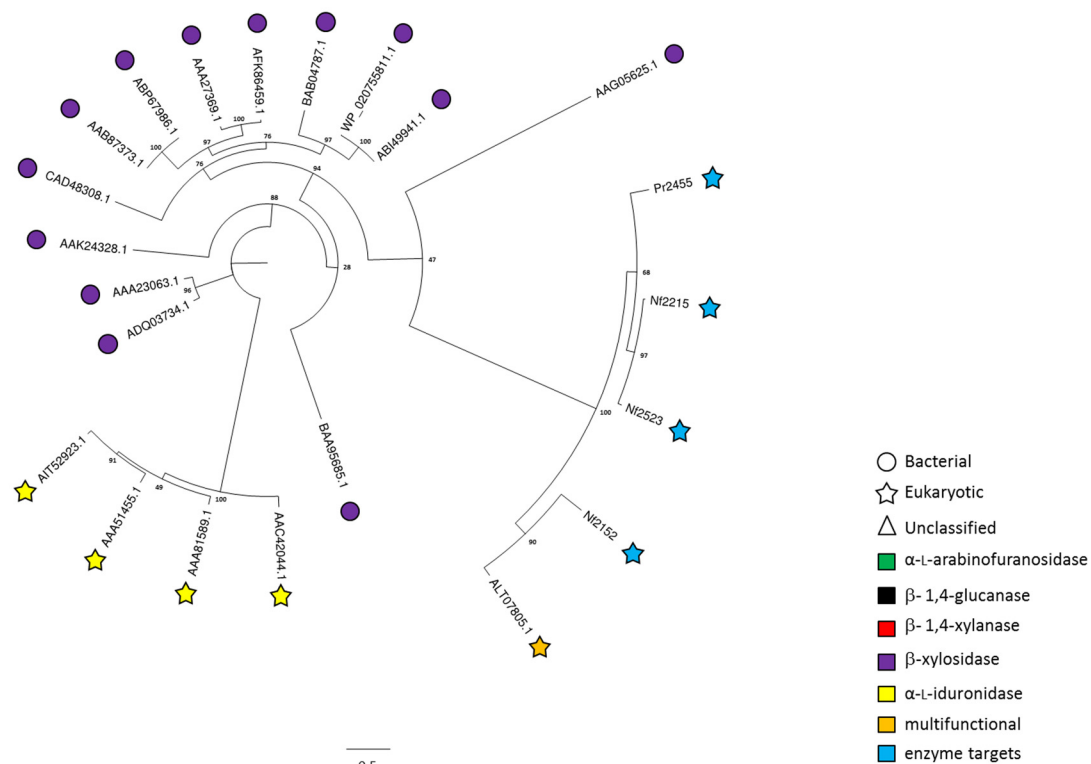
Genes encoding NF2152, NF2215, NF2523, and PR2455 were identified from the transcriptomes of anaerobic fungal cultures previously grown on barley straw. These gene products were used as query sequences to search for entries with homology in the NCBI non-redundant database using the algorithm BLASTP. The four proteins displayed low sequence identity to GH39s and GH51s, based on pfam annotation. This was consistent when aligned to all GH39 or GH51 entries in the CAZY database (14). NF2152 had 14.9% average identity against all GH39 entries, and 12.0% average identity against all GH51 entries. Therefore, to provide insight into the potential activities of their gene products, the four fungal sequences were aligned with characterized sequences from GH39 ( $n = 18$ ; Fig. 1A) and GH51 ( $n = 74$ ; Fig. 1B). In both trees, the fungal target sequences partitioned as single, distantly related clades. Comparisons with the distribution of GH39 activities, which include  $\alpha$ -L-iduronidase (EC 3.2.1.76) and  $\beta$ -xylosidase (EC 3.2.1.37) (14), revealed that the fungal sequences may have diverged from a common ancestor with the bacterial  $\beta$ -xylosidases and cluster with a “multifunctional” GH39 previously described from *Orpinomyces* sp. strain C1A, *Bgxg1* (24). Comparison with GH51 known activities, which include endoglucanase (EC 3.2.1.4), endo- $\beta$ -1,4-xylanase (EC 3.2.1.8),  $\beta$ -xylosidase (EC 3.2.1.37), and  $\alpha$ -L-arabinofuranosidase (ABF; EC 3.2.1.55) (14), indicated that the clade partitions from a region of eukaryotic ABFs. Based upon these comparisons, and their high degree of relatedness to *Bgxg1* (60.9–77.9%), these sequences were classified as a new subfamily of GH39s.

### Functional screening of NF2152

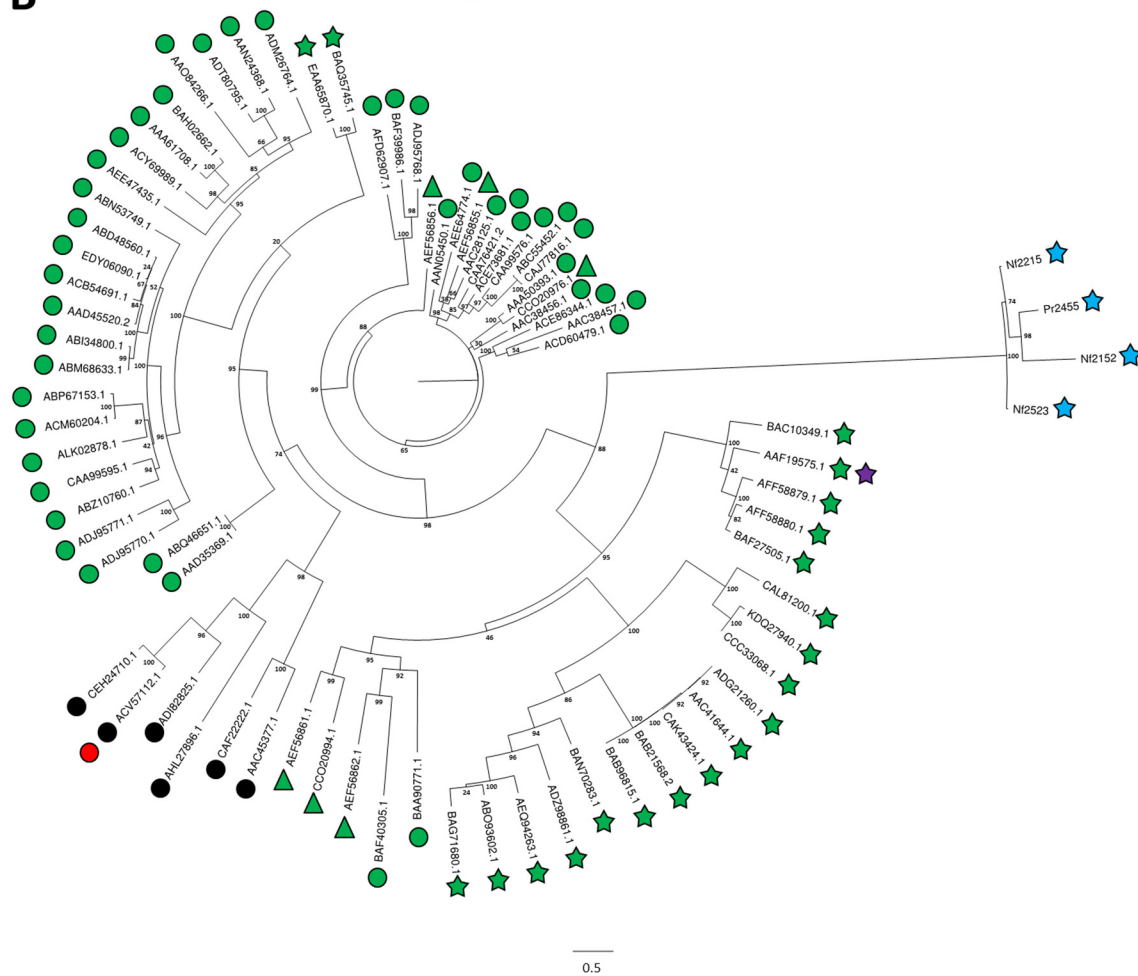
The N-terminal domain of the four GH39 genes were synthesized and recombinant protein was produced. Based upon purification yields and protein stability, NF2152 was selected for functional screening and preliminary enzyme characterizations by thin layer chromatography (TLC) on a variety of appropriate plant cell wall substrates, including SBA, RAX, wheat arabinoxylan,  $\beta$ -glucan, xyloglucan, gum arabic, galactomannan, pectic galactan, arabinogalactan, RG-I, arabinofuranosyl-

# Characterization of a novel glycoside hydrolase activity

**A**



**B**



xylobiose, arabinofuranosyl-xylotriase, PNP- $\alpha$ -L-arabinofuranoside, PNP- $\alpha$ -L-arabinopyranoside and beech wood xylan; activity was detected primarily on SBA (Fig. 2) and RAX (Fig. 3). Based upon the multifunctional activity reported for Bgxl (24), NF2152 was also screened against PNP- $\beta$ -D-glucopyranoside, PNP- $\beta$ -D-galactopyranoside, PNP- $\beta$ -D-xylopyranoside, and cellobiose under similar reaction conditions; however, no activity was observed. TLC analysis of the SBA digestion products revealed that NF2152 released a single product band that migrated slower than arabinose (Ara), xylose (Xyl),  $\alpha$ -1,5-arabinobiose (Ara<sub>2</sub>), and  $\alpha$ -1,5-arabinotriose (Ara<sub>3</sub>) (Fig. 2, A and B). These observations indicated that NF2152 is not a conventional ABF or xylosidase as predicted by the comparative analysis with other characterized GH39s and GH51s (Fig. 1).

### Chemical structure of the SBA product

To generate sufficient product for chemical characterization, a large-scale digestion of SBA was performed. Following ethanol precipitation, the soluble products were fractionated by size exclusion chromatography. Eluted fractions were screened by TLC to identify peak boundaries, and samples containing similar sized products were pooled and analyzed for purity by TLC (Fig. 2A).

To determine whether the NF2152 product had a degree of polymerization (DP) > 1, acid hydrolysis was performed. The hydrolyzed product migrated as a single band with a similar mobility to Ara on TLC (Fig. 2B) and eluted with an identical retention time as Ara when analyzed by high performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) (Fig. 2C). These results suggested that NF2152 generates an oligosaccharide that is solely composed of Ara.

Comparison with a commercially available  $\alpha$ -1,5-Ara<sub>2</sub> standard revealed that these arabinooligosaccharides displayed different mobility patterns in TLC plates (Fig. 2B). Additionally, a selection of GH51 and GH43  $\alpha$ -ABFs were unable to hydrolyze the SBA product (results not shown). This highlighted that the product has a differing chemistry, such as a modification, alternate stereochemistry (*i.e.*  $\beta$ ), or positional (*e.g.* 1  $\rightarrow$  2) linkage. Due to the limited availability of commercial standards, a  $\beta$ -1,2-arabinofuranosidase GH127 from *Bifidobacterium longum* (BlGH127) (25) was synthesized, purified, and used for enzymatic glycosequencing. Previously this enzyme was used to elucidate the structure of a  $\beta$ -1,2-arabinosyl oligosaccharide derived from an AGP glycan (25). Treatment with BlGH127 cleaved the NF2152 product completely into Ara (Fig. 2D, *inset*), establishing that the product is a pure oligosaccharide with a  $\beta$ -1,2-linkage. However, this digestion does not determine the DP of the product as the GH127 may act processively on a substrate with DP  $\geq$  2. Therefore, to determine the size of the arabinooligosaccharide, electrospray ionization mass spectrometry was performed (Fig. 2D). The product was determined to have a  $m/z$  ratio ((M + Na)<sup>+</sup>) of 305.1. This ratio

equates to a calculated mass of 282.1, which is the absolute mass of a pentose disaccharide, and confirms that the product is a pure  $\beta$ -1,2-Ara<sub>2</sub>. Additionally, the product released from an arabinan-derived tetrasaccharide was characterized (Fig. 2E). Enzymatic fingerprinting showed that this arabinotetraose contains a single  $\beta$ -L-arabinofuranosyl linkage and three  $\alpha$ -L-arabinofuranosyl linkages. Digestion with NF2152 releases a disaccharide product with similar mobility to the product released from SBA, and likewise, this product was cleaved to arabinose by the  $\beta$ -1,2-L-arabinosidase, BlGH127 (Fig. 2E). These results indicate that NF2152 must hydrolyze an  $\alpha$ -L-arabinofuranosyl linkage and is thus an  $\alpha$ -L-( $\beta$ -1,2)-arabinobiosidase.

### Characterization of a chemically distinct product from RAX

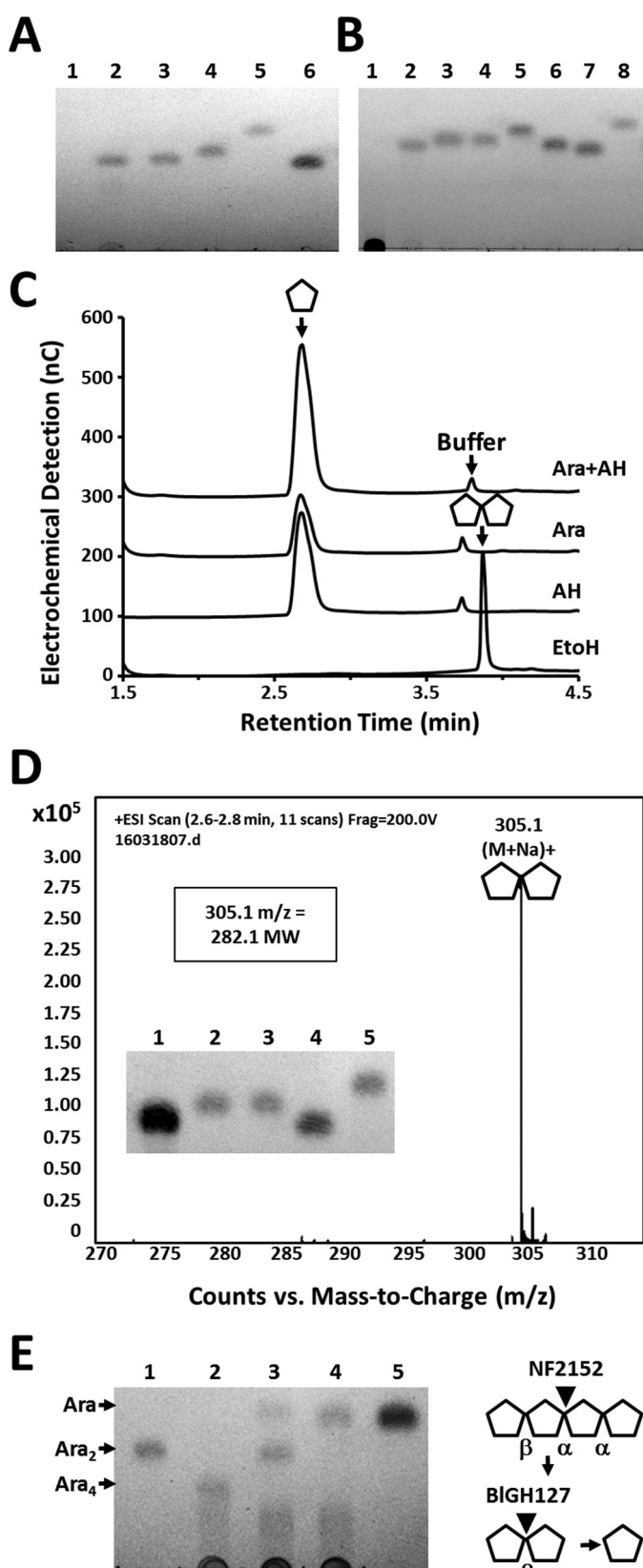
Digestion of RAX with NF2152 generated a second band with a slower mobility, which appeared to be the primary product of a second GH39 enzyme, NF2523 (Fig. 3A). The release of two structurally distinct products by NF2152; in addition to the unique product profile of NF2523, which only released one of these two products from RAX, highlighted that there are different substrates present in RAX and that this subfamily of fungal GH39 enzymes is polyspecific with specialized variants. Notably, very little product was detected when SBA was digested with NF2523. To determine whether the second product was an arabinooligosaccharide with a DP > 2, mass spectrometry was performed. Somewhat surprisingly, the  $m/z$  ratio of the product was ((M + Na)<sup>+</sup>) of 335.1 and its calculated mass was determined to be 312.1, which is equivalent to the molecular weight of a hexose-pentose disaccharide. Therefore, acid hydrolysis was performed and the products were analyzed by TLC and HPAEC-PAD (Fig. 3B). Consistent with the MS result, two distinct monosaccharides were detected: Ara and galactose (Gal). Ara is a main compositional sugar of SBA (88%) and RAX (Ara 38%); whereas, Gal is less common (SBA = 3%; RAX = none detected). In this regard, the presence of an Ara/Gal disaccharide suggests that it is a rare structure within both polysaccharides. Alternatively, Ara and Gal are commonly linked sugars in the arabinogalactan side chains of RG-I (26). However, digests of diverse pectic substrates with GH39 enzymes did not produce any detectable products. This suggests that NF2152 and NF2523 may be targeting rare substructures within SBA or AGPs glycans that co-purify with plant cell wall polysaccharides (25).

### Sequencing of the Gal/Ara product

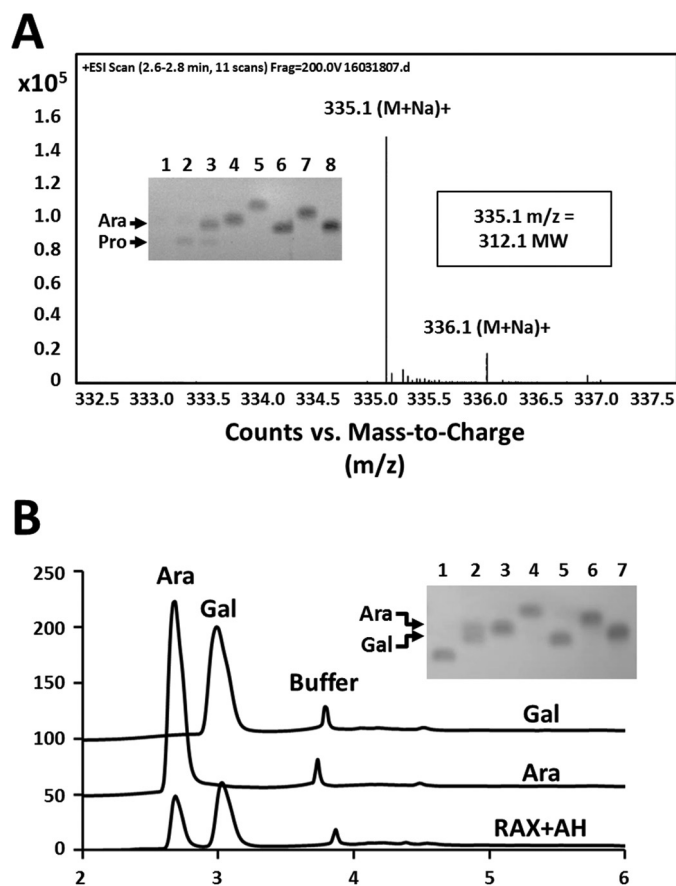
To determine whether Ara or Gal was positioned at the reducing end of the disaccharide, two complementary techniques were used: differential gas chromatography-mass spectrometry (GC-MS) and fluorophore-assisted carbohydrate electrophoresis (FACE). First, alditol acetates were generated from intact and hydrolyzed forms of the Gal/Ara disaccharide, and then visualized by GC-MS (Fig. 4A). This technique acety-

**Figure 1. Phylogenetic trees of four rumen fungal query sequences aligned with characterized sequences from related GH39 and GH51 families.** A, GH39 and B, GH51. Each sequence leaf represents the catalytic module from the associated ascension number. Taxonomic (bacterial, eukaryotic, and unclassified) and functional ( $\alpha$ -L-arabinofuranosidase,  $\beta$ -1,4-glucanase,  $\beta$ -1,4-xylanase,  $\beta$ -xylosidase,  $\alpha$ -L-iduronidase) differences are defined within the legend. Characterized activities and organism source are depicted with different symbol colors and shapes, respectively.

## Characterization of a novel glycoside hydrolase activity



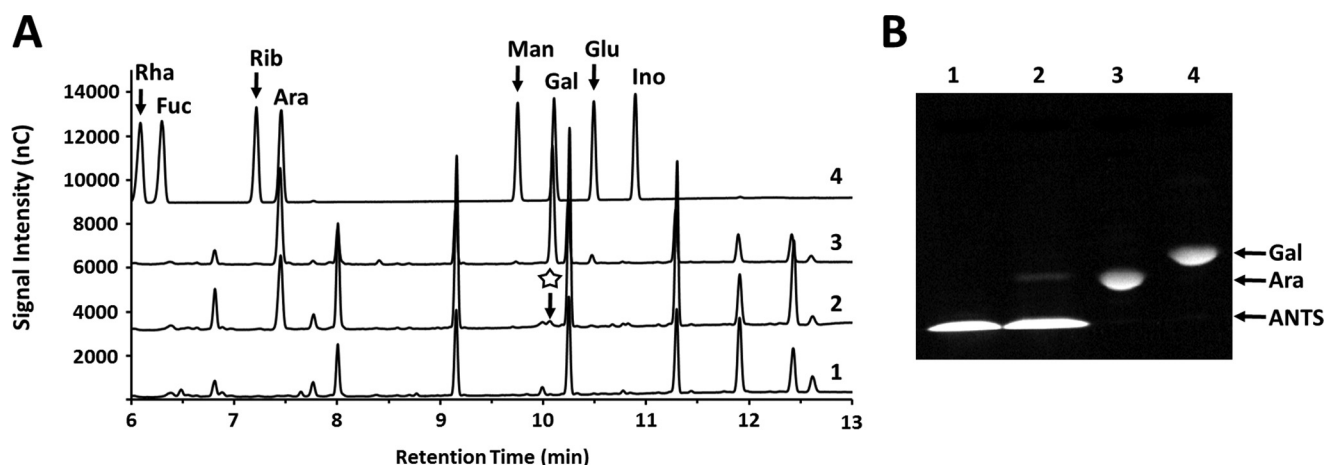
**Figure 2. Structural analysis of the NF2152 product released from SBA.** *A*, TLC analysis. Lane 1 = SBA; lane 2 = product before SEC purification; lane 3 = product after SEC purification; lane 4 = Ara; lane 5 = xylose; lane 6 = Gal. *B*, TLC of acid hydrolyzed P2 bio-gel column purified products. Lane 1 = SBA; lane 2 = product before acid hydrolysis; lane 3 = product after acid hydrolysis purification; lane 4 = Ara; lane 5 =  $\alpha$ -1,5-Ara<sub>2</sub>; lane 6 =  $\alpha$ -1,5-Ara<sub>3</sub>; lane 7 = Gal;



**Figure 3. Structural analysis of the NF2523 product released from RAX.** *A*, HR-MS spectrum of the Bio-Gel P-2 column purified NF2523 product released from RAX displaying measured 335.1 *m/z* ratio and calculated mass of hexose-pentose product (312.1). *Inset*: TLC displaying products from RAX digestions. Lane 1 = RAX; lane 2 = NF2523 digestion; lane 3 = NF2152 digestion; lane 4 = Ara; lane 5 = xylose; lane 6 = Gal; lane 7 =  $\alpha$ -1,5-Ara<sub>2</sub>; and lane 8 =  $\alpha$ -1,5-Ara<sub>3</sub>. Two major products are labeled. *B*, overlaid HPAEC-PAD chromatograms of Gal and Ara monosaccharides released by acid hydrolysis of the NF2523 product. *Inset*: TLC of the acid hydrolyzed NF2523 product. Lane 1 = purified product; lane 2 = acid hydrolysate; lane 3 = Ara; lane 4 = xylose; lane 5 = Gal; lane 6 =  $\alpha$ -1,5-Ara<sub>2</sub>; lane 7 =  $\alpha$ -1,5-Ara<sub>3</sub>.

lates the free reducing end; hence, the residue at the non-reducing end is protected until it is released by acid hydrolysis. In this manner, only the reducing sugar will be visible in the GC-MS chromatogram. In Fig. 4A, Ara was present before and after acid hydrolysis; however, Gal was only visible when derivatized after hydrolysis, indicating it was protected when the disaccharide is intact. This analysis revealed the product as Gal-Ara (*i.e.* Gal = non-reducing; Ara = reducing end), a result that was confirmed by labeling the reducing end of the disaccharide with the fluorogenic compound 8-aminonaphthalene-1,3,6-trisulfonic acid (ANTS). Following derivatization, the disaccharide was hydrolyzed into monosaccharides and analyzed by FACE (Fig. 4B).

and lane 8 = xylose. *C*, overlaid HPAEC-PAD chromatograms showing acid hydrolyzed products released from SBA. *D*, HR-MS spectrum displaying measured 305.1 *m/z* ratio and calculated mass of the Ara<sub>2</sub> product (282.1). *Inset*: GH127 digestion of Ara product. Lane 1 = NF2152 product; lane 2 = GH127 digest; lane 3 = Ara; lane 4 = Gal; and lane 5 =  $\alpha$ -1,5-Ara<sub>2</sub>. *E*, TLC of arabinotetraose derived from SBA (Ara<sub>4</sub>) digestion. Lane 1 =  $\beta$ -1,2-Ara<sub>2</sub> product; lane 2 = Ara<sub>4</sub> substrate; lane 3 = Ara<sub>4</sub> digested with NF2152; lane 4 = NF2152 digestion treated with BIGH127; and lane 5 = arabinose. *Right*, schematic of sequential digestion.



**Figure 4. Sequencing of the heterogeneous disaccharide released from RAX by differential labeling of the reducing end.** *A*, GC-MS chromatograms of the RAX product. The background reaction buffer is shown in 1. The disaccharide was reduced before (2) and after (3) acid hydrolysis, and then acetylated. Only Ara is visible in the non-hydrolyzed sample (the missing Gal peak is indicated with a star); whereas, both Ara and Gal are visible when the monosaccharides are generated before reduction and acetylation. Monosaccharide standards are shown in 4 and are labeled. *B*, FACE analysis of the ANTS-labeled disaccharide. The NF2523 product was labeled and then acid hydrolyzed. The products were compared with various standards. Lane 1 = ANTS; lane 2 = ANTS-labeled product hydrolyzate; lane 3 = ANTS-labeled Ara; lane 4 = ANTS-labeled Gal.

Visualization of the products revealed that only the Ara was labeled, and therefore, exposed at the reducing end of the intact disaccharide.

#### NMR analysis of the reducing end absolute configuration and linkage

To provide further insights into the structure of the two products, both were analyzed by  $^1\text{H}$  and  $^{13}\text{C}$  nuclear magnetic resonance (NMR) (Fig. 5, *A–D*, Table 1). Both samples generated background peaks resulting from ring dynamics that obscured their interpretation; therefore, they were reduced to alditols with sodium borohydride. The product released from SBA by NF2152 was determined to be  $\beta$ -arabinofuranose-(1–2)-arabitol (Fig. 5*E*). The chemical shifts and coupling constants of its non-reducing monosaccharide suggested a  $\beta$ -arabinofuranose (27). The arabitol portion and the linkage were confirmed by two-dimensional NMR spectra including correlation spectroscopy (COSY), heteronuclear single quantum correlation (HSQC), and heteronuclear multiple bond correlation (HMBC). The full structure was confirmed by two-dimensional NMR.

The reduced RAX product generated  $^1\text{H}$  and  $^{13}\text{C}$  chemical shifts and a coupling constant of the anomeric proton (4.0 Hz) of its non-reducing monosaccharide that matched those of  $\alpha$ -galactopyranose (28, 29). The NMR data for the arabitol portion are identical with those observed for the SBA product. The reduced RAX product generated by NF2523, therefore, was identified as  $\alpha$ -galactopyranose-(1–2)-arabitol (Fig. 5*F*). This ring configuration of Gal differs substantially from the furanose observed in the SBA product, which is structurally similar to galactofuranose (Fig. 5*G*).

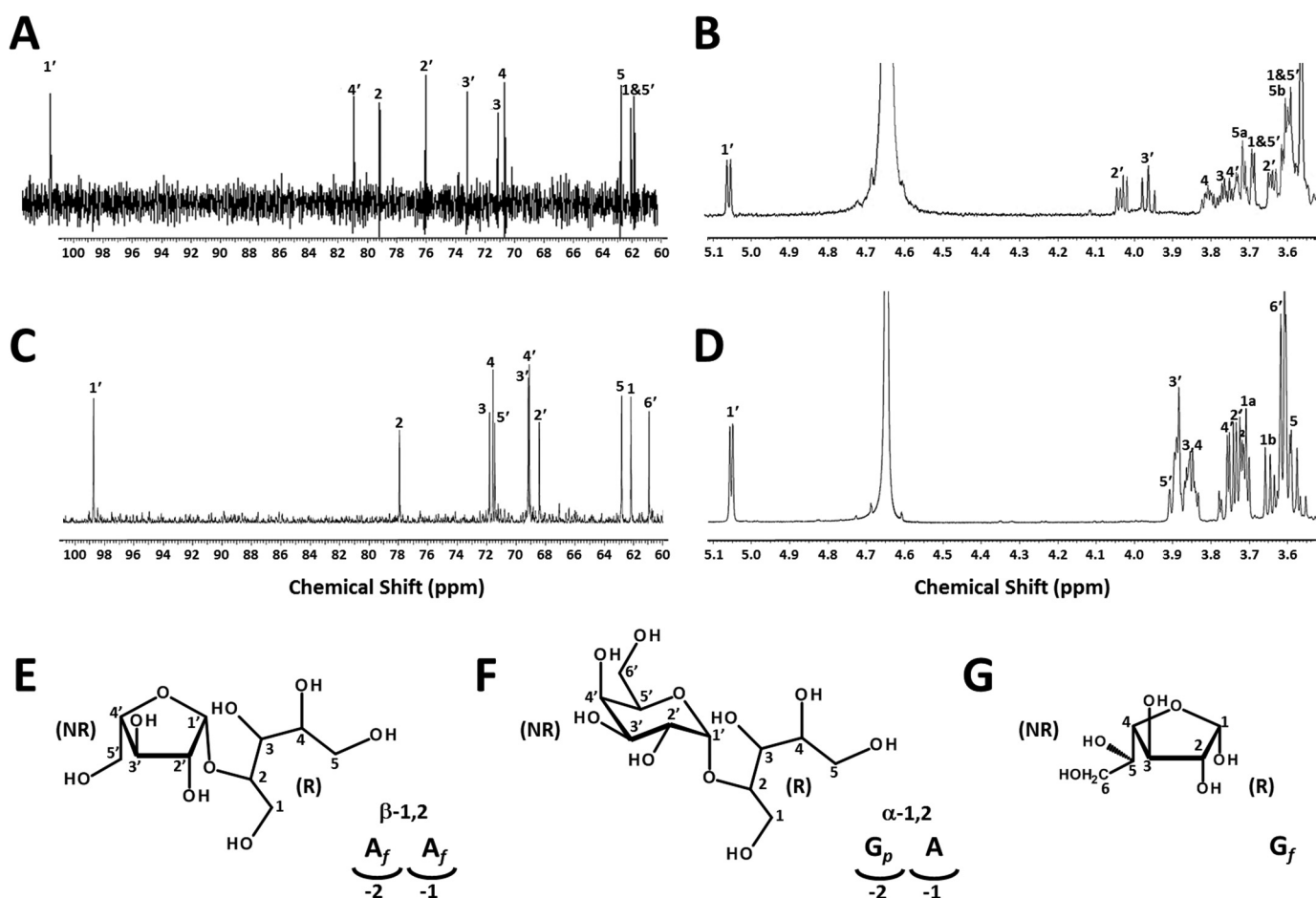
#### Three-dimensional structure of NF2152

To provide further insight into the molecular basis of substrate recognition in the –1 and 2 subsites we solved the structure of NF2152. The *de novo* structure of NF2152 was determined to a resolution of 1.75 Å by single wavelength anomalous

dispersion phasing using an iodide derivative ( $n = 13$ ) (Table 2). The protein adopted the canonical GH-A clan ( $\beta/\alpha$ )<sub>8</sub> TIM-barrel-fold with a fused, hybrid  $\beta$ -sandwich spanning residues Met<sup>29</sup>-Asp<sup>36</sup> and Thr<sup>321</sup>-Ala<sup>431</sup> (Fig. 6*A*). The N-terminal strand formed at one end of the  $\beta$ -sandwich and created a “closed” bimodular domain. DaliLite v.3 (30) searches revealed that NF2152 had the highest levels of structural similarity with GH39 (PDB code 4EKJ;  $Z_{\text{value}} = 25.1$ ; matched  $C_{\alpha} = 133$ ; r.m.s. deviation = 4.2 Å<sup>2</sup> (31)); GH51 (PDB code 2VRQ;  $Z_{\text{value}} = 25.8$ ; matched  $C_{\alpha} = 154$ ; r.m.s. deviation = 4.2 Å<sup>2</sup> (32)); and GH30 (PDB code 4QAW,  $Z_{\text{value}} = 25.8$ , matched  $C_{\alpha} = 201$ ; r.m.s. deviation = 4.2 Å<sup>2</sup>; (33)) families.

Inspection of the putative active site of NF2152 revealed electron density consistent with the presence of a molecule of bis-tris methane (Fig. 6*A*, *inset*). This molecule was a component of the crystallization solution and proved to be essential for the reproducible generation of high quality crystals. The bis-tris methane is tightly sandwiched between Trp<sup>118</sup> and Trp<sup>289</sup>, and forms several direct and water-mediated H-bonds. Despite multiple attempts to soak and co-crystallize NF2152 with  $\beta$ -1,2-Ara<sub>2</sub> we were unable to dislodge the buffer molecule from the active site, and other conditions failed to generate diffraction quality crystals. Therefore, to investigate the potential interactions between NF2152 and Ara in the –1 subsite we performed a structural superimposition with  $\alpha$ -L-Ara bound in the active site pocket of the GH51 ABF from *Thermotoga maritima* (TmGH51; PDB code 3UG4, Fig. 6*B* (34)). GH51s are retaining enzymes with experimentally determined catalytic residues. In TmGH51, the nucleophile is Glu<sup>281</sup> and the acid/base is Glu<sup>172</sup> (34), which overlay with Glu<sup>254</sup> and Glu<sup>155</sup>, respectively, in NF2152. Both of these residues are in reasonable proximity to the C1-OH of the Ara present in the overlaid TmGH51 structure to facilitate hydrolysis. To confirm catalytic function, Glu<sup>254</sup> and Glu<sup>155</sup> were mutated to glutamine residues and tested for activity on SBA. Mutations to these residues resulted in the complete loss of detectable product (Fig. 6*C*, *top*

## Characterization of a novel glycoside hydrolase activity



**Figure 5. NMR analysis and structural determination of the reduced NF2152 and NF2523 products.** A,  $^1\text{H}$  NMR and B,  $^{13}\text{C}$  NMR spectra of the purified NF2152 product. C,  $^1\text{H}$  NMR and D,  $^{13}\text{C}$  NMR spectra of the purified NF2523 product. E, structure of  $\beta$ -1,2-Ara<sub>2</sub> positioning within the active site of NF2152. F, structure of  $\alpha$ -1,2-Gal-Ara. Lower right: schematic representation of Gal-Ara positioning within the active site of NF2523. G, structure of  $\alpha$ -galactofuranose.

**Table 1**  
NMR  $^{13}\text{C}$  and  $^1\text{H}$  chemical shifts and coupling constants ( $\text{D}_2\text{O}$ , 25 °C)

No.	Chemical shifts		Coupling constants
	$^{13}\text{C}$	$^1\text{H}$	
	ppm		Hz
NF2152 sample from SBA			
1	61.897 or 62.132	a:3.696; b:3.589	m
2	79.274	3.630	dd (6.5; 3.5)
3	71.208	3.758	m
4	70.730	3.798	m
5	62.807	a:3.596; b:3.708	m
1'	101.759	5.049	d (5.0)
2'	76.148	4.022	dd (5.0; 8.0)
3'	73.310	3.954	t (8.0; 8.0)
4'	81.027	3.732	dt (8.0; 3.0; 3.0)
5'	61.897 or 62.132	a:3.696; b:3.589	m
NF2523 sample from RAX			
1	62.200	a: 3.706 b: 3.649	m
2	77.946	3.714	m
3	71.815	3.846 $\div$ 3.862	m
4	71.572	3.846 $\div$ 3.862	m
5	62.830	3.588 $\div$ 3.616	m
1'	98.739	5.051	d (4.0)
2'	68.430	3.732	m
3'	69.174	3.881	m
4'	69.106	3.744	m
5'	71.466	3.888	m
6'	60.971	3.602 $\div$ 3.616	m

panel). This suggested Glu<sup>155</sup> and Glu<sup>254</sup> are catalytic residues, whereas their spatial arrangement, indicated that NF2152 harnesses a retaining mechanism, consistent with other GH39s and GH51s. Previously, an unrelated retaining  $\beta$ -1,2-arabinobiosidase, BIGH121 was able to transglycosylate  $\beta$ -1,2-Ara<sub>2</sub> in the presence of primary alcohols (27), which is indicative of retaining mechanisms. When incubated with purified  $\beta$ -1,2-Ara<sub>2</sub> and MeOH or EtOH, NF2152 also generated products with shifts in mobility, which is consistent with transfer of the Ara<sub>2</sub>, through a double displacement mechanism (Fig. 6C).

Although the core catalytic residues are spatially conserved between NF2152 and TmGH51, closer inspection of the C2-OH of the Ara in the -1 subsite suggested that in this orientation the Ara is poised to form an interaction with Asn<sup>154/171</sup>. This interaction provides a recognition determinant *exo*-acting ABFs, such as TmGH51, which only removes single Ara decorations (34). In NF2152 this interaction with Asn<sup>154</sup> would preclude a conjugated  $\beta$ -1,2-Ara or  $\alpha$ -1,2-Gal from extending deeper into the pocket. Therefore, in this sub-family of anaerobic fungal GH39 enzymes it is probable that the Ara in the -1 adopts a different orientation. In support of this, the two tryptophan residues positioned at the mouth to the

**Table 2**  
NF2152 data collection and structure statistics

	Statistics
<b>Data collection</b>	
Beamline	Rigaku MicroMax-007HF
Wavelength (Å)	1.54178
Space group	P2 <sub>1</sub>
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	36.99, 104.79, 49.45
$\alpha$ , $\beta$ , $\gamma$ (°)	90.00, 110.06, 90.00
Resolution (Å) <sup>a</sup>	28.95–1.75 (1.795–1.75)
<i>R</i> <sub>merge</sub>	0.104 (0.381)
Wilson B-factor	19.11
( <i>I</i> / $\sigma$ <i>I</i> )	12.3 (2.8)
Completeness (%)	99.5 (98.8)
Redundancy	4.8 (3.1)
No. of reflections	170,864
No. unique	35,446
<b>Refinement</b>	
Resolution (Å)	1.75
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.156/0.194
No. of atoms	
Protein	3183
Ligand	13 (I), 12 (EDO), 13 (BTB)
Water	314
<i>B</i> -factors	
Protein	18.56
Ligand	50.06 (I), 22.87 (EDO), 20.89 (BTB)
Water	25.98
R.m.s. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	0.861
Ramachandran (%)	
Preferred	96.3
Allowed	3.7
Disallowed	0.0
<b>PDB code</b>	5U22

<sup>a</sup> Values for highest resolution shells are shown in parentheses.

active site cleft adopt strikingly different conformations in NF2152 and TmGH51 (Fig. 6B).

Blast searches of NF2152 determined that there are other related entries in the database, including bacterial entries from *Clostridia* spp. and *Fibrobacter* spp., suggesting that NF2152 and NF2523 may be representative members of a larger subfamily. To investigate the functional diversity of other fungal GH39s, digestions were performed on RAX with NF2215 and PR2455. These results demonstrated that they each have subtle differences in their product profiles when compared with NF2152 and NF2523 (Fig. 6D). These signature patterns occur despite strict conservation of residues lining the active site in the –1 and –2 subsites, which are predicted to be involved in substrate binding and catalysis (Fig. 6D). Four residue positions that displayed variations in primary structure were identified, including His<sup>52</sup>, Gln<sup>87</sup>, Glu<sup>110</sup>, and Val<sup>291</sup>. The conservation of these residues within the larger group of anaerobic fungal GH39 homologs was assessed by performing a cluster analysis with 12 sequences with Clustal Omega (36), and then mapped onto the surface of NF2152 using ConSurf (37) (Fig. 6E). Analysis of the active site confirmed that His<sup>52</sup>, Asn<sup>87</sup>, and Val<sup>291</sup> (*white*) and Glu<sup>110</sup> (*cyan*) represent potential hot spot sites for variation within the –2 subsite (Fig. 6E). In particular, the transition from Val<sup>291</sup> in NF2152 to an Asn in NF2523 and Thr in PR2455 may illuminate key molecular determinants for recognition of Gal at the non-reducing end  $\alpha$ -1,2-Gal-Ara product. An  $\beta$ -1,2-Ara<sub>2</sub> with an  $\alpha$ -L-Ara at the reducing end could be reasonably docked into the active site of NF2152, demonstrating that there is sufficient room to accommodate the product.

This interaction would require a ~20° clockwise rotation of the Ara in the –1 subsite. Importantly, His<sup>52</sup> (3.7 Å from the O4), Gln<sup>87</sup> (3.2 and 3.0 Å from the C2-OH and C3-OH, respectively), and Glu<sup>110</sup> (3.8 Å from the C5-OH) are disposed in reasonable proximity to interact with the side chains of Ara in the –2 subsite (Fig. 6F). In addition, Val<sup>291</sup> (3.8 Å from the C2-OH) may contribute to van der Waals interactions.

## Discussion

Discovery of new enzyme activities that improve the digestibility of recalcitrant plant cell wall polysaccharides are promising solutions for clean, sustainable industries. Additionally, biocatalysts with novel activities that operate as surgical tools hold promise for dissecting the structures of complex glycans or generating rare or commercially valuable carbohydrates. In this regard, the rumen anaerobic fungi represent an underexplored source of potentially unique enzymes. Here we have investigated the activities of GH39 enzymes from rumen fungi that were identified by their association with CBM13s.

Most commonly, CBM binding specificity parallels the catalytic activity of the appended enzyme module, and therefore, investigating uncharacterized enzymes associated with CBMs with diverse specificities should facilitate the discovery of new CAZyme families, and potentially, activities. Using this rationale, hypothetical proteins appended to predicted CBM13s (19) were identified within fungal transcriptomes from *N. frontalis* and *P. rhizinflatus*. These sequences were trimmed to an N-terminal fragment of sufficient length to encode a functional enzyme, and embedded into extracted datasets of characterized GH39 and GH51 sequences to generate “informed” phylogenetic trees. GH39 and GH51 are well characterized polyspecific enzyme families with activity on a wide range of substrates (14). In both trees the fungal sequences partition as a single clade, distantly related to bacterial  $\beta$ -xylosidases and ABFs, in GH39 and GH51, respectively (Fig. 1). These phylogenetic relationships suggest that GH39 enzymes may have diverged following horizontal gene transfer from a bacterial ancestor; and based upon sequence, they may be active in the hydrolysis of either  $\alpha$ -L-arabinofuranosyl or  $\beta$ -xylosyl substrates. However, with the contemporary perspective that many CAZyme families are polyspecific (15, 16), including GH39 and GH51, it has become apparent that accurate assignment of enzyme activity cannot rely solely on phylogenetic patterns and requires evidence-based biochemical characterization.

To investigate the function of these fungal GH39s, a synthetic form of NF2152 was produced and screened for activity on a variety of plant cell wall carbohydrates. NF2152 was shown to release an Ara<sub>2</sub> product from SBA (Fig. 2) and RAX (Fig. 3A). The release of similar products from two substrates isolated from different plant sources that display different structures suggests that there is a rare substrate consistent within both sources of plant polysaccharides.

SBA is an Ara-rich glycan found within the side chains of RG-I. It is composed of a  $\alpha$ -1,5-L-arabinofuranosyl backbone decorated with  $\alpha$ -1,2- and  $\alpha$ -1,3-L-arabinofuranosyl side chains. Ara is an abundant pentose in nature; and therefore, discovery of novel biocatalysts active on arabinans are promising tools for bioconversion industries. There are many enzymes

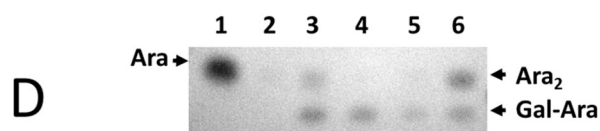
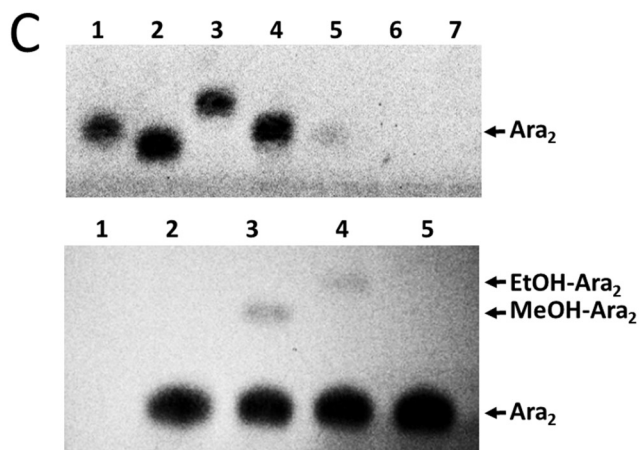
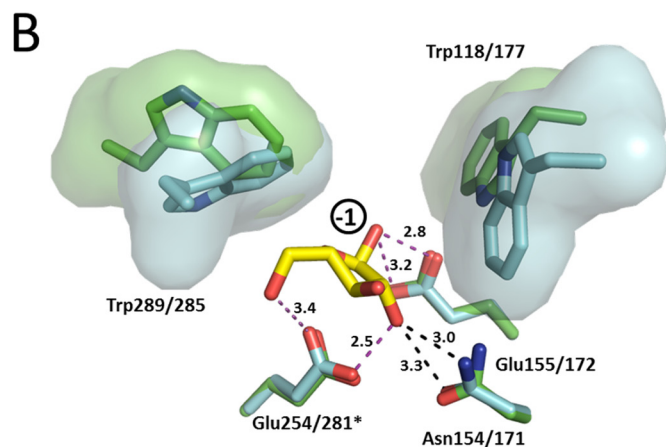
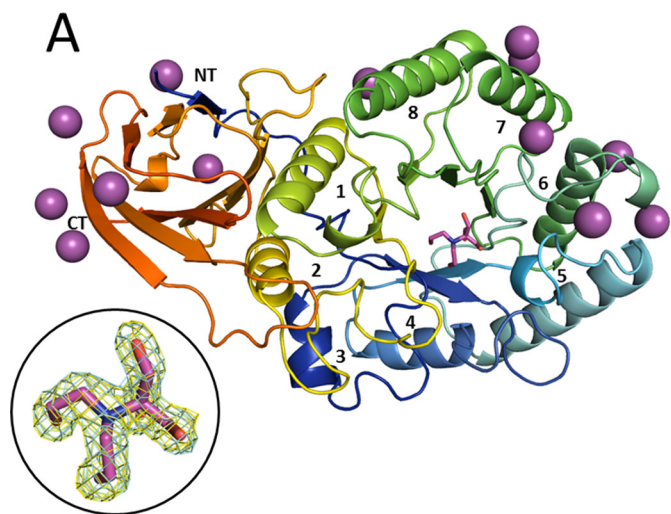


## Characterization of a novel glycoside hydrolase activity

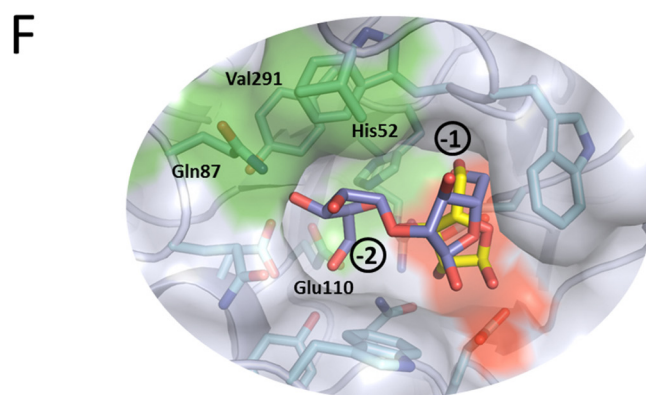
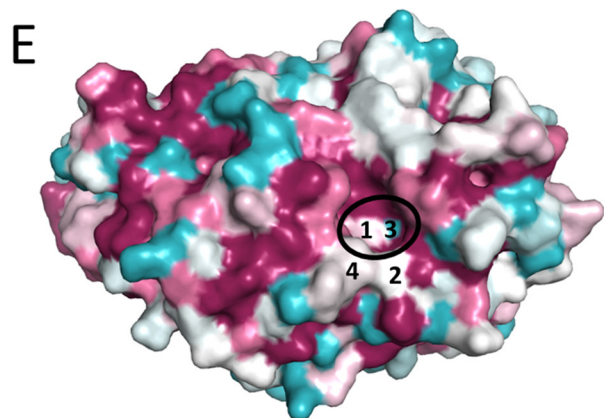
known to be active on SBA. Collectively these enzymes are referred to as arabinanases (ABNs), which are found in GH43 and GH93 families (38); and ABFs, which are found in GH3, GH43, GH51, GH54, GH62, GH93, and GH127 families (25, 27, 39). ABNs catalyze the hydrolysis of the  $\alpha$ -1,5-L-arabinofuranosyl backbone of plant cell wall arabinans, releasing arabinooligosaccharides and Ara. ABFs cleave arabinosyl decorations in an *exo*-fashion. Currently, several  $\alpha$ -1,5-arabinobiosidases have been reported that processively hydrolyze the backbone of arabinan (40). Recent studies have shown that the synergistic effect

of fungal enzyme mixtures supplemented with ABNs and ABFs improves the hydrolysis rate of plant biomass (41).

Arabinoxylans are an abundant polysaccharide within the plant cell wall (42). RAX possesses a D-xylosyl backbone decorated with  $\alpha$ -1,2- and  $\alpha$ -1,3-linked L-arabinosyl residues (42, 43). Digestion of RAX requires a combination of enzyme activities, including ABFs and xylanases, which are found mainly in GH10, GH11, and also in GH5, GH7, GH8, and GH43 (44). Multiple enzyme products that target the xylan backbone of arabinoxylan products are currently available indicating



	His52*	Tyr53*	Arg76*	Gln86*	Gln87*	Glu110*	Asp113*	Trp118	Tyr120*	Asn154	Glu155	Glu254	Trp288	Trp289*	Val291*
NF2152	H	Y	R	Q	Q	E	D	W	Y	N	E	F	W	W	V
NF2215	L	Y	R	Q	H	D	D	W	Y	N	E	E	W	W	V
NF2523	L	Y	R	Q	H	D	D	W	Y	N	E	E	W	W	N
PR2455	L	Y	R	Q	H	L	D	W	Y	N	E	E	W	W	T
BGXG1	L	Y	R	Q	H	D	D	W	Y	N	E	E	W	W	V
1px8	-	-	R	-	-	E	H	V	F	N	E	E	F	F	-
4obr	-	-	R	-	-	E	H	-	W	N	E	E	F	R	-



there is a market for biocatalysts that improve the digestion of RAX.

Using acid hydrolysis (Fig. 2, *B* and *C*), and diagnostic enzyme digests and HR-MS (Fig. 2, *D* and *E*), and diagnostic digestions of a purified arabinotetraose (Fig. 2*E*), the product released by NF2152 was determined to be  $\beta$ -1,2-*L*-Araf<sub>2</sub>. This underpinned that the target substrate is a rare, and potentially trace contaminant within SBA and RAX. Additionally, it reveals that NF2152 is not a conventional ABF or xylosidase as was suggested by phylogenetic analysis (Fig. 1). Recently, BIGH121 was reported to be the first described  $\beta$ -1,2-arabinobiosidase, which is active on AGP glycans that co-purify with SBA (27). Importantly, BIGH121 products are also hydrolyzed by BIGH127 (25). This suggests that GH39 enzymes, found in anaerobic fungi, and BIGH121, a family of bacterial enzymes, may have undergone functional convergence. To the best of our knowledge this small collection of enzymes represents some of the only known enzymes specific for the degradation of  $\beta$ -1,2-arabinofuranoside containing glycans.

The detection of a second product with lower mobility was observed when RAX was digested with NF2152 and NF2523 (Fig. 2 and 3). The composition of this heterogeneous disaccharide was determined to be Gal and Ara (Fig. 3*B*). To provide more insight into the product chemistry, the disaccharide was sequenced using differential alditol acetate and ANTS labeling, and analyzed by GC-MS (Fig. 4*A*) and FACE (Fig. 4*B*), respectively. Both techniques confirmed that Ara was positioned at the reducing end; and therefore, the product is a Gal-Ara disaccharide. Although NMR is the gold standard for these types of analyses, the development of differential labeling techniques can be a valuable approach for rapidly sequencing heterogeneous disaccharides if access to NMR infrastructure is lacking or product yields are limiting.

Preliminary NMR spectra of both products suffered from high backgrounds of acetate (reaction buffer) and multiple signals for the anomeric carbon, suggesting the carbohydrates were adopting several conformations. Therefore, large scale purifications of both disaccharides were performed and the products were reduced to arabitols to linearize the compounds. <sup>1</sup>H NMR (Fig. 5, *A* and *C*) and <sup>13</sup>C NMR (Fig. 5, *B* and *D*) analyses confirmed the order of sequence for both the Ara<sub>2</sub> and Gal-Ara, and that both products contained  $\beta$ -1,2-linkages (Fig.

5, *E* and *F*). Additionally, at the non-reducing end Ara is in the furanose configuration (which is consistent with the BIGH127 digest, Fig. 2*D*), and Gal is in a pyranose configuration. This latter observation was somewhat surprising as we had anticipated the Gal residue may be in the  $\alpha$ -D-galactofuranose configuration, which has an analogous ring structure to  $\beta$ -L-arabinofuranose and only differs by the presence of its C5-hydroxyl methyl group (Fig. 5*G*). This establishes that the NF2523 substrates are of plant origin, and not of fungal origin as galactofuranoses are common components of fungal cell walls (45). Furthermore, elucidation that Ara was positioned at the reducing end in both products indicates that there is a strict requirement for Ara in their  $-1$  subsites and they display plasticity in their  $-2$  subsites (Fig. 5, *E* and *F*).

The structure of NF2152 confirmed that it adopts the canonical TIM-barrel-fold of GH39s and share significant structural similarity with GH51s (Fig. 6*A*). The polyspecificity reported for these families, including ABFs, xylosidases, endoglucanases, and many others (Fig. 1) underscores the functional plasticity of this core-fold (14, 46). The position of the active site was highlighted by the presence of a bound bis-tris buffer molecule (Fig. 6*A*) and the superimposition of the Ara product from TmGH51 (Fig. 6*B*). Ara is orientated with its C1 exiting the pocket between two sequence conserved aromatics (*i.e.* Trp<sup>118</sup> and Trp<sup>289</sup>), and within suitable distances to the structurally conserved catalytic residues, Glu<sup>155</sup> and Glu<sup>254</sup>. The involvement of Glu<sup>155</sup> and Glu<sup>254</sup> in catalysis was confirmed by site-directed mutagenesis and their potential roles in a retaining mechanism confirmed by the transglycosylation of MeOH and EtOH (Fig. 6*C*). As these features are conserved in other enzymes that function as ABFs and xylosidases, the specificity of the fungal GH39 enzymes for  $\beta$ -1,2-linked Araf containing disaccharides is believed to result primarily from interactions within the  $-2$  subsite (Fig. 6*D*). This observation is in contrast to Bgxxg1, which was reported to hydrolyze monosaccharides from a variety of synthetic substrates and purified disaccharides (24). Mapping the distribution of conserved surface-exposed active site residues in fungal GH39 sequences onto the surface of NF2152 revealed four key residues, His<sup>52</sup>, Asn<sup>87</sup>, Glu<sup>110</sup>, and Val<sup>291</sup>, as being variable within the  $-2$  subsite (Fig. 6*E*). Importantly, they are also in suitable proximity for potential interactions with the Araf at the reducing end of  $\beta$ -1,2-Araf<sub>2</sub> (Fig. 6*F*).

**Figure 6. Structural basis of rumen fungal GH39 enzyme specificity.** *A*, X-ray structure of NF2152 (PDB code 5U22) displayed in schematic format with a rainbow color scheme (NT = blue; CT = red). Each  $\alpha/\beta$  pair are labeled 1 through 8. The 13 iodide atoms are shown as purple spheres and the bis-tris methane in magenta sticks. Inset: bound bis-tris methane represented as maximum-likelihood (69)/ $\sigma_a$  (35) weighted  $2F_{\text{obs}} - 2F_{\text{calc}}$  electron density map contoured at  $1.0\sigma$  ( $0.43 \text{ e}^-/\text{\AA}^3$ ; cyan), and  $F_{\text{obs}} - F_c$  map contoured at  $3.0\sigma$  ( $0.42 \text{ e}^-/\text{\AA}^3$ ; yellow). *B*, active site of NF2152 (cyan) superimposed with Ara (yellow sticks) from TmGH51 (PDB code 3UG4; green). Residues in contact with Ara in the  $-1$  position are labeled. Distances of potential hydrogen bonds are indicated with magenta dashed lines and the distances are labeled. Steric clash of the O<sub>2</sub> with Asn<sup>154</sup> are indicated with black dashed lines. Tryptophans at the mouth of the active site pocket are displayed as sticks with a solvent accessible surface. Labeling system is NF2152/TmGH51. *C*, validation of catalytic residues and transglycosylation. *Top panel*, TLC of site-directed mutagenesis. All digests were performed on SBA. Lane 1 = Ara; lane 2 = Gal; lane 3 =  $\alpha$ -1,5-Ara<sub>2</sub>; lane 4 =  $\alpha$ -1,5-Ara<sub>3</sub>; lane 5 = NF2152; lane 6 = NF2152-E155Q; lane 7 = NF2152-E254Q. *Bottom panel*, alignment of amino acids from NF2152, NF2215, NF2523, PR2455, and Bgxxg1 with GH39 structures (1px8,  $\beta$ -D-xylose complex; 4obr,  $\alpha$ -L-iduronic acid complex). Residues include those that line the surface of the  $-1$  and  $-2$  subsites of NF2152. \* indicates residues comprising the  $-2$  subsite. Green, divergent residues; red, catalytic residues; dash, not conserved; underlined, structurally conserved, but contributed from a different loop. *E*, ConSurf model of GH39 fungal homologs ( $n = 12$ ) mapped on the structure of NF2152. Magenta = conserved; cyan = variable. Four key residues are labeled: 1 = His<sup>52</sup>; 2 = Gln<sup>87</sup>; 3 = Glu<sup>110</sup>; and 4 = Val<sup>291</sup>. The  $-2$  subsite in the active site cleft is highlighted with a black ellipse. *F*, transparent surface model of NF2152 docked with a built  $\alpha$ -L-( $\beta$ -1,2)-Ara<sub>2</sub> ligand (purple sticks) overlaid with the superimposed Ara (yellow sticks) product from *A*. Residues of the active site pocket are shown as cyan sticks. The  $-1$  and  $-2$  subsites are labeled. The catalytic residue surfaces are shown in red, and divergent residues in the  $-2$  subsite are labeled and shown with green surfaces.

## Characterization of a novel glycoside hydrolase activity

These residues, therefore, may represent genetic markers for directing the discovery of other GH39 enzymes with unique specificities.

### Conclusion

Although many microbial ecosystems are currently being mined for beneficial enzyme activities (e.g. human distal gut) the rumen microbiota remains one of the most promising repositories of microbial enzymes active on plant cell wall polysaccharides (20). In this study, four rumen fungal genes, classified as GH39s, were selected for characterization based upon their expression and association with CBM13. CBM13s are a family with reported diversity in their binding specificities (19).

NF2152 is an  $\alpha$ -L-( $\beta$ -1,2)-arabinofuranobiosidase, and NF2523 a D-galacto-( $\alpha$ -1,2)-L-arabinosidase. To the best of our knowledge this is the first galacto-( $\alpha$ -1,2)-arabinosidase activity reported in the literature. Based upon its weak activity on the SBA oligosaccharide substrate (not shown) and the structural conservation of catalytic residues within its active site, NF2523 is most likely an  $\alpha$ -(D-galacto-( $\alpha$ -1,2)-L-arabino)-sidase. Other related GH39 enzymes, NF2215 and PR2455, display subtle differences in the generation of both products but overall have similar product profiles to NF2152 and NF2523, respectively. Both  $\beta$ -1,2-Araf<sub>2</sub> and  $\alpha$ -1,2-Gal-Ara are unconventional substructures within SBA and RAX, and may also represent products released from AGP glycans that co-purify with the polysaccharides.  $\beta$ -1,2-Linked arabinosyl products were previously reported to be released from SBA during the characterization of the phylogenetically unrelated  $\beta$ -1,2-arabinobiosidase BIGH121 (27).

Very little is known about the role of AGPs in plant biology and biomass conversion, mainly due to their extensive structural complexity. Recently, it was reported that an AGP may interact directly with the pectin and hemicellulose, which challenges the traditional view that these polysaccharides are independent networks (47). In this regard, enzymes that target discreet linkages within the arabinan or AGP glycan network, such as the fungal GH39 enzymes studied here, represent tools for helping to characterize complex arabinosyl-containing glycans; and potentially, may assist in their turnover during bioconversion processes.

### Experimental procedures

#### Phylogenetic trees

Characterized sequences from families GH51 and GH39 were extracted and then trimmed to functional modules (i.e. GH catalytic fragments) by dbCAN (48). Trimmed fragments were then merged with catalytic fragments of the fungal sequences, and aligned via multiple sequence comparison by log-expectation (49). Phylogenetic grouping was performed with FastTree 2 (50). ProtTest3 was used for the selection of the best-fit model used by FastTree (51). Finally, FigTree was used for tree creation ([tree.bio.ed.ac.uk/software/figtree](http://tree.bio.ed.ac.uk/software/figtree)).<sup>4</sup>

<sup>4</sup> Please note that the JBC is not responsible for the long-term archiving and maintenance of this site or any other third party hosted site.

### Gene synthesis and cloning

Sequences corresponding to the mature GH39 gene products (lacking signal peptides) of NF2152, NF2215, NF2523, and PR2455 (GenBank accession numbers: MF326649, MF326650, MF326651, and MF326652) were commercially synthesized (BioBasic Inc., Markham, Canada) codon optimized for expression in *E. coli* and cloned into the pET28a vector. For crystallization experiments the DNA sequence corresponding to residues 29–431 of NF2152 was PCR amplified and cloned into the NdeI and XhoI sites of pET28a to create pET28\_nf2152\_29–431. All constructs were confirmed by DNA sequencing (Eurofins Genomics, Louisville, KY).

### Expression of synthetic genes and purification of recombinant proteins

The pET28a constructs were transformed into *E. coli* BL21 Star (DE3) for recombinant protein production. Cells were grown at 37 °C to an OD = 600 nm of 0.8–1.0 in LB broth containing kanamycin (50  $\mu$ g ml<sup>-1</sup>). Cultures were cooled to 16 °C and gene expression was induced with 0.2 mM isopropyl 1-thio- $\beta$ -D-galactopyranoside overnight at 200 rpm. Overnight cultures were centrifuged at 6,500  $\times$  g for 10 min. Cells were resuspended in Binding Buffer (BB: 0.5 M NaCl, 20 mM Tris, pH 8.0) and lysed by sonication for 2 min of 1-s intervals of medium intensity sonic pulses at a power setting of 4.5 (Heat Systems Ultrasonics Model W-225 and probe). The cell lysate was clarified by centrifugation at 17,500  $\times$  g for 45 min and passed through a 0.45- $\mu$ m filter. The filtrate was loaded onto a nickel-nitrilotriacetic acid column and purified by immobilized metal affinity chromatography. Recombinant protein was eluted via a stepwise gradient of imidazole (5, 10, 30, 100, 200, and 500 mM) in BB. Fractions containing significant amounts of protein were pooled and buffer exchanged into storage buffer (NF2152, 20 mM Tris-HCl, pH 8.0, 500 mM NaCl; NF2152\_29–431, 20 mM NaPO<sub>4</sub>, pH 7.2, 100 mM NaCl; NF2215, NF2523, and PR2455, 20 mM Tris-HCl, pH 8.0, 150 mM NaCl). Following buffer exchange, samples were concentrated using a nitrogen-pressurized stirred ultrafiltration cell (Amicon) with a molecular mass cut-off of 5 kDa. Concentrated protein was filtered and passed through a HiPrep 16/60 Sephacryl S-200 HR size exclusion column (GE Healthcare) at a flow rate of 1.0 ml min<sup>-1</sup> in storage buffer. Pure fractions were pooled and concentrated. NF2152 was further buffer exchanged into 20 mM Tris-HCl (pH 8.0), 100 mM NaCl. Protein concentration was determined using the Beer-Lambert law and extinction coefficients determined with ProtParam (52).

### Enzyme characterization

The activity of NF2152 was screened on a variety of plant cell wall carbohydrates including: sugar beet arabinan (catalog number P-ARAB), wheat arabinoxylan (P-WAXYM), rye arabinoxylan (P-RAXY),  $\beta$ -glucan (P-BGBL), xyloglucan (P-XYGLN), galactomannan (P-GGM28), pectic galactan (P-PGAPT), arabinogalactan (P-ARGAL), rhamnogalacturonan-I (P-RHAM I), arabinofuranosyl-xylobiose (O-A3X), arabinofuranosyl-xylotriose (O-A2XX), PNP- $\alpha$ -L-arabinofuranoside (O-PNPAF) purchased from Megazyme International, PNP- $\alpha$ -L-arabinopyranoside (38018) from Glycosynth, gum

arabic (G9752), and beech wood xylan (X4252) from Sigma. The enzyme concentrations, buffer, and pH were optimized for digestion through empirical studies with different substrates, enzyme concentrations, and buffers (pH values ranging from 4.0 to 8.0). The final reaction mixture contained 5 mg ml<sup>-1</sup> of polysaccharide substrate or 1.5 mM oligosaccharide, 0.5 μM of each enzyme, and 20 mM sodium acetate buffer (pH 5.0). These reactions were incubated overnight at 37 °C with samples being removed at various time points. After incubation the samples were heat treated at 100 °C for 10 min to denature the enzyme and terminate the reaction followed by short centrifugation at 8,000 × g to pellet denatured protein from the product. Products were analyzed by TLC and HPAEC-PAD to characterize their chemistry and degree of polymerization.

### Thin layer chromatography

Digested samples (total 9 μl; spotted 3 times with 3 μl each time) were spotted onto TLC plates (TLC Silica Gel 60; EMD Millipore Corp.). The samples were dried between multiple rounds of spotting. Appropriate standards (6 μl of 1 mM concentration) were also included in each run. The samples were resolved using a mobile phase of 2:1:1 butanol:water:acetic acid, dried prior to visualization with an orcinol solution (70:3, acetic acid:sulfuric acid with 1% orcinol) and heating at 100 °C for 3–5 min.

### HPAEC-PAD of monosaccharide and oligosaccharide reaction products

HPAEC-PAD was performed with a Dionex ICS-3000 chromatography system (Thermo Scientific) equipped with an autosampler as well as a pulsed amperometric (PAD) detector. 10 μl of aqueous sample were injected onto an analytical (3 × 150 mm) CarboPac PA20 column (Thermo Scientific) and eluted at 0.4 ml min<sup>-1</sup> flow rate with a sodium acetate gradient (0 to 1 min, 0 mM; 1 to 18 min, 250–850 mM; 18 to 20 min, 850 mM; 20 to 30 min, 850–0 mM) in a constant background of 100 mM NaOH. The elution was monitored with a PAD detector (standard quadratic waveform).

### Ethanol precipitation

Precipitations were performed on digested products to increase their purity by separating small products (e.g. monosaccharides and disaccharides) from larger oligosaccharides and polysaccharides by their differential solubility. The digested products were dried by SpeedVac at ambient temperature or lyophilization, and then suspended in 95% ethanol (v/v, EtOH:H<sub>2</sub>O). After incubation on ice for 10 min the resuspension was mixed vigorously with a vortex to dissolve the small oligosaccharides and monosaccharides and then clarified by centrifugation at 14,000 × g for 10 min. The supernatant was removed and dried by SpeedVac. Purified carbohydrates were subsequently used directly or suspended in water at the desired concentrations for further analyses.

### Acid hydrolysis

30 μg of purified dried products were incubated with 200 μl of 2.0 M trifluoroacetic acid for 4 h at 100 °C. The reaction mixture was then dried to completion in a SpeedVac followed by

three washes with 100 μl of isopropyl alcohol. Released monosaccharides were analyzed by TLC and HPAEC-PAD.

### Molecular weight determination by mass spectrometry

A large scale (total reaction volume: 200 ml) digest was performed to generate milligram amounts of product with NF2152 and SBA or NF2523 with RAX. The products were further purified by ethanol precipitation (95% v/v EtOH:H<sub>2</sub>O) and Bio-Gel P-2 (Bio-Rad Laboratories) size exclusion chromatography at a flow rate of 0.17 ml min<sup>-1</sup> where distilled water was used as eluent. Extra fine (<45 μM) Bio-Gel P-2 has particle size beads with an exclusion limit of 100–1,800 Da. The elution peaks were screened for purity by TLC, pooled, and lyophilized. Mass spectrometry was carried out on the pooled fractions to determine the *m/z* and molecular weight of the product (Alberta Glycomics Centre, Department of Chemistry, University of Alberta, Edmonton, Canada). Electrospray ionization mass spectra were recorded on an Agilent Technologies 6220 TOF instrument. The sample was dissolved in methanol or a methanol:water mixture (1:1) and directly injected into the instrument (5 μl). The spectra were recorded in positive mode.

### Carbohydrate sequencing by GC-MS

The purified product was used for gas chromatography to identify the product as well as to recognize the reducing end of the product. Gas chromatographic analysis of mono- and disaccharides requires conversion of sugars into their volatile derivatives (53). Sugars (purified product) were first converted into alditol acetates, which involved reduction of sugars with sodium borohydride (NaBH<sub>4</sub>) following conversion of polyols to polyacetate esters. Fifty micrograms of products were reduced by 200 μl of NaBH<sub>4</sub> (10 mg/ml of NaBH<sub>4</sub> in 1 M NH<sub>4</sub>OH). The reduced sugar was then acid hydrolyzed by 200 μl of 2 M TFA followed by an incubation at 100 °C for 4 h. The reaction mixture was then dried to completion in a SpeedVac followed by three wash steps with 100 μl of isopropyl alcohol. Released monosaccharides were acetylated by the addition of 250 μl of acetic anhydride and dried on a SpeedVac to a volume of 200 μl. The resulting solution was transferred to a GC autosampler vial containing a 250–300 micro insert and injected into gas chromatograph (Hewlett Packard 5890) where a polar capillary GC column (SP2330) and flame ionization detector was used.

### Carbohydrate sequencing by FACE

FACE was performed to identify the reducing end of the digested products. Carbohydrates were fluorescently labeled at their reducing end using ANTS, and the resulting labeled sugars were separated in a high percentage (40%) polyacrylamide gel (54). 30 μg of each sample were dried and suspended by vortexing in 5 μl of fresh 0.15 M ANTS (dissolved in 15% acetic acid) and 5 μl of fresh 1 M 2-picoline borane (dissolve in 1 ml of DMSO). Samples were incubated overnight at 37 °C in a tube wrapped in foil. The labeled samples were then dried with a SpeedVac for 2–4 h or until completely dry. The labeled sugar was then acid hydrolyzed by the addition of 200 μl of 2 M TFA followed by incubation at 100 °C for 4 h. The reaction mixture was dried to completion in a SpeedVac followed by three wash

## Characterization of a novel glycoside hydrolase activity

steps with 100  $\mu\text{l}$  of isopropyl alcohol. The dried pellet was suspended in 25  $\mu\text{l}$  of FACE loading dye and run on gel immediately or stored at  $-20^\circ\text{C}$  wrapped in foil.

### Linkage and ring configuration by NMR

Spectra were measured with a Varian VNMR5-500 MHz in  $\text{D}_2\text{O}$  at  $25^\circ\text{C}$  (Department of Chemistry and Biochemistry, Concordia University, Montreal, Canada). Chemical shifts and coupling constants were interpreted basing on one-dimensional NMR ( $^1\text{H}$  and  $^{13}\text{C}$ ) as well as two-dimensional NMR homonuclear correlation spectroscopy and heteronuclear single quantum correlation. The observation that the anomeric proton (proton 1') had a coupling constant of 5.0 Hz was evidence for  $\beta$ -configuration of the Ara (55).

### Protein crystallization NF2153 and structure solution

Purified NF2152\_29–431 was buffer exchanged into 20 mM bis-tris (pH 6.0), and concentrated to 20  $\text{mg ml}^{-1}$ . Crystals of NF2152 were grown using the hanging drop vapor diffusion method at  $18^\circ\text{C}$  in 20% (w/v) PEG 3350, 0.1 M bis-tris (pH 6.0), and 0.15 M NaI with a drop ratio of 1:2 protein solution to mother liquor. Crystals were cryoprotected with 25% (w/v) ethylene glycol and flash cooled directly in the cryostream at 100 K. Diffraction data were collected on an instrument comprising a MicroMax MM-007HF X-ray generator coupled to a Dectris Pilatus 200K detector with VariMax HF Arc Sec Confocal optics and an Oxford Cryostream Crystream 800 cryocooler. Data were processed and scaled with HKL-3000R (56). An initial model was determined by single-wavelength anomalous dispersion using SHELXC/D/E (57, 58). Phases were improved with Phaser (59), followed by successive rounds of density modification with Parrot (60) and model building with Buccaneer (61) using the CCP4 Phaser SAD pipeline (62). With an estimated solvent fraction of 35%, 13 iodide sites were identified. The resulting phases were sufficient for Arp/Warp to build a virtually complete model (63). This model was then used for iterative rounds of manual correction with COOT (64) and refinement with PHENIX (65). Water molecules were added with COOT FINDWATERS and manually checked after refinement. Throughout, refinement procedures were monitored by flagging 5% of all observation as “free” (66). Model validation was performed with MolProbity (67). Data collection and processing statistics are shown in Table 2.

### Site-directed mutagenesis

Glu<sup>155</sup> and Glu<sup>254</sup> in NF2152 were predicted to be the acid/base and nucleophile, respectively, based upon superimposition with the GH51 from *T. maritima* (3UG4)(34). These residues were targeted for substitution to glutamine using site-directed mutagenesis with pET28\_nf2152\_29–431 for template as previously described (68). Mutations were confirmed by DNA sequencing (Eurofins Genomics). Once sequence confirmed, the mutant enzymes were tested for activity with appropriate substrates.

### Transglycosylation assay

Size exclusion purified  $\beta$ 1,2-Ara<sub>2</sub> was incubated at a concentration of 1 mM with 1  $\mu\text{M}$  NF2152 in 20 mM sodium acetate (pH

5.0) and 20% (v/v) methanol or ethanol. After overnight incubation at  $37^\circ\text{C}$  the samples were heat treated at  $100^\circ\text{C}$  for 10 min to terminate the reaction. Reaction products were visualized by TLC.

### Docking of $\beta$ -L-Araf-(1–2)- $\alpha$ -L-Araf with NF2153

PDB coordinates for a  $\beta$ -L-Araf-(1–2)- $\alpha$ -L-Araf molecule were generated with the GLYCAM Carbohydrate builder ([www.glycam.org](http://www.glycam.org)).<sup>4</sup> AutoDock Vina was used to dock the Ara<sub>2</sub> into the NF2152 active site pocket using a search area of 2,560  $\text{\AA}^3$  at an exhaustiveness setting of 8. The best fitting model had a predicted binding affinity of  $-6.5 \text{ kcal mol}^{-1}$ , with an upper and lower r.m.s. deviation bound of 2.7 and 1.3  $\text{\AA}$ , respectively.

*Author contributions*—D. R. J. performed enzyme characterization, product analysis, and solved the structure of NF2152; and assisted with figure generation and manuscript writing. M. S. U. performed enzyme screening, mutagenesis, and analysis of product structure; and assisted with writing and figure generation. R. J. G. generated fungal transcriptomes and gene sequences and assisted with protein crystallization. M. P. T. T. performed NMR experiments and determined structures of reduced SAB and RAX products. D. T. generated phylogenetic trees. A. B. B. and B. P. assisted with X-ray data collection and structure solution. J. B. generated the arabinotetraose. T. A. M. and R. F. helped design the study and identify target sequences. A. T. sequenced fungal transcripts, and helped in study design and NMR structure determination. L. B. S. helped in study design and interpretation of results. D. W. A. conceived of study, contributed to interpretation of the results, and wrote the initial draft of the paper. All authors reviewed the results and approved the final version of the manuscript.

*Acknowledgments*—We thank Professor Harry Gilbert (Newcastle University) for providing ABFs for diagnostic digestions of the SBA product that confirmed the products were  $\alpha$ -arabinosides. In addition, we thank the Alberta Glycomics Centre for assistance with the HR-MS and Darrell Vedres (AAFC) with the GC-MS analysis.

### References

1. Morgavi, D. P., Kelly, W. J., Janssen, P. H., and Attwood, G. T. (2013) Rumen microbial (meta)genomics and its application to ruminant production. *Animal* **7**, 184–201
2. Ribeiro, G. O., Gruninger, R., Badhan, A., and McAllister, T. A. (2016) Mining the rumen for fibrolytic feed enzymes. *Animal Front.* **6**, 20–26
3. Hess, M., Sczyrba, A., Egan, R., Kim, T. W., Chokhawala, H., Schroth, G., Luo, S., Clark, D. S., Chen, F., Zhang, T., Mackie, R. I., Pennacchio, L. A., Tringe, S. G., Visel, A., Woyke, T., Wang, Z., and Rubin, E. M. (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* **331**, 463–467
4. Qi, M., Wang, P., O'Toole, N., Barboza, P. S., Ungerfeld, E., Leigh, M. B., Selinger, L. B., Butler, G., Tsang, A., McAllister, T. A., and Forster, R. J. (2011) Snapshot of the eukaryotic gene expression in muskoxen rumen: a metatranscriptomic approach. *PLoS One* **6**, e20521
5. Dai, X., Tian, Y., Li, J., Luo, Y., Liu, D., Zheng, H., Wang, J., Dong, Z., Hu, S., and Huang, L. (2015) Metatranscriptomic analyses of plant cell wall polysaccharide degradation by microorganisms in the cow rumen. *Appl. Environ. Microbiol.* **81**, 1375–1386
6. Houston, K., Tucker, M. R., Chowdhury, J., Shirley, N., and Little, A. (2016) The plant cell wall: a complex and dynamic structure as revealed by the responses of genes under stress conditions. *Front. Plant Sci.* **7**, 984
7. Mohnen, D., Bar-Peled, M., and Somerville, C. (2008) Cell wall polysaccharide synthesis. *Biomass Recalcitrance: Deconstructing Plant Cell Wall*

- Bioenergy*, pp. 94–187, Blackwell Publishing, John Wiley and Sons, Hoboken, NJ
8. Bengtsson, S., Åman, P., and Andersson, R. (1992) Structural studies on water-soluble arabinoxylans in rye grain using enzymatic hydrolysis. *Carbohydr. Polymers* **17**, 277–284
  9. Jarvis, M., Briggs, S., and Knox, J. (2003) Intercellular adhesion and cell separation in plants. *Plant Cell Environ.* **26**, 977–989
  10. Atmodjo, M. A., Hao, Z., and Mohnen, D. (2013) Evolving views of pectin biosynthesis. *Annu. Rev. Plant Biol.* **64**, 747–779
  11. Ndeh, D., Rogowski, A., Cartmell, A., Luis, A. S., Baslé, A., Gray, J., Venditto, L., Briggs, J., Zhang, X., Labourel, A., Terrapon, N., Buffet, F., Nepogodiev, S., Xiao, Y., Field, R. A., *et al.* (2017) Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature* **544**, 65–70
  12. Jones, L., Milne, J. L., Ashford, D., and McQueen-Mason, S. J. (2003) Cell wall arabinan is essential for guard cell function. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 11783–11788
  13. Showalter, A. M. (2001) Arabinogalactan-proteins: structure, expression and function. *Cell. Mol. Life Sci.* **58**, 1399–1417
  14. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495
  15. Mewis, K., Lenfant, N., Lombard, V., and Henrissat, B. (2016) Dividing the large glycoside hydrolase family 43 into subfamilies: a motivation for detailed enzyme characterization. *Appl. Environ. Microbiol.* **82**, 1686–1692
  16. Aspeborg, H., Coutinho, P. M., Wang, Y., Brumer, H., 3rd, and Henrissat, B. (2012) Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol. Biol.* **12**, 186
  17. Davies, G., and Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases. *Structure* **3**, 853–859
  18. Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J.* **382**, 769–781
  19. Fujimoto, Z. (2013) Structure and function of carbohydrate-binding module families 13 and 42 of glycoside hydrolases, comprising a beta-trefoil fold. *Biosci. Biotechnol. Biochem.* **77**, 1363–1371
  20. Solomon, K. V., Haitjema, C. H., Henske, J. K., Gilmore, S. P., Borges-Rivera, D., Lipzen, A., Brewer, H. M., Purvine, S. O., Wright, A. T., Theodorou, M. K., Grigoriev, I. V., Regev, A., Thompson, D. A., and O'Malley, M. A. (2016) Early-branching gut fungi possess a large, comprehensive array of biomass-degrading enzymes. *Science* **351**, 1192–1195
  21. Badhan, A., Wang, Y., Gruninger, R., Patton, D., Powlowski, J., Tsang, A., and McAllister, T. (2014) Formulation of enzyme blends to maximize the hydrolysis of alkaline peroxide pretreated alfalfa hay and barley straw by rumen enzymes and commercial cellulases. *BMC Biotechnol.* **14**, 31
  22. Gruninger, R. J., Puniya, A. K., Callaghan, T. M., Edwards, J. E., Youssef, N., Dagar, S. S., Fliegerova, K., Griffith, G. W., Forster, R., Tsang, A., McAllister, T., and Elshahed, M. S. (2014) Anaerobic fungi (phylum *Neocallimastigomycota*): advances in understanding their taxonomy, life cycle, ecology, role and biotechnological potential. *FEMS Microbiol. Ecol.* **90**, 1–17
  23. Couger, M. B., Youssef, N. H., Struchtemeyer, C. G., Liggenstoffer, A. S., and Elshahed, M. S. (2015) Transcriptomic analysis of lignocellulosic biomass degradation by the anaerobic fungal isolate *Orpinomyces* sp. strain C1A. *Biotechnol. Biofuels* **8**, 208
  24. Morrison, J. M., Elshahed, M. S., and Youssef, N. (2016) A multifunctional GH39 glycoside hydrolase from the anaerobic gut fungus *Orpinomyces* sp. strain C1A. *Peer J.* **4**, e2289
  25. Fujita, K., Takashi, Y., Obuchi, E., Kitahara, K., and Sugauma, T. (2014) Characterization of a novel  $\beta$ -L-arabinofuranosidase in *Bifidobacterium longum* functional elucidation of a DUF1680 protein family member. *J. Biol. Chem.* **289**, 5240–5249
  26. Yapo, B. M. (2011) Rhamnolacturonan-I: a structurally puzzling and functionally versatile polysaccharide from plant cell walls and mucilages. *Polymer Rev.* **51**, 391–413
  27. Fujita, K., Sakamoto, S., Ono, Y., Wakao, M., Suda, Y., Kitahara, K., and Sugauma, T. (2011) Molecular cloning and characterization of a  $\beta$ -L-arabinobiosidase in *Bifidobacterium longum* that belongs to a novel glycoside hydrolase family. *J. Biol. Chem.* **286**, 5143–5150
  28. Bock, K., Pedersen, C., and Pedersen, H. (1984) Carbon-13 nuclear magnetic resonance data for oligosaccharides. *Adv. Carbohydr. Chem. Biochem.* **42**, 193–225
  29. Duss, J., Gottfredsen, C. H., and Bock, K. (2000) Carbohydrate structural determination by NMR spectroscopy: modern methods and limitations. *Chem. Rev.* **100**, 4589–4614
  30. Holm, L., and Rosenström, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549
  31. Santos, C. R., Polo, C. C., Corrêa, J. M., Simão Rde, C., Seixas, F. A., and Murakami, M. T. (2012) The accessory domain changes the accessibility and molecular topography of the catalytic interface in monomeric GH39 beta-xylosidases. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 1339–1345
  32. Paës, G., Skov, L. K., O'Donohue, M. J., Rémond, C., Kastrop, J. S., Gajhede, M., and Mirza, O. (2008) The structure of the complex between a branched pentasaccharide and *Thermobacillus xylanilyticus* GH-51 arabinofuranosidase reveals xylan-binding determinants and induced fit. *Biochemistry* **47**, 7441–7451
  33. Sainz-Polo, M. A., Valenzuela, S. V., González, B., Pastor, F. I., and Sanz-Aparicio, J. (2014) Structural analysis of glucuronoxylan-specific Xyn30D and its attached CBM35 domain gives insights into the role of modularity in specificity. *J. Biol. Chem.* **289**, 31088–31101
  34. Im, D. H., Kimura, K., Hayasaka, F., Tanaka, T., Noguchi, M., Kobayashi, A., Shoda, S., Miyazaki, K., Wakagi, T., and Tushinobu, S. (2012) Crystal structures of glycoside hydrolase family 51  $\alpha$ -L-arabinofuranosidase from *Thermotoga maritima*. *Biosci. Biotechnol. Biochem.* **76**, 423–428
  35. Read, R. J. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. Sect. A* **42**, 140–149
  36. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Systems Biol.* **7**, 539
  37. Goldenberg, O., Erez, E., Nimrod, G., and Ben-Tal, N. (2009) The ConSurf-DB: pre-calculated evolutionary conservation profiles of protein structures. *Nucleic Acids Res.* **37**, D323–D327
  38. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* **37**, D233–D238
  39. Shi, H., Ding, H., Huang, Y., Wang, L., Zhang, Y., Li, X., and Wang, F. (2014) Expression and characterization of a GH43 endo-arabinanase from *Thermotoga thermarum*. *BMC Biotechnol.* **14**, 35
  40. Sakamoto, T., and Thibault, J.-F. (2001) Exo-arabinanase of *Penicillium chrysogenum* able to release arabinobiose from  $\alpha$ -1,5-L-arabinan. *Appl. Environ. Microbiol.* **67**, 3319–3321
  41. Santos, C. R., Polo, C. C., Costa, M. C., Nascimento, A. F., Meza, A. N., Cota, J., Hoffmam, Z. B., Honorato, R. V., Oliveira, P. S., Goldman, G. H., *et al.* (2014) Mechanistic strategies for catalysis adopted by evolutionary distinct family 43 arabinanases. *J. Biol. Chem.* **289**, 7362–7373
  42. Vinkx, C., and Delcour, J. (1996) Rye (*Secale cereale* L.) Arabinoxylans: a critical review. *J. Cereal Sci.* **24**, 1–14
  43. Vinkx, C., Reynaert, H., Grobet, P., and Delcour, J. (1993) Physicochemical and functional properties of rye nonstarch polysaccharides: V. variability in the structure of water-soluble arabinoxylans. *Cereal Chem.* **70**, 311–311
  44. Collins, T., Gerday, C., and Feller, G. (2005) Xylanases, xylanase families and extremophilic xylanases. *FEMS Microbiol. Rev.* **29**, 3–23
  45. Tefsen, B., Ram, A. F., van Die, I., and Routier, F. H. (2012) Galactofuranose in eukaryotes: aspects of biosynthesis and functional impact. *Glycobiology* **22**, 456–469
  46. Tóth-Petróczy, A., and Tawfik, D. S. (2014) The robustness and innovability of protein folds. *Curr. Opin. Struct. Biol.* **26**, 131–138
  47. Tan, L., Eberhard, S., Pattathil, S., Warder, C., Glushka, J., Yuan, C., Hao, Z., Zhu, X., Avci, U., Miller, J. S., Baldwin, D., Pham, C., Orlando, R., Darvill, A., Hahn, M. G., Kieliszewski, M. J., and Mohnen, D. (2013) An *Arabidopsis* cell wall proteoglycan consists of pectin and arabinoxylan covalently linked to an arabinogalactan protein. *Plant Cell* **25**, 270–287
  48. Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012) dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **40**, W445–W451

## Characterization of a novel glycoside hydrolase activity

49. Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797
50. Price, M. N., Dehal, P. S., and Arkin, A. P. (2010) FastTree 2: approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490
51. Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2011) ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164–1165
52. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. e., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005) Protein identification and analysis tools on the ExPASy server. *Methods Mol. Biol.* **112**, 531–552
53. Blakeney, A. B., Harris, P. J., Henry, R. J., and Stone, B. A. (1983) A simple and rapid preparation of alditol acetates for monosaccharide analysis. *Carbohydr. Res.* **113**, 291–299
54. Gao, N. (2005) Fluorophore-assisted carbohydrate electrophoresis: a sensitive and accurate method for the direct analysis of dolichol pyrophosphate-linked oligosaccharides in cell cultures and tissues. *Methods* **35**, 323–327
55. Joseleau, J.-P., Chambat, G., Vignon, M., and Barnoud, F. (1977) Chemical and <sup>13</sup>C NMR studies of two arabinans from the inner bark of young stems of *Rosa Glauca*. *Carbohydr. Res.* **58**, 165–175
56. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326
57. Sheldrick, G. M. (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 479–485
58. Schneider, T. R., and Sheldrick, G. M. (2002) Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1772–1779
59. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storz, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674
60. Cowtan, K. (2010) Recent developments in classical density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 470–478
61. Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011
62. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763
63. Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nature Protoc.* **3**, 1171–1179
64. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
65. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221
66. Brünger, A. T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475
67. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21
68. McLean, R., Hobbs, J. K., Suits, M. D., Tuomivaara, S. T., Jones, D. R., Boraston, A. B., and Abbott, D. W. (2015) Functional analyses of resurrected and contemporary enzymes illuminate an evolutionary path for the emergence of exolysis in polysaccharide lyase family 2. *J. Biol. Chem.* **290**, 21231–21243
69. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255
70. Rambaut, A. (2014) *FigTree – Tree Figure Drawing Tool*, Version 1.4.2, Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom
71. Woods Group (2017) *Carbohydrate Builder – GLYCAM Web*, Complex Carbohydrate Research Center, University of Georgia, Athens, GA

**Discovery and characterization of family 39 glycoside hydrolases from rumen anaerobic fungi with polyspecific activity on rare arabinosyl substrates**

Darryl R. Jones, Muhammed Salah Uddin, Robert J. Gruninger, Thi Thanh My Pham, Dallas Thomas, Alisdair B. Boraston, Jonathan Briggs, Benjamin Pluvinae, Tim A. McAllister, Robert J. Forster, Adrian Tsang, L. Brent Selinger and D. Wade Abbott

*J. Biol. Chem.* 2017, 292:12606-12620.

doi: 10.1074/jbc.M117.789008 originally published online June 6, 2017

---

Access the most updated version of this article at doi: [10.1074/jbc.M117.789008](https://doi.org/10.1074/jbc.M117.789008)

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

This article cites 68 references, 14 of which can be accessed free at <http://www.jbc.org/content/292/30/12606.full.html#ref-list-1>