CHARACTERIZATION OF NFGHX: A FOUNDING MEMBER OF A NEW GLYCOSIDE HYDROLASE FAMILY FROM THE ANAEROBIC FUNGUS *NEOCALLIMASTIX FRONTALIS*

MUHAMMED SALAH UDDIN
B.Sc., University of Dhaka, January 2010

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfilment of the
Requirements for the Degree

MASTER OF SCIENCE

Department of Biological Sciences
University of Lethbridge
Lethbridge, Alberta, Canada

CHARACTERIZATION OF NFGHX A FOUNDING MEMBER OF A NEW GLYCOSIDE HYDROLASE FAMILY FROM THE ANAEROBIC FUNGUS *NEOCALLIMASTIX FRONTALIS*

MUHAMMED SALAH UDDIN

Date of Defense: August 30th, 2016

Dr. D. Wade Abbott Co-supervisor	Research Scientist	Ph.D.
Dr. L. Brent Selinger Co-supervisor	Professor	Ph.D.
Dr. Tony Russell Thesis Examination Committee Member	Assistant Professor	Ph.D.
Dr. Nehal Kumar Thakor Thesis Examination Committee Member	Assistant Professor	Ph.D.
Dr. Harry Brumer External Examiner University of British Columbia Vancouver, Canada	Professor	Ph.D.
Dr. Robert Laird Chair, Thesis Examination Committee	Associate Professor	Ph.D.

Dedication

This work is dedicated to my beloved wife Razia Sultana - thank you for your love, sacrifice, support, inspiration and understanding. I promise we will not be apart like this in future; to my younger brother and best friend Md. Jashim Uddin - for your impetus words and suggestions in all of my difficulties; my adoring mother Mrs. Sakila Khatun - for your encouragement, prayers and support; and my caring father Mr. Alauddin Ahmed - for your guidance, motivation and prayers. Thank you all for your unconditional love, intangible stimulus and support that I received each and every day from thousands mile away all through this program.

Above all, I dedicate this work to the almighty Allah, most merciful beneficent creator who has enabled me to perform this research work.

I love you all.

Abstract

The discovery and characterization of enzymes that catalyze new reactions, and increase the rate or efficiency of bioconversion are promising lines of biotechnological research. In this regard the rumen anaerobic fungi represent an underexploited source of novel carbohydrate active enzymes that are active on recalcitrant plant cell wall polysaccharides. In this study, four rumen fungal genes (Nf2152, Nf2215, Nf2523 and Pr2455) were identified as functional CAZymes that represent founding members of a new GH family, referred to here as 'GHX'. Two members of GHX, Nf2152 and Nf2523, were functionally characterized, and it was discovered that they have two different activities. Nf2152 releases a single product (β-1,2-arabinobiose) from sugar beet arabinan (SBA), and β-1,2-arabinobiose and a galactoarabinose with an unknown linkage from rye arabinoxylan (RAX). Nf2523 exclusively produces galactoarabinose from RAX. To the best of my knowledge, this represents the first reported galactoarabinobiosidase. Both β-1,2-arabinobiose and galactoarabinose are not conventional structures within SBA and RAX, and in fact are more likely released from the glycan side chains of arabinogalactan proteins (AGPs) that co-purify with the polysaccharides. In this regard, the GHXs studied here represent valuable tools for sequencing diverse AGP glycans and assisting in their bioconversion for industrial applications, such as animal production.

Acknowledgements

I would like to convey my gratitude to my co-supervisor Dr. D Wade Abbott for his borderless time, guidance, expertise and lab resources in the completion of this project. I am indebted to him for this research and study opportunity. Thank you for being an awesome mentor and a generous guide. Without your continuous support and supervision, it would not be possible for me to complete this research project.

I would also like to thank my co-supervisor Dr. L Brent Selinger for his time, direction, insight and expertise throughout my research project. To my committee members, Drs. Nehal Kumar Thakor and Tony Russell, thank you for your guidance and expertise during and outside committee meetings. To Dr. Harry Brumer, thank you for acting as my external examiner.

I would like to thank all of the Abbott lab members, past and present, who have been of help to me in different ways: Dr. Darryl Jones, thank you for all of your time and guidance to enhance my laboratory skills and knowledge; Dr. Julie Grondin thank you for all of your direction anytime I asked for help. Richard McLean and Anuoluwapo Anele, thanks for the company in lab on the evening and weekends, the good laughs while working at protein lab and for enthusiastically listening to all of my 'newest' findings. Thank you Dr. Carolyn Amundsen for your prompt response anytime I requested for new substrates. Thanks Adam Smith for helping me on the labelling protocol. Thank you Jaclyn Macmillan, Marshall Smith, Leeann Klassen, Jeffrey Tingley, Ben Vuong for being lovely team members. Also all the past members: Benjamin J. Farnell, Kaitlyn Shearer,

Justin Yamashita, Hannah Dyer, Alvin Lee, Jordan Henrikssen, Erin Kelly, Janelle Keys, Amanda Hofer, Adam Sebzda, and Jessa Drury -thank you. I would also like to thank Paul Moote of the Inglis Lab for all of his encouraging words.

I would like to thank Virgil R. Bremer (Elanco Animal Health), Adrian Tsang (Concordia University), and Tim A. McAllister (Lethbridge Research and Development Centre) as leads on the Elanco project. Thanks to other collaborators from Lethbridge Research and Development Centre on this project: Dr. Robert Gruninger for his expertise advice to solve purification issue, Dallas Thomas for the phylogenetic tree, Darrell Vedres for gas chromatography. Thanks also to Dr. Daniel Hernandez Armada (Alberta Glycomics Centre) for the mass spectrometry analysis and Thi Thanh My Pham (Concordia University) for the NMR analysis.

Thanks to all at the School of Graduate Studies, University of Lethbridge for their assistance all through my program. Also, thanks to all at the Department of Biological Sciences, University of Lethbridge who played a role in my progress and supported me during this program. Special thanks to Sheila Matson for lots of help. You are all greatly appreciated.

Finally, I would like to express my utmost gratitude to my parents- Alauddin Ahmed and Sakila khatun, my wife – Razia Sultana, my younger brother - Jashim Uddin, also all of my friends and relatives for their constant inspiration and blessing.

Conflict of Interest

This project and my stipend were funded by Elanco Animal Health (Division of Eli Lily), and they have an invested interest in this project. An innovation disclosure document has been filed and a provisional patent, with co-inventers from Agriculture and Agri-Food Canada, and Concordia University, is being prepared to protect the intellectual property surrounding the activity and applications of GHXs. Currently this body of the work is un-published. Following finalization of this provisional patent it will be submitted for publication.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ABF- Arabinofuranosidase
ABN- Arabinanases
AGP- Arabinogalactan proteins
ANTS- 8-Aminonaphthalene-1,3,6-trisulfonic acid
Bl- *Bifidobacterium longum*
CAZy- Carbohydrate active enzyme database
CAZymes- Carbohydrate active enzymes
CBM- Carbohydrate-binding modules
cDNA- Complementary DNA
CoMPASS- CAZyme Multifunctional Predictive Analyses using Sequence Similarity
Da- Dalton
DP- Degree of polymerization
FACE- Fluorophore assisted carbohydrate electrophoresis
GH- Glycoside hydrolase(s)
Glu- Glutamic acid
GT- Glycosyl transferases
HG- Homogalacturonan
HPAEC-PAD- High performance anion exchange chromatography with pulsed amperometric detection
IMAC- Immobilized metal affinity chromatography
IPTG- Isopropyl β-D–thiogalactopyranoside
Kan- Kanamycin
LB- Luria Broth
MW- Molecular weight
Nf- *Neocallimastix frontalis*
NfGHX- Enzymes of the glycoside hydrolases GHX of *Neocallimastix frontalis*
NR- Non-reducing end
OD- Optical density
PCR- Polymerase chain reaction
PDB- Protein data bank
Pr- *Piromyces rhizinflatus*
RAX- Rye arabinoxylan
RG-I- Rhamnogalacturonan I
RG-II- Rhamnogalacturonan II
SBA- Sugar beet arabinan
SDS-PAGE- sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC- size exclusion chromatography
TLC- thin layer chromatography
Tx-*Thermobacillus xylanilyticus*

Chapter 1

Literature review

1.1 Rumen Microbial Ecosystem:

Over 10,000 years ago groups of humans began the transition from nomadic hunter gatherers to settled agricultural societies [1]; this agricultural revolution allowed for the development of society as it is enjoyed today. In part, this transition was aided by the domestication of livestock, which provided a stable source of meat and dairy products for these emerging societies. Historically, many of these domesticated animal species were ruminants, and today this trend continues. Ruminants are herbivorous animals that are able to digest large amounts of plant material due to special adaptations of their digestive tract (i.e. rumen) [2] and the unique microbial community that colonizes this organ (i.e. rumen microbiota; Figure 1.1)[3-5]. The name 'ruminant' refers to the rumen, which is an anaerobic compartment where ingested plant fibres are saccharified and fermented into small chain fatty acids [6, 7]. This process is of enormous significance to mankind as it enables the conversion of chemical energy stored within plant polysaccharides into high protein food products, such as milk and meat [3, 8, 9].

Microbial adhesion to and colonization of masticated plant tissues are major contributing factors for the efficient degradation of plant cell wall polysaccharides [10]. The chemical composition and physical structure of the plant material [11] are also important factors that affect digestion. In low quality forages, less than 50% of the cell wall fraction is digested and utilized by the ruminant host [12, 13]. Moreover, there are

some rare substrates like recalcitrant glycoproteins that that are difficult to digest due to their structural complexity [14].

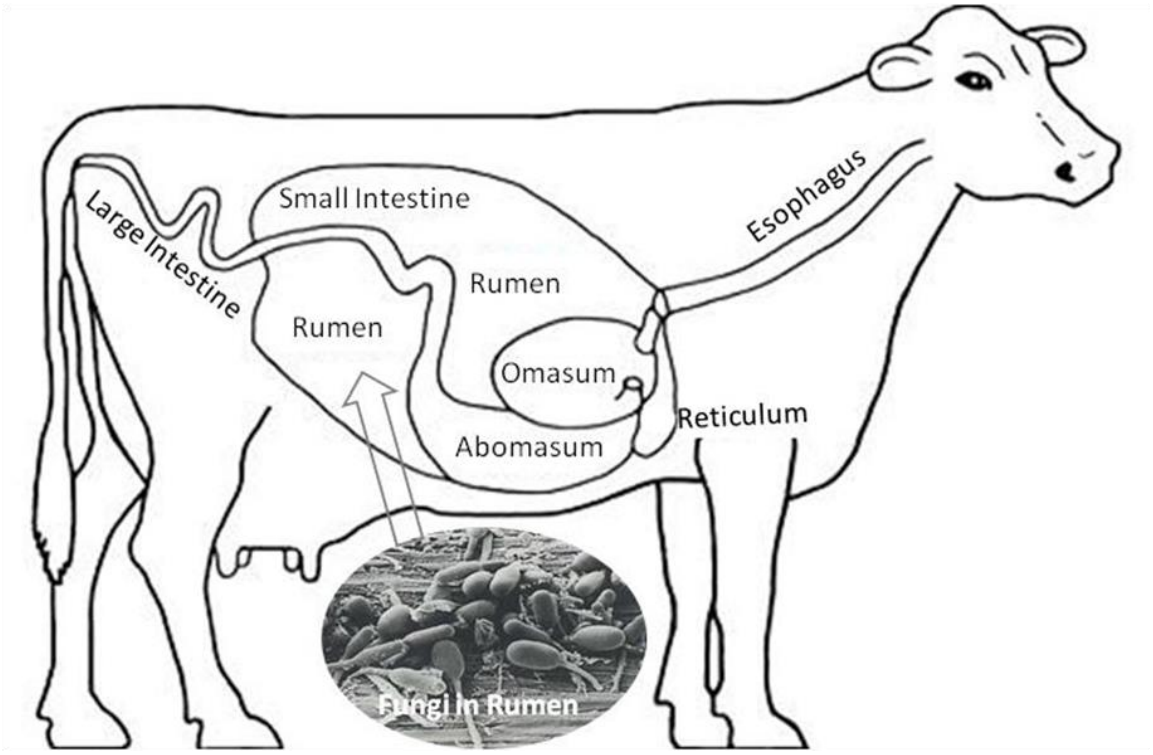


Figure 1.1: Ruminant digestive system. The ruminant digestive tract consists of the mouth, esophagus, a complex four-compartment stomach, small intestine and large intestine. The rumen is the largest of four compartments and where fermentation occurs. The reticulum, or "honeycomb," is the organ responsible for rumination (or cud chewing) and trapping hard, indigestible substances like rocks and nails that may be ingested by accident while the bovine is grazing. The omasum is the third compartment of the stomach. It has leaf-like folds and acts as a gateway to the abomasum. The omasum allows well broken down digesta to pass through into abomasum while redirecting larger material back to rumen and reticulum. The abomasum, also known as the "true stomach," is very similar to human stomach as it is responsible for producing acids and enzymes to break down proteins, and it sends the chyme to the small intestine. The rumen harbours a large number of microbes, including anaerobic bacteria, fungi and protozoa that play an important role in plant fibre digestion. This figure is adapted from [15].

1.2 Rumen Microbiota:

The rumen contains large numbers of bacteria (10^{10}-10^{11}ml^{-1}), anaerobic fungi (10^{3}-10^{6} ml^{-1}), ciliate protozoa (10^{4}-10^{6} ml^{-1}), and bacteriophages (10^{7}-10^{9}ml^{-1}) [9, 16, 17]. Rumen bacteria and anaerobic fungi are central to plant fibre digestion.

1.2.1 Rumen Bacteria

Historically the identity of rumen bacteria and composition of rumen communities has depended upon culturing techniques for isolation and characterization. In recent years, our knowledge of rumen bacterial diversity has vastly increased due to improvements in sequencing technologies [3, 17-21]. Although 16S rRNA gene sequences provide a census of bacteria present in the community, they provide limited information about their biology. As a result, genomics, metagenomics, and transcriptomics have emerged as potent techniques for investigating the metabolic potential of microbes [3, 17]. The majority of rumen bacteria are from the following phyla: *Firmicutes* (41-43%), *Bacteriodetes* (45-50%), *Proteobacteria* (5-6%), and *Actinobacteria* (1-2%). The ratios between the different phyla vary between individuals and with age of the host animal [3, 4, 22]. These bacteria play a major role in the degradation of cellulose, xylan, starch and inulin by a range of Carbohydrate Active enZymes (CAZymes).

1.2.2 Rumen Anaerobic Fungi

Anaerobic fungi act in a synergistic role in the digestion of lignocellulosic and cellulosic substrates within the rumen by invasive rhizoidal growth that physically disrupts recalcitrant tissues [23-26]. Zoospores are the mobile phase of the fungal life cycle. They preferentially colonize lignin-rich regions of the plant cell wall and upon germination,

solubilize these regions to a greater extent than rumen bacteria [27]. The activities of rumen fungi allow other ruminal microbes greater access to the anatomically restricted and indigestible portions of the plant [11, 28].

The degradation of recalcitrant lignocellulosic material by anaerobic fungi is facilitated by the production of a range of potent CAZymes [24]. Comparative analysis of fungal genomes [29] and metatranscriptomes [30] has revealed that fungi exhibit tremendous diversity in the number and variety of CAZymes produced.This genetic reservoir holds vast potential for biotechnological applications. Studies *in vitro* show that rumen fluid supplemented with select combinations of lignocellulolytic enzymes significantly boosts the release of cellobiose, glucose, xylose from straw and forage [31, 32]. In the following sections, I will describe the structure of these polysaccharides (section 1.3), provide an introduction into their chemical composition (section 1.4), and introduce the enzymes that catalytically modify their structure (section 1.5).

1.3 Plant fibre:

Plant fibre is comprised almost entirely of cellulose, hemicellulose, and pectin [33]. These polysaccharides are often tightly packed, contain many diverse sugar residues, and are branched with a variety of complex structures (Figure 1.2). The plant cell wall is divided into two main components referred to as the ‘primary’ and ‘secondary’ wall [34]. A primary wall is defined as a flexible extracellular matrix that is deposited while the cell is still enlarging, whereas a secondary wall is deposited on the primary wall when expansion ceases, and can be seen in sections as a structurally distinct layer (or layers), generally much thicker than the primary wall [35, 36].

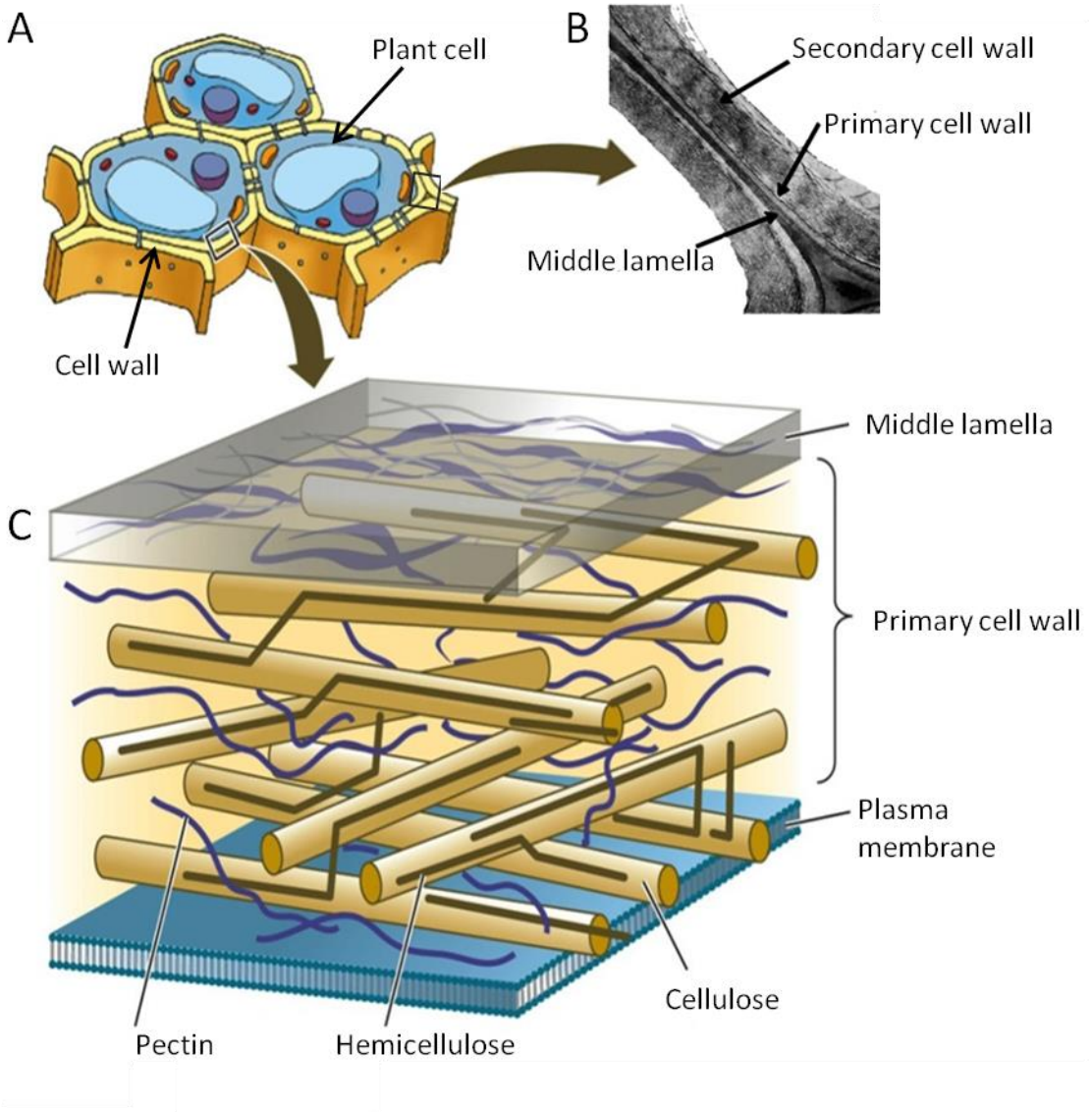


Figure 1.2: Model of plant cell wall structure. (A) Cross section of several plant cells showing the protective cell wall surrounding the cell outside of the plasma membrane. (B) A magnified cross section showing primary and secondary cell walls along with middle lamella, which is a thin sticky, amorphous layer between two adjacent cells. The secondary wall is formed later in plant development and internal to the primary wall. (C) Schematic representation of the major structural polysaccharide components of the primary cell wall, including cellulose, hemicellulose and pectin. The cellulose microfibrils are linked via hemicellulosic tethers to form the cellulose-hemicellulose network, which is embedded in the pectin matrix. This picture has been adapted from a publically accessible website [37].

Both the primary and secondary walls are predominantly composed of cellulose, hemicellulose and pectin [35, 38]. Vascular tissue secondary walls also contain an extensive network of aromatic compounds called lignin that provides further structural support [39].

Cellulose is composed of linear polymers of β-1,4-linked D-glucose and is a component of both the primary and secondary plant cell wall [40]. It is a highly crystalline polysaccharide that provides load-bearing properties of the cell wall [41]. Hemicellulose is a group of branched polysaccharides that are classified according to the main sugar found in the backbone of the polymer (e.g. xylan = β-1,4-linked D-xylose; mannan = β-1,4-linked D-mannose). The backbone of hemicellulose can display many decorations composed of monomers such as D-galactose, D-xylose, L-arabinose, and D-glucuronic acid. Variations in structure are also present in hemicelluloses. Glucomannans for example, have a backbone of randomly dispersed β-1,4-linked glucose and β-1-4-linked mannose [38]. Alternatively, arabinoxylans such as rye arabinoxylan (RAX), consist of a polymer chain of β-1,4-linked D-xylose units, many of which are 2- or 3-substituted, or 2,3-disubstituted by α-L-arabinose residues [42].

Pectin is a plant cell wall structural polysaccharide within the primary cell wall and the middle lamella, which punctuates the junctions between primary walls of neighboring cells, and participates in intercellular connections [43]. Pectin is a highly complex polysaccharide that is enriched in D-galacturonic acid. Within the plant cell wall, pectin is divided into three classes of distinct pectic polysaccharides: homogalacturonan (HG), rhamnogalacturonan-I (RG-I) and rhamnogalacturonan-II (RG-II) that vary in size,

branching, and function [44]. For example the sidechains of RG-I can be heavily decorated with galactan and arabinan [45], such as sugar beet arabinan (SBA), which is composed of a α-1,5-L-arabinofuranosyl backbone decorated with α-1,2- and α-1,3-L-arabinofuranosyl side chains.

Lignin is a network of aromatic compounds primarily composed of phenyl propane units. The main building blocks of lignin are the hydroxycinnamyl alcohols (or monolignols) coniferyl alcohol and sinapyl alcohol, with minor amounts of p-coumaryl alcohol [46]. The lignin network is built with chemically diverse and poorly reactive linkages and a variety of monomer units which precludes the ability of any single enzyme to recognize and degrade it [47]. It is deposited predominantly in the secondary thickened cell walls, making them rigid and impervious to water.

In addition to structural polysaccharides, complex carbohydrates present on the surface of plant glycoproteins are also a component of the plant cell wall. Arabinogalactan-proteins (AGPs) are heavily glycosylated glycoproteins (90-98% w/w) enriched in the amino acids: hydroxyproline/proline, alanine, serine/threonine found on the plasma membrane and in the cell walls of a diverse array of plant species (Figure 1.3; [48, 49]). AGPs are thought to have important roles in various aspects of plant growth and development [50, 51].

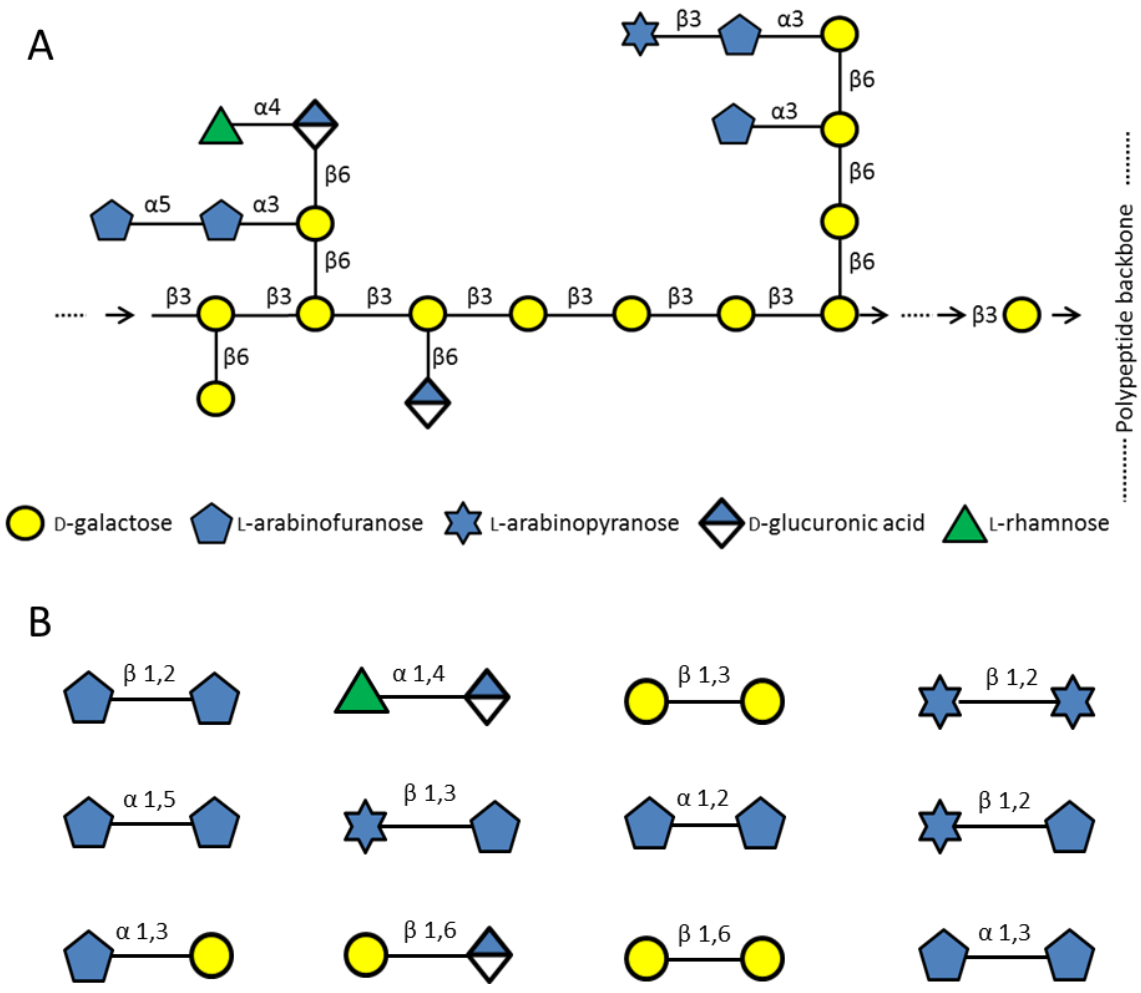


Figure 1.3: Schematic representation of AGP glycan structure. (A) Model of a common AGP glycan. Backbone of β-1,3- galactan, which is often substituted at O6 with side chains of β-1,6-galactan decorated further with arabinose, and less frequently with rhamnose, and (methyl)-glucuronic acid. (B) Representative carbohydrate linkages found within the glycosylation of AGP. The arabinogalactan model is adapted from [50, 52].

1.4 Carbohydrates:

Carbohydrates are biological molecules, consisting of carbon (C), hydrogen (H) and oxygen (O) atoms, which perform numerous functions that are important to living organisms. Carbohydrates are classified by their chemistry (e.g. chain length, composition, and stereochemical orientation), degree of polymerization (DP), and the position (e.g. 1→4) and stereochemistry of linkage (i.e. α or β). Carbohydrates are well recognized as the most complex biological molecules in nature because of this vast structural diversity.

The major classes of carbohydrates include simple sugars (i.e. monosaccharides = DP1; or disaccharides = DP2), oligosaccharides (i.e. short chain carbohydrates; DP 3-9) and polysaccharides (i.e. long chain carbohydrates; DP >9)[53]. Monosaccharides are single sugar aldehyde or ketone derivatives of straight chain alcohols, which contain at least three carbon atoms; examples include D-glucose, D-xylose, D-galactose, L-arabinose. Oligosaccharides and polysaccharides are combinations of monosaccharides that are covalently linked by glycosidic bonds. Polysaccharides can either be branched such as in RAX and SBA, or linear such as in cellulose. The absolute configuration of monosaccharides (i.e. 'D' or 'L') is made according to the orientation of the asymmetric carbon furthest from the carbonyl group in each enantiomer [54]. In a standard Fischer projection, if the hydroxyl group is on the right then the molecule is a D-sugar (Figure 1.4); otherwise, it is an L-sugar.

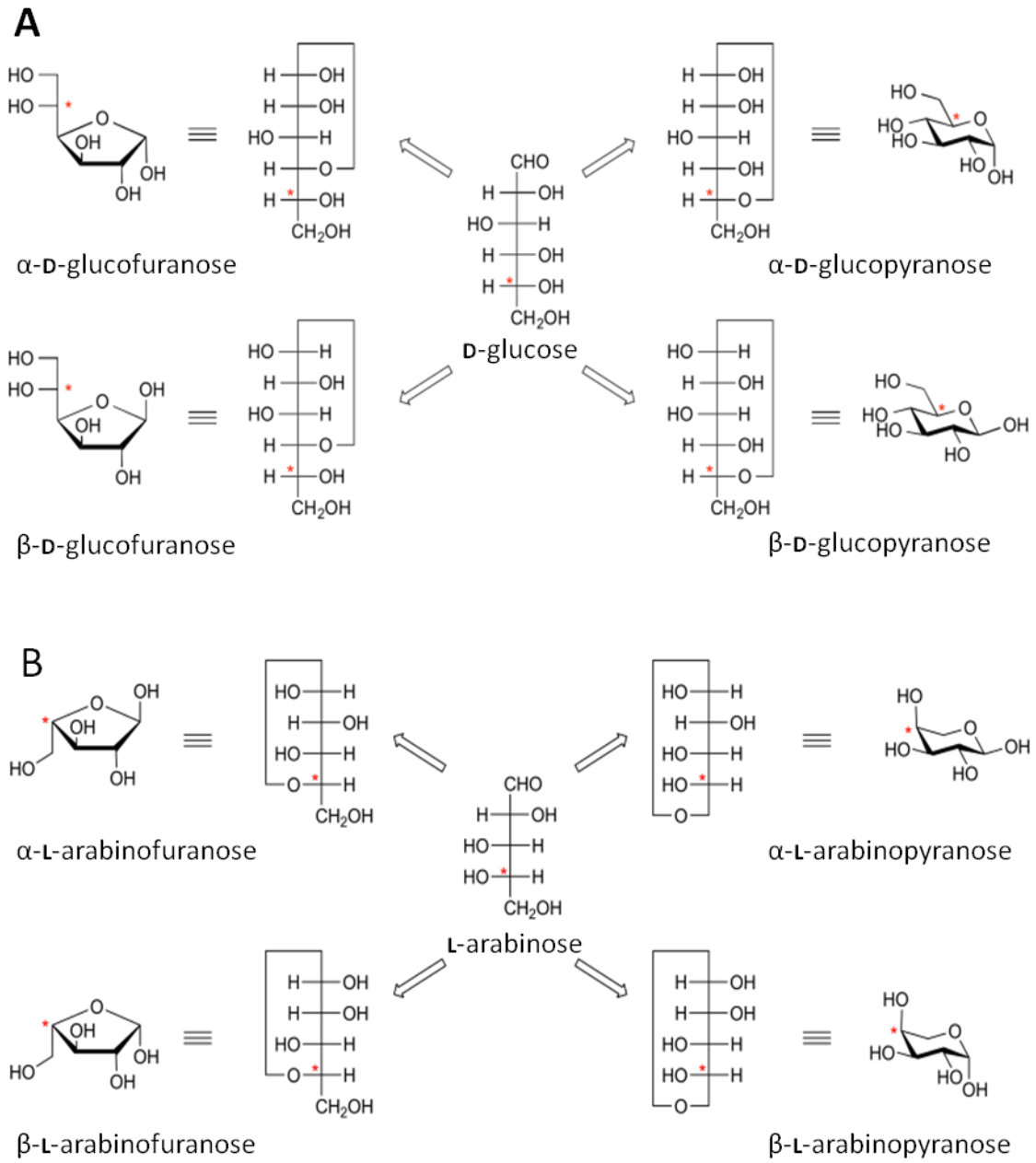


Figure 1.4: Ring configuration dynamics of monosaccharides. (A) Glucose is a hexose sugar that can adopt furanose (left) or pyranose (right) configurations. Glucofuranose configurations are possible through C-1—O—C-4 ring closure and glucopyranose via a C-1—O—C-5 ring closure. Alternative anomeric sterochemistries are also possible. (B) Arabinose is a pentose sugar, which is dynamic in solution and can adopt several alternate configurations. Arabinose converts into arabinofuranose (left) through C-1—O—C-4 ring closure and arabinopyranose (right) via C-1—O—C-5 ring closure. Alternative anomeric sterochemistries are also possible. These figures have been adapted from CAZypedia [55].

When cyclized into rings, monosaccharides acquire an additional asymmetric center derived from the carbonyl carbon atom, which is termed as anomeric carbon, C-1 (Figure 1.4). The anomeric hydroxyl group can undertake two possible orientations, which can result in two different stereoisomers during cyclization. When the orientation of the anomeric carbon and the stereogenic center furthest from the anomeric carbon are on the same plane of the sugar, the monosaccharide is defined as the β-anomer. When the orientations are on opposite sides, the monosaccharide is defined as the α-anomer [56, 57].

In solution, monosaccharides exist as an equilibrium mixture of acyclic and cyclic forms. Five and six-membered rings are most frequently formed from acyclic monosaccharides [54, 56]. Generally, ketohexoses form furan-like (i.e. five-membered) rings via a C-2—O—C-5 ring closure commonly referred to as the 'furanose' conformation. Aldohexoses can also form furanoses through C-1—O—C-4 ring closure, but more commonly are observed to cyclize into pyran-like (i.e. six-membered) rings via a C-1—O—C-5 ring closure referred to as the 'pyranose' conformation (Figure 1.4).

1.5 CAZymes:

The CAZy database (www.cazy.org) classifies CAZymes into protein families based on their sequence relatedness [58]. Importantly, however, this classification does not always equate into functional relatedness as many families (e.g. GH39 and GH51) have now been defined that contain different enzyme specificities [59, 60]. As of June 2016, CAZy categorized 135 families of glycoside hydrolases (GHs), 99 families of glycosyl transferases (GTs), 16 families of carbohydrate esterases (CEs), 24 families of

polysaccharides lyases (PLs), and 13 families of auxiliary activities. Among them, plant cell wall degrading enzymes have received special attention as they can target recalcitrant or rate-limiting substrates within the plant cell wall which make excellent targets for use in biomass conversion [61].

CAZymes represent several classes of enzymes using distinct mechanisms, including: Glycoside Hydrolases (GH), which cleave glycosidic linkages by the addition of water [62]; Glycosyl transferases (GTs), which catalyse the formation of glycosidic bonds [63]; Polysaccharide Lyases (PL), which cleave uronic acid-containing polysaccharides via a β-elimination mechanism to generate an unsaturated hexenuronic acid residue and a new reducing end [64]; Carbohydrate Esterases (CE), which catalyse the de-O- or de-N-acylation of substituted saccharides; and Auxiliary Activities (AA), which are families of enzymes that perform redox reactions on crystalline polysaccharides [65]. Carbohydrate-binding modules (CBMs) bind to a carbohydrate ligand and direct or concentrate the catalytic machinery onto its substrate, thus enhancing the catalytic efficiency of the carbohydrate-active enzyme [66]. Most often in nature, the binding specificity CBMs reflects the substrate specificity of the parent enzyme.

1.5.1 Glycoside hydrolases:

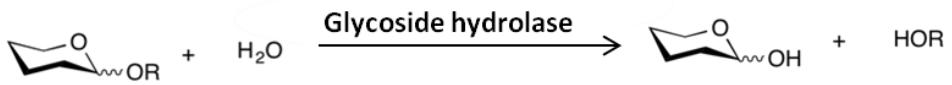
The hydrolysis of a glycosidic linkage between two or more carbohydrates, or a carbohydrate and non-carbohydrate adduct, leading to the generation of a new reducing end (a sugar hemiacetal or hemiketal) and leaving group is catalyzed by enzymes known as GHs [62, 67] (Figure 1.5A). Hydrolysis of the glycosidic bond is typically catalyzed by two amino acid residues of the enzyme, namely a general acid, which is the proton

donor, and a nucleophile/base. The reaction can occur with either inversion (Figure 1.5B) or retention (Figure 1.5C) of the anomeric configuration at C1 depending on the spatial position of the catalytic residues in the active site [67, 68].

Inverting GHs carry out hydrolysis of the glycosidic bond via a one step, single-displacement mechanism resulting in inversion of the anomeric configuration (i.e. α→β or β→α; Figure 1.5B). These enzymes contain a general acid and general base, which are most commonly glutamic or aspartic acid residues [67, 69] that are typically located 6-11 Å apart [62]. Alternatively, retaining GHs carry out hydrolysis through a double-displacement mechanism (Figure 1.5C) where one residue plays the role of a nucleophile whilst the other residue acts as general acid / base [67, 68]. The carboxylic acid residues (typically glutamic or aspartic acid) for retaining enzymes are commonly located 5.5 Å apart [62]. Products of retaining reactions maintain the same anomeric configuration of the substrate (i.e. α→α or β→β). Both the inverting and retaining mechanisms involve oxocarbenium ion-like transition states [67, 69].

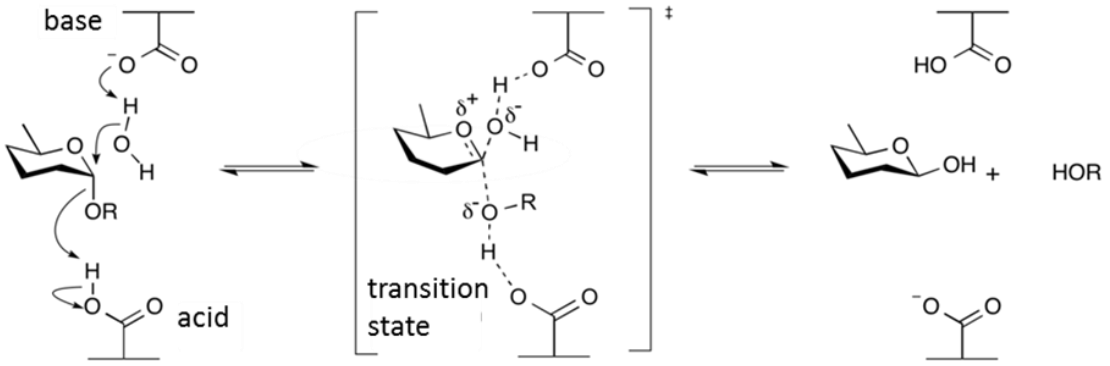
Names of GHs of a particular enzyme family can be abbreviated based on the details of the family they originate from and/or the microbial species. For example, glycoside hydrolases of family 43 can be described as GH43. GH43 enzymes originating from *Cellvibrio japonicus* are described in the short form *Cj*GH43 [70]. Glycoside hydrolases can also be named after the substrate that they act upon. For example, xylanases catalyze the cleavage of the xylosyl residue within the homopolymer xylan.

A



B

Inverting mechanism for an α-glycosidase



C

Retaining mechanism for an α-glycosidase

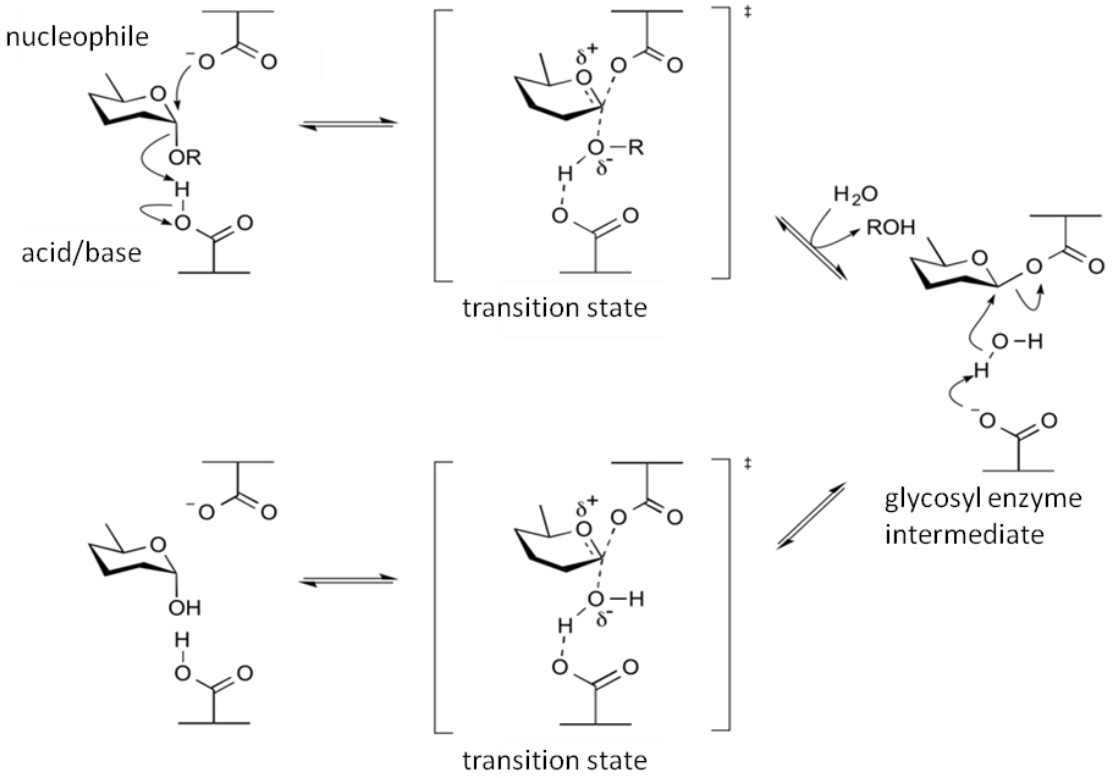


Figure 1.5: Schematic depiction of glycoside hydrolase mechanisms. (A) General model of hydrolysis by glycoside hydrolases. (B) Schematic depiction of an α-glycosidase single displacement inverting mechanism. In inverting GHs, the product has an inverted anomeric configuration compared to the substrate (e.g. α→β). The protonation of the anomeric carbon is accomplished by the catalytic acid while the catalytic base facilitates its attack on the anomeric carbon by removing a proton from a water molecule. (C) Schematic depiction of double displacement retaining mechanism for an α-glycosidase resulting a product with the same stereochemistry as the substrate (e.g. α→α). In retaining GHs, hydrolysis occurs through a two-step, double-displacement mechanism involving a covalent glycosyl-enzyme intermediate. In the first step, one residue functions as an acid catalyst, protonates the glycosidic oxygen, and facilitates departure of the leaving group while the other residue acts as a nucleophile, attacking the anomeric centre to release the aglycon and form a glycosyl enzyme intermediate. In the second step of deglycosylation, the deprotonated acid/base acts as a base catalyst to activate a water molecule that carries out a nucleophilic attack on the glycosyl–enzyme intermediate. These figures have been adapted from CAZypedia [71].

1.6 Approaches to discover new CAZyme activities:

The unexplored microbial diversity within the rumen microenvironment represents a source of potentially novel enzymatic activities and metabolic pathways. To date most enzyme bioprospecting studies have focused on rumen bacteria and methanogenic archaea, with the ecology and function of eukaryotic protozoa and fungi being far less studied. A current view is that ruminal anaerobic fungi represent more promising systems for enzyme discovery as they are more adept at digesting insoluble and recalcitrant substrates than ruminal bacteria [27, 72]. Recently, the analysis of the genome sequence of an anaerobic fungus (*Orpinomyces* sp. strain C1A) provided valuable insight into the diversity of anaerobic fungal CAZymes [73]. Many of its CAZymes appear to have been acquired by horizontal gene transfer from rumen bacteria. The cellulolytic machinery of anaerobic fungi consists of not only free enzymes but also high molecular weight extracellular multi-enzyme complexes called cellulosomes [7, 74, 75]. The above insights have made the anaerobic fungi potential candidates as a source of highly efficient, novel CAZymes for industry applications.

In contrast to genomic approaches, transcriptomics and metatranscriptomics reveal the most highly expressed genes in pure culture and complex communities, respectively. A study comparing the metatranscriptome and metagenome of the human gut revealed that 41% of the transcripts were not directly linked to the metagenome [76]. Metatranscriptome profiling provides a snapshot of the composition and relative abundance of actively transcribed genes. Furthermore, a second metatranscriptome study [77] found that the rumen CAZyme profile of dairy cows show distinct differences

compared to the rumen CAZyme profile derived from metagenomic analysis [29]. Such findings suggest that metatranscriptomics may enhance the discovery of functional enzymes [76]. Metatranscriptomics have also revealed that rumen fungi produce a large portion of the CAZymes required for fibre digestion. Taken together, results from these studies suggest that the metatranscriptomic approach provides a more direct path than metagenomics for identifying fibrolytic enzymes induced by discreet carbon sources.

In order to discover new GH families, fungal genes encoding 'hypothetical proteins' appended to CBMs can be identified. CBMs play significant biological roles in targeting appended catalytic modules to their dedicated substrates within complex macromolecular structures such as the plant cell wall [78] and promote the association of the full length enzyme with its substrate. CBMs potentiate the activity of their cognate catalytic module against insoluble substrates [79], as well as complete cell walls [80, 81]. The polyspecific CBM family like CBM13 usually adopts a β-trefoil structure with three potential binding sites (alpha, beta and gamma) for a variety of small sugars, xylooligosaccharides, and xylan polymers [82]. The objective of this study is to use CBM13s to direct the discovery new GH families (i.e. GHX) from the differential transcriptomes of ruminal anaerobic fungi when grown on glucose, xylan, and barley straw.

1.7 General Hypothesis:

I hypothesize that polyspecific CBM families (e.g. CBM13) that are appended to sequences of unknown function can be used to direct the discovery of novel enzyme activities from large sequence datasets, such as the transcriptomes of anaerobic rumen fungi cultured on barley straw.

1.8 Main Objectives:

1. To use CBMs to identify CAZymes, that are catalytically active on plant cell wall polysaccharides, from fungal transcriptomes.

2. To identify new glycoside hydrolase families (GHX).

3. To characterize the substrate specificity of GHX.

4. To characterize the products released by GHX.

Chapter 2

Material and Methods:

2.1 Selection of target sequences using CBM:

Four species of anaerobic fungi (*Neocallimastix frontalis, Piromyces rhizinflatus, Anaeromyces mucronatus, Orpinomyces joyonii*) were cultured in vitro in the presence of barley straw and glucose. The RNA from these cultures was extracted, converted into cDNA and the transcriptomes of each was generated (this work was completed previous to my MSc project at the Lethbridge Research and Development Centre). From these fungal transcriptomes, hypothetical proteins that were appended to polyspecific CBM families including CBM1, CBM6, CBM10, CBM13, and CBM35 were identified (Table 2.4). Target GHX sequences were selected, from these hypothetical proteins appended to CBMs, by analyzing with different bio-informatics programs, including dbCAN, SignalP 4.1, InterPro, Phyre 2, PSI-BLAST, ClustalW2, Pymol.

2.2 Phylogenetic Trees:

Phylogenetic trees were created with GHX sequences embedded with the characterized GH51 and GH39 sequences from CAZy database using CoMPASS (CAZyme Multifunctional Predictive Analyses using Sequence Similarity), a fully-automated pipeline established at Lethbridge Research and Development Centre to predict CAZyme function using their phylogenetic relationships. Firstly, all the CAZy sequences were extracted from selected families (GH51-characterized and GH39-characterized). dbCAN was utilized for carbohydrate enzyme identification and trimming into functional modules boundaries (i.e. GH catalytic fragments) [83], and then merged with catalytic

fragments of GHX sequences, which were determined by InterPro [84]. Alignment and phylogenetic grouping was performed via multiple sequence comparison by log-expectation [85] and Randomized Axelerated Maximum Likelihood (RAxML) [86] respectively. ProtTest3 was used for the selection of the best-fit model used by RAxML [87]. Finally, FigTree was used for phylogenetic tree creation.

2.3 GHX gene synthesis and cloning:

Selected GHX gene sequences (Nf2152, Nf2215, Nf2523 and Pr2455; Table 2.1) were commercially synthesized by BioBasic Canada Inc with codon optimization for expression in *Escherichia coli* BL21 (DE3). Synthesized gene templates were subsequently dissected into functional catalytic and CBM fragments using polymerase chain reaction. Primers were designed to target modular boundaries and introduce restriction sites that do not cleave internally within the gene to ensure directional cloning (see Table 2.2 for list of primers). The PCR amplified coding sequences were then purified, enzymatically restricted with NdeI and XhoI (based on the engineered restriction sites), and ligated into similarly digested pET28a vector [88]. The *E.coli* expression vector pET28a contains an N-terminal Hexa-histidine tag for purification by immobilized metal ion affinity chromatography (IMAC) [89]. The ligated DNA was transformed into *E.coli* DH5α and grown overnight on LB plates containing kanamycin (Kan) (50 µg ml^{-1}). Transformants were selected and grown up in 10 mL LB-Kan medium overnight. The following day the plasmid DNA was purified (E.Z.N.A Kit-Omega Biotek, Cat # D6942-02) from the overnight cultures, and screened for successful ligation by diagnostic PCR and restriction

enzyme digestions. Sequence confirmed clones were used for recombinant protein production [90].

2.4 Over expression and purification of the recombinant proteins:

The purified plasmids were transformed into *E.coli* BL21 (DE3) (Novagen, EMD Millipore, Cat #69450-3) for recombinant protein production. The growth medium was supplemented with kanamycin (50 µg ml^{-1}) to select for transformed cells. After inoculation, the cultures were grown at 37°C to an OD 600 nm of 0.8-1.0 at which point the temperature was lowered to 16°C for one hour prior to induction with IPTG at a final concentration of 0.2 mM. Cell cultures were continuously shaken at 200 rpm and with 16°C overnight. The following day the culture was pelleted via centrifugation at 6,500 x g for 10 min. The bacterial pellet was suspended in Binding Buffer (BB: 0.5 M NaCl, 20mM Tris, pH 8.0) and lysed by sonication for 2 min of 1 sec 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 x g for 45 min and the supernatant was syringe filtered using a MW cut-off of 0.25 µm. The filtrate was loaded onto a sepharose-Ni^{2+}NTA column for purification by immobilized metal affinity chromatography. Immobilized protein was washed with BB to reduce the amount of non-specific binding interactions. The target protein was eluted via a step-wise gradient of imidazole (5, 10, 30, 100, 200, 500 mM) in BB. All eluted fractions were collected and analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) for purity. Coomassie Brilliant Blue (0.2% Coomassie, 40% Methanol, 10% Acetic Acid) was used to stain the gel. Fractions containing significant portions of protein were pooled

and dialyzed against 2 x 4.0 L 20 mM Tris-HCl (pH 8.0), 500 mM NaCl. Following dialysis, samples were concentrated using a nitrogen pressurized stirred ultrafiltration cell (Amicon) with a molecular weight cut-off of 5 kDa. Once concentrated, the protein was filtered and further purified by size exclusion chromatography (HiPrep 16/60 Sephacryl S-200 HR; GE Healthcare). The column was run at a flow rate of 1.0 ml min^{-1} in 500 mM NaCl, 20 mM Tris-HCl (pH 8.0). The collected fractions were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis and pooled based on purity. Pooled fractions were dialyzed again against 20 mM Tris-HCl (pH 8.0), 100 mM NaCl. After dialysis, pooled fractions were concentrated to a desired concentration depending on the analysis to follow. Concentration was determined using the calculated extinction coefficients as determined with ProtParam [91].

2.5 Enzyme characterization:

The activity of the purified GHXs was screened on a variety of plant cell wall carbohydrates including: sugar beet arabinan (Cat# P-ARAB), wheat arabinoxylan (P-WAXYM), rye arabinoxylan (P-RAXY), β-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), *P*NP-α-L-arabinofuranoside (O-PNPAF) purchased from Megazyme International Ireland, *P*NP-α-L-arabinopyranoside (38018) from Glycosynth, and beech wood xylan (X4252) from Sigma. The enzyme concentration, buffer and pH were optimized for digestion through empirical studies with different substrates, enzyme concentrations and buffers (pHs ranging from 4.0-8.0). The final

reaction mixture contained 5 mg mL^{-1} of substrate, 0.5 µM of 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 minutes to denature the enzyme and terminate the reaction followed by short centrifugation at 8,000 x g to pellet denatured protein from the product. Products were analyzed by thin layer chromatography (TLC) and high performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) to characterize their chemistry and degree of polymerization.

2.6 Site-directed mutagenesis:

Using Phyre2 molecular modeling [92], Glu112 and Glu211 of Nf2152 were predicted to be the acid/base and nucleophile, respectively based upon superimposition with the GH51 from *Thermobacillus xylanilyticus* (2VRQ) [93]. These residues were targeted for substitution to glutamine using site-directed mutagenesis [94]. Mutagenic primer sets (Table 2.3) were extended with KOD polymerase (Novagen, cat no. 71086) using the pET28a-GHX catalytic fragment plasmid as a template. The entire reaction mixture was digested with DpnI (New England BioLabs, cat no. R0176) and then transformed into *E. coli* DH5α competent cells and plated on LB Kan agar plates. Plasmids were extracted by plasmid extraction kit (Omega, cat no. D6945) from overnight colonies and positive mutant clones were verified by sequencing. Once sequence confirmed, the mutant enzymes were tested for activity with appropriate substrates.

2.7 Thin Layer Chromatography:

Digested samples (total 9 µL; spotted thrice with 3 µL each time) were spotted onto TLC plates (TLC Silica gel 60; EMD Millipore Corporation). 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 minutes [70].

2.8 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 was injected onto an analytical (3 X 150 mm) CarboPac PA20 column (Thermo Scientific) and eluted at 0.4 mL min^{-1} 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). Data were collected using the Chromeleon^{R} chromatography management system (Dionex) via a Chromeleon^{R} server (Dionex). HPAEC-PAD was employed to profile the reaction products based on known carbohydrate standards [95].

2.9 Ethanol Precipitation:

Ethanol precipitation was 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 first dried with a speed-vac at ambient temperature and then suspended in 95% ethanol. The reaction mixture was incubated on ice for 10 minutes. After the incubation period, it was mixed vigorously with a vortex to dissolve the small oligosaccharides and monosaccharides and then clarified by centrifugation at 14,000 x g for 10 min. The supernatant was removed and dried by speed-vac. Purified carbohydrates were subsequently used directly or suspended in water at the desired concentrations for further analyses.

2.10 Acid hydrolysis:

Thirty micrograms of purified dried GHX 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 speed-vac followed by three wash steps with 100 µL isopropyl alcohol. Released monosaccharides were analyzed by TLC and HPAEC-PAD.

2.11 Mass spectrometry:

Prior to performing the mass spectrometry analysis, a large scale (total reaction volume was 200 mL) digest was performed to generate mg amounts of GHX products. Following ethanol precipitation, the products were further purified by P2 Bio-gel (Bio-Rad) size exclusion chromatography at a flow rate of 0.17 mL min^{-1} where distilled water was used as eluent. Extra fine (< 45 µM) particle size beads were used for the Bio-gel P2 gel, which has an exclusion limit of 100-1,800 Da. The elution peaks were screened by TLC, pooled, and lyophilized. Mass spectrometry was carried out by the Alberta Glycomics Centre on the pooled fractions to determine the m/z and molecular weight of the product. Electrospray-ionization mass spectra were recorded on an Agilent Technologies 6220

TOF instrument. The sample was dissolved in methanol or a methanol-water mixture (methanol:water 1:1) and directly injected into the instrument (5 uL). The spectra were recorded in positive mode.

2.12 Gas Chromatography:

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 [96, 97]. Sugars (purified product) were first converted into alditol acetates, which involved reduction of sugars with sodium borohydride ($NaBH_4$) following conversion of polyols to polyacetate esters [98, 99]. Fifty micrograms of products were reduced by 200 µL $NaBH_4$ (10 mg/mL $NaBH_4$ in 1M NH_4OH). The reduced sugar was then acid hydrolysed 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 speed-vac followed by three wash steps with 100 µL isopropyl alcohol. Released monosaccharides were acetylated by the addition of 250 µL of acetic anhydride and dried on a speed-vac to a volume of 200 µL. The resulting solution was transferred to a GC auto sampler vial containing a 250-300 micro insert and injected into gas chromatograph (Hewlett Packard 5890) where polar capillary GC column (SP2330) and flame ionization detector was used.

2.13 Fluorophore-assisted carbohydrate electrophoresis:

Fluorophore-assisted carbohydrate electrophoresis (FACE) was performed to identify the reducing end of the digested products. Carbohydrates were fluorescently labelled at their reducing end using 8-aminonaphthalene-1,3,6-trisulphonic acid (ANTS), and the

resulting labelled sugars were separated in a high percentage (40%) polyacrylamide gel [100]. 30 µg of each sample were dried and suspended by vortexing in 5 µL fresh 0.15 M ANTS (dissolved in 15% acetic acid) solution and 5 µL fresh 1 M 2-picoline borane (dissolve in 1 mL of DMSO) solution. Samples were incubated overnight at 37 °C in a tube wrapped in foil. The labelled samples were then dried with a speed-vac for 2-4 h or until completely dry. The labelled sugar was then acid hydrolysed by the addition of 200 µL of 2 M TFA followed by an incubation at 100° C for 4 h. The reaction mixture was dried to completion in a speed-vac followed by three wash steps with 100 µL isopropyl alcohol. The dried pellet was suspended in 25 µl FACE loading dye and run on gel immediately or stored at -20 °C wrapped in foil.

Table 2.1: List of gene sequences:

Gene	Source Organism	Sequence
Nf2215	*N. frontalis**	AVNKLTVDCGNVLRNATHCANGSLYGIIENVPPENQYKSLVDDIHPYVMRNPARGSYGNQHPFGDAIAVAKRL SRTPGALVSVDLADILPYWPYQWPGMQKWLDEVRSFIRDKKASGLKNWYGYEIWNEAENTWRDSNGYNFLD FWKETYKVIREMDPDEKIIGPCDGLYSEQRMRPFLEYCKKNDVMPDIMCWHELMGIELIPGRYKAYRELERSLGI KELPITINEYCDIDHDLEGQPGSSARFIGKFERYKIESALISWWFVPQPGHLGSLLATDSKKGAGWWFYKWYGD MTGSMVNVKPPNDNSKLIDGAACVDEKEKYVSFIFGGPNDGTVSANFVNLPSFIGDVANVKFEKIDWKHKDAE SSGPNTVFEKKYAVSNGQISITLNGMNASSGYRLYITKGDNSGTIEPPSEVTTLIIPDGTYRIINRKSGKALAVANDS NADAANVVQATYSGKKSQQWEISEDGGYKLLNVNANKVLDMDRESKEDGGNALIWRNTGNINQRWEIEDA GDGYVYIINFNSGKLLDVENGSTTDGGNVIQWRQNNGANQQWQIIPIGSDTPVTPQKPTPAPAKTTTKTNSAA PTSSKNTCSEAILKLGYKCCSKDCGIIYTDDDGTWGYENDEWCGCNAKASCPAAIVSQGYKCCSESNCNVYETD ESGKWGIENDDWCGISDKC
Nf2523	*N. frontalis*	AVNKLTVDCGNVLRNATHCANGSLYGIIENVPPENQYKSLVDDIHPYVMRNPARGSYGNQHPFGDAIAVAKRL SRTPGALVSVDLADILPYWPYQWPGMQKWLDEVRSFIRDKKASGLKNWYGYEIWNEAENTWRDSNGYNFLD FWKETYKVIREMDPDEKIIGPCDGLYSEERMRPFLQYCLDNDVMPDIMCWHELMGVELIPRRYKAYRELEASLG IKELPITINEYCDIKHELEGQPGSSARFIGKFERYKIESALISWWFNNLPGHLGSLLATDSEKGAGWYFYKWYSDM TGEMVYVKPPNDNSILVDGAASIDSKKEYISFIFGGPNDGTINASFINIPDFIGDVAKVKFEKIDWVSKDTISNGPN TVFEKQYAVVDGQMTIGLKNCNASSGYRLYITKGDPSGEIEESSEIPTIIIPDGTYRIINRNSGKALSVVNDSKNNA ASVVQYTYSGKTSQQWKIVEDGGYRLTNVNSDKVLDMNGASLEDGGNAIIWPDIAGPNQRWSIENAGDGYV YIINFNSNKVLDVANVSKEEGGNVHQWTITNGANQQWQLIPIGSDTPLDLTPRKTTTTKKTTSKPAPTPSKESCS QAILNQGYKCCSNDCIIIFTDDDGT
Pr2455	*P. rhizinflatus**	ATHVLVVDCDNRLRNATHCANGSLYGLIETVPPDNKFQALVDDLHPFVFRNPARGYPGTQHPFGGAIPVAERLS KTPGALVSVLLSDILPYWPYQWPGMDKWLEEVRKFILDKKASGLTNWYGLEIWNEAENTWRDSNGYTFIEMW KKTYDLIRELDPDVKIIGPCDGLYSEQRLRPFFEYCKKNNCLPDIMCWHELMGIELVTPRMKAYRKFEKEMGIPEL PVTINEYCDIKHEFEGQPGSNARFIGKFERLKVDSALISWWFTPQPGHLGSLLATDTEKGAGWYFHKWYGDMT GDMLYVKPPNENSNLVDGAACIDSSQKYISFIFGGPNDGTINASFINLPDFIGDIAKVKFEKIDWISKDTVSSGPTT VFEKEYRVVDGQLAIIINNCNATSGYRIYITKGDGKSGSIEPPSKVTTLIIPDGTYKLINRHSGLALSLKNGSNAIQSA DTNDKSQQWTISDDGGYRLLNVEAGKVLDVEGDSKNDGGNVIVWRDQSSFNQRWSIENTDDGYVVFINLSS GKVLDVEYESVSDGANIHQWTNRSAPNQQWKLVPVGDTKPLTTKTKTKITTTTTTQKTTPTFVAENSSCYESILK QGYKCCSPGCIIYYTDDDGTWGVEKGEWCGCPSVTSDSEKCSASILAQGYSCCKSNNCNVFAEDDDGRWGIEN NAWCGISDSC
Nf2152	*N. frontalis*	AVSQLTVDCNAKIRRATHCASGAHYGLIENVPKDYKSLVAPLNINVMRAPARAGNGRQQPIGDVIKVAQRLKES PGARVTIELADILPGWPYRWPGIQTWFNEIRSFINDKKKSGLTNFYGNEIWNEPDVTWKDSNGLSFNQMWKQ TYDLLRQIDPNEKIIGPSFSWYEENKMKNFLQFSKQNNCLPDIIAWHELSGIDGVSSHFRSYRNLEKSLGISERPITI NEYCDENHDLEGQPGSSARFIGRFERYKVDSGMITWWFVPHPGRLGSLLASDTQKGAGWYFYKWYGDMTG DMVSVSPPNENSKLIDGAASVDASAQYVSFIFGGPNDGSVKANFKNLPSFLGSSAHVKVEKIDWKSKDTPSNGP NTIFEKNYSISNGQISVDLSGTNASSGYRIYITKADGSSNTGNNGNDQPAPSNSSSNGERYKIINRYTNRVLAVEN DSTANNANVLQWGDNGSSGQQWVVAKEGDQYRITSYDTNKALDVSGRSTANGGNVIIYDDHAQGNQRWK FIDAGDGYVIIENVNSGKVLDVDNASKEYGANIMQWNKNGSTNQQWKLVPLNPPAVTTTIKTTVTKSAQSTN NASSCSAKILSQGYKCCKEGCVVVYTDDDGTWGV
BlGH127	*B.longum**	MNVTITSPFWKRRRDQIVESVIPYQWGVMNDEIDTTVPDDPAGNQLADSKSHAVANLKVAAGELDDEFHGM VFQDSDVYKWLEEAAYALAYHPDPELKALCDRTVDLIARAQQSDGYLDTPYQIKSGVWADRPRFSLIQQSHEM YVMGHYIEAAVAYHQVTGNEQALEVAKKMADCLDANFGPEEGKIHGADGHPEIELALAKLYEETGEKRYLTLSQ YLIDVRGQDPQFYAKQLKAMNGDNIFHDLGFYKPTYFQAAEPVRDQQTADGHAVRVGYLCTGVAHVGRLLGD QGLIDTAKRFWKNIVTRRMYVTGAIGSTHVGESFTYDYDLPNDTMYGETCASVAMSMFAQQMLDLEPKGEYA DVLEKELFNGSIAGISLDGKQYYYVNALETTPDGLDNPDRHHVLSHRVDWFGCACCPANIARLIASVDRYIYTER DGGKTVLSHQFIANTAEFASGLTVEQRSNFPWDGHVEYTVSLPASATDSSVRFGLRIPGWSRGSYTLTVNGKPA VGSLEDGFVYLVVNAGDTLEIALELDMSVKFVRANSRVRSDAGQVAVMRGPLVYCAEQVDNPGDLWNYRLA DGVTGADAAVAFQADLLGGVDTVDLPAVREHADEDDAPLYVDADEPRAGEPATLRLVPYYSWANREIGEMRV FQRR

* *N. frontalis = Neocallimastix frontalis*; *P. rhizinflatus= Piromyces rhizinflatus; B.longum= Bifidobacterium longum*

Table 2.2: List of primers used to amplify the catalytic fragments:

Primer Name	Primer Sequence (5' - 3')
Nf2215_For	GTGCCGCGCGGCAGC
Nf2215_Rev	TATATATACTCGAGCTAGCCTTTGGTGATGTACAGGCGG
Nf2523_For	GTGCCGCGCGGCAGC
Nf2523_Rev	TATATATACTCGAGCTAACCTTTAGTGATATACAGGCGGTAACCGG
Pr2455_For	GTGCCGCGCGGCAGC
Pr2455_Rev	TATATATACTCGAGCTACTTTGGTGATGTAGATACGGTAACCGG
Nf2152_For	TATATACATATGGTCAGCCAGCTGACTGTAGATTGC
Nf2152_Rev	TATATACTCGAGCTACGCTTTGGTAATGTAAATGCGGTAAC

Table 2.3: List of primers used for NfGHX site-directed mutagenesis:

Primer Name	Primer Sequence (5' - 3')
Glu112Gln_For	GCAACGAGATCTGGAACCAGCCGGATGTGACTTGG
Glu112Gln_Rev	CCAAGTCACATCCGGCTGGTTCCAGATCTCGTTGC
Glu211Gln_For	GTCCGATCACTATCAACCAATACTGCGACGAAAAC
Glu211Gln_Rev	GTTTTCGTCGCAGTATTGGTTGATAGTGATCGGAC

Table 2.4: Total identified sequences from anaerobic fungal transcriptomes that contain CBMs.

CBM Family	Total Proteins	Unknown Domains
CBM6, CBM35	41	30
CBM13	93	65
CBM1, CBM10, CBM29	726	307
Other CBMs	2288	724
Total	3148	1126

Chapter 3

Results

3.1 Target gene selection and synthesis:

3.1.1 Identifying novel enzymes associated with CBM13s

Four species of ruminal anaerobic fungi (*Neocallimastix frontalis, Piromyces rhizinflatus, Anaeromyces mucronatus, Orpinomyces joyonii*) were cultured *in vitro* in the presence of barley straw and glucose. The RNA from these cultures was extracted, converted into cDNA and the transcriptomes of each was generated. In the preliminary analysis 3,148 hypothetical proteins that were (1) upregulated in the presence of barley straw, (2) contained unknown coding regions larger than 300 amino acids, and (3) were appended to CBMs from families CBM1, CBM6, CBM10, CBM13, and CBM35 were identified (Table 3.1). These CBM families were selected because they have polyspecific binding specificities [66, 78, 101-103] and the majority of them are known to interact with cellulose and hemicellulose [78, 101, 102]. Of these sequences, 1,126 were appended to dockerin and / or CBM10 domains (Table 3.1), suggesting that they may be components of cellulosomes involved in the breakdown of recalcitrant polysaccharide substrates [104, 105]. These CBM and dockerin domain-containing proteins of unknown function therefore represent a potential treasure trove of novel fibrolytic enzymes. In this study, four CBM13 containing sequences with predicted catalytic modules distantly related to GH39 and GH51, named Nf2152, Nf2215, Nf2523 and Pr2455, were selected for biochemical characterization. If functional, these enzymes would be founding members of a novel GH family, referred to here as GHX.

3.1.2 Comparative analyses of GHX sequences with characterized sequences from GH51 and GH39.

The automated pipeline 'CoMPASS' (CAZyme Multifunctional Predictive Analyses using Sequence Similarity) developed at the Lethbridge Research Station [106] was used to create phylogenetic trees. Two trees were constructed with functionally characterized (i.e. with validated EC numbers) enzymes from family GH51 (n =76; Figure 3.1A) or GH39 (n = 16; Figure 3.1B) and the four trimmed GHX sequences. In both trees the GHX sequences partition as a single clade that is distantly related to the characterized sequences. Comparisons with the distribution of GH51 known activities, which include endoglucanase, endo-β-1,4-xylanase, β-xylosidase and α-L-arabinofuranosidase (ABF); reveals that the GHX clade branches out from a region of bacterial ABF, suggesting they may have resulted by horizontal gene transfer. By comparison, the GHX sequences diverge much earlier in the GH39 tree. GH39 is also a polyspecific family containing α-L-iduronidase and β-xylosidase activities (Figure 3.1B). Intriguingly, in this analysis, GHX sequences display a distant relatedness to bacterial β-xylosidases. Based upon sequence, therefore, the GHX enzymes appear to be members of a novel GH family potentially involved in the hydrolysis of α-L-arabinofuranose or β-xylose substrates. Both of these pentoses are abundant in arabinoxylans.

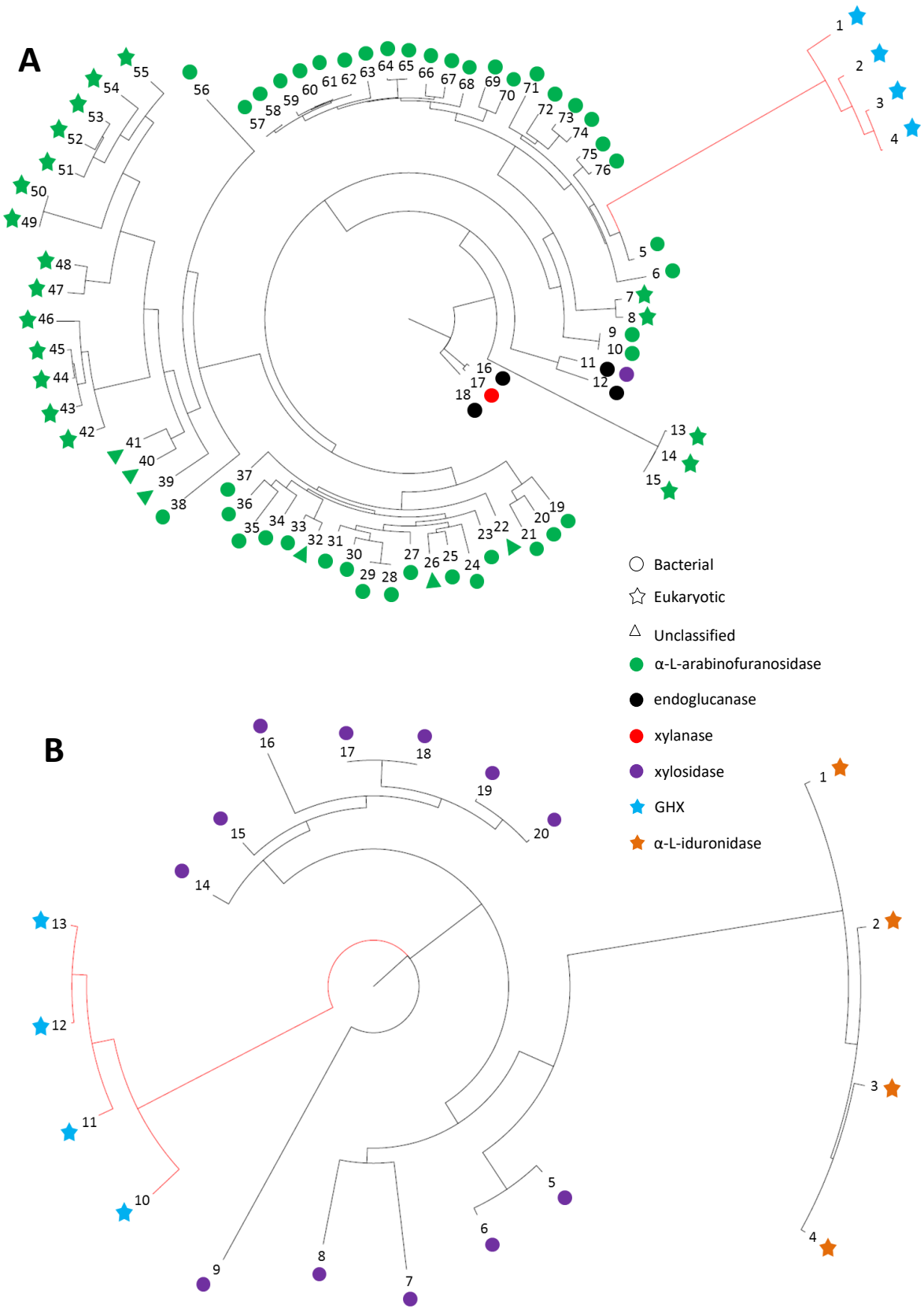

A
B
Bacterial
Eukaryotic
Unclassified
α-L-arabinofuranosidase
endoglucanase
xylanase
xylosidase
GHX
α-L-iduronidase

Figure 3.1: Phylogenetic trees of GHX sequences aligned with characterized sequences from (A) GH51 and (B) GH39. Phylogenetic trees were created using the CoMPASS pipeline [106]. Each sequence leaf represents the catalytic module from each sequence entry. Characterized activities and organism source are depicted with different symbol colours and shapes, respectively. Accession numbers of the characterized sequences are leveled with numbers.

3.1 (A):

SL	Accession number	SL	Accession number	SL	Accession number	SL	Accession number
1	Nf2152	20	ADJ95768.1	39	AEF56862.1	58	ABI34800.1
2	Pr2455	21	BAF39986.1	40	AEF56861.1	59	ACB54691.1
3	Nf2523	22	AEF56856.1	41	CCO20994.1	60	ABD48560.1
4	Nf2215	23	AAN05450.1	42	AAF19575.1	61	EDY06090.1
5	ACY69989.1	24	AEE64774.1	43	AFF58879.1	62	AAD45520.1
6	AEE47435.1	25	AAC28125.1	44	BAF27505.1	63	ABN53749.1
7	EAA65870.1	26	AEF56855.1	45	AFF58880.1	64	ABP67153.1
8	BAQ35745.1	27	CAA76421.1	46	BAC10349.1	65	ACM60204.1
9	AAD35369.1	28	ABC55452.1	47	CAL81200.1	66	ABZ10760.1
10	ABQ46651.1	29	CAJ77816.1	48	CCC33068.1	67	CAA99595.1
11	CAF22222.1	30	CAA99576.1	49	BAB96815.1	68	ALK02878.1
12	AAC45377.1	31	ACE73681.1	50	CAK43424.1	69	ADJ95771.1
13	BAB21568.1	32	CCO20976.1	51	AEQ94263.1	70	ADJ95770.1
14	ADG21260.1	33	AAA50393.1	52	BAG71680.1	71	AAO84266.1
15	AAC41644.1	34	ACD60479.1	53	ABO93602.1	72	ADM26764.1
16	CEH24710.1	35	AAC38457.1	54	ADZ98861.1	73	ADT80795.1
17	ACV57112.1	36	ACE86344.1	55	BAN70283.1	74	AAN24368.1
18	ADI82825.1	37	AAC38456.1	56	BAA90771.1	75	AAA61708.1
19	AFD62907.1	38	BAF40305.1	57	ABM68633.1	76	BAH02662.1

3.1 (B):

SL	Accession number	SL	Accession number	SL	Accession number	SL	Accession number
1	AAC42044.1	6	AAA23063.1	11	Pr2455	16	CAD48308.1
2	AAA81589.1	7	BAA95685.1	12	Nf2215	17	AAB87373.1
3	AAA51455.1	8	AAK24328.1	13	Nf2523	18	ABP67986.1
4	AIT52923.1	9	AAG05625.1	14	BAB04787.1	19	AFK86459.1
5	ADQ03734.1	10	Nf2152	15	ABI49941.1	20	AAA27369.1

3.2 Production of GHX and screening of substrate activity:

3.2.1 Molecular engineering and production of Nf2152

All four GHX sequences were commercially synthesized and subcloned into the expression vector pET28a (BioBasic Inc); however, only Nf2152 and Nf2523 were examined further in this project. Based upon expression trials with a variety of modular versions of Nf2152, an optimized construct that purified at high yields was engineered (Figure 3.2A). The construct consists of linker 1 + catalytic fragment + linker 2 spanning amino acids 29 to 430 with an estimated molecular weight of 45.1 kDa [91] (Figure 3.2A). Recombinant Nf2152 reproducibly purified at approximately 20 mg L^{-1} and was stable in 100 mM NaCl following IMAC and size exclusion chromatography (Figure 3.2B & 3.2C).

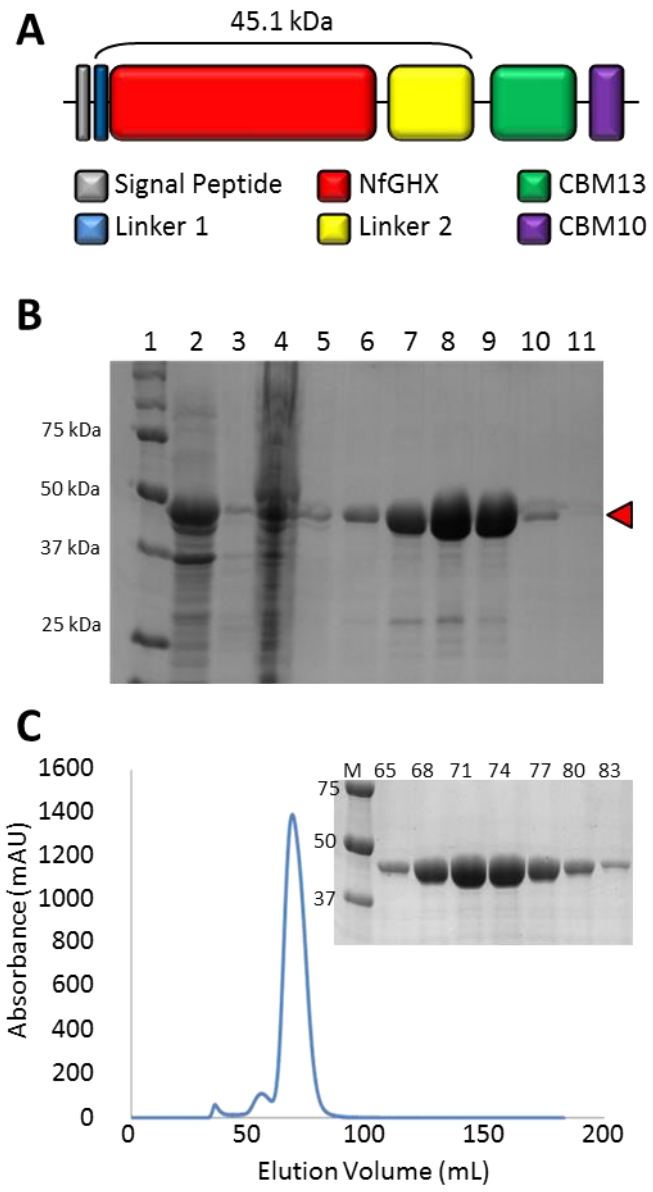
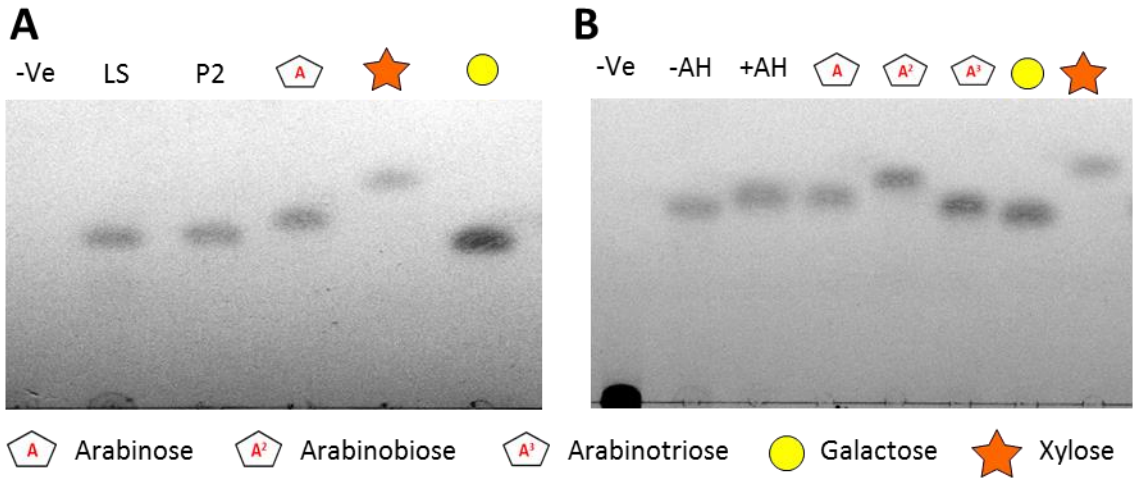
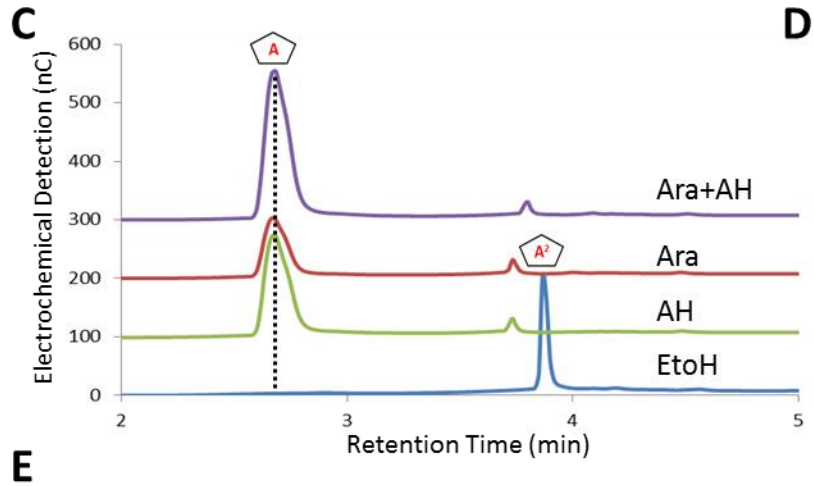


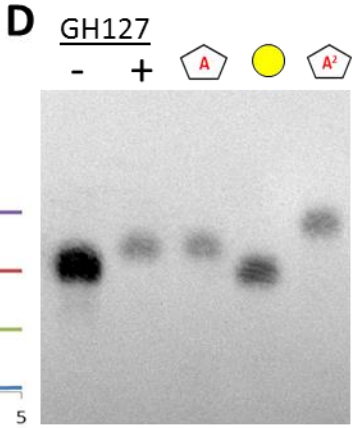
Figure 3.2: Structure of the Nf2152 expression construct and purification of recombinant Nf2152. (A) Modular structure of the full length Nf2152. Different modules are depicted with different colours (see legend). The NfGHX functional fragment (Linker 1–NfGHX–Linker 2) was subcloned into the bacterial vector pET28a for expression and purification of protein by IMAC chromatography. **(B)** SDS gel showing the purification of Nf2152 by IMAC. Lane (1): Marker, (2): Soluble, (3): Insoluble, and (4): IMAC Flow Through. The remaining lanes display fractions from an imidazole gradient of (5): 0 mM, (6): 5 mM, (7): 10 mM, (8): 30 mM, (9): 100 mM, (10): 200 mM, and (11): 500mM. **(C)** Chromatogram of SEC purified NfGHX. The inset is a SDS gel of collected fractions (mL).

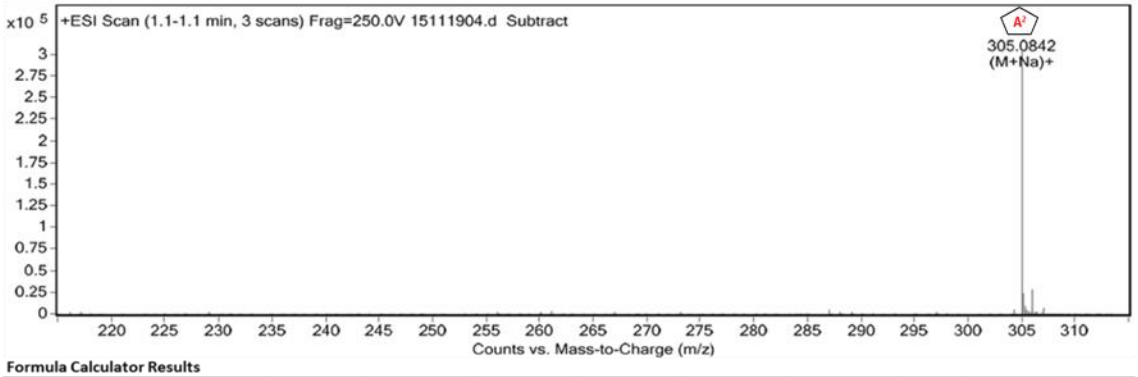
3.2.2 Substrate screening of Nf2152

The activity of purified Nf2152 was screened by thin layer chromatography on a variety of plant cell wall carbohydrates, including SBA, RAX, wheat arabinoxylan, β-glucan, xyloglucan, galactomannan, pectic galactan, arabinogalactan, RG-I, arabinofuranosyl-xylobiose, arabinofuranosyl-xylotriose, *P*NP-α-L-arabinofuranoside, *P*NP-α-L-arabinopyranoside and beech wood xylan. Activity was detected primarily on SBA and RAX. The optimized digestion reaction for large scale production contained 5 mg mL^{-1} of SBA, 0.5 µM of enzyme and 20 mM sodium acetate buffer (pH 5.0). TLC analysis of the digestion products demonstrates that Nf2152 releases a single product band that migrates slower than arabinose and xylose, α-1,5-arabinobiose and α-1,5-arabinotriose (Figure 3.3A). This observation indicates that the protein is a functional enzyme; however, it is not an arabinofuranosidase or xylosidase as suggested by the comparative analysis with GH51 (Figure 3.1A) and GH39, respectively (Figure 3.1B).









Formula Calculator Results

Formula	Ion Species	Mass	Calc. Mass	m/z	Calc. m/z	Diff (mDa)	Diff (ppm)	DBE	Ion	Score
C10 H18 O9	C10 H18 Na O9	282.095	282.0951	305.0842	305.0843	0.1	0.35	2	(M+Na)+	95.08

Figure 3.3: Structural analysis of the Nf2152 product released from SBA. (A) TLC analysis of the Nf2152 product released from SBA before (LS) and after (P2) SEC purification. Samples are compared with arabinose, xylose, and galactose standards. **(B)** Acid hydrolysis of P2 bio-gel column purified products. Appropriate standards are labeled. **(C)** Overlaid HPAEC-PAD chromatograms showing acid hydrolyzed products released from SBA. Blue = ETOH soluble sample, green = acid hydrolysate, red = arabinose standard, and purple = acid hydrolysate spiked with arabinose. **(D)** Digestion of product with GH127 (β1,2-L-arabinofuranosidase). Appropriate standards are labeled. **(E)** HR-MS spectrum displaying measured m/z ratio and calculated mass of product.

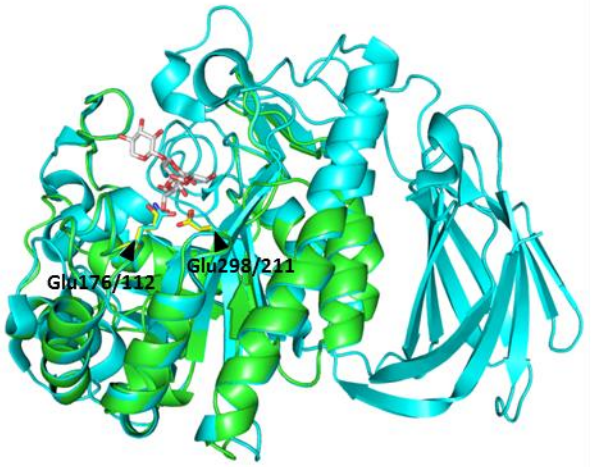
3.3 Site-directed mutagenesis:

Phyre2 homology modeling [92] reveals that Nf2152 displays the most structural similarity with the GH51 from *Thermobacillus xylanilyticus* (TxGH51; PMID: 2VRQ; [93]). GH51s are retaining enzymes with experimentally determined catalytic residues [93, 107]. In TxGH51, the nucleophile is Glu298 and the acid/base is Glu176. Structural superimposition of Nf2152 with TxGH51 suggested that these catalytic residues are spatially conserved between the two enzymes and align with Glu112 and Glu211 in Nf2152 (Figure 3.4B). Therefore, to determine if these two amino acids were essential for the catalytic function of Nf2152 both were mutated to glutamine residues. The mutant enzymes were sequence verified, produced, purified, and tested for activity on SBA. Mutations to both residues, which was about 5.5 A apart from each other, resulted in the complete loss of detectable product (Figure 3.4C), suggesting that they are catalytic in function and that GHXs deploy a retaining mechanism.

A

```
Nf2152      ----------------------------ATHCAS-GAHYGLIENVP------KDYKSLV   24
TxGH51      MNVASRVVVNADRVKGTINRNIYGHFSEHLGRCIYEGLWVGEDSPIPNTNGIRNDVLEAL   60
                                         :*   *   *  . :*      :*  . :

Nf2152      APLNINVMRAPARA---------GNGRQQ--------PIGDVIKVAQRLKESPGARVTIE   67
TxGH51      KQMKIPVLRWPGGCFADEYHWKDGVGPREKRKRMVNTHWGGVIEN-----NHFGTHEFMM   115
              ::* *:* *. .          * * ::          *.**:      :  *::  :

Nf2152      LADILPGWPYRWP--------GIQTWFNEI-----RSFINDKKKSGLTNFYGNEIWNEPD   114
TxGH51      LCELLGCEPYISGNVGSGTVQEMSEWVEYITFDGESPMANWRRENGREKPWRIKYWGVGN   175
            *.::*   **            :. *.: *       : * :::.*  : :  : *.  :

Nf2152      VTWKDSNGLS---FNQMWKQTYDLLRQIDPNEKIIGPSFSWYEENKMKNFLQFSKQNNCL   171
TxGH51      ENWGCGGNMRAEYYADLYRQFQTYLRNY--------------GDNKLHKI--------ACG   214
             .*  ...:    : ::::*    **:                :**::::         *

Nf2152      PDI--IAWHEL------SGIDGVSSHFRSYRNLEKSLGISERPITINEYCDENHDLEGQP   223
TxGH51      ANTADYHWTEVLMKQAAPFMHGLSLHYYT-------------------------------   243
             :     * *:        :.*:* *: :

Nf2152      GSSARFIGRFERYKVDSG--MITWWFVPHPGRLGSLLAS---------DTQKGAGW---Y   269
TxGH51      -----VPGPWEKKGPATGFTTDEWWVTLKKALFMDRLVTKHSAIMDVYDPDKRIDLIVDE   298
                 . * :*:    :*     **.. : . : . *.:         * :*  .

Nf2152      FYKWYGDMTG-------------------------------------------------   279
TxGH51      WGTWYDVEPGTNPGFLYQQNSIRDALVAGATLHIFHRHCDRVRMANIAQLVNVLQSVILT   358
            : .**.   *

Nf2152      -------------------------------------------------------------   279
TxGH51      EGERMLLTPTYHVFNMFKVHQDAELLDTWESVERTGPEGELPKVSVSASRAADGKIHISL   418

Nf2152      -------------------------------------------------------------   279
TxGH51      CNLDFETGASVDIELRGLNGGVSATGTTLTSGRIDGHNTFDEPERVKPAPFRDFKLEGGH   478

Nf2152      ------------------  279
TxGH51      LNASLPPMSVTVLELTAG  496
```

B



C

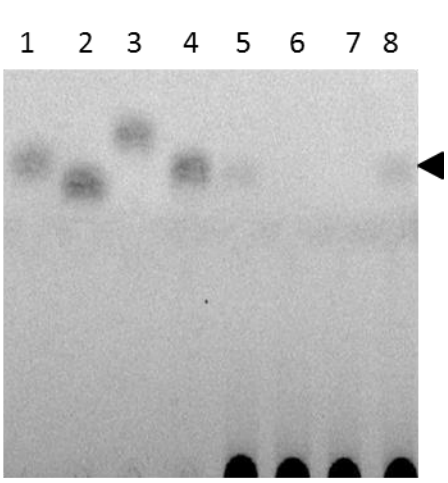


Figure 3.4: Identification of catalytic residues in Nf2152. (A) Multiple sequence alignment of the amino acid sequence of Nf2152 (catalytic fragment) with TxGH51 from *Thermobacillus xylanilyticus* **(B)** Superimposition of the Phyre2 derived [92] homology model of Nf2152 (green) with the TxGH51 from *Thermobacillus xylanilyticus* (cyan; PDB ID 2VRQ; [93]). Both enzymes are represented in a cartoon format. TxGH51 was used as the threading template. The potential catalytic amino acid residues Glu112 (176 TxGH51) and Glu211 (298 TxGH51) in Nf2152 are labeled. The bound arabinosyl-xylotriose substrate of TxGH51 is shown as grey sticks. **(C)** TLC of digested SBA with wild-type and mutant enzymes. Lane (1) arabinose, (2) galactose, (3) α1,5-arabinobiose, (4) α1,5-arabinotriose, and digestions of SBA with (5) wild type NfGHX, (6) E176Q-NfGHX, (7) E298Q-NfGHX, and (8) wild type NfGHX. Black arrow indicates released product from the wild-type enzyme.

3.4 Structural analysis of the Nf2152 product:

3.4.1. Product purification

In order to characterize the structure of the product, 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 (data not shown), and samples containing similar sized products were pooled and analyzed for purity by TLC. The primary peak (Fractions #169-190) was selected for structural analysis (Figure 3.3A).

3.4.2 Acid hydrolysis of product

To determine if the product had a DP > 1, acid hydrolysis was performed. This treatment generates a single band that co-migrates with arabinose when analyzed by TLC (Figure 3.3B) and eluted with an identical retention time to arabinose when analyzed by HPAEC-PAD (Figure 3.3C). These results suggest that Nf2152 generates an oligosaccharide that is solely composed of arabinose.

3.4.3 Determining the arabinooligosaccharide linkage and degree of polymerization

Comparison with a commercially available α1,5-arabinobiose standard reveals that these two arabinooligosaccharides display different mobility patterns (Figure 3.3B). This result suggests the product must have a differing chemistry, such as a modification or alternate stereochemical (i.e. β) or positional (e.g. 1→2) linkage. Due to the limited availability of commercial arabinooligosaccharide standards, a β1,2-arabinofuranosidase GH127 [108] was synthesized, purified, and used for enzymatic glycosequencing. Previously this enzyme was used to elucidate the structure of an β1,2-arabinosyl

oligosaccharide derived from an AGP [108]. Treatment with the GH127 cleaved the Nf2152 product completely into arabinose (Figure 3.3D), indicating that the product is a pure oligosaccharide with β1,2-linkages. This digestion does not determine the DP of the product; however, as the GH127 could act processively. To determine the size of the arabinooligosaccharide; therefore, electrospray-ionization mass spectrometry was performed (Alberta Glycomics Centre). The product was determined to have a m/z ratio [(M+Na)+] of 305.1. This ratio equates into a calculated mass of 282.1, which is the mass of a pentose disaccharide (Figure 3.3E).

3.5 Structural characterization of a second product generated by GHXs:

3.5.1 Detection of a distinct product

The functional screening of Nf2152 generated a second band from RAX with a slower mobility (Figure 3.5A). Interestingly, this band appears to be the primary product of a second GHX enzyme, Nf2523 (Figure 3.5A). This result indicates that (1) different substrates are present in RAX, (2) Nf2152 releases two structurally distinct products, and (3) there is functional specialization between members of GHX. To determine if the second product was an arabinofuranooligosaccharide with a higher DP, mass spectrometry was performed (Alberta Glycomics Centre). Surprisingly, the calculated mass was observed to be 312.1, which is equivalent to the molecular weight of a hexose-pentose disaccharide (Figure 3.5B).

3.5.2 Determining the Composition of a heterogeneous disaccharide product

To determine the structure of the hexose-pentose disaccharide, acid hydrolysis was performed. When analyzed by TLC (Figure 3.5C) and HPAEC-PAD (Figure 3.5D), two

distinct monosaccharides were visible: arabinose and galactose. Arabinose is a main compositional sugar of SBA (88%) and RAX (Ara 38%); whereas, galactose is less common (SBA = 3%; RAX = none detected). In this regard, the presence of an arabinose-galactose disaccharide product, suggests that it is a rare structure within both polysaccharides and it may represent a potentially valuable carbohydrate reagent. Alternatively, arabinose and galactose are commonly linked sugars in the arabinogalactan sidechains of RG-I [109]. However, digestion of various forms of these substrates did not produce detectable products. This suggests that the substrates of Nf2152 and Nf2523 may by targeting another structure within the glycans of AGPs. Further insights into the sequence of this heterogeneous disaccharide would help to identify the chemistry of the natural substrate present within the RAX preparation.

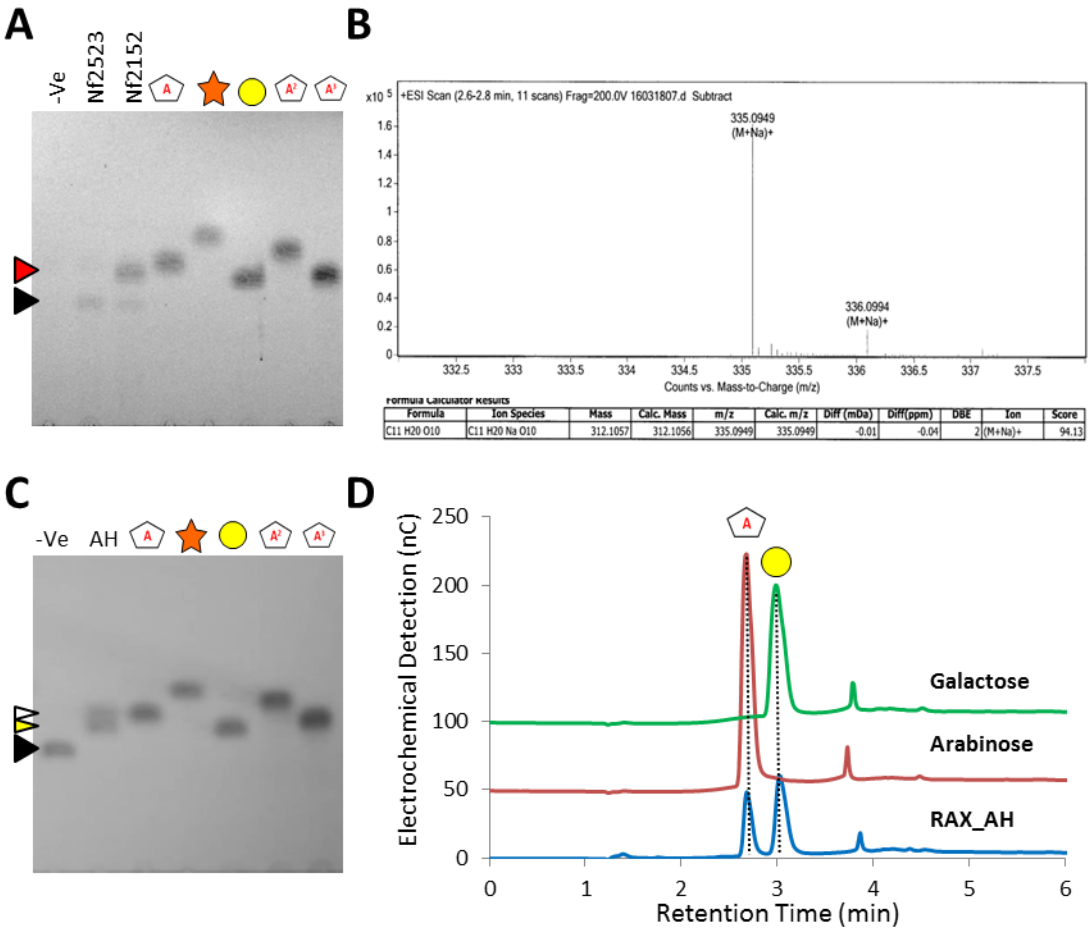


Figure 3.5: Structural analysis of the Nf2152 and N2523 products released from RAX. **(A)** Nf2523 releases a single product (black arrow) that displays lower mobility than the arabinobiose released by Nf2152 (red arrow). When RAX is digested by Nf2152 both bands are visible. **(B)** HR-MS spectrum of the P2 bio-gel column purified Nf2523 product released from RAX. **(C)** TLC of acid hydrolyzed RAX product. The unhydrolyzed sample is indicated with a black arrow; the galactose and arabinose are indicated with a yellow and white arrow, respectively. **(D)** HPAEC-PAD chromatograms of the galactose and arabinose monosaccharides released by acid hydrolysis of the Nf2523 generated product.

3.5.3 Sequencing the galactose/arabinose Nf2523 product

In order to determine whether arabinose or galactose was positioned at the reducing end of the disaccharide two complementary techniques were performed. First, alditol derivatized profiles of the intact product and the product after it was hydrolyzed to arabinose and galactose, were visualized using GC-MS. This technique only reduces and acetylates the reducing end. In this way the carbohydrate at the non-reducing end would be protected from derivatization (until it is released by acid hydrolysis) and therefore is not visible in the GC-MS chromatogram. In Figure 3.6A, arabinose is acetylated in both conditions; however, galactose is only visible if derivative after hydrolysis, indicating it is protected from reduction when the disaccharide is intact. This analysis indicates that the product is a galactoarabinose (galactose = non-reducing; arabinose = reducing end).

To confirm this result a second technique was conducted. The reducing end of the disaccharide was labeled by reductive amidation with the fluorogenic compound ANTS. Following derivatization, the disaccharide was hydrolyzed into monosaccharides and analyzed by FACE (Figure 3.6B). Visualization of the products reveals that arabinose is the labeled sugar, and therefore, must be exposed at the reducing end in the starting material. To the best of my knowledge, this is the first report of galactoarabinobioside releasing enzyme.

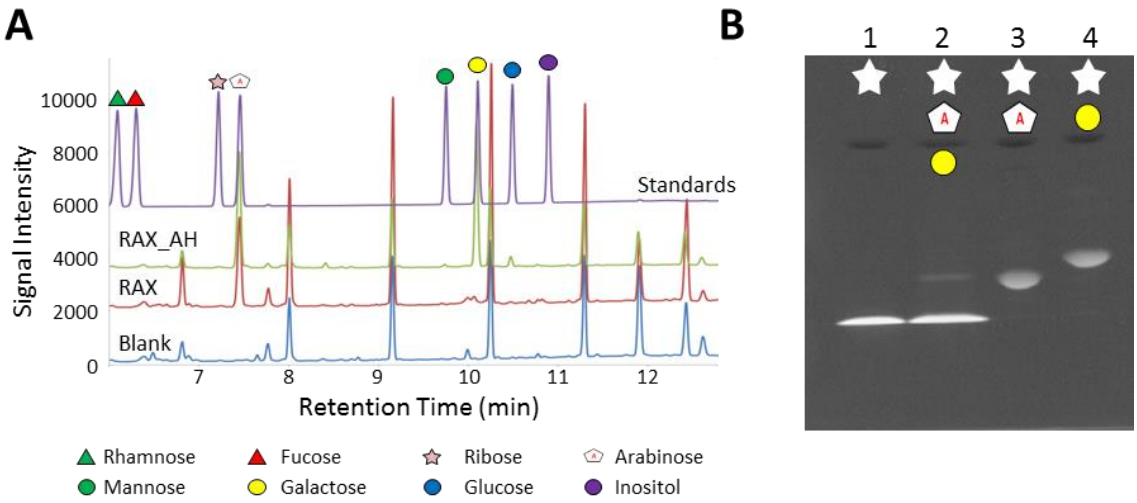


Figure 3.6: Sequencing of the heterogeneous disaccharide released from RAX by labeling and visualization of the reducing end. (A) GC-MS chromatogram of the RAX product. The background reaction buffer is shown in blue. The disaccharide was reduced before (red trace) and after (green trace) acid hydrolysis, and then acetylated. Only arabinose is detected in the non-hydrolyzed sample, whereas both arabinose and galactose are visible when the monosaccharides are generated before reduction and derivatization. The monosaccharide standards are shown in purple and are labeled with symbols. **(B)** FACE analysis of the ANTS labeled disaccharide. The Nf2523 product was labeled and then acid hydrolyzed. The products were compared to various standards. Lane (1): ANTS, (2): acid hydrolyzed disaccharide, (3) arabinose, and (4) galactose. The detection of arabinose in the disaccharide sample confirms that arabinose is positioned at the reducing end. ANTS = white star, other symbols as in 3.6A.

Chapter 4

Discussion

Discovering new enzyme activities that improve the digestibility of recalcitrant plant cell wall polysaccharides is a promising solution for livestock production. In this regard, the rumen anaerobic fungi (Figure 1.1) represent an unexploited source of potentially novel and unique enzymatic activities. Enzymes that improve upon 'rate-limiting' reactions in the saccharification of complex carbohydrates within the rumen hold promise for increasing the efficiency of ruminant digestion. Within this thesis, I have explored the potential of a new family of enzymes, entitled GHX, found within *N. frontalis* and *P. rhizinflatus* to hydrolyze substrates found within the glycans of AGPs that co-purify with common cell wall polysaccharides.

4.1 Discovery of novel GHX family using CBMs and phylogenetic tree:

Enzymes that participate in the deconstruction of plant cell walls often contain one or more noncatalytic CBMs. CBMs play fundamental roles in targeting their appended catalytic modules to or concentrating them on the surface of dedicated substrates [81, 110]. Most commonly, CBM binding specificity reflects the catalytic specificity of the parent enzyme, and therefore, investigating uncharacterized enzymes that are appended to CBMs from polyspecific families, may allow the discovery of a wide range of activities [78, 101, 102]. My original hypothesis was supported by this study, in that hypothetical proteins upregulated in the presence of barley straw (preliminary data) and that contained CBM13s were used successfully to select functional target genes for biochemical characterization.

To gain insights into the potential function of four GHX sequences, Nf2152, Nf2215, Nf2523 and Pr2455 (Table 2.1), these sequences were embedded into extracted datasets of characterized GH51 and GH39 sequences and phylogenetic trees were created using the fully automated pipeline CoMPASS [106]. In both trees the GHX sequences partition as a single clade that is distantly related to the characterized sequences (Figure 3.1A & 3.1B). Based upon their distributions, the GHX enzymes appear to be members of a novel GH family potentially involved in the hydrolysis of α-L-arabinofuranose or β-xylose substrates.

4.2 Substrate specificity and product profile of Nf2152

To further investigate the function of a representative GHX, recombinant Nf2152 was produced (Figure 3.2) and screened for activity on a variety of plant cell wall carbohydrates. Nf2152 was shown to release an arabinobiose product from SBA (Figure 3.3). Based upon structural superimpositions with TxGH51 (Figure 3.4A), site directed mutagenesis of Nf2152 (Figure 3.4B) and the distance between the catalytic residues it is evident that Nf2152 deploys a retaining mechanism during catalysis. Surprisingly, Nf2152 also generated a similar product from RAX (Figure 3.5). SBA and RAX are isolated from different plant sources and display very different chemistries.

SBA is a polysaccharide constituent of pectin. It is composed of a α-1,5-L-arabinofuranosyl backbone decorated with α-1,2- and α-1,3-L-arabinofuranosyl side chains. Arabinose is the second most abundant pentose in nature, and therefore, discovery of novel biocatalysts active on arabinans are promising tools for various bioconversion industries, including livestock production. There are many enzymes

known to be active on SBA. Collectively these enzymes are referred to as arabinanases (ABNs), which are found in GH43 and GH93 families [111]; and arabinofuranosidases (ABFs), which are found in GH3, GH43, GH51, GH54, GH62, GH93 and GH127 families [108, 112, 113]. ABNs catalyze the hydrolysis of the α-1,5-linked L-arabinofuranosyl backbone of plant cell wall arabinans, releasing arabinooligosaccharides and arabinose. ABFs cleave arabinosyl decorations in an exolytic fashion[114, 115]. Currently, several α1,5-arabinobiosidases have been reported that processively hydrolyze the backbone of arabinan [116-118]. Recent studies have shown that the synergistic effect of fungal enzyme mixtures supplemented with ABNs and ABFs have improved the hydrolysis rate of plant biomass [119-121].

RAX is an abundant polysaccharide within the plant cell wall, and a primary component of many livestock feeds / forages [122]. It possesses a D-xylosyl backbone decorated with α-1,2- and α-1,3- linked L-arabinosyl residues [122, 123]. 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 [124]. Multiple enzyme products that target the xylan backbone of arabinoxylan products, such as Danisco Xylanase and Pentopan Mono BG are currently available, indicating there is a market available for improved biocatalysts that improve the digestion of RAX.

Structural analysis of the arabinobiose product released by Nf2152 from SBA (Figure 3.4A) and RAX (Figure 3.5A) revealed that the product does not appear to be an expected component of either polysaccharide. This finding indicates that Nf2152 is not a common ABF or xylosidase as was suggested by the comparative analysis with GH51

(Figure 3.1A) and GH39 (Figure 3.1B). When analyzed by TLC, the product migrates slower than arabinose, xylose, and α1,5-arabinobiose; similar to α1,5-arabinotriose; and slightly faster than galactose (Figure 3.4A). Acid hydrolysis confirmed that the product is a pure arabinooligosaccharide (Figure 3.4B & 3.4C); however, it does not reveal the 1) DP of the sugar, 2) if there are acid labile modifications present, or 3) absolute ring configuration of the carbohydrate (Figure 1.4)

To determine the stereochemistry of the linkage and ring configuration of the arabinose located at the non-reducing end, the product was digested with BlGH127, an exolytic β-1,2-L–arabinofuranosidase (Figure 3.4D). This evidence combined with the HR-MS data (Figure 3.4E), determined that the product is β-1,2-arabinobiose (Figure: 4.1A), and therefore, Nf2152 could be classified as a β-1,2-arabinobiosidase (EC #3.2.1.-; GH121) [112]. Recently, BlGH121 was the first reported β-1,2-arabinobiosidase, which is also active on SBA [112] and its products are also hydrolyzed by BlGH127 [108]. This small collection of enzymes represents some of the only known enzymes specific for the degradation of β1,2-arabinosyl containing AGP glycans. Sequence comparisons (data not shown) reveal that BlGH121 is not related to Nf2152 at the sequence level, and therefore, Nf2152 represents the founding member of a new GH family. Moreover, BlGH121 was of bacterial origin where as Nf2152 is of eukaryotic origin, so GHX represents a novel family of β1,2-arabinobiosidases of fungal origin.

4.3 Substrate specificity and product profile of Nf2523

Closer examination of the RAX digestions with Nf2152 revealed that there was a second product generated that displayed lower mobility than β-1,2-arabinobiose (Fig. 3.5A). This result suggests that Nf2152 may have a side-activity or that it may be a β-1,2-endoarabinase that cleaves β-1,2-arabinan chains present only in the RAX preparation. Intriguingly in screening reactions with other GHXs, it was observed that this second band was the exclusive product of Nf2523 (Figure 3.5A). Therefore, large scale digests of RAX were performed to generate preparative amounts of the alternative product.

HR-MS analysis of the Nf2523 product revealed that it was a disaccharide containing one hexose and one pentose subunit (312.1 Da; Figure 3.5B). To determine its composition, acid hydrolysis was performed demonstrating that it consists of galactose and arabinose (Figure 3.5C & 3.5D). This was a surprising result as the composition of commercially prepared RAX is not reported to contain galactose [125]. Indeed, both the β-1,2-arabinobiose and galactose-arabinose disaccharide are not known to exist in SBA or RAX; however, they are likely part of the larger glycan network of the plant cell wall. In particular, these enzymes may be specific for the complex arabinosyl/galactosyl containing glycans of AGPs [50].

In order to provide more insight into the product profiles of these two enzymes, the Nf2523 product was sequenced using GC-MS (Figure 3.6A) and fluorescent derivatization (Figure 3.6B) techniques. Both results confirmed that the galactose is positioned at the non-reducing end and the arabinose is the reducing sugar (Figure

4.1B). Interestingly, the presence of a galactose at the non-reducing end would suggest that the arabinose at the non-reducing end of the β1,2-arabinobiose structure may adopt a pyranose configuration (6-dehydroxymethyl D-galactose in a $^4C_1$ configuration) most commonly seen in AGP glycans (Figure 1.4). The hydrolysis of the Nf2152 product by the BlGH127 β-L-ABF (EC#3.2.1.185), however, confirms that the arabinose must be in the furanose configuration. Alternatively, the galactose in the Nf2523 product could be in the α-D-galactofuranose configuration, which has an identical ring structure as β-L-arabinose (Figure 4.1C). Regardless, these findings indicate that these two GHX enzymes have a strict requirement for arabinose (presumably in the furanose configuration by BlGH121 [112]) in their -1 subsites and display plasticity in their -2 subsites (Figure 4.1A & 4.1B). Further structural analysis of the Nf2523 product will be required to elucidate its complete stereochemistry. This event will also highlight the potential differential recognition determinants of Nf2152 and Nf2523 for β-1,2-arabinobiose and galactoarabinose, respectively.

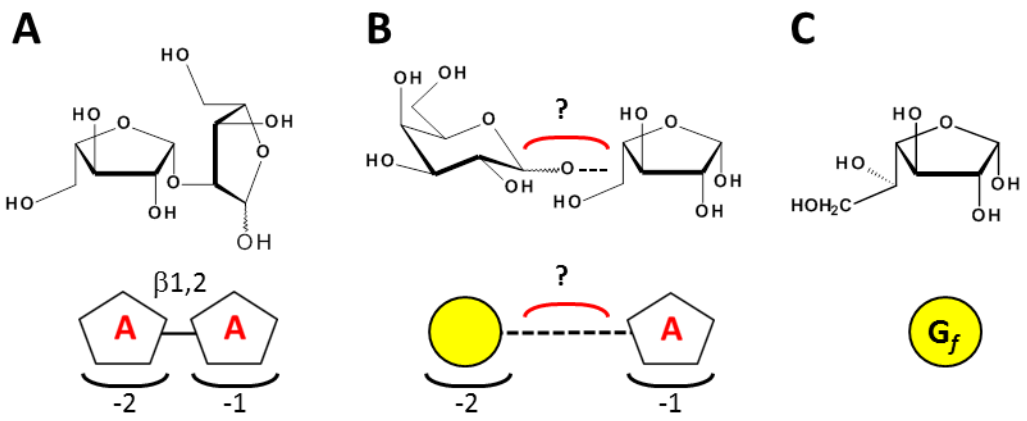


Figure 4.1: Chemical structures of disaccharides produced by GHXs. (A) β1,2-arabinobiose. The reducing end is represented as a furanose similar to what was reported for the complete *O*-methylated structure produced by BlGH121 [112]. **(B)** The sequence of the Nf2523 structure is known; however, the linkage has not yet been determined. **(C)** The rare galactofuranose configuration, which is structurally similar to β-L-arabinofuranose. For (A) and (B), the reducing end sugars are positioned in the -1 subsite; the non-reducing end sugars are positioned in the -2 subsite. Symbols: white pentagon with 'A' = arabinofuranose; yellow circle = galactopyranose; yellow circle with G_f = galactofuranose.

Chapter 5

Conclusions and Future Directions

5.1 Conclusions:

The changing landscape of animal agriculture will require the development of safe, environmentally sustainable, and effective solutions that can be implemented on-farm. In this regard, the discovery and characterization of enzymes for in-feed applications that catalyze novel reactions, and increase the rate or efficiency of bioconversion are promising lines of biotechnological research. Although many microbial ecosystems are currently being mined for beneficial enzyme activities (e.g. human distal gut; [126]) the rumen microbiota remains one of the most promising repositories of microbial enzymes active on plant cell wall polysaccharides [3, 10]. For example, the rumen microecosystem of domesticated cattle is highly tuned for converting lignocellulosic biomass into high quality meat and milk protein for human consumption. In addition to the abundance and diversity of bacterial species that colonize the cow rumen, the presence of rumen anaerobic fungi represents an underexploited source of novel CAZymes that are active on recalcitrant plant cell wall polysaccharides.

In this thesis, four rumen fungal genes (Nf2152, Nf2215, Nf2523 and Pr2455; Table 2.1) belonging to a new GH family (Figure 3.1), referred to as GHX, were selected for investigation because they were: (1) previously observed to be upregulated in the presence of barley straw, (2) they were appended to a CBM13 (a CBM family with a reported diversity of binding specificities), and (3) they displayed distant homology to GH39 and GH51 (Figure 3.1), which suggested they were functional CAZymes. Two of

these GHXs, Nf2152 and Nf2523, were selected for recombinant production (Figure 3.2) and biochemical characterization (i.e. substrate and product profiles).

GHXs were determined to be founding members of a new GH family that release two different disaccharides from SBA and RAX. Nf2152 releases β1,2-arabinobiose from SBA (Figure 3.3 & 3.4), and β1,2-arabinobiose and an galactoarabinose from RAX (Figure 3.5). The galactose-arabinose was exclusively produced by Nf2523 from RAX (Figure 3.5). Derivatization of the reducing end of the galactoarabinose product determined that it is a galactoarabinose disaccharide (Figure 3.6) with an unknown linkage. To the best of my knowledge this is the first reported galactoarabinobiosidase activity. Both β1,2-arabinobiose and galactoarabinose are not common within either SBA and RAX, and in fact are more likely released from AGP glycans that co-purify with the polysaccharides. A similar result was reported for SBA during the characterization of the unrelated β1,2-arabinobiosidase BlGH121 [112]. Interestingly, if the galactose residue in the galactoarabinose product adopts an α-galactofuranosyl configuration there would be substantial structural similarity between the two products apart from the C5 hydroxymethyl of galactose (Figure 4.1).

Very little is known about the role of AGPs in plant biology and biomass conversion, mainly due to their extensive structural complexity. In this regard, the GHXs studied here represent valuable tools for helping to sequence diverse AGP glycans and assisting in their biological turnover during digestion.

5.2 Future Directions:

This project has set the stage for future research and enzyme applications. The first goal should be to determine the linkage of the galactoarabinose. During my degree the 1,2-arabinose product was analyzed by three different NMR spectroscopists (University of Alberta, University of Lethbridge, and Concordia University). Unfortunately, the spectrum was too convoluted to resolve. It was concluded that the arabinose at the reducing end was dynamic and may be adopting multiple conformations (e.g. α-furanose, β-furanose, α-pyranose, and β-pyranose). To resolve this issue both disaccharides could be reduced to prevent re-cyclization, which would simplify the spectrum.

Alternatively, x-ray crystallographic structures with GHXs in combination with substrates and or products could be attempted. A high resolution product complex would also provide insights in the linkage between the -1 and -2 subsite (Figure 4.1), and the configuration of the galactosyl residue in the -2 subsite (Figure 4.1).

The long-term goal of this project should be to investigate the impact of GHXs on the digestibility of barley straw *in vitro*, and if appropriate, *in vivo*. Regardless of its outcome for improving digestibility of barley straw, this project has established that enzyme discovery directed by the detection of CBMs appended to hypothetical proteins is a valuable approach for streamlining enzyme discovery and may help to direct the discovery of other GH families and novel enzyme activities.

References

1. Diamond, J., *Evolution, consequences and future of plant and animal domestication.* Nature, 2002. **418**(6898): p. 700-707.

2. Reece, W.O., *Functional anatomy and physiology of domestic animals*. 2009: John Wiley & Sons.

3. Flint, H.J., et al., *Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis.* Nat Rev Microbiol, 2008. **6**(2): p. 121-31.

4. Jami, E. and I. Mizrahi, *Composition and similarity of bovine rumen microbiota across individual animals.* PLoS One, 2012. **7**(3): p. e33306.

5. Selinger, L., C. Forsberg, and K.-J. Cheng, *The rumen: a unique source of enzymes for enhancing livestock production.* Anaerobe, 1996. **2**(5): p. 263-284.

6. Mackie, R.I., *Mutualistic fermentative digestion in the gastrointestinal tract: diversity and evolution.* Integr Comp Biol, 2002. **42**(2): p. 319-26.

7. Krause, D.O., et al., *Opportunities to improve fiber degradation in the rumen: microbiology, ecology, and genomics.* FEMS Microbiology Reviews, 2003. **27**(5): p. 663-693.

8. Chaucheyras-Durand, F., et al., *Live yeasts enhance fibre degradation in the cow rumen through an increase in plant substrate colonization by fibrolytic bacteria and fungi.* J Appl Microbiol, 2016. **120**(3): p. 560-70.

9. Bath, C., et al., *The symbiotic rumen microbiome and cattle performance: a brief review.* Animal Production Science, 2013.

10. Flint, H.J., et al., *Microbial degradation of complex carbohydrates in the gut.* Gut Microbes, 2012. **3**(4): p. 289-306.

11. Varga, G.A. and E.S. Kolver, *Microbial and animal limitations to fiber digestion and utilization.* The Journal of nutrition, 1997. **127**(5): p. 819S-823S.

12. McCartney, D., et al., *Review: The composition and availability of straw and chaff from small grain cereals for beef cattle in western Canada.* Canadian journal of animal science, 2006. **86**(4): p. 443-455.

13. Badhan, A., et al., *Improvement in Saccharification Yield of Mixed Rumen Enzymes by Identification of Recalcitrant Cell Wall Constituents Using Enzyme Fingerprinting.* BioMed research international, 2015. **2015**.

14. Lu, C.D., J.R. Kawas, and O.G. Mahgoub, *Fibre digestion and utilization in goats.* Small Ruminant Research, 2005. **60**(1): p. 45-52.

15. *Ruminant's digestive system*. Available from: http://www.freechoiceminerals.com/The-Rumen.

16. Mackie, R., C. McSweeney, and A. Klieve, *4 Microbial Ecology of the Ovine Rumen.* Sheep Nutrition, 2002: p. 71.

17. Creevey, C.J., et al., *Determining the culturability of the rumen bacterial microbiome.* Microbial biotechnology, 2014. **7**(5): p. 467-479.

18. Brulc, J.M., et al., *Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases.* Proceedings of the National Academy of Sciences, 2009. **106**(6): p. 1948-1953.

19. de Menezes, A.B., et al., *Microbiome analysis of dairy cows fed pasture or total mixed ration diets.* FEMS microbiology ecology, 2011. **78**(2): p. 256-265.

20. Pitta, D.W., et al., *Rumen bacterial diversity dynamics associated with changing from bermudagrass hay to grazed winter wheat diets.* Microbial ecology, 2010. **59**(3): p. 511-522.

21. Robert, W., *The bacterial community composition of the bovine rumen detected using pyrosequencing of 16S rRNA genes.* Metagenomics, 2012. **May 1;2012**.

22. Jami, E., et al., *Exploring the bovine rumen bacterial community from birth to adulthood.* ISME J, 2013. **7**(6): p. 1069-79.

23. Akin, D., et al., *Physical degradation of lignified stem tissues by ruminal fungi.* Applied and environmental microbiology, 1989. **55**(3): p. 611-616.

24. Gruninger, R.J., et al., *Anaerobic fungi (phylum Neocallimastigomycota): advances in understanding their taxonomy, life cycle, ecology, role and biotechnological potential.* FEMS microbiology ecology, 2014. **90**(1): p. 1-17.

25. Lee, S., J. Ha, and K.-J. Cheng, *Relative contributions of bacteria, protozoa, and fungi to in vitro degradation of orchard grass cell walls and their interactions.* Applied and Environmental Microbiology, 2000. **66**(9): p. 3807-3813.

26. Leis, S., et al., *Finding a robust strain for biomethanation: anaerobic fungi (Neocallimastigomycota) from the Alpine ibex (Capra ibex) and their associated methanogens.* Anaerobe, 2014. **29**: p. 34-43.

27. Rezaeian, M., G.W. Beakes, and D.S. Parker, *Distribution and estimation of anaerobic zoosporic fungi along the digestive tracts of sheep.* Mycological research, 2004. **108**(10): p. 1227-1233.

28. Lowe, S.E., M. Theodorou, and A. Trinci, *Cellulases and xylanase of an anaerobic rumen fungus grown on wheat straw, wheat straw holocellulose, cellulose, and xylan.* Applied and Environmental Microbiology, 1987. **53**(6): p. 1216-1223.

29. Hess, M., et al., *Metagenomic discovery of biomass-degrading genes and genomes from cow rumen.* Science, 2011. **331**(6016): p. 463-467.

30. Qi, M., et al., *Snapshot of the eukaryotic gene expression in muskoxen rumen—a metatranscriptomic approach.* PLoS One, 2011. **6**(5): p. e20521.

31. Badhan, A., et al., *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, 2014. **14**: p. 31.

32. Paul, S.S., et al., *Effect of anaerobic fungi on in vitro feed digestion by mixed rumen microflora of buffalo.* Reproduction Nutrition Development, 2004. **44**(4): p. 313-319.

33. Houston, K., et al., *The Plant Cell Wall: A Complex and Dynamic Structure As Revealed by the Responses of Genes under Stress Conditions.* Frontiers in Plant Science, 2016. **7**.

34. Burton, R.A., M.J. Gidley, and G.B. Fincher, *Heterogeneity in the chemistry, structure and function of plant cell walls.* Nat Chem Biol, 2010. **6**(10): p. 724-32.

35. Cosgrove, D.J., *Growth of the plant cell wall.* Nature reviews molecular cell biology, 2005. **6**(11): p. 850-861.

36. Harris, P.J., *Primary and secondary plant cell walls: a comparative overview.* New Zealand Journal of Forestry Science, 2006. **36**(1): p. 36.

37. *Plant cell wall structure*. Available from: http://biology-forums.com/index.php?action=gallery;sa=view;id=5421

38. Gilbert, H.J., *The biochemistry and structural biology of plant cell wall deconstruction.* Plant Physiol, 2010. **153**(2): p. 444-55.

39. Martone, P.T., et al., *Discovery of lignin in seaweed reveals convergent evolution of cell-wall architecture.* Current Biology, 2009. **19**(2): p. 169-175.

40. Mohnen, D., M. Bar-Peled, and C. Somerville, *Cell wall polysaccharide synthesis.* Biomass Recalcitrance: Deconstructing Plant Cell Wall Bioenergy, 2008: p. 94-187.

41. Ryden, P., et al., *Tensile properties of Arabidopsis cell walls depend on both a xyloglucan cross-linked microfibrillar network and rhamnogalacturonan II-borate complexes.* Plant physiology, 2003. **132**(2): p. 1033-1040.

42. Bengtsson, S., P. Åman, and R. Andersson, *Structural studies on water-soluble arabinoxylans in rye grain using enzymatic hydrolysis.* Carbohydrate polymers, 1992. **17**(4): p. 277-284.

43. Jarvis, M., S. Briggs, and J. Knox, *Intercellular adhesion and cell separation in plants.* Plant, Cell & Environment, 2003. **26**(7): p. 977-989.

44. Atmodjo, M.A., Z. Hao, and D. Mohnen, *Evolving views of pectin biosynthesis.* Annual Review of Plant Biology, 2013. **64**: p. 747-779.

45. Jones, L., et al., *Cell wall arabinan is essential for guard cell function.* Proceedings of the National Academy of Sciences, 2003. **100**(20): p. 11783-11788.

46. Vanholme, R., et al., *Lignin biosynthesis and structure.* Plant physiology, 2010. **153**(3): p. 895-905.

47. Cesarino, I., et al., *An overview of lignin metabolism and its effect on biomass recalcitrance.* Brazilian Journal of Botany, 2012. **35**(4): p. 303-311.

48. Liang, Y., et al., *Identification and characterization of in vitro galactosyltransferase activities involved in arabinogalactan-protein glycosylation in tobacco and Arabidopsis.* Plant physiology, 2010. **154**(2): p. 632-642.

49. LiTan, A.M., et al., *Arabinogalactan-proteins and the research challenges for these enigmatic plant cell surface proteoglycans.* Current challenges in plant cell walls, 2007. **3**: p. 161.

50. Knoch, E., A. Dilokpimol, and N. Geshi, *Arabinogalactan proteins: focus on carbohydrate active enzymes.* Front Plant Sci, 2014. **5**: p. 198.

51. Showalter, A.M., *Arabinogalactan-proteins: structure, expression and function.* Cell Mol Life Sci, 2001. **58**(10): p. 1399-417.

52. Tan, L., et al., *Plant O-hydroxyproline arabinogalactans are composed of repeating trigalactosyl subunits with short bifurcated side chains.* Journal of Biological Chemistry, 2010. **285**(32): p. 24575-24583.

53. Cummings, J. and A. Stephen, *Carbohydrate terminology and classification.* European Journal of Clinical Nutrition, 2007. **61**: p. S5-S18.

54. Stick, R.V. and S. Williams, *Carbohydrates: the essential molecules of life*. 2010: Elsevier.

55. *Withers, S. "Anomeric centre (alpha and beta)" in CAZypedia, available at URL https://www.cazypedia.org/index.php/Anomeric_centre_(alpha_and_beta), accessed 10 June 2016. .*

56. Bertozzi, C.R. and D. Rabuka, *Structural basis of glycan diversity.* 2009.

57. Pigman, W. and E. Anet, *4. MUTAROTATIONS AND ACTIONS OF ACIDS AND BASES.* THE CARBOHYDRATES 2E V1A: Chemistry and Biochemistry, 2012: p. 165.

58. Lombard, V., et al., *The carbohydrate-active enzymes database (CAZy) in 2013.* Nucleic Acids Res, 2014. **42**(Database issue): p. D490-5.

59. Eckert, K. and E. Schneider, *A thermoacidophilic endoglucanase (CelB) from Alicyclobacillus acidocaldarius displays high sequence similarity to arabinofuranosidases belonging to family 51 of glycoside hydrolases.* European Journal of Biochemistry, 2003. **270**(17): p. 3593-3602.

60. Brian Rempel, S.W. *"Glycoside Hydrolase Family 39" in CAZypedia, available at URL https://www.cazypedia.org/index.php/Glycoside_Hydrolase_Family_39, accessed 10 June 2016*.

61. Zhao, Z., et al., *Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi.* BMC Genomics, 2013. **14**: p. 274.

62. Davies, G. and B. Henrissat, *Structures and mechanisms of glycosyl hydrolases.* Structure, 1995. **3**(9): p. 853-9.

63. Lairson, L., et al., *Glycosyltransferases: structures, functions, and mechanisms.* Biochemistry, 2008. **77**(1): p. 521.

64. Yip, V.L. and S.G. Withers, *Breakdown of oligosaccharides by the process of elimination.* Curr Opin Chem Biol, 2006. **10**(2): p. 147-55.

65. Levasseur, A., et al., *Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes.* Biotechnology for biofuels, 2013. **6**(1): p. 1.

66. Boraston, A.B., et al., *Carbohydrate-binding modules: fine-tuning polysaccharide recognition.* Biochem J, 2004. **382**(Pt 3): p. 769-81.

67. McCarter, J.D. and S.G. Withers, *Mechanisms of enzymatic glycoside hydrolysis.* Curr Opin Struct Biol, 1994. **4**(6): p. 885-92.

68. Koshland, D.E., *Stereochemistry and the mechanism of enzymatic reactions.* Biological Reviews, 1953. **28**(4): p. 416-436.

69. Rye, C.S. and S.G. Withers, *Glycosidase mechanisms.* Curr Opin Chem Biol, 2000. **4**(5): p. 573-80.

70. Cartmell, A., et al., *The structure and function of an arabinan-specific α-1, 2-arabinofuranosidase identified from screening the activities of bacterial GH43 glycoside hydrolases.* Journal of Biological Chemistry, 2011. **286**(17): p. 15483-15495.

71. *Withers, S. and Williams, S. "Glycoside hydrolases" in CAZypedia, available at URL https://www.cazypedia.org/index.php/Glycoside_hydrolases , accessed 10 June 2016.*

72. Morrison, J.M., M.S. Elshahed, and N. Youssef, *A multifunctional GH39 glycoside hydrolase from the anaerobic gut fungus Orpinomyces sp. strain C1A.* PeerJ Preprints, 2016. **4**: p. e2076v1.

73. Youssef, N.H., et al., *The genome of the anaerobic fungus Orpinomyces sp. strain C1A reveals the unique evolutionary history of a remarkable plant biomass degrader.* Appl Environ Microbiol, 2013. **79**(15): p. 4620-34.

74. Joblin, K., et al., *Ruminal fungi for increasing forage intake and animal productivity.* Sustainable improvement nations, rome of animal production and health, 2010: p. 129-36.

75. Wilson, C.A. and T.M. Wood, *The anaerobic fungus Neocallimastix frontalis: isolation and properties of a cellulosome-type enzyme fraction with the capacity to solubilize hydrogen-bond-ordered cellulose.* Applied microbiology and biotechnology, 1992. **37**(1): p. 125-129.

76. Franzosa, E.A., et al., *Relating the metatranscriptome and metagenome of the human gut.* Proc Natl Acad Sci U S A, 2014. **111**(22): p. E2329-38.

77. Dai, X., et al., *Metatranscriptomic analyses of plant cell wall polysaccharide degradation by microorganisms in the cow rumen.* Applied and environmental microbiology, 2015. **81**(4): p. 1375-1386.

78. Abbott, D.W., et al., *Analysis of the structural and functional diversity of plant cell wall specific family 6 carbohydrate binding modules.* Biochemistry, 2009. **48**(43): p. 10395-10404.

79. Tomme, P., et al., *Studies of the cellulolytic system of Trichoderma reesei QM 9414.* European Journal of Biochemistry, 1988. **170**(3): p. 575-581.

80. Gilbert, H.J., J.P. Knox, and A.B. Boraston, *Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules.* Current opinion in structural biology, 2013. **23**(5): p. 669-677.

81. Hervé, C., et al., *Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects.* Proceedings of the National Academy of Sciences, 2010. **107**(34): p. 15293-15298.

82. Scharpf, M., et al., *Site-specific characterization of the association of xylooligosaccharides with the CBM13 lectin-like xylan binding domain from Streptomyces lividans xylanase 10A by NMR spectroscopy.* Biochemistry, 2002. **41**(13): p. 4255-63.

83. Yin, Y., et al., *dbCAN: a web resource for automated carbohydrate-active enzyme annotation.* Nucleic Acids Res, 2012. **40**(Web Server issue): p. W445-51.

84. Hunter, S., et al., *InterPro: the integrative protein signature database.* Nucleic acids research, 2009. **37**(suppl 1): p. D211-D215.

85. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic Acids Res, 2004. **32**(5): p. 1792-7.

86. Stamatakis, A., *RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies.* Bioinformatics, 2014. **30**(9): p. 1312-3.

87. Darriba, D., et al., *ProtTest 3: fast selection of best-fit models of protein evolution.* Bioinformatics, 2011. **27**(8): p. 1164-1165.

88. Abbott, D.W. and A.B. Boraston, *A family 2 pectate lyase displays a rare fold and transition metal-assisted beta-elimination.* J Biol Chem, 2007. **282**(48): p. 35328-36.

89. Cheung, R.C.F., J.H. Wong, and T.B. Ng, *Immobilized metal ion affinity chromatography: a review on its applications.* Applied microbiology and biotechnology, 2012. **96**(6): p. 1411-1420.

90. Structural Genomics, C., et al., *Protein production and purification.* Nat Methods, 2008. **5**(2): p. 135-46.

91. Gasteiger, E., et al., *Protein identification and analysis tools on the ExPASy server*. 2005: Springer.

92. Kelley, L.A., et al., *The Phyre2 web portal for protein modeling, prediction and analysis.* Nature protocols, 2015. **10**(6): p. 845-858.

93. Paës, G., et al., *The Structure of the Complex between a Branched Pentasaccharide and Thermobacillus xylanilyticus GH-51 Arabinofuranosidase Reveals Xylan-Binding Determinants and Induced Fit†‡.* Biochemistry, 2008. **47**(28): p. 7441-7451.

94. McLean, R., et al., *Functional analyses of resurrected and contemporary enzymes illuminate an evolutionary path for the emergence of exolysis in polysaccharide lyase family 2.* Journal of Biological Chemistry, 2015. **290**(35): p. 21231-21243.

95. Corradini, C., A. Cavazza, and C. Bignardi, *High-performance anion-exchange chromatography coupled with pulsed electrochemical detection as a powerful tool to evaluate carbohydrates of food interest: principles and applications.* International Journal of Carbohydrate Chemistry, 2012. **2012**.

96. Blakeney, A.B., et al., *A simple and rapid preparation of alditol acetates for monosaccharide analysis.* Carbohydrate Research, 1983. **113**(2): p. 291-299.

97. Brunton, N., T. Gormley, and B. Murray, *Use of the alditol acetate derivatisation for the analysis of reducing sugars in potato tubers.* Food chemistry, 2007. **104**(1): p. 398-402.

98. Griggs, L.J., et al., *Identification and quantitation of alditol acetates of neutral and amino sugars from mucins by automated gas-liquid chromatography.* Analytical biochemistry, 1971. **43**(2): p. 369-381.

99. Kim, J.H., et al., *Analysis of neutral sugars by gas-liquid chromatography of alditol acetates: application to thyrotropic hormone and other glycoproteins.* Analytical biochemistry, 1967. **20**(2): p. 258-274.

100. Gao, N., *Fluorophore-assisted carbohydrate electrophoresis: a sensitive and accurate method for the direct analysis of dolichol pyrophosphate-linked oligosaccharides in cell cultures and tissues.* Methods, 2005. **35**(4): p. 323-327.

101. Abbott, D.W., J.M. Eirín-López, and A.B. Boraston, *Insight into ligand diversity and novel biological roles for family 32 carbohydrate-binding modules.* Molecular biology and evolution, 2008. **25**(1): p. 155-167.

102. Correia, M.A., et al., *Signature active site architectures illuminate the molecular basis for ligand specificity in family 35 carbohydrate binding module.* Biochemistry, 2010. **49**(29): p. 6193-6205.

103. Shoseyov, O., Z. Shani, and I. Levy, *Carbohydrate binding modules: biochemical properties and novel applications.* Microbiology and Molecular Biology Reviews, 2006. **70**(2): p. 283-295.

104. Fontes, C.M. and H.J. Gilbert, *Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates.* Annual review of biochemistry, 2010. **79**: p. 655-681.

105. Bayer, E.A., et al., *The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides.* Annu. Rev. Microbiol., 2004. **58**: p. 521-554.

106. *CoMPASS: An automated pipeline for the extraction, trimming, and alignment of modular carbohydrate active enzyme sequences to direct the discovery of new enzyme activities within polyspecific families and meta-datasets. In preparation.*

107. Shallom, D., et al., *Detailed kinetic analysis and identification of the nucleophile in α-L-arabinofuranosidase from Geobacillus stearothermophilus T-6, a family 51 glycoside hydrolase.* Journal of Biological Chemistry, 2002. **277**(46): p. 43667-43673.

108. Fujita, K., et al., *Characterization of a Novel β-l-Arabinofuranosidase in Bifidobacterium longum FUNCTIONAL ELUCIDATION OF A DUF1680 PROTEIN FAMILY MEMBER.* Journal of Biological Chemistry, 2014. **289**(8): p. 5240-5249.

109. Yapo, B.M., *Rhamnogalacturonan-I: a structurally puzzling and functionally versatile polysaccharide from plant cell walls and mucilages.* Polymer Reviews, 2011. **51**(4): p. 391-413.

110. Cuskin, F., et al., *How nature can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic specificity.* Proceedings of the National Academy of Sciences, 2012. **109**(51): p. 20889-20894.

111. Cantarel, B.L., et al., *The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics.* Nucleic acids research, 2009. **37**(suppl 1): p. D233-D238.

112. Fujita, K., et al., *Molecular cloning and characterization of a β-L-arabinobiosidase in Bifidobacterium longum that belongs to a novel glycoside hydrolase family.* Journal of Biological Chemistry, 2011. **286**(7): p. 5143-5150.

113. Shi, H., et al., *Expression and characterization of a GH43 endo-arabinanase from Thermotoga thermarum.* BMC biotechnology, 2014. **14**(1): p. 1.

114. Fujimoto, Z., H. Ichinose, and S. Kaneko, *Crystallization and preliminary crystallographic analysis of exo-α-1, 5-l-arabinofuranosidase from Streptomyces avermitilis NBRC14893.* Acta Crystallographica Section F: Structural Biology and Crystallization Communications, 2008. **64**(11): p. 1007-1009.

115. Kosugi, A., K. Murashima, and R.H. Doi, *Characterization of two noncellulosomal subunits, ArfA and BgaA, from Clostridium cellulovorans that cooperate with the cellulosome in plant cell wall degradation.* Journal of bacteriology, 2002. **184**(24): p. 6859-6865.

116. Sakamoto, T. and J.-F. Thibault, *Exo-arabinanase of Penicillium chrysogenum able to release arabinobiose from α-1, 5-L-arabinan.* Applied and environmental microbiology, 2001. **67**(7): p. 3319-3321.

117. Carapito, R., et al., *Molecular basis of arabinobio-hydrolase activity in phytopathogenic fungi crystal structure and catalytic mechanism of Fusarium graminearum GH93 exo-α-L-arabinanase.* Journal of Biological Chemistry, 2009. **284**(18): p. 12285-12296.

118. Sakamoto, T., et al., *Molecular characterization of a Penicillium chrysogenum exo-1, 5-α-L-arabinanase that is structurally distinct from other arabinan-degrading enzymes.* FEBS letters, 2004. **560**(1-3): p. 199-204.

119. da Silva Delabona, P., et al., *Understanding the cellulolytic system of Trichoderma harzianum P49P11 and enhancing saccharification of pretreated sugarcane bagasse by supplementation with pectinase and α-L-arabinofuranosidase.* Bioresource technology, 2013. **131**: p. 500-507.

120. Gao, D., et al., *Strategy for identification of novel fungal and bacterial glycosyl hydrolase hybrid mixtures that can efficiently saccharify pretreated lignocellulosic biomass.* BioEnergy Research, 2010. **3**(1): p. 67-81.

121. Santos, C.R., et al., *Mechanistic strategies for catalysis adopted by evolutionary distinct family 43 arabinanases.* Journal of Biological Chemistry, 2014. **289**(11): p. 7362-7373.

122. Vinkx, C. and J. Delcour, *Rye (Secale cerealeL.) Arabinoxylans: A Critical Review.* Journal of Cereal Science, 1996. **24**(1): p. 1-14.

123. Vinkx, C., et al., *Physicochemical and functional properties of rye nonstarch polysaccharides. V. Variability in the structure of water-soluble arabinoxylans.* Cereal Chemistry, 1993. **70**: p. 311-311.

124. Collins, T., C. Gerday, and G. Feller, *Xylanases, xylanase families and extremophilic xylanases.* FEMS microbiology reviews, 2005. **29**(1): p. 3-23.

125. *Rye Arabinoxylan from Megazyme*. Available from: https://secure.megazyme.com/Arabinoxylan-Rye-Flour.

126. Martens, E.C., et al., *Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts.* PLoS Biol, 2011. **9**(12): p. e1001221.