Discovery and Biochemical Characterization of New CAZymes from PULs and Metagenome Sequences

by

Fakhria Muhammad Razeq

A thesis submitted in conformity with the requirements for the degree of
Master of Applied Science
Department of Chemical Engineering and Applied Chemistry
University of Toronto

© Copyright by Fakhria Muhammad Razeq 2017
Discovery and Biochemical Characterization of New CAZymes from PULs and Metagenome Sequences

Fakhria Muhammad Razeq
Master of Applied Science
Department of Chemical Engineering and Applied Chemistry
University of Toronto
2017

Abstract

(methyl)-Glucuronoxylan, a main hemicellulose in hardwood trees, has a broad range of potential applications, ranging from biofuels to bio-based coatings and packaging films. It is substituted with (methyl)glucuronic ((Me)GlcpA) and acetic acids, which affect polymer solubility, rheology, interaction with other polymers, and accessibility of the polymer to enzymes. To date, there are no known esterases that are able to deacetylate internal xylose residues that are both 3-O-acetylated and 2-O substituted with (Me)GlcpA (Xyl-3-Ac-2MeGlcpA). In this study a protein of unknown function, which was named FjoAcXE, was produced and characterized to be an acetyl xylan esterase capable of deacetylating xylose residues at all positions, including Xyl-3-Ac-2MeGlcpA. Additionally, FjoAcXE was shown to display synergism with α-glucuronidases from glycoside hydrolase (GH) 67 and GH115 families. Synergistic action between FjoAcXE with α-glucuronidases was further tested with three recombinantly produced α-glucuronidases from metagenomic sequences of the Canadian beaver (Castor Canadensis) droppings enriched on poplar hydrolysate.
Acknowledgments

I would like to thank my supervisor, Prof. Emma Master for all her support and patient guidance. Emma, thank you for being an amazing mentor and always encouraging me to achieve more. I really appreciate that you give your students intellectual freedom to develop as independent researchers, while also providing an incredible amount of support every step of the way.

Thank you to Dr. Weijun Wang and Dr. Thu Vuong for all your help and always being available to answer many of my questions, I have learned a lot from both of you.

Also, I would like to thank all members of the Master lab for providing a fun and friendly environment for learning and research. A special thanks to Mabel T. Wong, Dr. Ruoyu Yan, Dr. Edita Jurak, and Dr. Marie Couturier for all their helpful discussions, ideas and willingness to share their knowledge and expertise. Also, thank you to our lab manager, Susie Susilawati, for always making sure students in Biozone have everything they need to do their research. Thank you to Weijun Gao and Dean Robson for their help with sequence download from CAZy database.

Finally, thank you to my family for all their support and always believing in me. I could not have done this without you.
Table of Content

A note on saccharide nomenclature ........................................................................ vi
List of Abbreviations ............................................................................................... vii
List of Tables ............................................................................................................ viii
List of Figures ........................................................................................................... viii
List of Appendices .................................................................................................. ix

1. Research Hypothesis and Objectives .................................................................. 2
   Chapter II: FjoAcXE ............................................................................................. 2
   Chapter III: METAGH115 .................................................................................... 3

2. Chapter I: Literature review .................................................................................... 4
   2.1. Lignocellulosic biomass .................................................................................. 4
   2.2. Hemicelluloses ............................................................................................... 5
       2.2.1. Xylan .................................................................................................. 6
       2.2.1.1. Structure ......................................................................................... 6
       2.2.1.2. Applications ................................................................................... 8
   2.3. Enzymatic treatment as a tool for hemicellulose recovery and broader use ...... 9
       2.3.1. Carbohydrate active enzymes (CAZymes) ............................................. 9
       2.3.2. Xylan active carbohydrate esterases ...................................................... 10
       2.3.3. α-glucuronidases from GH115 family .................................................. 12
   2.4. PUL database: a source for discovery of new and novel CAZymes ............... 16

3. Chapter II: Production and characterization of acetyl xylan esterase FjoAcXE .... 19
   3.1. Introduction ................................................................................................... 19
   3.2. Materials and Methods ................................................................................ 19
       3.2.1. Materials ............................................................................................. 19
       3.2.2. Candidate selection ............................................................................ 21
       3.2.3. Gene synthesis and molecular cloning .................................................. 21
       3.2.4. Protein expression and purification ....................................................... 22
       3.2.5. In-gel trypsin digest and mass spectrometry ........................................ 23
       3.2.6. Enzyme assays ..................................................................................... 24
       3.2.7. Enzyme kinetics ................................................................................... 27
       3.2.8. Synergism of FjoAcXE with α-glucuronidases ..................................... 27
       3.2.9. 1H-NMR analysis of FjoAcXE towards acetylated xylo-oligosaccharides 28
       3.2.10. Predicted structural model of FjoAcXE ................................................ 28
   3.3. Results and Discussion ................................................................................... 29
       3.3.1. Candidate selection and recombinant protein production .................... 29
       3.3.2. Activity screen ..................................................................................... 31
       3.3.3. Activity of FjoAcXE on synthetic and natural substrates ....................... 34
       3.3.4. Effect of pH, temperature, metal ions, and chemical reagents .............. 35
       3.3.5. Enzyme kinetics ................................................................................... 37
       3.3.6. FjoAcXE’s synergistic action with α-glucuronidase ............................... 38
       3.3.7. 1H-NMR analysis of FjoAcXE towards acetylated xylo-oligosaccharides 40
       3.3.8. Sequence analysis and structural model of FjoAcXE ............................... 41
4. Chapter 2: Production and characterization of diverse GH115 α-glucuronidases from metagenome sequences

4.1. Introduction

4.2. Materials and Methods

4.2.1. Materials

4.2.2. Candidate selection and phylogenetics

4.2.3. Gene synthesis and molecular cloning

4.2.4. Protein expression and purification

4.2.5. In-gel trypsin digest and mass spectrometry

4.2.6. Activity towards aaldouronic acid mixture

4.2.7. Activity towards pNP substrates

4.2.8. Hydrolysis of polymeric substrates

4.2.9. Optimum pH screen for META1, 2, and 4

4.2.10. Activity of META1 and META2 on Udnm2XX and XUdnm2 XX

4.2.11. Synergism between METAs and FjoAcXE

4.2.12. 1H-NMR analysis towards acetylated saccharides

4.3. Results and discussion

4.3.1. Candidate selection, phylogenetics and sequence analysis

4.3.2. Protein expression and purification

4.3.3. META1 and 2 show α-glucuronidase activity, while META4 shows an esterase activity

4.3.4. Activity screen on pNP substrates showed META4 is an esterase

4.3.5. Hydrolysis of polymeric substrates

4.3.6. Optimum pH screen for META1, 2, and 4

4.3.7. META1 and META2 are active on Udnm2XX and XUdnm2 XX

4.3.8. Synergism between METAs and FjoAcXE and confirmation by 1H-NMR analysis

5. Engineering significance

6. Conclusion and future considerations

References

Appendices
A note on saccharide nomenclature

- The term oligosaccharide refers to 3 to 9 covalently linked monosaccharide units, and a polysaccharide is more than 9 units long.
- Oligosaccharides in this document are referred to by their sequences: $U^{4m2}XX$ and $XU^{4m2}XX$ represent aldotetraouronic and aldopentaouronic acids, respectively, where X represents an unsubstituted xylose residue and U represents a xylose residue that is 2-0 substituted with (4-0-methyl)glucuronic acid (as defined by 4m2 superscript).
- Side groups and linkages in sugars (ex. $\beta$-D-(1→4)-Xylp or $\alpha$-L-(1→4)-Araf) are defined as follows: the first letter is anomeric configuration ($\alpha$ or $\beta$), followed by absolute configuration (D or L), then glycosidic linkage type (1→2 or 1→3 or 1→4), then the residue (ex. xylose (Xyl) or arabinose (Ara)), and then the ring form (pyranose (p) or furanose (f)).
- Brackets around 4-0-(methyl)glucuronic acid ((Me)GlcP A) indicate that $\alpha$-1→2-glucuronic acid may or may not be O-methylated at the position 4.
- Xylopyranose (Xylp) residues may be single or di-acetylated at O-2 and/or O-3 positions or 3-O-acetylated and 2-O substituted with (Me)GlcP A and in this document the following designations are used:
  - Xyl-2Ac: 2-O-acetylated Xylp
  - Xyl-3Ac: 2-O-acetylated Xylp
  - Xyl-2,3Ac: 2,3-di-O-acetylated Xylp
  - Xyl-3-Ac-2MeGlcP A: 3-O-acetylated Xylp 2-O substituted with (Me)GlcP A
- Chemical structures of most polysaccharides used in this study are shown in Appendix G.
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-MUA</td>
<td>4-methylumbelliferyl acetate</td>
</tr>
<tr>
<td>4-MU</td>
<td>4-methylumbelliferone</td>
</tr>
<tr>
<td>AcE</td>
<td>Acetyl esterase</td>
</tr>
<tr>
<td>AcXE</td>
<td>Acetyl xylan esterase</td>
</tr>
<tr>
<td>AG</td>
<td>Arabinogalactan</td>
</tr>
<tr>
<td>AGE</td>
<td>Affinity gel electrophoresis</td>
</tr>
<tr>
<td>AGX</td>
<td>Arabinoglucuronoxylan</td>
</tr>
<tr>
<td>Araβ</td>
<td>α-L-arabinofuranosyl</td>
</tr>
<tr>
<td>BEX</td>
<td>Beechwood xylan</td>
</tr>
<tr>
<td>CMC</td>
<td>Carboxymethyl cellulose</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>GAX</td>
<td>Glucuronoxylan</td>
</tr>
<tr>
<td>GX</td>
<td>Glucuronoxylan</td>
</tr>
<tr>
<td>EC</td>
<td>Enzyme commission</td>
</tr>
<tr>
<td>ECF-σ/anti-σ</td>
<td>extracytoplasmic function sigma/anti-sigma factor pairs</td>
</tr>
<tr>
<td>HTCS</td>
<td>Hybrid two-component system</td>
</tr>
<tr>
<td>LCCs</td>
<td>Lignin carbohydrate complexes</td>
</tr>
<tr>
<td>(Me)GlcPβA</td>
<td>collective term for α-1→2 glucuronic and 4-O-methyl α-1→2 glucuronic acids</td>
</tr>
<tr>
<td>MGX</td>
<td>4-O-Methyl glucuronoxylan</td>
</tr>
<tr>
<td>pNP</td>
<td>p-nitrophenol</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PUL</td>
<td>Polysaccharide Utilization Loci</td>
</tr>
<tr>
<td>OSX</td>
<td>Oat spelt xylan</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulfate</td>
</tr>
<tr>
<td>Sus</td>
<td>Starch utilization system</td>
</tr>
<tr>
<td>XG</td>
<td>Xyloglucan</td>
</tr>
<tr>
<td>XOS</td>
<td>Xylo-oligosaccharides</td>
</tr>
<tr>
<td>Xylβ</td>
<td>Xylopyranose</td>
</tr>
</tbody>
</table>
List of Tables

Table 1: List of characterized GH115 (as of April 2017) and their reported properties. 14
Table 2: Kinetic parameters of FjoAcXE were tested on 4-MUA and pNP-acetate. Data were fitted to the Michaelis-Menten equation. 38
Table 3: List of FjoAcXE homologs. In bold are selected candidates for future characterization. 42

List of Figures

Figure 1: Schematic diagram of a plant fiber. 4
Figure 2: Structural representation of various types of xylans depending on their botanical source. 7
Figure 3: Xylan active CAZymes. 9
Figure 4: Catalytic mechanism of a typical serine-type acetyl esterase (adapted from Biely, 2012). 11
Figure 5: Bacteroides thetaiomicron starch utilization system (Sus) showing (A) gene organization, and (B) a model of concerted action of Sus in starch degradation. 17
Figure 6: PUL20 Flavobacterium johnsoniae showing possible functions of predicted CAZymes (as indicated by BLAST searches) 30
Figure 7: Purified FjoAcXE is approximately 45.2 kDa. 30
Figure 8: PAHBAH reducing-end assay was used to test for hydrolytic activity of FjoAcXE on 0.5% (w/v) polysaccharides. 31
Figure 9: Affinity gel electrophoresis. 32
Figure 10: A general screen of FjoAcXE on pNP substrates showed that its active on pNP-acetate. 33
Figure 11: A screen of pNP alkyl esters showed that FjoAcXE is active on short chain alkyl esters (<C4) suggesting that it is likely an esterase and not a lipase. 34
Figure 12: Activity of FjoAcXE on synthetic and natural substrates. 35
Figure 13: General biochemical properties of FjoAcXE. A) Optimum pH and B) pH stability. The samples were tested using 100mM Tris, 50 mM MES, 50mM acetic acid and 50mM sodium acetate trihydrate universal buffer with pH range of 3.5 to 9.5. C) Optimum temperature and D) thermal stability. E) Effects of metal ions and F) effect of chemical reagents on activity on the activity of FjoAcXE. 36
Figure 14: Measurement of glucuronic and acetic acids released when tested on 1% Ac-XOS. 39
Figure 15: 1H-NMR spectrum of FjoAcXE tested on 1% (w/v) acetylated xylo-oligosaccharides. 40
Figure 16: FjoAcXE (Fjoh_3879) has GDSL-like motif, conserved SGNH residues (red circles) and catalytic Ser-His-Aps triad (green triangle). 41
Figure 17: Structural model of FjoAcXE based on c4xhA, c2o14A, d1snra2 and c2w9xA as templates. 43
Figure 18: Phylogenetic tree for GH115 family. 51
Figure 19: Alignment of META candidates with all characterized GH115s showing GH115 domain. 52
Figure 20: Recombinantly expressed META1, 2 and 4 were pure and of expected size 53
Figure 21: META1 and META2 are active on aldouronic acid mixture, while META4 showed no activity. ................................................................. 54
Figure 22: pNP screen showed META4 has activity on pNP-acetate and pNP-hexanoate, and META1 and 2 do not show activity on any of the tested substrates.............. 55
Figure 23: Activity of META1, META2 and META4 on 1% (w/v) beechwood xylan (AGX), oat spelt xylan (OSX), spruce arabinoglucuronoxylan (AGX) and acacia gum arabinogalactan (AG).................................................................................. 57
Figure 24: META1 and 2 have optimum pH of 6.0, while META4 has optimum pH of 8.0........................................................................................................ 58
Figure 25: META1 and META2 are active on both terminally (U^4m2 XX) and internally (XU^4m2 XX) substituted xylo-oligosaccharides. ............................................. 59
Figure 26: Synergism between FjoAcXE and α-glucuronidases on 1% (w/v) of hot water extracted mixed hardwood samples .................................................................................. 60
Figure 27: ^1H-NMR spectrum of FjoAcXE tested on 1% (w/v) hot water extracted mixed hardwood...................................................................................... 62

List of Appendices

Appendix A: Experimentally validated PULs containing at least two of the CAZyme families of interest (as of April 2015) .............................................................................. 73
Appendix B: META2 and META3 (in red) are part of PUL117 and PUL97, respectively, from beaver droppings microcosms enriched on poplar hydrolysate (BD-PH). ...... 73
Appendix C: Production and characterization of unks from PUL73 ........................................ 74
Appendix D: Predicted secondary structure of FjoAcXE using the PSIPRED protein sequence analysis workbench (McGuffin, Bryson, and Jones 2000) ............... 77
Appendix E: Summary and sequences for the candidates selected from PULDB ............. 78
Appendix F: Summary and sequences for the candidates selected from metagenomes . 79
Appendix G: Polysaccharide structures.................................................................................. 80
Introduction

The inevitable decline in oil reserves, combined with concerns regarding climate change have contributed to the growing interest in renewable and sustainable resources, which could replace petroleum-based products. It has been estimated that land plants produce $1.7-2.0 \times 10^{11}$ tonnes of biomass worldwide each year, 70% of which comprises plant cell walls (i.e. lignocellulosic biomass); only 2% of which is utilized for production of biomaterials and biofuels (Pauly and Keegstra 2008). Lignocellulosic biomass consists of cellulose, hemicellulose, and lignin, and represents a sustainable source for production of fuels, polymeric materials and chemical feedstocks. In addition to their abundance, lignocellulosic biomass are non-edible, and their utilization will not take away from starch-based feedstocks that could be used for animal or human consumption. While a lot of research effort in the past couple of decades has been put towards bioconversion of lignocellulose to biofuels (ex. bioethanol) (Valdivia et al. 2016), there is critical need for new biotechnologies that would help facilitate production and synthesis of high-value chemicals and biomaterials from forestry and agricultural residues.

Hemicelluloses represent the second main polysaccharide in lignocellulosic biomass after cellulose. Reported applications of hemicelluloses range from rheology modifiers and packaging films, to hydrogels as well as nutrient additives to food and feed (Ebringerová 2006). Despite their broad applied potential, hemicelluloses remain relatively underutilized, in part due to the diversity and heterogeneity of hemicellulose structures.

Glucuronoxylan, as a predominant hemicellulose, is the main component of secondary cell wall in hardwood. Depending on botanical source and extraction method, xylans may be acetylated and have various side groups (ex. (methyl)glucuronic acid). These side groups affect the several properties of the polymer, such as water solubility, rheology, association with other biopolymers, and nutrient value in food and health products. In order to produce biopolymers with reproducible properties, it is therefore necessary to have tools to selectively tune hemicellulose side group chemistry. One way that can be done is through enzymatic modification using Carbohydrate Active Enzymes (CAZymes). Structural diversity of hemicellulose requires concerted action of many different CAZymes to work together to modify the polymer. Polysaccharide utilization loci (PULs) are set of physically linked and functionally related genes that encode for CAZymes that work together to modify and degrade a specific polysaccharide. While some CAZymes found on PULs have a predicted or known function,
there are other open reading frames that are predicted to encode proteins with no known function; such proteins with unknown function are often called “unks”.

The main objective of this study is to identify and characterize uncharted CAZymes from PULs and metagenomes that can be used for site-selective modification of plant polymers, allowing us to expand their utility and improve their performance in biomaterials, and reduce our dependency on petroleum-based products. Through bioinformatics analysis, an unk that was later named FjoAcXE, was chosen for expression and biochemical characterization, and was shown to be an acetyl xylan esterase. There are currently no known esterases that are able to access 3-O-acetylated Xylp that is also substituted at the 2-O position with (Me)GlcpA (i.e., Xyl-3-Ac-2MeGlcpA residues), presumably due to the steric hindrance. Recently, AnCE16 from Aspergillus niger was shown to deacetylate Xyl-3-Ac-2MeGlcpA residue at non-reducing end of acetylated xylo-oligosaccharides (Neumüller et al. 2015). However, given that acetyl groups at position 3 on non-reducing end tend to migrate to position 4 (Vladimír Puchart and Biely 2014), it is possible that AnCE16 does not truly deacetylate xylose residues that also have (Me)GlcpA at the vicinal position. In this study, FjoAcXE was shown to deacetylate Xyl-3-Ac-2MeGlcpA residues, and by doing so allowed α-glucuronidases from glycoside hydrolase families GH67 and GH115 to access previously inaccessible (Me)GlcpA. To our knowledge, this is the first report of an esterase that is able to substantially deacetylate internal Xyl-3-Ac-2MeGlcpA residues of xylan substrates. Details of the biochemical characterization of FjoAcXE and its synergism with GH115 are presented in Chapter II of this thesis. Action and substrate specificity of GH115 family was further explored with and without FjoAcXE by selecting four GH115 candidates from metagenomic sequences derived from enrichments of the digestive system of Canadian beaver (Castor Canadensis). Corresponding results are presented in Chapter III.

1. Research Hypothesis and Objectives

Chapter II: FjoAcXE

Hypothesis

- Proteins of unknown function from polysaccharide utilization loci present new and potentially novel classes of CAZymes
• Unknown proteins from PULs containing mostly xylan active enzyme will be involved in xylan modification/degradation

Objectives

• Select unknown protein candidates from PULs containing xylan active CAZymes of interest
• Produce and biochemically characterize selected candidates
• Assess synergistic interaction between the candidate and another CAZyme family found on the same PUL (ex. GH115)
• Assign function and predict homologs of the selected unknown candidate(s)

Chapter III: METAGH115

Hypothesis

• α-glucuronidases from family GH115 with novel activities, such as ability to act on xylose residue 3- O-acetylated and 2- O substituted with (4- O-methyl)glucuronic acid, will be found on unique clades of the GH115 phylogenetic tree containing no characterized members
• Metagenomic sequences from beaver droppings enriched on simple (cellulose) and complex substrate (poplar hydrolysate) will encode distinct GH115 α-glucuronidase activities

Objectives

• Select predicted GH115 α-glucuronidases from metagenomic sequences
• Produce and biochemically characterize selected candidates
• Assess synergistic interaction with FjoAcXE
• Test polyspecificity of GH115 family by looking at their ability to act on different types of substrates
2. Chapter I: Literature review

2.1. Lignocellulosic biomass

Plant cell walls play a crucial role in plant development, provide structural and mechanical support for plant tissues, and protect plants against biotic and abiotic stress. Cell walls are composed of the middle lamella, primary and secondary cell walls (Figure 1), with primary cell wall being ubiquitous to all cells, whereas secondary cell wall is found in certain types of tissue, including wood (Chundawat et al. 2011). The main polysaccharide components of cell walls are cellulose (35-50%), hemicellulose (15-35%) and lignin (10-35%), which together represent a bulk of lignocellulosic biomass (Chen 2014). Pectin is another major polysaccharide component of the cell wall, mostly found in the middle lamella and primary cell wall, and plays important role in cell adhesion and cell wall hydration. Composition and role of pectin is reviewed elsewhere (Xiao and Anderson 2013), and not discussed here further. Other minor components of cell walls include structural proteins (3-10%), lipids (1-5%), soluble sugars (1-10%) and minerals (5-10%) (Chundawat et al. 2011).

![Figure 1: Schematic diagram of a plant fiber.](image)

Cellulose is one of the main structural components of the cell wall and consists of long linear polymer of β-1→4-linked-D-glucopyranose units that can reach several thousand glucose monomers in length (Limayem and Ricke 2012). The absence of side chains allows extensive hydrogen bonding and interactions through van der Waals forces between glucose chains, allowing formation of microfibrils (Guerriero et al. 2016). Hemicellulose is a β-1→4-linked heteropolymer of pentose and hexose sugars, and may also contain uronic acids (Limayem and
Ricke 2012). Hemicellulose is discussed in more detail in the next section. Finally, lignin is the third major component of lignocellulose, and is composed of cross-linked phenyl-propanoid subunits, leading to p-hydroxyphenyl (H), guaicyl (G), and syringyl (S) lignin sub-structures (Christopher, Yao, and Ji 2014). Lignin moieties are thought to form covalent bonds with side chains of branched hemicelluloses, forming lignin carbohydrate complexes (LCCs) (Chundawat et al. 2011). The current models of cell walls suggest that cellulose microfibrils are surrounded by a hemicellulose and lignin matrix, with hemicellulose forming hydrogen bonds with the surface of cellulose microfibrils (Somerville et al. 2004).

Lignocellulosic biomass composition varies depending on the plant source of the feedstock, which can be divided into the following categories: 1) agricultural residue from crops or non-edible agricultural products, 2) forestry residues from logging, 3) fast-growing grasses, 4) fast-growing trees (ex. hardwoods and softwoods), 5) municipal plant-derived waste (Amarasekara 2014). The recalcitrant nature of lignocellulose, combined with variation in its composition within and between plant species, makes utilization challenging. However, recent progress in enzymatic modification of lignocellulosic biomass, combined with its sustainability, abundance and relatively low cost, makes lignocellulosic biomass a valuable source of biopolymers and chemical feedstocks.

2.2. Hemicelluloses

Hemicelluloses are non-cellulose heteropolymers of pentose and/or hexose sugars characterized by the presence of equatorial β-(1→4)-linked backbone and with average degree of polymerization of 80-200 units (Scheller and Ulvskov 2010; Peng et al. 2012). Based on their composition and structure, hemicelluloses can be grouped into four general categories: xylans, mannans and glucomannans, mixed linkage β-glucans, and xyloglucans (Ebringerová, Hromádková, and Heinze 2005; Scheller and Ulvskov 2010). The presence of extensive amounts of side groups in hemicellulose prevents formation of extensive crystalline structures; hemicelluloses may also bridge neighboring cellulose fibrils thereby promoting flexibility of plant cell walls during cell expansion (Marriott, Gómez, and McQueen-Mason 2015).

Hemicellulose composition and branching frequency varies depending on the plant species, cell type and stages of the plant development. In dicots (hardwoods and herbaceous plants) xyloglucan is a major component of primary cell walls (20-25%), and (methyl)glucuronoxylan (MGX) is a major component of secondary cell wall (20-30%) (Scheller and Ulvskov 2010). In grasses, glucuronoroarabinoxylan represents 20-40% of
polysaccharides in primary cell wall and up to 50% in secondary cell wall, while a minor component of β-(1→3, 1→4)-glucan is only found in primary cell walls of grasses (2-15%) (Scheller and Ulvskov 2010; Peng et al. 2012). Conifers contain mostly xyloglucan (10%) in their primary cell wall, and glucuronoarabinoxylan (5-15%) and galactoglucomannan (10-30%) in their secondary cell walls (Scheller and Ulvskov 2010).

2.2.1. Xylan

2.2.1.1. Structure

Xylan is β-(1→4)-linked polymer of β-xylopyranose units, which are usually partially substituted with specific sugars and O-acetyl groups depending on the botanical source. Xylans account for 30% of cell wall material in annual plants, 15-30% in hardwoods and 7-10% in softwoods (Walia et al. 2017). Xylans can be found as homopolymers with β-1→3 or mixed β-1→3/1→4-glycosidic linkages in red seaweed and green algae or heteroxylans with β-1→4 in higher plants (Ebringerová and Heinze 2000). Heteroxylans include glucuronoxylans (GX), arabinoglucuronoxylan (AGX), glucuronoarabinoxylan (GAX), arabinoxylan (AX), and complex heteroxylan (Ebringerova, Hromadkova, and Heinze 2005). As Figure 2 shows, xylan backbone may be mono-substituted at O-2 or O-3 position or di-substituted at O-2 and O-3 positions of xylopyranosyl residues with acetyl groups and/or α-L-arabinofuranosyl (Araf, which may be further substituted with ferulic and p-coumaric acids), and/or mono-substituted at O-2 with (4-O-methyl)glucuronic acid ((Me)GlcP) (Scheller and Ulvskov 2010).
Figure 2: Structural representation of various types of xylans depending on their botanical source.

Harwood: Glucuronoxylan (GX)

Softwood: Arabinoglucuronoxylan (AGX)

Grasses and cereals: Glucuronoarabinoxylan (GAX)

Cereal grains: Arabinoxylan (AX)
2.2.1.2. Applications

Xylitol, a low-caloric sweetener, and bioethanol are currently two of the primary xylan based products. However, their production requires complex and energy intensive processes that must begin by breaking down the polymer into monosaccharides. Recent research interest has been focused on production of high-value bioproducts from xylans that would take advantage of innate physicochemical properties of the polymer and/or oligomers, such as low oxygen permeability and ability to adhere to cellulose fibres. Xylan polymer is promising resource for production of surfactants, foam, gels, packaging films and coatings, while xylo-oligosaccharides (XOS) can be used in food and health products (prebiotics) (Hansen and Plackett 2008; Deutschmann and Dekker 2012; Aachary and Prapulla 2011).

The pattern of side groups linked to the xylan backbone affects xylan solubility, interaction with other polymers, accessibility of the polymer to enzymes, solution behavior, and other polymer properties, such as tensile strength, crystallinity and oxygen permeability (Ebringerová and Heinze 2000; Heikkinen et al. 2013). Packaging films produced from unmodified xylan are usually brittle but addition of other polymeric materials, such as chitosan or cellulose derived products (ex. sorbitol), can improve and enhance the final structure (Deutschmann and Dekker 2012). Moreover, xylans can be chemically modified through esterification and etherification to introduce hydrophobic moieties (Littunen et al. 2017). Additional modifications, such as acetylation can also improve polymer properties. For example, xylan acetate showed comparable or better material properties compared to other major bio-plastics, such as cellulose acetate and poly(lactic acid) in terms of strength, Young’s modulus, elongation to break, moisture uptake in water or at 97% room humidity, and transparency (Stepan 2013). Recently, enzymatically debranched wheat arabinoxylan was grafted to glycidyl methacrylate, creating a copolymer that could be absorbed to cellulose (Littunen et al. 2017). Therefore, ability to introduce specific modifications to the xylan backbone and/or side groups is important for both the recovery (e.g., precipitation from water suspensions) and production of high-value bioproducts that could compete with or replace petroleum-based products.
2.3. Enzymatic treatment as a tool for hemicellulose recovery and broader use

2.3.1. Carbohydrate active enzymes (CAZymes)

Carbohydrate-active enzymes (CAZymes) are involved in assemblage, breakdown and modification of carbohydrates. Traditional enzyme classification based on substrate specificity using enzyme commission (EC) number presents a challenge for enzymes that show broad specificities and it does not account for structural features of CAZymes. In 1999 a curated CAZyme classification database (http://www.cazy.org) was established that classified CAZymes based on their sequence similarity and presence of common structural folds (Lombard et al. 2014). Currently, CAZymes are classified into 5 enzyme classes and one associated module: glycoside hydrolase (GH), glycosyl transferase (GT), polysaccharide lyase (PL), carbohydrate esterase (CE), auxiliary activities (AA), and carbohydrate-binding module (CBM).

![Diagram of CAZymes](image)

**Figure 3: Xylan active CAZymes.**

GH and CE families play important roles in main and side-chain modification of xylan polymers (Figure 3). For example, endo-xylanases (EC 3.2.1.8) from families GH8, 10 and 11 are involved in hydrolysis of internal β-1→4-xylose bond of xylan backbone (Peter Biely, Singh, and Puchart 2016). Others, such as β-xylosidases (EC 3.2.1.37) from families GH43 and GH54 remove D-xylose from non-reducing end of xylans. Side groups, such as Araf can be removed by α-arabinofuranosidases (EC 3.2.1.55) from GH43 and GH62 families (Saha 2000), (Me)GlcP A can be removed by α-glucuronidase.
from GH67 and GH115 families (Wang et al. 2016), and acetyl groups can be removed by acetyl xylan esterases (EC 3.1.1.72) from CE1-7, 12 and 16 families (Adesioye et al. 2016).

### 2.3.2. Xylan active carbohydrate esterases

Xylans are often $O$-acetylated at position 2 and/or 3 of the xylose residue but acetylation at other positions are possible depending on species, and spontaneous migration of acetyl groups to the neighboring free hydroxyls have been observed (Teleman et al. 2000). Xylan acetylation is more common in hardwood and grasses than in softwood. For example, every 7 out of 10 xylose units of 4-O-methyl-D-glucuronoxylan are acetylated (Alalouf et al. 2011). Alkaline pretreatment of xylan removes acetyl groups but if non-alkaline pretreatment methods, such as steam explosion, are preferred, then acetyl groups must be removed enzymatically because acetylation presents a steric hindrance for glycoside hydrolases, which are necessary when the objective is to produce mono- or oligosaccharides (Martínez-Martínez et al. 2008). Additionally, no acetylation or complete acetylation reduces solubility in polar solvents, while partial acetylation increases water solubility (Pawar et al. 2013). The presence of acetyl groups on xylan also has been shown to decrease its absorption to cellulose, which is an important consideration for production of biocomposites (Kabel et al. 2007).

Carbohydrate esterases (CE), more specifically acetyl xylan esterases (AcXEs, EC 3.1.1.72) belonging to families CE1-7, 12 and 16, have been reported to be involved in deacetylation of xylo-oligosaccharides and high molecular weight xylan, where enzymes from CE1, 3, 4, 5 and 6 are able to deacetylate and precipitate xylan (Adesioye et al. 2016; Nakamura, Nascimento, and Polikarpov 2017; Peter Biely et al. 2014). Other CEs are involved de-$O$ or de-$N$-acylation of variety of different substrates, including pectin, rhamnogalacturonan, and/or chitin.

Some CE families, such as CE2, 3, 6 and CE12 contain SGNH domain, which is a derivative of the GDSL family of proteins (Pfam accession PF00657) (Nakamura, Nascimento, and Polikarpov 2017; Molgaard, Kauppinen, and Larsen 2000). The SGNH family is found in both prokaryotes and eukaryotes and displays a wide range of hydrolytic functions such as lipase, carbohydrate esterase, thioesterase, protease,
arylesterase, acyltranferease, and lysophospholipase activities (Asler et al. 2010).

**Figure 4: Catalytic mechanism of a typical serine-type acetyl esterase (adapted from Biely, 2012).**

The SGNH-hydrolase family (Pfam clan CL0264) was first identified as a new family of lipolytic enzymes about 20 years ago, based on the presence of four conserved sequence blocks: S (Ser), G (Gly), N (Asn), and H (His), in blocks I, II, III and V, respectively (Upton 1995). Block IV, upon more comprehensive sequence comparisons, is not conserved as originally was proposed by Upton and Buckley (Upton 1995). The Ser in block I is one of the active sites and part of GDS(L) motif found near N-terminal, which is distinct from the classical GxSxG motif typically found in lipases/esterases near the middle of the protein (Akoh et al. 2004). Gly in block II acts as a proton donor for the oxyanion hole, which is a positively charged pocket that stabilizes negatively charged tetrahedral intermediates during catalysis (Alalouf et al. 2011). The Asn in block III is part of GXND motif and also part of the oxyanion hole (Mølgaard, Kauppinen, and Larsen 2000). The His in block V is part of the DXXHP motif, which together with D (Asp) make up second and third member of catalytic serine-histidine-aspartic acid triad (Mølgaard, Kauppinen, and Larsen 2000). More specifically, the His residue in block V deprotonates the hydroxyl group of Ser in block I, allowing Ser to serves as the nucleophile and a protein donor to the oxyanion hole consisting of Gly and Asn residues in blocks II and III, respectively (Akoh et al. 2004).

A catalytic mechanism of a typical serine-type AcXE can be described as show in Figure 4 (Biely 2012). In the presence of substrate, serine is activated by formation of
Ser-His-Asp catalytic triad, which allows serine to attack the carbonyl group of the substrate and form a tetrahedral species with help of the oxyanion hole (Williams et al. 2014). This is followed by the formation of covalent acyl-enzyme intermediate, while the product is released. The acetate group is then hydrolysed from serine, thereby regenerating serine to its protonated state to act on the next substrate molecule.

All AcXEs, with the exception of CE4, are serine esterases with serine being part of catalytic Ser-His diad or Ser-His-Asp/Glu triad (Biely 2012). CE4 is aspartic esterase and the only metallo-dependent AcXEs, requiring Co$^{2+}$ or Zn$^{2+}$ cations for activity of their His-His-Asp catalytic triad (Adesioye et al. 2016). In terms of regioselectivity, most AcXEs are able to deacetylate both positions 2 and 3, as well as doubly acetylated 2,3 Xylp residues. Some AcXEs may preferentially deacetylate position 2 or 3, while others target both positions simultaneously (Biely et al. 2004). In xylosides and xylo-oligosaccharides, acetyl groups can migrate between positions 2 and 3, and in the case of non-reducing Xylp residue, position 4 can also be acetylated (Busse-Wicher et al. 2016; Busse-Wicher et al. 2014; Peter Biely, Singh, and Puchart 2016). This phenomena makes positional specificity studies challenging, but using monoacetyl derivatives prepared in absence of water and short reaction time can prevent acetyl migration, allowing assignment of positional specificity for the enzymes (Biely et al. 2004; Biely 2012).

In glucuronoxylan, the 3-O acetylated Xylp may also be 2-O substituted with (Me)Glc$p$A. Recently, a CE16 has been shown to act on 3-O acetylated Xylp 2-O substituted with (Me)Glc$p$A at the non-reducing end; however, the rate of deacetylation was slow (Neumüller et al. 2015). There are currently no known esterases that are able to remove 3-O acetyl group from internal Xylp residue that is also 2-O substituted with (Me)Glc$p$A (Biely, Singh, and Puchart 2016).

2.3.3. α-glucuronidases from GH115 family

Presence of (Me)Glc$p$A affects polymer solubility, reduces accessibility of xylanases to the β-1→4-bond of the xylan backbone; (Me)Glc$p$A has also been shown to cross-link to lignin, hindering efficient hydrolysis of xylans (Siika-aho et al. 1994; Das, Das, and Mukherjee 1984). 4-O-Methylglucuronoxylan is the major hemicellulose present in hardwood and on average contains one (Me)Glc$p$A substitution per ten
xylopyranosyl residues (Teleman et al. 2000). MeGlcA α-1→2-linked to Xylp residue at position 2 is one of the most stable glycosidic linkages that persists even after acid hydrolysis (Peter Biely, Singh, and Puchart 2016). α-Glucuronidases (EC 3.2.1.131/139) from families GH67 and GH115 are able to hydrolyze α-1→2 linked (Me)GlcA from Xylp residues. However, GH67 family can only remove (Me)GlcA from the non-reducing end of corresponding oligosaccharides. By contrast, members of GH115 family are the only known CAZymes that can remove (Me)GlcA from both terminal and internal Xylp residues of xylo-oligosaccharides and polymeric substrates (Peter Biely, Singh, and Puchart 2016). CAZy family GH4 also contains α-glucuronidases but they are not able to act on XOS or polymeric xylan (Ryabova et al. 2009).

The ability of GH115 α-glucuronidases to act on polymeric substrates and remove (Me)GlcA from both internally and terminally substituted residues makes them ideal for xylan side chain modification. This unique activity, along with evolutionally distinct sequence from GH4 and 67 families, allowed classification of GH115 as a new family of α-glucuronidase (Ryabova et al. 2009). The first characterized α-glucuronidase was from Thermoascus aurantiacus and showed activity on xylan and xylo-oligosaccharides from larchwood (Khandke, Vithayathil, and Murthy 1989); however, the protein sample was recovered from culture filtrate and requires further investigation with more extensively purified protein. Of the 485 GH115 sequences listed in the CAZy database, 8 members have been biochemically characterized (Table 1). All characterized GH115 to date are active on glucuronoxylan, except for BtGH115 that is active on arabinogalactan from acacia gum, which contains a β-1→3-linked galactose backbone with β-1→6-linked galactose branches decorated with up to 17% MeGlcA α-1→6 substitutions (Aalbers et al. 2015). This finding indicates that GH115 is a polyspecific family (i.e contains at least two EC numbers). The optimal pH of characterized GH115s ranges from between 4.4 to 7 (Table 1). Recently characterized alkaliphilic GH115 from Amphilbacillus xylanus, AxyAgu115A, showed improved activity under alkaline conditions, increasing release of (Me)GlcA by 30% when the reaction was performed at pH11.0 rather than pH4.0 (Yan et al. 2017). Continued discovery and characterization of new α-glucuronidases with unique biochemical properties is expected to improve recovery and utilization of xylans.
Table 1: List of characterized GH115 (as of April 2017) and their reported properties.

<table>
<thead>
<tr>
<th>ID</th>
<th>Species</th>
<th>GenBank ID</th>
<th>Structure</th>
<th>MW kDa</th>
<th># of domains</th>
<th>Optimum pH</th>
<th>Notes</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bacterial sources</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BoAgu115A</td>
<td><em>Bacteroides ovatus</em> ATCC 8483</td>
<td>EDO10816.1</td>
<td>4C90, 4C91</td>
<td>85&lt;sup&gt;a&lt;/sup&gt; (199&lt;sup&gt;b&lt;/sup&gt;, dimer)</td>
<td>4</td>
<td>7</td>
<td></td>
<td>Rogowski et al., 2013</td>
</tr>
<tr>
<td>BtGH115A</td>
<td><em>Bacteroides thetaiotaomicron</em> VPI-5482</td>
<td>AAO78064.1</td>
<td>5BY3</td>
<td>90&lt;sup&gt;c&lt;/sup&gt; (monomer&lt;sup&gt;d&lt;/sup&gt;)</td>
<td>4</td>
<td>5.5</td>
<td>Unique substrate: active on arabinogalactan</td>
<td>Aalbers et al., 2015</td>
</tr>
<tr>
<td>SdeAgu115A</td>
<td><em>Saccharophagus degradans</em> 2-40</td>
<td>ABD81015.1</td>
<td>4ZMH</td>
<td>? (189&lt;sup&gt;b&lt;/sup&gt;, dimer)</td>
<td>5</td>
<td>6.5</td>
<td>between domains C and D</td>
<td>Wang et al., 2016</td>
</tr>
<tr>
<td>Pjd2Agu115A</td>
<td><em>Paenibacillus</em> sp. JDR-2 (Pjdr2)</td>
<td>ACT04581.1</td>
<td>100&lt;sup&gt;a&lt;/sup&gt;</td>
<td>5</td>
<td></td>
<td></td>
<td>Shows better performance under alkaline condition</td>
<td>Rhee et al., 2016</td>
</tr>
<tr>
<td>AxyAgu115A</td>
<td><em>Amphibacillus</em> xylanus NBRC 15112</td>
<td>BAM48432.1</td>
<td>110&lt;sup&gt;a&lt;/sup&gt;</td>
<td>5</td>
<td>7</td>
<td></td>
<td></td>
<td>Yan et al., 2017</td>
</tr>
<tr>
<td>SpAgu115A</td>
<td><em>Streptomyces pristinaespiralis</em> ATCC 25486</td>
<td>EDY63299</td>
<td>crystal</td>
<td>111&lt;sup&gt;a&lt;/sup&gt;</td>
<td>5</td>
<td></td>
<td></td>
<td>Fujimoto et al., 2011</td>
</tr>
<tr>
<td><strong>Fungal sources</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ScAgu115A</td>
<td><em>Schizosaccharomyces pombe</em> FGSC</td>
<td>ADV52250.1</td>
<td>38548</td>
<td>125&lt;sup&gt;a&lt;/sup&gt; (330-400&lt;sup&gt;b&lt;/sup&gt;, ?)</td>
<td>5</td>
<td>5.8</td>
<td></td>
<td>Tenkanen and Siika-Aho, 2000</td>
</tr>
<tr>
<td>PsAgu115/SsAgu115</td>
<td><em>Scheffersomyces stipitis</em> CBS 6054</td>
<td>ABN67901.2</td>
<td></td>
<td>120&lt;sup&gt;a&lt;/sup&gt;</td>
<td>5</td>
<td>4.4</td>
<td></td>
<td>Ryabova et al., 2009; Bajwa et al., 2016</td>
</tr>
</tbody>
</table>

<sup>a</sup> = determined by SDS-PAGE
<sup>b</sup> = determined by gel filtration
<sup>c</sup> = size exclusion chromatography
The first crystal structure of GH115 was obtained from *Bacteroides ovatus*, BoAgu115A (PDB code 4C90, 4C91), and was shown to be a four domain structure consisting of N-terminal β-strand/α-helix domain I (or domain A), catalytic TIM barrel (8 stranded α/β fold) domain II (or domain B), five-helical bundle domain III (or domain C), and C-terminal β-sandwich domain IV (or domain D) (Rogowski et al. 2013; Wang et al. 2016). BoAgu115A forms dimer in solution with domains C and D involved in inter-chain protein dimerization (Rogowski et al. 2013). Crystal structure from *Bacteroides thetataomicron*, BtGH115A (PDB code 5BY3) showed the same four-domain architecture as BoAgu115A, except for positioning of domain D, which appears to have undergone domain-swapping phenomenon and is folded back atop of the catalytic domain (Aalbers et al. 2015). Swapping of the domain D is likely involved in dimerization of BoAgu115, which is does not occur in BtAgu115A as it is a monomer in its native form (Aalbers et al. 2015). Recently, a five-domain structure was resolved from the marine bacterium *Saccharophagus degradans*, SdeAgu115A (PDB code 4ZMH), which has an additional C⁺ insertion between domains C and D (Wang et al. 2016). The C⁺ domain was shown to be involved in substrate binding and common in other members of GH115 family. Residues E216 and D335 have been identified as essential residues in SdeAgu115A (Wang et al. 2016), and are conserved in all of the characterized GH115s to date.

The mechanism of action of α-glucuronidases has been described for ScAgu115A from *Schizophyllum commune* and PsAgu115 from *Pichia stipitis* (Kolenová et al. 2010). They employ an inverting mechanism similar to the GH67 family, where the glycosidic bond is hydrolysed through inversion of the anomeric configuration releasing (Me)Glc pA as its β-anomer (Biely et al. 2000; Kolenová et al. 2010). The difference between GH67 and GH115 is the topology of the active site, with GH67 likely having a deeper substrate binding site pocket that is able to accommodate only short XOS with (Me)Glc pA linked to the non-reducing Xyl p (Kolenová et al. 2010). GH115 α-glucuronidases recognize (Me)Glc pA and part of the xylan chain (Kolenová et al. 2010), and as a result, a substituent in the vicinity of GH115 active site is likely to introduce steric hindrance. For example, (Me)Glc pA substitution on Xyl p residues that are acetylation on the vicinal hydroxyl group is an especially recalcitrant glucuronosyl substitution (Biely, Singh, and
Puchart 2016). Notably, 10-18% of (methyl)glucuronoxylan in hardwood contains 3-O acetylated Xylp 2-O substituted with (Me)GlcA (Monclaro and Ferreira Filho 2013). Accordingly, the discovery of an esterase and/or α-glucuronidase that can act on this type of substitution will help to improve xylan side-chain modification, enabling effective utilization of the polymer in production of high-value bioproducts.

2.4. PUL database: a source for discovery of new and novel CAZymes

Animal gut microbiota is constantly presented with a wide range of different polysaccharides composed of structurally diverse sugar moieties of various degree of polymerization, with different side chains and glycosidic linkages. Breakdown and utilization of such substrates requires concerted action of large repertoire of different CAZymes. Members of the dominant gut-associated bacterial phyla, Bacteroidetes, have evolved a multiprotein system for degradation of complex glycans called polysaccharide utilization loci (PULs) (Martens et al. 2009). PULs are clusters of physically linked genes whose corresponding proteins work to together to detect and import polysaccharide to be used as carbon and energy sources (Bjursell, Martens, and Gordon 2006; Grondin et al. 2017).

The Bacteroides thetaiomicron starch utilization system (Sus) was the first described PUL and consists of eight genes, namely susRABCDEFG (Figure 5) (Reeves, Wang, and Salyers 1997; Martens, Chiang, and Gordon 2008). The outer membrane proteins SusD, and possibly SusE and SusF, bind starch to the surface of the bacterium, where SusG α-amylase hydrolyses starch into oligosaccharides. Oligosaccharides are then imported into the periplasm by the TonB-dependent transporter SusC, where they are broken down into monosaccharides by neopullunase (SusA) and α-glucosidase (SusB) (Martens et al. 2009). The products of degradation are able to regulate PUL expression through a regulator of transcriptional activation (SusR), located in the cytoplasmic membrane. PULs often contain other inner membrane sensor regulators similar to SusR, such as extracytoplasmic function sigma/anti-sigma (ECF-σ/anti-σ) factor pairs or hybrid two-component system (HTCS) (Martens, Chiang, and Gordon 2008). Each PUL characterized to date contain a pair of SusC-like and SusD-like
proteins, and these SusCD-like pairs can be used as markers for prediction of new PULs (Terrapon et al. 2015).

A) 

B) 

Figure 5: *Bacteroides thetaiomicron* starch utilization system (Sus) showing (A) gene organization, and (B) a model of concerted action of Sus in starch degradation. The outer membrane proteins SusD (and possibly SusE and SusF) bind starch, then α-amylase (SusG) hydrolyses starch into oligosaccharides, which are then imported into the periplasm by the TonB-dependent transporter SusC. The oligosaccharides are broken down into monosaccharides by neopullulanase (SusA) and α-glucosidase (SusB). The products of degradation are able to regulate PUL expression through a regulator of transcriptional activation (SusR), located in the cytoplasmic membrane (Martens et al. 2009).

Metagenomic studies are a rich source for discovery of new PULs capable of degrading predominantly recalcitrant polysaccharides due to the presence of different combinations of CAZyme families (Pope et al. 2010; Pope et al. 2012; Rosewarne, Pope, and Cheung 2014). Moreover, PULs may serve a source for discovery of new CAZymes with new or novel functions. Recently, whole genome profiling and gene inactivation experiments have revealed the identity of new PULs and their specificity for various polysaccharides (Martens, Chiang, and Gordon 2008; Martens et al. 2011). For example, Martens *et al.* showed that *Bacteroides ovatus* is capable of degrading various different hemicelluloses because of the presence of several unique PULs that are absent in *B.*
*thetaiomicron* (Martens et al. 2011). On the other hand, *B. thetaiomicron* is capable of degrading host mucin O-glycan, which *B. ovatus* is not able to degrade.

Other recent studies have shown that characterization of all components of a PUL can uncover new enzyme combinations for degradation of a specific polysaccharide. For example, Larsbrink *et al.* showed a mechanism by which a single PUL in *B. ovatus* is capable of degrading xyloglucan into individual monosaccharides (Larsbrink et al. 2014). In another study, Cuskin *et al.* showed that three PULs encoded by *B. thetaiomicron* are able to completely breakdown a complex yeast mannann through a “selfish mechanism”, in which products of polysaccharide breakdown are not shared with any other species (Cuskin et al. 2015).
3. Chapter II: Production and characterization of acetyl xylan esterase FjoAcXE

3.1. Introduction

Modification and utilization of structurally diverse polysaccharides with different side groups and glycosidic linkages requires concerted action of large repertoire of different CAZymes. The PUL database (PULDB, www.cazy.org/PULDB) contains experimentally validated PULs from literature as well as almost 4000 predicted PULs (Terrapon et al. 2015), and therefore serves as a valuable source for discovering new and novel CAZymes, including esterases. The PULDB mostly contains known or predicted CAZymes; however, there are many proteins that have no known or predicted function. These unknown proteins may be part of uncharted clades or part of entirely new family of CAZymes with new or novel function.

A protein of unknown function from Flavobacterium johnsoniae UW 101 was chosen for characterization from a PUL containing xylan active CAZymes, such as glycoside hydrolases from family GH43 and GH115, and carbohydrate esterases from families 1 and 6. This unknown protein, later renamed FjoAcXE, has Lipase_GDSL_2 Pfam domain (PF13472), which is SGNH hydrolase-type esterase domain belonging to GDSL-like Lipase/Acylhydrolase family. FjoAcXE was able to deacetylate xylo-oligosaccharide residues at all positions, including those that have (Me)Glc pA at the vicinal position. Additionally, FjoAcXE significantly enhanced the activity of a xylan α-1→2-glucuronidase from family GH115, suggesting synergistic action between the two enzymes.

3.2. Materials and Methods

3.2.1. Materials

Polysaccharides were purchased from the following sources: 4-0-methylglucuronoxylan (Sigma, M5144), oat spelt xylan (Sigma, X0627), carboxymethylcellulose (Megazyme, P-CMC4M), β-glucan (low viscosity; from barley;
Megazyme, P-BGBL), starch (from corn; Sigma-Aldrich, S4126), pectin (from apple; Sigma, 76282), arabinoxylan (high viscosity; from wheat; Megazyme, P-WAXYH), galactan (Aldrich; 851396), arabinoxylan (from wheat; Megazyme, P-WAXYH), glucomannan (low viscosity; from konjac; Megazyme, P-GLCML), galactomannan (from guar; GD28; Megazyme, enzyme modified), and xyloglucan (amyloid, from tamarind seed; Megazyme, P-XYGLN).

Aliphatic p-nitrophenol (pNP) substrates were purchased from the following sources: p-nitrophenyl acetate (C_2) (Sigma, Cat# N8130), p-nitrophenyl butyrate (C_4) (Sigma, Cat#N9876), p-nitrophenyl hexanoate (C_6) (TCI-EP, Cat#H0484), p-nitrophenyl octanoate (C_8) (Sigma, Cat.#21742), p-nitrophenyl decanoate (C_10) (Sigma, Cat#N0252), p-nitrophenyl dodecanoate (C_12) (Sigma, Cat#61716), p-nitrophenyl myristate (C_{14}) (Sigma, Cat#70124), and p-nitrophenyl palmitate (C_{16}) (Sigma, Cat#N2752).

Carbohydrate containing pNP substrates were purchased from the following sources: pNP-α-L-arabinopyranoside (Sigma, Cat#N3512), pNP-β-D-cellobioside (Sigma, Cat# N5759), pNP-α-L-fucopyranoside (Sigma, Cat#N3628), pNP-β-D-glucopyranoside (Sigma, Cat#N7006), pNP-β-D-lactopyranoside (Sigma, Cat#N1752), pNP-β-D-mannopyranoside (Sigma, Cat#N1268), pNP-β-D-xylopyranoside (Sigma, Cat#N2132), pNP-acetate (Sigma, Cat#N8130), pNP-benzoate (Alfa Aesar, Cat#L16024), pNP-hexanoate (TCI-EP, Cat#H0484), pNP-octanoate (Sigma, Cat#21742), pNP-phenyl ether (Aldrich, Cat#N21901).

4-methylumbelliferyl acetate (4-MUA) was purchased from Sigma (Cat#M0883) and 4-methylumbelliferone (4-MU) was purchased from Aldrich (Cat#M1381).

Acetylated monosaccharide substrates were purchased from the following sources: β-D-glucose pentaacetate (Cat. #285943) from Aldrich (St. Louis, MO, USA), β-D-xylopyranose tetraacetate (Cat. #TX534) and 1,2,3,4-tetra-O-acetyl-L-rhamnopyranose (Cat. #TH212) from LC Scientific Inc. (Toronto, ON, Canada). Acetylated xylo-oligosaccharides (Ac-XOS) from Eucalyptus were a generous gift from Prof. Maija Tenkanen (University of Helsinki, Finland).

AxyAguGH115 (GH115) α-glucuronidase from *Amphibacillus xylanus* was purified as previously described (Yan et al. 2017). The CE6 acetylxylan esterase from
Orpinomyces sp. (E-AXEAO) was purchased from Megazyme; GH67 α-glucuronidase from Cellvibrio japonicus (PRO-E0069) was purchased from PROZOMIX Ltd.

3.2.2. Candidate selection

The polysaccharide-utilization loci database (PULDB, www.cazy.org/PULDB, experimentally validated) was searched for polysaccharide-utilization loci (PUL) that contained at least two of the following xylan active CAZyme families of interest: GH8, GH10, GH11, GH43, GH54, GH115, CE2, CE3, and CE6. From the obtained list of PULs, all proteins of unknown function were extracted and analyzed using Signal P4.1 (Petersen et al., 2011); those containing a predicted signal sequence for secretion were further analyzed based on presence of Pfam domains, sequence length, and the type of CAZymes present on the corresponding PUL.

3.2.3. Gene synthesis and molecular cloning

The gene encoding FjoAcXE (PULDB ID: Fjoh_3879; GenBank ID: ABQ06890.1) lacking the predicted signal sequence (residues 1 – 21) was codon optimized for expression in Escherichia coli K12 using IDT Codon Optimization Tool (www.idtdna.com/CodonOpt). Fifteen base pair extensions homologous to the p15TV-L vector (GenBank ID: EF456736.1) (T7, 5’-TTGTATTTCCAGGGC and T7term, 5’-CAAGCTTCGTCATCA) were added to each end of the sequence and the gene was synthesized as gBlock® Gene Fragments (IDT). gBlock fragments were PCR amplified with T7 and T7term primers using 2xKAPA HiFi HotStart ready mix (KAPA Biosystems) and then processed using the Nucleospin® Gel and PCR clean-up kit (Macherey-Nagel). The p15TV-L expression vector was digested with BSERI (NEB) and the PCR product was cloned into the vector using the In-Fusion® HD EcoDry™ Cloning Kit (Clontech). The resulting plasmid construction was transformed into E. coli HST08 Stellar™ Competent Cells (Clontech), and the sequence was verified using DNA sequencing service at the Center of Applied Genomics at the SickKids Hospital in Toronto.
3.2.4. Protein expression and purification

E. coli BL21(λDE3) codon plus strain harboring p15TV-L-FjoAcXE was propagated at 37°C in 8 L of Luria-Bertani (LB) Broth-Miller (BioShop) supplemented with 33 µg/mL chloramphenicol and 100 µg/mL ampicillin until the OD$_{600}$ reached 0.6-0.8. The cultures were cooled on ice for 5 min, additional 33 µg/mL chloramphenicol and 100 µg/mL ampicillin were added, and recombinant expression was induced overnight at 16°C with 0.5 mM isopropyl β-D-1-thiogalactopyranoside. The cells were harvested by centrifugation at 8,967xg (Beckman Coulter, JLA-8.1000) for 15 min at 4°C and the pellet (approx. 25 g fresh weight) was frozen at -80°C. The pellet was then suspended in binding buffer (50 mM HEPES pH 7.5, 300 mM NaCl, 5% glycerol and 5 mM imidazole) and the cells were disrupted by sonication (100 amplitude, 5 s ON and 5 s OFF for 20 min). Cell debris was removed by centrifugation at 27,167xg (Beckman Coulter, JA-25.5 rotor) for 15 min at 4°C and supernatant was filtered through Acrodisc® Syringe Filters with 0.45 µm Supor® membrane (Pall Corporation).

The sample was loaded onto 5 mL HisTrap HP (GE Healthcare) pre-equilibrated with binding buffer. A combination of step-wise and gradient elutions using elution buffer (50 mM HEPES pH 7.5, 300 mM NaCl, 5% glycerol and 300 mM imidazole) was performed on Biologic DuoFlow™ chromatography system (BioRad) where the column was washed with 4 column volumes (CV) of 100% binding buffer, and then increasingly substituted with elution buffer as follows: 5% elution buffer for 4 CV, 10% elution buffer for 6 CV and then gradient elution from 10% to 100% elution buffer over 12 CV. All steps were done at 1 ml/min, protein elution was monitored at A$_{280nm}$ and 2 ml fractions were collected throughout the entire run. The resulting fractions were analyzed by 12% SDS-PAGE; selected fractions were pooled and then exchanged to 25 mM HEPES pH8.0 using 10kDA Jumbosep™ centrifugal devices (Pall Corporation). The sample was then further purified using a 1.3 ml UNO™ Q ion exchange column (BioRad) pre-equilibrated with 25 mM HEPES pH8.0 (Buffer A). Following a wash using 11.5 CV of Buffer A, a step-wise gradient elution was performed at 1 ml/min on Biologic DuoFlow™ using 25 mM HEPES pH8.0 with 1M NaCl as elution buffer (Buffer B), with 0% to 50% Buffer B over 15.4 CV, then up to 100% Buffer B over 6 CV. The samples were collected as 1 ml fractions throughout the entire run and purity was checked with 12% SDS-PAGE. The
purified sample was exchanged to 10 mM HEPES pH 7.5 containing 300 mM NaCl
before being flash frozen in liquid nitrogen and stored at -80°C.

Protein concentration was measured using the Bradford assay with bovine serum albumin as a standard (Bradford, 1976).

3.2.5. In-gel trypsin digest and mass spectrometry

The identity of purified FjoAcXE was tested using in-gel trypsin digest (Shevchenko et al. 2006). Briefly, 2 µg of the protein was run on 12% SDS-PAGE and stained with Brilliant Blue R stain. The gel was washed for several hours with MiliQ water and the protein band of approximately 45 kDa was excised into ~ 1 mm cubes, followed by addition of 500 µL of acetonitrile. The sample was incubated for 10 min at room temperature (RT) and then all the liquid was removed, and 50 µL of 10 mM dithiothreitol (DTT) in 100 mM ammonium bicarbonate pH 8.5 was added and incubated at 56°C for 30 min. Another 500 µL of acetonitrile was added and incubated for 10 min at RT, all the liquid was removed, and then 50µL of 55mM iodoacetamide in 100mM ammonium bicarbonate pH 8.5 was added and incubated 20min at RT in the dark. Another 500µL of acetonitrile was added and incubated at RT until gel pieces turned white. All the liquid was removed and gel was covered with 50µL of trypsin buffer (13ng/uL trypsin in 10mM ammonium bicarbonate, pH 8.5, and 10% acetonitrile) for 30 min on ice. Additional 50µL of trypsin buffer was added and incubated for 90 min on ice. Then, 20µL of 10mM ammonium bicarbonate pH 8.5 was added and incubated at 37°C overnight.

Following in-gel trypsin digest, the samples were cleaned-up using OMIX C18 pipette tips (Agilent) as per manufacturer’s specifications. The samples were dried using a speedvac and re-suspended in 100 µL of 0.1% formic acid, and analyzed on an LTQ XL mass spectrometer (ThermoFisher) at the BioZone Mass Spectrometry Facility (University of Toronto). Briefly, peptides were separated using an easy-nLC-1000 (ThermoFisher) equipped with a 10.5 cm PicoTip Emitter Silica Tip packed in-house with C18 media coupled online to a Q-Exactive mass spectrometer (ThermoFisher) using a gradient of 0-95 % acetonitrile over 70 min. Data was queried against BL21 K12 E. coli database and the FjoAcXE sequence; peptide matches were identified using
3.2.6. Enzyme assays

Unless otherwise stated, all reactions were performed in triplicate and errors indicate standard deviation.

**PAHBAH reducing-end assay**

Reactions (50 µL reaction volume) comprised 50 mM HEPES (pH 8.0) containing 0.5% (w/v) of each polysaccharide: beechwood xylan, oat spelt xylan, carboxymethylcellulose, β-glucan, starch, pectin, wheat arabinoxylan, arabinan, glucomannan, galactomannan, xyloglucan, and arabinogalactan. Reactions were initiated by adding 5 µg of FjoAcXE, and continued for 16 hrs at 30°C. Blank samples replaced the enzyme with water. After the incubation period, 50 µl of the reaction was combined with one part of 5% (w/v) p-hydroxybenzoic acid hydrazide (PAHBAH) and four parts of 0.5M NaOH (Lever 1972). The samples were incubated at 70°C for 30 min. The samples were cooled to a room temperature and absorbances were measured at 410 nm. D-glucose was used as representative reducing sugar (10-150 nmoles) for the standard curve.

**Affinity gel electrophoresis**

Binding of FjoAcXE was tested by native polyacrylamide gel electrophoresis with 5 µg of the protein loaded onto gels containing 0.01% 4-O-methylglucuronoxylan, oat spelt xylan, carboxymethylcellulose, beta-glucan, starch, pectin, arabinoxylan, galactan, arabinan, glucomannan, galactomannan, and xyloglucan as described previously (Foumani et al. 2015).

**pNP substrates**

Reactions (200 µL reaction volume) comprised 50 mM HEPES (pH 8.0) and 2 mM of pNP-α-L-arabinopyranoside, pNP-β-D-celllobioside, pNP-α-L-fucopyranoside, pNP-β-D-glucopyranoside, pNP-β-D-lactopyranoside, pNP-β-D-mannopyranoside, pNP-β-D-xylopyranoside, pNP-acetate, and were initiated by adding 5 µg of FjoAcXE. Reactions continued for 2 hrs at 30°C, and absorbance was measured continuously at 410 nm. Reaction mixtures without FjoAcXE were used as a blank for each substrate.
pNP assay using aliphatic substrates

C_2 to C_{12} pNP substrates were dissolved in 100% DMSO, and C_{14} and C_{16} substrates were dissolved in 1:1 (v/v) of isopropanol:acetonitrile. Reactions (200 μL reaction volume) comprised 50 mM HEPES (pH 7.0) and 2 mM substrate, and were initiated by adding 0.5 μg of FjoAcXE. Reactions were conducted at pH 7.0 as the rate of non-enzymatic desterification of pNP substrates increases at higher pH and gives a high background. Reactions continued for 2 hrs at 30°C, and absorbance was measured continuously at 410 nm. Reaction mixtures without FjoAcXE were used as a blank for each substrate.

Acetic acid release from acetylated saccharides

Reactions (30 μL reaction volume) comprised 50 mM HEPES (pH 8.0) and 0.1% (w/v) of β-D-glucose pentaacetate (final concentration 2.5 mM), β-D-xylopyranose tetraacetate (final concentration 3.14 mM), or 1,2,3,4-tetra-O-acetyl-L-rhamnopyranose (final concentration 3 mM). Reactions were initiated by adding 0.5 μg of FjoAcXE, and continued for 20 min at 30°C. The reaction was stopped by boiling the samples at 100°C for 10 min. The samples were spun down and release of acetic acid was measured using Acetic Acid kit (K-ACETRM, Megazyme) following manufacturer’s specifications. Reaction mixtures without FjoAcXE were used as a blank for each substrate.

Optimum pH and pH stability

The pH optimum of FjoAcXE was tested using 100 mM Tris, 50 mM MES, 50 mM acetic acid and 50 mM sodium acetate trihydrate buffer with pH range of 3.5 to 9.5. Reactions (200 μL reaction volume) containing 0.5 mM 4-MUA were initiated by adding 5 μg of FjoAcXE, and continued for 20 min at 30°C. Absorbance was measured at 354 nm and the reaction mixture without FjoAcXE was used as a blank.

To test the effect of different pH on enzyme stability, reactions (40 μL reaction volume) containing 4 μg of FjoAcXE was incubated in 100 mM Tris, 50 mM MES, 50 mM acetic acid and 50 mM sodium acetate trihydrate buffer of different pH, ranging from 3.5 to 9.5 for 16 hrs on ice at 4°C. After the incubation, residual activity was tested.
in reactions (200 µL reaction volume) containing 0.5 mM 4-MUAc in 50 mM HEPES (pH8.0). Reactions were initiated by adding 1 µg of FjoAcXE from each treatment, and continued for 20 min at 30°C. Absorbance was measured continuously at 354 nm and the reaction mixture without FjoAcXE was used as a blank.

**Optimum temperature and thermostability**

The optimum temperature of FjoAcXE was tested at 20, 30, 35, 40, 45, 50, 55, 60, 65 and 70°C. Reactions (200 µL reaction volume) containing 0.5 mM 4-MUA in 50 mM HEPES (pH8.0) were initiated by adding 0.1 µg of FjoAcXE, and continued for 20 min. Absorbance was measured at 354 nm and the reaction mixture without FjoAcXE was used as a blank.

To test thermostability, reactions (40 µL reaction volume) containing 4 µg of FjoAcXE was incubated in 50 mM HEPES (pH7.0) for 10 min, 20 min, 40 min, 2 hrs, 16 hrs at 20, 30, or 40°C or for 5, 10, 15, 30, 60 min at 50, 60, or 70°C. After the incubation, residual activity was tested in reactions (200 µL reaction volume) containing 0.5 mM 4-MUAc in 50 mM HEPES (pH8.0). Reactions were initiated by adding 1 µg of FjoAcXE from each treatment, and continued for 20 min at 30°C. Absorbance was measured continuously at 354 nm and the reaction mixture without FjoAcXE was used as a blank.

**Effect of metal ions**

Metal-free apoenzyme of FjoAcXE was prepared as described in Wang et al. (2010). Briefly, 0.05 g of CHELEX 100 (Sigma) was added to 1 mg of FjoAcXE and incubated at room temperature for 20 min. The effect of metal ions was tested in reactions (200 µL reaction volume) containing 0.5 mM 4-MUA in 50 mM HEPES (pH8.0). Reactions were initiated by adding 5 µg of FjoAcXE, and continued for 20 min at 30°C in the presence of 1 mM of the following metal ions: Ag^{2+}, Ca^{2+}, Cd^{2+}, Co^{2+}, Cu^{2+}, Fe^{3+}, Mg^{2+}, Mn^{2+}, Ni^{2+}, and Zn^{2+} (all as chloride salts, except for Ag^{2+}, which was a nitrate). CHELEX treated (T_only) and untreated (U_only) FjoAcXE was tested with and without 20 mM EDTA as a control. Absorbance was measured at 354 nm and the reaction mixture without FjoAcXE was used as a blank.
Effect of chemical reagents

The effect of chemical reagents was tested in reactions (200 µL reaction volume) containing 0.5 mM 4-MUAc in 50 mM HEPES (pH8.0). Reactions were initiated by adding 0.5 µg of FjoAcXE, and continued for 20 min at 30°C in the presence of 30% (v/v) ethanol, 30% (v/v) isopropanol, 1% (w/v) SDS, 1% (w/v) dithiothreitol (DTT), 1% (v/v) Tween-20, 1% (v/v) Tween-80, and 1% (v/v) TritonX-100. Absorbance was measured at 354 nm and the reaction mixture without FjoAcXE was used as a blank.

3.2.7. Enzyme kinetics

Kinetic parameters of FjoAcXE were determined using 4-MUA and pNP-acetate. Reactions (200 µl reaction volume) contained approximately 0.02 µg (4.41x10^{-7} µmol) of FjoAcXE in 50 mM HEPES (pH8.0) and final of 10% DMSO. Reactions were initiated by adding 0.01 to 1.5 mM of 4-MUA, and absorbance at 354 nm was read continuously for 40 min at 30°C. Reactions (200 µl reaction volume) contained approximately 0.06 µg (1.32x10^{-6} µmol) of FjoAcXE in 50 mM HEPES (pH8.0). Reactions were initiated by adding 0.05 mM to 10 mM pNP-acetate, and absorbance at 410 nm was read continuously for 40 min at 30°C.

Reaction mixtures without FjoAcXE were used as a blank for each substrate. Kinetic parameters were calculated using Michaelis-Menten equation and Graphpad Prism 5 software (La Jolla, CA, USA). Unless otherwise stated, all reactions were performed in triplicate and errors indicate standard deviation.

3.2.8. Synergism of FjoAcXE with α-glucuronidases

FjoAcXE, GH67 (PROZOMIX), AxyGH115, and CE6 (Megazyme) were tested alone and as pairs of carbohydrate esterase and glucuronidase activities. Reactions (30 µL reaction volume) comprised 1% (w/v) Ac-XOS in 50 mM HEPES (pH7.0). Reactions were initiated by adding 0.5 µg of each enzyme, and continued for 20 min at 30°C. Reactions were stopped by boiling the samples at 100°C for 10 min. The samples were spun down and release of acetic acid and D-glucuronic acid were measured using
Megazyme’s Acetic Acid (K-ACETRM) and D-glucuronic acid/D-galacturonic acid (K-URONIC) assay kits, respectively, following manufacturer’s specifications. Reaction mixture without the enzymes was used as a blank.

3.2.9. $^1$H-NMR analysis of FjoAcXE towards acetylated xylo-oligosaccharides

One-dimensional $^1$H-NMR was performed to analyze products generated following treatment of acetylated xylo-oligosaccharides with FjoAcXE after 1 hr and 24 hrs of incubation. Briefly, reaction mixtures (600 µL) comprised 1% (w/v) acetylated xylo-oligosaccharides, 20 mM 3-(N-morpholino)propanesulfonic acid (MOPS) buffer pH 7.5, and 5 µg of FjoAcXE. MOPS buffer was used because its methylene group can serve as an internal standard for NMR analysis and reactions were conducted at pH 7.5 as it is an optimal pH for source organism of FjoAcXE. The reaction mixture without the enzymes was used as a control. After the incubation, the samples were filtered through Acrodisc® Syringe Filters with 0.2 µm Supor® membrane (Pall Corporation) and lyophilized. The samples were dissolved in 600 µL D$_2$O and transferred into 5 mm NMR tubes (Norell) for analysis on Agilent DD2 700 MHz spectrometer equipped with triple resonance HCN cold probe with a scan number of 64, relaxation delay of 1 s and acquisition time of 4.5 s. The data were obtained using VnmrJ 4.0 (Agilent) and analyzed with MestReNova 10.0 (Mestrelab Research). The methylene protons at position 2 of MOPS at 2.076 ppm (Ulrich et al. 2008) and D$_2$O peak at 4.790 were used as internal standards.

3.2.10. Predicted structural model of FjoAcXE

Structural model of FjoAcXE was built using PHYRE2.0 (Kelley et al., 2015) using c4xhA, c2o14A, d1snra2 and c2w9xA as templates. The model was displayed with PyMOLv1.7.4.5 Edu (PyMOL Molecular Graphics System, Schrödinger, LLC).
3.3. Results and Discussion

3.3.1. Candidate selection and recombinant protein production

Search of PULDB for PULs containing at least two xylan active CAZyme families of interest (GH8, GH10, GH11, GH43, GH54, GH115, CE2, CE3, and CE6) revealed several interesting PULs (Appendix A). Four unknown proteins (unks) were selected from these results based on the presence of Pfam domains of interest, presence of signal sequence, sequence length and presence of interesting CAZymes on a given PUL (see Appendix E for summary and sequences of the selected candidates). Three out of four proteins, which we named unk1, unk2 and unk3, were part of PUL73 from *Bacteroides ovatus* ATCC 8483. Their production and characterization is described in Appendix C and is not discussed further in this chapter. The fourth protein, “unk4” or FjoAcXE from *Flavobacterium johnsoniae* UW 101, was found on PUL20 containing GH43, GH115, and CE6 families and so was selected for recombinant expression and biochemical characterization (Figure 6).

PUL20 has been experimentally validated and predicted to be involved in hemicellulose utilization (McBride *et al.*, 2009). Terrapon *et al.* (2015) predicted PUL23, which is an extension of experimentally validated PUL20, and together these PULs contain four proteins of unknown function, including FjoAcXE. Two of the proteins of unknown function have no predicted signal sequence (as determined by SignalP4.0 server), while the third unknown protein has a predicted signal sequence but it has no known Pfam domains. FjoAcXE has predicted signal sequence (residues 1-21) and Lipase_GDSL_2 Pfam domain (PF13472), which is an SGNH hydrolase-type esterase domain belonging to GDSL-like Lipase/Acylhydrolase family.
Figure 6: PUL20 *Flavobacterium johnsoniae* showing possible functions of predicted CAZymes (as indicated by BLAST searches). Proteins of unknown function (unk) are in grey. The red arrow indicates unk (renamed FjoAcXE), selected for this study. Hybrid two-component system is responsible for PUL transcriptional activation and Starch utilization system C- and D-like proteins (SusC and SusD) are involved in binding and import of a substrate. Carbohydrate active families are labeled as follows: glycoside hydrolase (GH) and carbohydrate esterase (CE). Vertical slash between two modules represents multidomain CAZymes, and the arrows show whether the gene is found on the sense (†) and antisense (‡) strand.

FjoAcXE was recombinantly overexpressed and purified from cell lysate by Ni-NTA affinity chromatography using 6x-Histag at the N-terminus of the protein. A second purification step with anion exchange chromatography was used to improve purity of the final protein. A total of 150 mg of FjoAcXE was purified from 25 g of fresh cell pellet, and the protein had ~ 98 % purity as assessed by ImageJ 1.50i (Schneider *et al.*, 2012) with molecular mass of 45.2 kDa (Figure 7) and theoretical pI of 6.4. FjoAcXE was identified correctly by in-gel trypsin digest and mass spectrometry with coverage from N to C terminus.

Figure 7: Purified FjoAcXE is approximately 45.2 kDa.
3.3.2. Activity screen

Initial screening of FjoAcXE was done using a wide range of polysaccharides and pNP substrates. Hydrolytic activity of FjoAcXE was tested using the PAHBAH reducing-end assay and 12 different polysaccharides (Figure 8). Low levels of reducing sugars were detected for some of the tested substrates. It is not clear if this is a product of an artifact or FjoAcXE displays hydrolytic activity. It has been suggested that some carbohydrate esterases show substrate specificity towards polysaccharide backbone, however this has not been proven (Biely et al. 2014).

![Figure 8: PAHBAH reducing-end assay was used to test for hydrolytic activity of FjoAcXE on 0.5% (w/v) polysaccharides.](image)

Lack of FjoAcXE binding to the 12 polysaccharides was confirmed by affinity gel electrophoresis (Figure 9). A native polyacrylamide gel containing either no substrate or 0.01% of a polysaccharide was loaded with 5 µg of FjoAcXE. A distance travelled by the protein on the gel not containing a polysaccharide was compared to the ones that contained a polysaccharide. If binding occurs, the protein will not travel far, if at all, through the gel. FjoAcXE did not bind to any of the substrates tested. BSA was used as
non-binding control, and a no substrate control was run with each batch of gels, so each row of Figure 9 should be read separately.

**Figure 9: Affinity gel electrophoresis.** 5 µg of FjoAcXE was run on 7.5% (w/v) native polyacrylamide gel (25 mM Tris, 250 mM glycine buffer (pH8.3) containing 0.01% of each substrate for 2 hrs at 90V and stained with Coomassie Blue G. Bovine serum albumin (BSA) is used as a non-binding control. MGX = 4-O-Methylglucuronoxylan; OSX = oat spelt xylan; CMC = carboxymethylcellulose, WAX = wheat arabinoxylan.
Initial screen screen using pNP substrates showed that FjoAcXE had esterase activity as observed by the release of pNP from pNP acetate (Figure 10).

![Graph showing pNP release vs. 2 mM pNP substrate](image)

**Figure 10: A general screen of FjoAcXE on pNP substrates showed that its active on pNP-acetate.** In 200 μL reaction volume, 5 μg of FjoAcXE in 50 mM HEPES (pH 8.0) was tested on 2 mM of each substrate. Absorbances were measured after 2 hrs incubation at 30°C.

An esterase activity was further tested on eight pNP alkyl esters, with chain length of C2 to C16. FjoAcXE showed highest activity on C₂, followed by C₄, and showed no activity on long-chain alkyl esters (C₆-C₁₆), consistent with esterase rather than lipase activity (Figure 11).
Figure 11: A screen of pNP alkyl esters showed that FjoAcXE is active on short chain alkyl esters (<C4) suggesting that it is likely an esterase and not a lipase. No significant activity was observed when tested on C6-C16. In 200 μL reaction volume, 0.5 μg of FjoAcXE in 50 mM HEPES (pH8.0) was tested on 2 mM of each substrate. Absorbances at 410 nm were measured after 2 hrs incubation at 30°C. pNP acetate (C2), pNP butyrate (C4), pNP hexanoate (C6), pNP octanoate (C8), pNP decanoate (C10), pNP dodecanoate (C12), pNP myristate (C14), and pNP palmitate (C16).

3.3.3. Activity of FjoAcXE on synthetic and natural substrates

FjoAcXE was active on acetylated monosaccharides (glucose, xylose and rhamnose) and acetylated xylo-oligosaccharides (Ac-XOS), indicating that FjoAcXE has low susceptibility to steric hindrance as it can access acetyl groups on substrates of various residue types and degree of polymerization (Figure 12). Ability to access acetyl groups from different residues has been shown in other esterases. For example, an acetyl esterase from T. reesei was shown to be active on acetylated glucose and xylose, and was able to release 50% of the acetyl groups on both substrates (Poutanen and Sundberg 1988). Generally, AcEs cannot deacetylate oligomeric or polymeric xylans, while AcXE can (Adesioye et al. 2016). Given FjoAcXE activity on oligomeric xylans, it is likely that this enzyme is an acetyl xylan esterase.
3.3.4. **Effect of pH, temperature, metal ions, and chemical reagents**

4-MUA was used to determine the effect of pH, temperature, metal ions and chemical reagents on activity of FjoAcXE. FjoAcXE activity was optimum at pH 8.0 (Figure 13A); however, FjoAcXE showed broad pH stability and retained over 90% of activity at pH 7.0 to pH 9.5, and almost 70% of activity after 16 hrs at pH 3.5 to pH 6.5 (Figure 13B). FjoAcXE displayed a highest activity at 20°C to 45°C (Figure 13C), and retained 100% of its initial activity after 16 hrs at 20°C to 40°C; only 50% activity was retained after 10 min at 50°C and no residual activity was detected after 5 min at 60°C (Figure 13D).

The optimum pH, temperature and thermostability of FjoAcXE is similar to previously reported SGHN hydrolases, such as Alr1529 from *Anabaena* sp. (Bakshy, Gummadi, and Manoj 2009), Sm23 from *Sinorhizobium meliloti* (Hwang et al. 2010), LI22 from *Listeria innocua* (Kim et al. 2012), and Est24 from *Sinorhizobium meliloti* (Bae et al. 2013). Most SGHN hydrolases lose most of their activity below pH 6.0. However, FjoAcXE shows significant pH stability, retaining over 80% of its activity after 16 hrs pre-incubation at pH4.5-9.5, while some other SGHN hydrolases, such as Est19

![Figure 12: Activity of FjoAcXE on synthetic and natural substrates.](image)
from *Bacillus* sp., show substantial loss of activity after 1 hr pre-incubation at pH below 6.0 and above 10.0 (Yu et al. 2016).

**Figure 13: General biochemical properties of FjoAcXE.** A) Optimum pH and B) pH stability. The samples were tested using 100mM Tris, 50 mM MES, 50mM acetic acid and 50mM sodium acetate trihydrate universal buffer with pH range of 3.5 to 9.5. C) Optimum temperature and D) thermal stability. E) Effects of metal ions and F) effect of chemical reagents on activity on the activity of FjoAcXE. For all samples, relative or residual activities were measured using 0.5 mM 4-MUA in 50 mM HEPES (pH8.0) at for 20 min at 30°C. Absorbance at 354 nm was measured and reaction mixture without the protein was used as a blank.
Of the metal ions tested, none had significant effect on FjoAcXE activity (Figure 13E), which is consistent with the observation that most carbohydrate esterases, except CE4, are not metal-ion dependent (Adesioye et al. 2016). FjoAcXE showed significant sensitivity towards 1% SDS and 1% DTT, showing almost no activity after 20 min incubation (Figure 13F). Similarly, addition of 30% ethanol and isopropanol resulted in over 60% loss of activity. Other SGNH hydrolases, such as Est24 retained 75% of its activity when treated with 5% SDS (Bae et al. 2013), while LI22 showed less sensitivity to 30% ethanol compared to FjoAcXE, retaining over 75% of its activity (Kim et al. 2012). On the other hand, FjoAcXE activity in the presence of surfactants is higher than in some other SGNH hydrolases. Addition of 1% Tween-20, Tween-80 and Triton-X-100 resulted in only 30% loss of activity of FjoAcXE, while Sm23 was shown to lose over 70% activity (Hwang et al. 2010).

Overall, FjoAcXE shows significant pH stability and stability to certain chemical surfactants, which are important considerations for certain industrial applications, for example use in detergents.

3.3.5. Enzyme kinetics

Kinetic parameters of FjoAcXE were determined using 4-MUA and pNP-acetate. Due to the instability of pNP-acetate at high pH, the reaction was initiated by addition of the substrate. Reactions containing 4-MUA contained 10% DMSO as 4-MUA is not soluble in aqueous solution at concentrations above 1.5 mM. Therefore, kinetic parameters could not be calculated for substrate concentration above 1.5 mM of 4-MUA. The enzyme dose was chosen to ensure a linear relationship between reaction time and product release (Mai-Gisondi and Master 2017). FjoAcXE showed similar specific activity towards both 4-MUA and pNP-acetate (Table 2). FjoAcXE activity towards pNP-acetate was 160.7 U/mg, while for Est19 SGNH hydrolase it was 277.78 U/mg (with 1 U being the amount of enzyme that catalyses the reaction of 1 µmol of substrate per min) (Yu et al. 2016).
Table 2: Kinetic parameters of FjoAcXE were tested on 4-MUA and pNP-acetate. Data were fitted to the Michaelis-Menten equation.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Specific activity (µmol/min/mg)</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$K_m$ (mM)</th>
<th>$k_{cat}/K_m$ (s$^{-1}$mM$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-MUA</td>
<td>126.3±5.9</td>
<td>95.1</td>
<td>0.829±0.07</td>
<td>114.7</td>
</tr>
<tr>
<td>pNP-acetate</td>
<td>160.7±4.0</td>
<td>120.8</td>
<td>0.944±0.08</td>
<td>127.9</td>
</tr>
</tbody>
</table>

3.3.6. FjoAcXE’s synergistic action with α-glucuronidase

Action of FjoAcXE towards the 3-O-acetyl group on Xylp also 2-O substituted with (Me)GlcPA was further explored through synergistic action of FjoAcXE with a GH115 α-glucuronidase. CE6 and GH67 were used for comparison. The reaction condition was chosen in relation to the general properties of each enzyme, where 0.5 µg of each enzyme was incubated in 50 mM HEPES pH7.0 at 30°C for 20 min. In the presence of FjoAcXE, GH115 releases almost 9-fold more (Me)GlcPA compared when GH115 is tested alone on acetylated xylo-oligosaccharides (Figure 14). GH67 also shows increase in activity in the presence of FjoAcXE, where release of (Me)GlcPA is 5 times higher than when GH67 is tested alone. Higher release of (Me)GlcPA from GH115 in the presence of FjoAcXE compared to GH67 suggested that some of (Me)GlcPA are released from internal residues. On the other hand, CE6 does not show the same synergistic activity with GH115 as there is no increased release of (Me)GlcPA.
Figure 14: Measurement of glucuronic and acetic acids released when tested on 1% Ac-XOS. In 30 µL reaction, 0.5 µg of each protein was incubated in 50 mM HEPES, pH7.0 for 20 min at 30°C, reaction was stopped by incubating the samples at 100°C for 10 min. Glucuronic acid released was measured using the α-glucuronidase kit (Megazyme) and acetic acid released was measured using the acetic acid kit (Megazyme) Axy = AxyAguGH115; CE6 (E-AXEAO; Megazyme), GH67 (PRO-E0069; PROZOMIX).

The vicinal 3-O acetylated Xylp position provides steric hindrance to GH67 and GH115 and prevents them from accessing (Me)GlcP residue for cleavage, significantly reducing efficiency of α-glucuronidases. Ability of an esterase to act on 3-O acetylated Xylp 2-O substituted with (Me)GlcP at the non-reducing end of acetylated xylo-oligosaccharides has been reported previously (Neumüller et al., 2015), however this is the first report of an esterase that is able to access internal 3-O acetylated Xylp 2-O substituted with (Me)GlcP and showed synergistic action between an esterase and α-glucuronidase.
3.3.7. $^1$H-NMR analysis of FjoAcXE towards acetylated xylo-oligosaccharides

$^1$H-NMR was used to examine positional specificity of FjoAcXE on acetylated xylo-oligosaccharides. The peak assignments were done as described previously (Uhliariková et al. 2013; Neumüller et al. 2015; Vladimir Puchart et al. 2016). The change in signal intensity in the regions between 5.4-4.4 ppm corresponding to acetylated xylopyranosyl (Xylp) residues in the anomeric region of the spectrum, and 2.30-2.15 ppm region corresponding to the acetyl group methyl protons were used to assign proton chemical shifts, as reported in Uhliariková et al. (2013).

![Diagram of xylo-oligosaccharides](image)

Figure 15: $^1$H-NMR spectrum of FjoAcXE tested on 1% (w/v) acetylated xylo-oligosaccharides. Red arrow shows unique activity of FjoAcXE to deacetylate 3-O-acetylated Xylp 2-O-substituted with (Me)GlcA. Tested on 1% acetylated xylo-oligosaccharides in 20 mM MOPS pH7.5 for 1hr and 24 hrs at room temperature, followed by lyophilization and suspension in 0.6 ml D$_2$O. Panel A shows the changes of signal intensity of acetylated Xylp residues in the anomeric region of the spectrum, and Panel B shows the changes of intensity of the methyl protons of the acetyl groups (Uhliariková et al., 2013).

FjoAcXE is active on 2- and 3-monoacetylated, and 2,3-di-O-acetylated Xylp residues (Figure 15). FjoAcXE was also active on the 3-O-acetyl group on Xylp residues
which 2-\textit{O}-substituted with (Me)Glc\textsubscript{p}A, as shown by the change in intensity of the signal corresponding to 2.28 and 5.12 ppm (Figure 14). Thus, FjoAcXE is not affected by a steric hindrance presented by (Me)Glc\textsubscript{p}A, and able to act on both mono- and di-acetylated Xyl\textsubscript{p} residues.

3.3.8. Sequence analysis and structural model of FjoAcXE

Sequence alignment of FjoAcXE using MAFFT v7.017 (Katoh et al., 2002) with some members of GDSL- and SGNH-hydrolases of known function (Akoh et al., 2004), showed that FjoAcXE has SGNH conserved residues and putative Ser-His-Asp catalytic triad at positions 196, 374 and 371 (mature sequence), respectively (Figure 16).

![Sequence alignment and structural model of FjoAcXE](image)

**Figure 16: FjoAcXE (Fjoh\_3879) has GDSL-like motif, conserved SGNH residues (red circles) and catalytic Ser-His-Asp triad (green triangle).**

FjoAcXE has less than 20\% sequence identity with all other unknown proteins listed on PULDB and less than 30\% sequence identity with all of CE families listed on CAZy database (including non-classified CE family). BLASTp search of FjoAcXE against GeneBank query showed that the top 100 hits have 44\% to 74\% sequence identity to FjoAcXE and are listed as lipases, GDSL family lipases, SGNH hydrolases or hypothetical proteins. However, none of these proteins have been biochemically characterized, and therefore may have new function that has never been reported before. Table 3 summarizes predicted FjoAcXE homologs, which can be biochemically characterized to expand our understanding of substrate specificity and mechanism of action of these uncharacterized group of proteins, and decide if they should be classified as a new CE family.
Table 3: List of FjoAcXE homologs. In bold are selected candidates for future characterization.

<table>
<thead>
<tr>
<th>IDs</th>
<th>Sequences identity with Fjoh_3879 (%)</th>
<th>Species</th>
<th>Signal Seq (gram neg)</th>
<th>Pfam domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>WP_012025856.1</td>
<td>100</td>
<td>Flavobacterium johnsoniae</td>
<td>YES</td>
<td>Lipase_GDSL_2 (PF13472), SGNH_hydro</td>
</tr>
<tr>
<td>K392DRAFT_2214</td>
<td>98.1</td>
<td>Flavobacterium sp. 4058</td>
<td>YES</td>
<td>Lipase_GDSL_2 (PF13472), SGNH_hydro</td>
</tr>
<tr>
<td>WP_035683315.1</td>
<td>73.9</td>
<td>Flavobacterium reichenbachii</td>
<td>YES</td>
<td>[Flavobacterium reichenbachii]</td>
</tr>
<tr>
<td>WP_057120560.1</td>
<td>72.9</td>
<td>Flavobacterium sp. Root901</td>
<td>YES</td>
<td>GDSL family lipase (Flavobacterium sp. Root901)</td>
</tr>
<tr>
<td>Ga0070134_102117</td>
<td>69.9</td>
<td>Flavobacterium sp?</td>
<td>No</td>
<td>Lipase_GDSL_2 (PF13472), SGNH_hydro</td>
</tr>
<tr>
<td>AOC96206.1 (BB050_03116)</td>
<td>69.7</td>
<td>Flavobacterium johnsoniae</td>
<td>No</td>
<td>lipase (Flavobacterium sp. MEBO61)</td>
</tr>
<tr>
<td>WP_042562834.1</td>
<td>69.7</td>
<td>Flavobacterium sp. MEBO61</td>
<td>No</td>
<td>Lipase_GDSL_2 (PF13472), SGNH_hydro</td>
</tr>
<tr>
<td>SCZ00891 (IP94DRAFT_0083)</td>
<td>69.4</td>
<td>Flavobacterium anhuiense CGMCC</td>
<td>No</td>
<td>Lipase_GDSL_2 (PF13472), SGNH_hydro</td>
</tr>
<tr>
<td>EJG00033.1 (FF52_18088)</td>
<td>69.4</td>
<td>Flavobacterium sp. F52</td>
<td>No</td>
<td>Lipase_GDSL_2 (PF13472), SGNH_hydro</td>
</tr>
<tr>
<td>WP_057118157.1</td>
<td>68.9</td>
<td>Flavobacterium sp. Root935</td>
<td>No</td>
<td>Lipase_GDSL_2 (PF13472), SGNH_hydro</td>
</tr>
<tr>
<td>HMPREFG0204_3619</td>
<td>57</td>
<td>Chryseobacterium gleum ATCC 35910</td>
<td>No</td>
<td>Lipase_GDSL_2 (PF13472), SGNH_hydro</td>
</tr>
<tr>
<td>WP_062669866.1</td>
<td>55.3</td>
<td>Chryseobacterium sp. GSE06</td>
<td>No</td>
<td>Lipase_GDSL_2 (PF13472), SGNH_hydro</td>
</tr>
<tr>
<td>WP_010663709.1</td>
<td>53.9</td>
<td>Marinilabilia salmonicolor</td>
<td>No</td>
<td>Lipase_GDSL_2 (PF13472), SGNH_hydro</td>
</tr>
<tr>
<td>WP_002978728.1</td>
<td>53.8</td>
<td>Chryseobacterium gleum</td>
<td>No</td>
<td>Lipase_GDSL_2 (PF13472), SGNH_hydro</td>
</tr>
<tr>
<td>WP_040994537.1</td>
<td>53.5</td>
<td>Chryseobacterium oranimense</td>
<td>No</td>
<td>Lipase_GDSL_2 (PF13472), SGNH_hydro</td>
</tr>
<tr>
<td>WP_047423182.1 (EW79DRAFT_2436)</td>
<td>53.3</td>
<td>Chryseobacterium sp. YR480</td>
<td>No</td>
<td>Lipase_GDSL_2 (PF13472), SGNH_hydro</td>
</tr>
<tr>
<td>WP_062696765.1</td>
<td>53.1</td>
<td>Chryseobacterium indologenes</td>
<td>No</td>
<td>Lipase_GDSL_2 (PF13472), SGNH_hydro</td>
</tr>
<tr>
<td>WP_010422885.1 (ATH1DRAFT_02676)</td>
<td>52.5</td>
<td>Anaerophaga thermohalophila</td>
<td>No</td>
<td>Lipase_GDSL_2 (PF13472), SGNH_hydro</td>
</tr>
<tr>
<td>ADF53345.1 (ZPR_3026)</td>
<td>52.5</td>
<td>Zunongwangia profunda SM-A87</td>
<td>No</td>
<td>Lipase_GDSL_2 (PF13472), SGNH_hydro</td>
</tr>
<tr>
<td>KUR67571.1</td>
<td>52.5</td>
<td>Rhodothermaceae bacterium RA</td>
<td>No</td>
<td>Lipase_GDSL_2 (PF13472), SGNH_hydro</td>
</tr>
<tr>
<td>ACY49515.1 (Rmar_2643)</td>
<td>52.4</td>
<td>Rhodothermus marinus DSM 4252</td>
<td>No</td>
<td>lipolytic protein G-D-S-L family (Rhodothermus marinus DSM 4252)</td>
</tr>
<tr>
<td>ACY49515.1 (Rmar_2643)</td>
<td>52.4</td>
<td>Rhodothermus marinus DSM 4252</td>
<td>Yes</td>
<td>Lipase_GDSL_2 (PF13472), SGNH_hydro</td>
</tr>
<tr>
<td>WP_027586161.1 (T426DRAFT_00687)</td>
<td>51.6</td>
<td>Prolibacter bellariivorans</td>
<td>No</td>
<td>Lipase_GDSL_2 (PF13472), SGNH_hydro</td>
</tr>
<tr>
<td>BAP32337.1 (CHO_3300)</td>
<td>51.3</td>
<td>Chryseobacterium sp. StrR126</td>
<td>No</td>
<td>Lipase_GDSL_2 (PF13472), SGNH_hydro</td>
</tr>
<tr>
<td>WP_016777808.1 (QOSDRAFT_02396)</td>
<td>51.3</td>
<td>Anaerophaga thermohalophila</td>
<td>No</td>
<td>Lipase_GDSL_2 (PF13472), SGNH_hydro</td>
</tr>
<tr>
<td>GAT62571.1</td>
<td>50.7</td>
<td>Paludibacter jiangxiensis</td>
<td>No</td>
<td>Lipase_GDSL_2 (PF13472), SGNH_hydro</td>
</tr>
<tr>
<td>MROS_2085</td>
<td>50.1</td>
<td>Alkaliflexus imshenetskii DSM 15055</td>
<td>No</td>
<td>Lipase_GDSL_2 (PF13472), SGNH_hydro</td>
</tr>
<tr>
<td>AlkimDRAFT_0871</td>
<td>47</td>
<td>AlkimDRAFT_0871</td>
<td>Yes</td>
<td>Lipase_GDSL_2 (PF13472), SGNH_hydro</td>
</tr>
</tbody>
</table>

Due to the low sequence similarity between SGNH hydrolases and their wide substrate specificity, prediction based on phylogenetic relationship is not reliable. However, there is high structural homology between SGNH hydrolase members. The SGNH family members are characterized by the presence of central five stranded parallel β-sheet and lack of the nucleophilic elbow motif (Martinez-Martinez et al. 2008). For example, rhamnogalacturonan acetyl esterase (RGAE) from *Aspergillus aculeatus* from CE12 contains conserved SGNH sequence block, Ser-His-Asn catalytic triad and adopts α/β/α fold structure with five-stranded parallel β-sheet surrounded by α helices (Mølgaard, Kauppinen, and Larsen 2000). Several plant lipases/esterases from SGNH family have been characterized and shown to be involved in plant development and defense response against environmental stresses but characterization of bacterial SGNH
hydrolases is limited (Yu et al. 2016).

The three-dimensional structure of FjoAcXE has not been resolved yet but protein modeling using PHYRE2.0 (Figure 17) showed that FjoAcXE has predicted 3D structure similar to CE2 or CE12 families (100% confidence in the model but 13-18% sequence similarity). Predicted secondary structure of FjoAcXE is shown in Appendix D. FjoAcXE appears to have α/β/α fold structure with five-stranded parallel β-sheet surrounded by α helices, as well as CBM-like structure at the N-terminus. Members of CE2 family typically have two domains, comprised of N-terminal jellyroll domain and C-terminal SGNH domain (Nakamura, Nascimento, and Polikarpov 2017), which matches the predicted structure of FjoAcXE.

Figure 17: Structural model of FjoAcXE based on c4xhA, c2o14A, d1snra2 and c2w9xA as templates. The final model shows predicted loop regions in green, α-helices in red, β-sheets in yellow and predicted active residues as part of Ser-His-Asp catalytic triad in gray. The model was built with PHYRE2 (82% modeled at >90% confidence) and displayed with PyMOLv1.74.
The jelly-roll domain may be involved in carbohydrate binding (Till et al. 2013; Montanier et al. 2009). Similarly, SGNH RGAE from *Bacillus halodurans* belonging to CE12 was shown to have a putative substrate binding region at the N-terminus with immunoglobulin-like β sandwich fold, which could represent cellulose or xylan binding-like domain (Navarro-Fernández et al. 2008). Although there is no consensus sequence for this binding domain aside from the presence of aromatic residues (tyrosines and tryptophans), it is possible that FjohAcE has a similar N-terminal binding domain. However, affinity gel electrophoresis (AGE) using wide range of substrates did not show any binding capacity for FjoAcXE (Figure 9).
4. Chapter 2: Production and characterization of diverse GH115 α-glucuronidases from metagenome sequences

4.1. Introduction

Metagenomic analysis of microbial communities is a source for discovery of new and novel CAZymes relevant for modification and/or breakdown of lignocellulosic biomass. Comparative analysis of microbial communities from grass-fed mammalian herbivores, such as cow (Hess et al. 2011), with wood-degrading organisms, such as termites (Warnecke et al. 2007) could provide insight on how these microbial communities differ and help to identify relevant biomass degrading enzymes. Additionally, enrichments of these microbial communities on specific substrates could help to select for CAZymes that are particularly important for deconstruction of a given polysaccharide.

Microbial communities from the Canadian beaver (Castor Canadensis) droppings and North American moose (Alces americanus) rumen enriched on cellulose and poplar hydrolysate were shown to display substrate-driven convergence (Wong et al. 2016). This indicates that the composition of microbial communities shift in response to the type of lignocellulosic material available to them, making studies like this especially valuable source for discovery of new and novel CAZymes.

Four predicted α-glucuronidases from family GH115 were selected from beaver droppings enriched on cellulose and poplar hydrolysate, and were recombinantly produced and biochemically characterized in terms of specific activity and synergistic action with FjoAcXE.

4.2. Materials and Methods

4.2.1. Materials

Aldouronic acid mixture of aldotriouronic, aldotetraouronic and aldopentaouronic acids (2:2:1) was purchased from Megazyme (Cat# O-AMX). MeGlcpA-substituted xylo-oligosaccharides (U^{4m2}XX and XU^{4m2}XX) were a generous gift from Dr. Thu Vuong (University of Toronto, Canada). Beechwood xylan (X0627), oat spelt xylan (X0627),
gum arabic from acacia tree (G9752) were purchased from Sigma. Spruce
arabinoglucuronoxylan was a generous gift from Professors Paul Gatenholm and
Professor Guillermo Toriz (Chalmers University of Technology, Sweden). Hot water
extracted samples from mixed hardwood was kindly provided by Professor Bradley
Saville (University of Toronto). All other substrates and reagents are as described in
Section 4.2.1.

AxyAguGH115A α-glucuronidase from *Amphibacillus xylanus* (BAM48432.1) and FjoAcXE from *Flavobacterium johnsoniae* UW101 (ABQ06890.1) were purified as
previously described (Yan et al. 2017) and as in Section 4.2.4.

### 4.2.2. Candidate selection and phylogenetics

Metagenomic sequences from microcosms from anaerobic granules (AG),
Canadian beaver (*Castor Canadensis*) droppings (BD), and North American moose
(*Alces americanus*) rumen (MR), enriched on cellulose (C) and polar hydrolase (PH)
(Wong et al. 2016) were searched for predicted α-glucuronidase belonging to family
GH115. The obtained 26 sequences were sorted based on the following criteria: 1) evidence of secretion as predicted by Signal P4.1 (Petersen et al. 2011); 2) presence of
conserved E216 and D335 residues (Wang et al. 2016); 3) presence of multimodular
domains; 4) number of GH115 domains within the sequence (Wang et al. 2016); 5) whether or not they are part of PUL; 6) uniqueness of the phylogenetic clades.

Phylogenetic analysis was performed as follows: all GH115 sequences listed on
CAZy database (http://www.cazy.org) were downloaded (as of April 7, 2017) and
combined with the metagenomic sequences and aligned with MAFFT v.7.0.17 using
FFT-NS-i x1000 algorithm (Katoh et al. 2002). The alignment was inspected, incomplete
or problematic sequences were removed, and remaining sequences were re-aligned with
MAFFT using the same settings. The tree was built using the Neighbor-Joining method
combined with the Jukes-Cantor genetic distance model with 500 bootstrap replicates
using Geneious v8.1.9 (Kearse et al. 2012). The tree was visualized and color-coded
using interactive Tree of Life v3.4 (Letunic and Bork 2016).
4.2.3. Gene synthesis and molecular cloning

The genes encoding selected candidates and lacking the predicted signal sequence were codon optimized for expression in *Escherichia coli* K12 using IDT Codon Optimization Tool (www.idtdna.com/CodonOpt). Fifteen base pair extensions homologous to the p15TV-L vector (GenBank ID: EF456736.1) (T7, 5’-TTGTATTTCCAGGGC and T7term, 5’-CAAGCTTCGTCATCA) were added to each end of the sequence, and depending on the size of the genes, they were synthesized as single, two or three gBlock® Gene Fragments (IDT) with 20 base pair internal overlaps. The gene fragments were assembled and cloned into p15TV-L vector using Gibson Assembly® Master Mix (E2611S, New England Biolabs Inc.), following manufacturer’s specification. The product of the reaction was transformed into *E. coli* HST08 Stellar™ Competent Cells (Clontech), and the sequence was verified using DNA sequencing service at the Center of Applied Genomics at the SickKids Hospital in Toronto.

4.2.4. Protein expression and purification

BD-PH_46267_3 (META1), BD-PH_45034_39 (META2), BD-PH_44413_18 (META3) and BD-C_16082_12 (META4) were expressed and purified as described in Section 4.2.4 with a few modifications. After the proteins were expressed, initial purification was done by HisTrap HP, META1 was further purified using UNO™Q ion exchange column, followed by another round of purification with HisTrap HP. META2 was further purified using size exclusion column (Supradex 200 16/60 µL). META4 did not require additional purification after the first round of purification on HisTrap HP.

4.2.5. In-gel trypsin digest and mass spectrometry

The identities of purified proteins were tested by in-gel trypsin digest and mass spectrometry as described in Section 4.2.5.

4.2.6. Activity towards aldouronic acid mixture

Initial testing of predicted α-glucuronidases META1, META2 and META4 were done using mixture of aldotriouronic (UX), aldotetraouronic (UXX) and
aldopentaouronic (UXXX) acids (2:2:1) (Megazyme). Reactions (50 µL reaction volume) containing 20 mM of aldouronic acid mixture in 200 mM universal buffer (67 mM H₃BO₃, 67 mM H₃PO₄, 67 mM CH₃COOH) (pH 6.0) were initiated by adding 5 µg of each protein, and continued for 20 min at 30°C. Reactions were done in duplicates. The samples were spun down and released D-glucoronic acid was measured using the D-glucuronic acid/D-galacturonic acid assay kit (Megazyme), following the manufacturer’s specifications. Reaction mixtures without enzyme served as blanks.

4.2.7. Activity towards pNP substrates

Reactions (200 µL reaction volume) comprised 50 mM HEPES (pH 6.0) and 2 mM of pNP-α-L-arabinopyranoside, pNP-β-D-celllobioside, pNP-α-L-fucopyranoside, pNP-β-D-glucopyranoside, pNP-β-D-lactopyranoside, pNP-β-D-mannopyranoside, pNP-β-D-xylopyranoside, pNP-acetate and pNP-acetate. Reactions were initiated by adding 5 µg of each protein, and continued for 2 hrs at 30°C, and absorbance was measured continuously at 410 nm. Reaction mixture without the enzymes was used as a blank.

4.2.8. Hydrolysis of polymeric substrates

Spruce arabinoglucuronoxylan (AGX), beechwood xylan (BEX), oat spelt xylan (OSX), and acacia gum arabinogalactan (AG) were used for initial testing of activity of META1, META2 and META4. Reactions (50 µL reaction volume) containing 1% (w/v) of each substrate in 50 mM HEPES (pH 7.0) were initiated by adding 1 µg of each protein, and continued for 16 hrs at 30°C. The samples were spun down and released D-glucoronic acid was measured using the D-glucuronic acid/D-galacturonic acid assay kit (Megazyme), following the manufacturer’s specifications. The reaction mixtures without enzyme were used as blanks.

4.2.9. Optimum pH screen for META1, 2, and 4

The pH optima of META1 and META2 were tested using 200 mM universal buffer with pH range of 3.5 to 10. Reactions (20 µL reaction volume) containing 16 mM of aldouronic acid mixture were initiated by adding 1 µg of each protein, and continued
for 20 min at 30°C. The samples were spun down and released D-glucoronic acid was measured using the D-glucuronic acid/D-galacturonic acid assay kit (Megazyme), following the manufacturer’s specifications. Reaction mixtures without enzyme were used as blanks.

For META4, reactions (200 µL reaction volume) containing 0.5 mM 4-MUA were initiated by adding 5 µg of the protein, and continued for 20 min at 30°C. Absorbance was measured at 354 nm and the reaction mixture without enzyme was used as a blank.

4.2.10. Activity of META1 and META2 on U^{4m2}XX and XU^{4m2} XX

Activity of META1 and META2 was tested on pure U^{4m2}XX and XU^{4m2} XX to test for substrate specificity. Reactions (30 µL reaction volume) comprised 50 mM HEPES (pH6.0) and 2 mM of U^{4m2}XX and XU^{4m2} XX. Reactions were initiated by adding 1 µg of each protein. Release of MeGlcPA was measured after 20 min and 4 hrs of incubation at 30°C using the D-glucuronic acid/D-galacturonic acid assay kit (Megazyme), following the manufacturer’s specifications. Reaction mixtures without enzyme were used as blanks.

4.2.11. Synergism between METAs and FjoAcXE

The impact of FjoAcXE on the activity of META1, META2 and AxyGH115 was tested using mixed xylo-oligosaccharides recovered from hot water extracted samples from mixed hardwood. Reaction mixtures (400 µL) comprised 1% (w/v) mixed xylo-oligosaccharides in 50 mM HEPES buffer (pH 7.0) with 10 µg of each protein. The reaction continued for 20 hrs at 30°C and gentle shaking. META4 was tested with and without FjoAcXE, as well as in combination with META1 and META2, as described above. The reaction mixture without enzyme was used as a control. The samples were spun down and 10 µL of each reaction was used to measure release of D-glucuronic acid using the D-glucuronic acid/D-galacturonic acid assay kit (Megazyme), following the manufacturer’s specifications.
4.2.12. $^1$H-NMR analysis towards acetylated saccharides

The remaining 390 µL reaction from Section 5.2.8 were used to perform one-dimensional $^1$H-NMR analysis. Briefly, the samples were filtered through Acrodisc® Syringe Filters with 0.2 µm Supor® membrane (Pall Corporation) and lyophilized. The samples were dissolved in 300 µL D$_2$O and transferred into 3 mm NMR tubes (Norell) for analysis on Agilent DD2 700 MHz spectrometer equipped with triple resonance HCN cold probe with a scan number of 64, relaxation delay of 1 s and acquisition time of 4.5 s. The data were obtained using VnmrJ 4.0 (Agilent) and analyzed with MestReNova 10.0 (Mestrelab Research). The D$_2$O peak at 4.790 was used as internal standard.

4.3. Results and discussion

4.3.1. Candidate selection, phylogenetics and sequence analysis

The criteria for selection of the candidates for production and characterization was as follows: evidence of secretion, presence of conserved E216 residue, presence of multimodular domains, number of GH115 domains within the sequence, whether or not they are part of PUL, and uniqueness of the phylogenetic clades. Using the criteria above, four predicted GH115 sequences from beaver droppings (BD) microcosms enriched on cellulose (C) and polar hydrolase (PH) were selected: BD-PH_46267_3 (META1), BD-PH_45034_39 (META2), BD-PH_44413_18 (META3) and BD-C_16082_12 (META4). See Appendix F for summary and sequences of the selected candidates.

A phylogenetic tree was constructed for all predicted and characterized GH115 sequences listed on CAZy database alongside with the metagenomic sequences for predicted GH115 sequences from Wong et al. (2016) (Figure 18). All four selected metagenomic candidates fall into unique clades that do not have previously characterized members. Additionally, all of the selected candidates had conserved E216 and D335 residues, which were shown to be essential for GH115 activity (Wang et al. 2016). Because these candidates are from metagenomes, their source organism is unknown, and therefore both Gram-negative and Gram-positive bacteria were considered when
predicting the presence of a signal sequence. META1, 2 and 4 had signal sequence and META3 did not when they were considered to be Gram-negative. However, when META3 was considered to be Gram-positive, a signal sequence was predicted. Due to the discrepancy in the signal sequence prediction, preference was given to Gram-negative bacteria as the gene encoding for META3 was predicted to come from Proteobacteria (a gram negative bacteria) (personal communication with Mabel T. Wong; Wong et al., 2016).

Figure 18: Phylogenetic tree for GH115 family. Previously characterized members are in red and candidates from metagenomic sequences are in blue. Sequences were aligned with MAFFT v.7.0.17 using FFT-NS-i x1000 algorithm and the tree was built using the Neighbor-Joining method combined with the Jukes-Cantor genetic distance model with 500 bootstrap replicates using Geneious v8.1.9. The tree was visualized and color-coded using interactive Tree of Life v3.4.

Sequence analysis of the candidates (Figure 19) showed that META1 is predicted five-domain α-glucuronidase due to the presence of C+ domain, which was shown to be important in substrate binding (Wang et al. 2016). META2 and META3 are four domain
proteins, lacking the C+ domain. However, META2 has around 600 additional amino acids following domain D. A protein domain identifier, IntroProScan (Quevillon et al. 2005), predicted this additional domain of META2 to be a galactose-binding domain-like sequence (IPR008979). It is also possible that META2 could represent a 5+ domain GH115 but involvement of the additional domain in substrate binding and its exact function has not been characterized. META2 is part of predicted PUL (BD-PH_PUL117, see Appendix B for PUL modularity), which based on the types of CAZymes present, could be active on galactoglucomannan or other complex hemicelluloses.

![Alignment of META candidates with all characterized GH115s showing GH115 domain.](image)

**Figure 19:** Alignment of META candidates with all characterized GH115s showing GH115 domain. META1 is five domain protein, META2 and 3 are four domains proteins, and META4 is missing D domain and only has 3 domains. Sequences were aligned with MAFFT v.7.0.17 using FFT-NS-i x1000 algorithm and displayed using Geneious v8.1.9.

META3 is also part of PUL (BD-PH_PUL97, Appendix B), which contains ten proteins of unknown function, and therefore it is hard to predict the possible function of this PUL. META3 has an additional domain following domain D (Figure 19) and is predicted to be an endo-xylanase from GH10 family (Pfam ID: PF00331). However, META3 could not be expressed and produced in *E. coli*, despite several attempts, and not discussed here further. Interestingly, META4 has only three domains (A, B and C), which has not been observed in previously characterized GH115s.

### 4.3.2. Protein expression and purification

META1, META2 and META4 were recombinantly expressed in *E. coli* and purified from 4L cultures using 6x-Histag at the N-terminus of the protein. Protein yields were 0.15 mg, 2.4 mg and 2.2 mg, respectively, where purity was greater than 95% as
assessed by SDS-PAGE (Figure 20). All proteins were identified correctly by in-gel trypsin digest and mass spectrometry with coverage from N to C terminus.

Figure 20: Recombinantly expressed META1, 2 and 4 were pure and of expected size. Protein sizes in brackets were calculated based on the amino acid sequence. 1 µg of each protein was loaded onto 10% SDS-PAGE.

4.3.3. META1 and 2 show α-glucuronidase activity, while META4 shows an esterase activity

Initial screening of META1, META2 and META4 was done using mixture of aldotriouronic (UX), aldotetrauronic (UXX) and aldopentaouronic (UXXX) acids (2:2:1) (O-AMX, Megazyme). META1 and META2 show activity on the aldouronic mixture as detected by the release of glucuronic acid (Figure 21), indicating that both are α-glucuronidases. META1 released approximately 45% more glucuronic acid than META2 under tested conditions. META4 showed no detectable activity on aldouronic acid mixture.
Figure 21: META1 and META2 are active on aldouronic acid mixture, while META4 showed no activity. The structure of aldouronic mixture is shown with GlcpA residue in blue. In 50 µL reaction, 5 µg of each protein tested on 20 mM aldouronic acids mixture of UX, UXX and UXXX (Megazyme) in 200 mM universal buffer pH 6.0 (67 mM boric/ 67 mM phosphoric / 67 mM acetic acids) for for 20 min at 30°C. Release of GlcpA was measured using K-URONIC kit from Megazyme.

4.3.4. Activity screen on pNP substrates showed META4 is an esterase

A more general screen using pNP substrates was done for META1, META2 and META4 using 2 mM of different pNP saccharides and pNP alkyl esters (Figure 22). As expected, META1 and META2 did not show activity on any of the tested pNP substrates. However, the META4 sample showed activity on pNP-acetate and pNP-hexanoate, suggesting that the enzyme may have esterase activity; alternatively, the protein sample may be contaminated with low levels of esterase encoded by the E. coli expression host. Notably, sequence alignment with previously characterized GH115 (Figure 19) showed that META4 is missing domain D.
Figure 22: pNP screen showed META4 has activity on pNP-acetate and pNP-hexanoate, and META1 and 2 do not show activity on any of the tested substrates. In 200 µL reaction volume, 5 µg of each protein was tested on 2 mM of each substrate in 50 mM HEPES (pH6.0). After 2 hrs incubation at 30°C, absorbance were measured at 410 nm.

Catalytic residues of previously characterized GH115 are found in domain B but domain D is predicted to be important for protein dimerization (Wang et al. 2016). Protein dimerization also appears important for enzyme activity and substrate preference. For example, the active site of the enzyme is predicted to form at the subunit interphase of the two subunits; dimerization may also lead to strong substrate binding (Marianayagam, Sunde, and Matthews 2004). The other domains (A, B and C) of META4 have predicted three-dimensional structure (model not shown) similar to that of BtGH115A (Aalbers et al. 2015), which is an α-glucuronidase with specificity for decorated arabinogalactans. Even though META4 has less than 18% sequence similarity with all of the previously characterized GH115s, and is part of a large uncharacterized GH115 clade (Figure 18), its predicted structural similarity to members of GH115 family supports META4 membership to family GH115 and rather than a CE family, for example. It has been proposed that certain CE families may have evolved from GH
families (Biely et al. 2014). However, additional protein preparations and mutagenesis will be required to clarify whether META4 truly displays esterase activity.

4.3.5. Hydrolysis of polymeric substrates

Activity of META1, META2 and META4 on polymeric substrates were tested using spruce arabinoglucuronoxylan (AGX), beechwood xylan (BEX), oat spelt xylan (OSX), and acacia gum arabinogalactan (AG). AGX, BEX and OSX were used as representatives of softwood, hardwood and cereal xylans (see Figure 2 and Appendix G for the structures), respectively, while AG was used because a member of the GH115 family has been shown to be active on arabinogalactans (Aalbers et al. 2015). AGX is a more complex substrate because of the presence of arabinofuranose in addition to (Me)GlcA substitutions. AGX contains on average one (Me)GlcA per every 5 or 6 Xylp, while glucuronoxylan from BEX contains one (Me)GlcA per every 4 to 16 Xylp (Ebringerová, Hromádková, and Heinze 2005). OSX is (glucurono)arabinoxylan that contains one (Me)GlcA per every 24 Xylp (Kormelink and Voragen 1993).

META1 showed highest activity on AGX, followed by BEX and OSX (Figure 23); it was not active on AG. Whereas META2 only showed low activity on BEX, META4 was not active on any of the tested substrates. Higher activity of META1 on AGX compared to BEX can be explained by higher (Me)GlcA present in AGX, and low activity of META1 on OSX likely reflects the lower (Me)GlcA content, as well as low substrate solubility. The low activity of META2 towards BEX and no activity on AGX and OSX suggest that the presence of arabinofuranose side groups in AGX and OSX introduce steric hindrance to the enzyme. Additionally, the significantly lower activity of META2 on BEX compared to META1 suggests that META2 prefers short oligosaccharides. The impact of the additional domain on the C-terminal of META2 remains unclear (Figure 19).
Figure 23: Activity of META1, META2 and META4 on 1% (w/v) beechwood xylan (AGX), oat spelt xylan (OSX), spruce arabinoglucuronoxylan (AGX) and acacia gum arabinogalactan (AG). In 50 µL reaction volume containing 50 mM HEPES pH 7.0, 1 µg of each protein was tested 16 hrs at 30°C.

4.3.6. Optimum pH screen for META1, 2, and 4

Aldouronic acid mixtures were used to measure the optimum pH of META1 and META2. By contrast, 4-MUA was used to measure the optimum pH of META4 as 4-MUA is more stable than pNP-acetate at alkaline conditions. The pH optimum of both META1 and META2 was pH 6.0, similar to previously characterized GH115s (Figure 24). The pH optimum of META4 was pH 8.0, similar to FjoAcXE and other esterases. It should be noted that even though 4-MUA is more stable than pNP-acetate, there is still some degree of autohydrolysis of the substrate at pH 8.0 and higher. Hence, the significant drop in activity of META4 above pH 9.0 may be an artifact of the high background absorbance at this pH condition.
META1 and META2 were tested on 16 mM aldouronic acid mixture for 2 hrs incubation at 30°C. Release of glucuronic acid (GlcA) was measured using K-URONIC kit from Megazyme. META4 is not active on aldouronic acid mixture, so 1 µg of META4 was tested on 5 mM 4-MUA for 30 min at 30°C, absorbance was measured at 354 nm.

4.3.7. META1 and META2 are active on \( U^{4m2}XX \) and \( XU^{4m2}XX \)

Substrate specificity of META1 and META2 towards terminally and internally substituted xylo-oligosaccharides were tested on aldotetraouronic \( (U^{4m2}XX) \) and aldopentaouronic \( (XU^{4m2}XX) \) acids. Aldotriouronic \( (U^{4m2}X) \) in its pure form was not available for this test. \( \alpha \)-Glucuronidases from family GH67 act only on terminally substituted (Me)Glcp residues, while \( \alpha \)-glucuronidases from family GH115 are able to act on both internally and terminally substituted (Me)Glcp residues. After 20 min, both META1 and META2 showed preference for \( XU^{4m2}XX \); however, nearly equal extents of (Me)Glcp release from \( U^{4m2}XX \) and \( XU^{4m2}XX \) was observed after 4 hrs. This result illustrates the preference of META1 and META2 for internally substituted xylo-oligosaccharides (Figure 25), consistent with activities reported for family GH115 \( \alpha \)-glucuronidases rather than GH67 \( \alpha \)-glucuronidases.
Figure 25: META1 and META2 are active on both terminally (U\textsuperscript{4m2XX}) and internally (XU\textsuperscript{4m2XX}) substituted xylo-oligosaccharides. In 30 µL reaction volume, 1 µg of each protein was tested on 2 mM of U\textsuperscript{4m2XX} and XU\textsuperscript{4m2XX} in 50 mM HEPES (pH6.0). Release of MeGlc\textsubscript{p}A was measured after 20 min and 4 hrs of incubation at 30°C.

4.3.8. Synergism between METAs and FjoAcXE and confirmation by \textsuperscript{1}H-NMR analysis

To further explore synergism of FjoAcXE (from Section 4) with members of GH115 family, FjoAcXE was tested with META1, META2 and META4 using mixed xylo-oligosaccharides recovered from hot water extracted mixed hardwood. This substrate represents an industrially relevant, minimally processed glucuronoxylan fraction. The exact composition of this substrate is not known but is predicted to be mixture of acetylated glucuronoxylan, and acetylated xylo-oligosaccharides with (Me)Glc\textsubscript{p}A substitutions, similar to acetylated xylo-oligosaccharides used for synergism studies reported in Section 4.
When tested on their own, META1 and META2 release modest amounts of (Me)Glc\(_p\)A (Figure 26). However, in the presence of FjoAcXE, a 4-fold and 3-fold increase in (Me)Glc\(_p\)A release is observed for META1 and META2, respectively. Similar results were obtained for AxyAgu115A tested in the presence of FjoAcXE, where the increase in (Me)Glc\(_p\)A release was 3.5-fold, compared to when AxyAgu115A was used alone. Because META4 showed an esterase activity, it was tested on combination with META1, META2 and AxyAgu115A but there was no synergism observed.

The activity measurements were confirmed with \(^1\)H-NMR, which showed that FjoAcXE is able to remove acetyl groups from all of the position on Xyl\(_p\) as observed by the disappearance of peaks corresponding to methyl protons of the acetyl groups: Xyl-3Ac-2MeGlc\(_p\)A (2.26 ppm), Xyl-2Ac (2.14 ppm), Xyl-3Ac (2.13 ppm) and Xyl2,3Ac (2.09 ppm), (Figure 27). The peak corresponding to H1 proton of MeGlc\(_p\)A (5.32 ppm) decreases slightly in the presence of AxyAguGH115 but when FjoAcXE is also added,
that peak disappears completely, indicating an effective removal of most (if not all) of MeGlc₆A. A very modest decrease in the same peak was observed when META1 was combined with FjoAcXE and possibly the same is true for META2. Consistent with the measured release of soluble MeGlc₆A (Figure 26), the ¹H-NMR analysis showed that META1 and META2 are not as active as AxyAgu115A as measured by the release of (Me)Glc₆A at the end of the reaction. Additionally, ability of FjoAcXE to deacetylate O-3 position of 3-O-acetylated Xylp that is also 2-O substituted with (Me)Glc₆A was confirmed as observed by disappearance of a peak at 5.10 ppm corresponding to H3 proton of Xyl-3Ac-2MeGlc₆A. These observations match results obtained in Chapter II.
Figure 27: $^1$H-NMR spectrum of FjoAcXE tested on 1% (w/v) hot water extracted mixed hardwood. Red arrow shows unique activity of FjoAcXE to deacetylate 3-O-acetylated Xylp 2-O-substituted with (Me)GlcA. Blue arrow shows synergistic effect of FjoAcXE with AxyAgu115A. Tested in 50 mM HEPES pH7.0 for 20 hrs at 30°C, followed by lyophilization and suspension in 0.3 ml D$_2$O.
5. Engineering significance

Xylans, and hemicellulose in general, have a great potential for production of biofuel and high-value bioproducts. However, their application is hindered by their structural diversity and complexity. Enzymatic modification of xylan and its oligosaccharides can help to improve their utilization by controlling their physico-chemical properties. FjoAcXE with its ability to deacetylate xylose residues at all positions, and as a result, its synergism with α-glucuronidases, can be used to promote access of xylanases to the xylan backbone and promote saccharification and control polymer solubility.

Debranching enzymes from this study have potential to alter xylan structure, and as a result modify xylan solubility, rheology and its ability to interact and bind to other biopolymers, such as cellulose, allowing production of high-value bioproducts, while reducing our dependency on petroleum derived products.

6. Conclusion and future considerations

In this study, a protein of unknown function from the polysaccharide utilization loci database (www.cazy.org/PULDB), was recombinantly produced and characterized to be an acetyl xylan esterase. Search for FjoAcXE homologs revealed a large number of uncharacterized proteins, which may suggest that they belong to a new group of carbohydrate esterases. A novel activity not previously observed in other carbohydrate esterases was observed, as FjoAcXE was able to deacetylate 3-O-acetylated internal Xylp residues that are also 2-O substituted with (4-O-methyl)glucuronic acid. Deacetylation of the 3-O position of Xylp allowed α-glucuronidases from families GH67 and GH115 to access previously inaccessible (4-O-methyl)glucuronic acid at the 2-O position on the same Xylp residue. Additionally, three α-glucuronidases from metagenomic sequences from beaver droppings enriched on cellulose and poplar hydrolysate were produced and characterized, and an α-glucuronidase with unexpected esterase activity was discovered.

Although glucuronoxylan, a main hemicellulose from hardwood, can be effectively debranched by the enzymes reported in this study, there are other xylans and
hemicelluloses that have not been tested with these enzymes. For example, xylans from softwood and cereals, aside from (Me)Glc\(_p\)A, may be also be substituted with \(\alpha\)-L-arabinofuranosyl (Araf) groups. Araf groups occurs more often at O-3 position of Xyl\(_p\) but O-2 and O-2,3 may be substituted as well. Testing for synergistic action between FjoAcXE and \(\alpha\)-L-arabinofuranosidase (ex. from family GH43 or GH62) on residues that are 2-O-acetylated and 3-O substituted with Araf will provide more information on the range of substrates FjoAcXE is active on. Additionally, members of SGNH hydrolase family usually display substrate promiscuity, and FjoAcXE as a member of this family might have other activities in addition to acetyl esterase activity. Therefore, additional tests for feruloyl esterase (removes ferulate group) and glucuronyl esterase (able to act on ester linkages between (Me)Glc\(_p\)A and lignin) activities need to be done. Determination of the FjoAcXE crystal structure may provide insight into the catalytic mechanism of the enzyme and help with classification of FjoAcXE into existing or new CE family.

Finally, META1, META2 and META4 \(\alpha\)-glucuronidases need to be biochemically characterized further in terms of their pH and thermostability and kinetics parameters. In particular, further characterization of the META4 esterase activity may provide insight into polyspecificity within the GH115 family.
References


from Trichoderma Reesei RUT C-30.” *Enzyme and Microbial Technology* 16 (9): 813–19. doi:10.1016/0141-0229(94)90041-8.


Williams, Allison H., Frederic J. Veyrier, Mathilde Bonis, Yann Michaud, Bertrand


## Appendices

### Appendix A: Experimentally validated PULs containing at least two of the CAZyme families of interest (as of April 2015).
Three unks from PUL73 and one unk from PUL20 were selected for characterization (in light green). Unk, proteins of unknown function; HTCS, hybrid two-component system; SusC and SusD, starch utilization system C- and D-like proteins (SusC and SusD); CE, carbohydrate esterase; GH, glycoside hydrolase (GH); CBM, carbohydrate binding module; PL, polysaccharide lyase. Vertical slash between two modules that encode for multimodule CAZymes, and the arrows show whether the gene is found on the sense (↑) and antisense (↓) strand.

<table>
<thead>
<tr>
<th>PUL#</th>
<th>Modularity as per PULDB</th>
<th>CAZyme families of interest</th>
<th># of unks</th>
<th>Signal Seq</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>HTCS ▶ unk ▶ unk ▶ SusC ▶ SusD ▶ SusD ▶ unk ▶ GH10 ▶ PL9 ▶ GH115 ▶ GH2 ▶</td>
<td>GH10 GH115</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>unk ▶ unk ▶ HTCS CBM35</td>
<td>GH ▶ SusC ▶ SusD ▶ CE12 ▶ CE4 ▶ GH43 ▶ GH115 ▶</td>
<td>GH43 GH115</td>
<td>2</td>
</tr>
<tr>
<td>44</td>
<td>unk ▶ HTCS ▶ GH43 ▶ unk ▶ unk ▶ unk ▶ unk ▶ SusC ▶ SusD ▶ GH115 ▶ unk ▶ GH105 ▶ unk ▶ GH43 ▶ unk ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ unk ▶ GH2 ▶ GH130 ▶ GH2 ▶ GH105 ▶ GH105 ▶ unk ▶ unk ▶</td>
<td>GH43 GH115</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>73</td>
<td>unk ▶ unk ▶ GH43 ▶ GH3 ▶ unk ▶ GH43 ▶ GH31 ▶ GH97 ▶ CBM6</td>
<td>GH43 ▶ SusC ▶ SusD ▶ SusD ▶ unk ▶ GH10 ▶ GH30_8 ▶ CBM35 ▶ GH98 ▶ GH115 ▶ CE1 ▶ CE6 ▶ CBM6 ▶ GH43 ▶ ▶ HTCS ▶ GH95 ▶ unk ▶ unk ▶ HTCS ▶ GH30 ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ unk ▶ ▶ unk ▶ GH115 ▶ unk</td>
<td>GH10 GH43 GH115 CE6</td>
<td>10</td>
</tr>
<tr>
<td>20</td>
<td>GH3 SusC ▶ SusD ▶ GH43 ▶ unk ▶ CE1 ▶ CE6 ▶ GH115 ▶ unk ▶ GH43 ▶ GH3 ▶ GH97 ▶ HTCS ▶ SusC ▶ SusD ▶</td>
<td>GH115 GH43 CE6</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>93</td>
<td>HTCS SusC ▶ SusD ▶ unk ▶ CBM4</td>
<td>GH10 ▶ unk ▶ unk ▶ GH10 ▶ GH43 ▶ GH67 ▶</td>
<td>GH43 GH10</td>
<td>3</td>
</tr>
<tr>
<td>77</td>
<td>GH106 ▶ GH28 ▶ unk ▶ GH28 ▶ GH28 ▶ CE12 ▶ GH2 GH42</td>
<td>GH43 ▶ GH28 ▶ CE4 ▶ GH28 ▶ ▶ GH2 ▶ CBM35 ▶ GH27 ▶ unk ▶ unk ▶ GH35 ▶ unk ▶ unk ▶ unk ▶ unk ▶ SusC ▶ SusD ▶ unk ▶ unk ▶ SusC ▶ unk ▶ PL9_1 ▶ unk ▶ unk ▶ unk ▶ CE12 ▶ GH105 PL11_1 ▶ GH105 ▶ unk ▶ unk ▶ HTCS ▶ CE4 ▶ CE6 ▶ ▶ GH2 ▶ HTCS PL9_1 ▶</td>
<td>GH43 CE6</td>
<td>12</td>
</tr>
<tr>
<td>19</td>
<td>GH3 ▶ HTCS ▶ SusC ▶ SusD ▶ unk ▶ GH10 ▶ ▶ GH16 ▶ GH8 ▶ unk ▶</td>
<td>GH10 GH8</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

### Appendix B: META2 and META3 (in red) are part of PUL117 and PUL97, respectively, from beaver droppings microcosms enriched on poplar hydrolysate (BD-PH).
Appendix C: Production and characterization of unks from PUL73

PUL73 has been experimentally validated and was predicted to be active on wheat arabinoxylan and oat spelt xylan (Martens et al. 2011). BACOVA_03430 (or unk1) has domain of unknown function (DUF1735; PF08522 - this domain of unknown function is found in a number of bacterial proteins including acylhydrolases) and has completed 3D structure (PDB code 3N91). Blast searches indicate that unk1 could be an adhesin. BACOVA_03440 (or unk2) has no known Pfam domains. BACOVA_03447 (or unk3) is a hypothetical protein and predicted 3D model shows similarity with CMB6 (it is also possible that unk3 sequence is partial/incomplete). The gene synthesis, cloning and protein expression and purification of unk1, 2 and 3 were done as described for FjoAcXE in Section 4. Unk2 could not be recombinantly expressed in E. coli despite several trials (using different induction temperatures (16°C and 37°C) and different IPTG concentrations (0.1 mM, 1 mM and 5 mM). The obtained unk1 and unk3 proteins were >90% pure (Fig C1).

**Figure C1.** Unk1 and unk3 were purified with HisTrap column and polished with UNOQ anion exchange column. Calculated sizes: unk1 is around 38.4 kDA and unk3 is around 16.2 kDA (unk3 runs around 14 kDA on SDS-PAGE gel).

Initial screen on pNP and polysaccharide substrates using Nelson Somogyi assay (Figure C3) showed some activity on the tested substrates but activity could not be assigned with a confidence. Unk1 and Unk3 need to be characterized further.
**Figure C2. Unk1 and unk3 screen on pNP substrates and polysaccharides.**

A) **pNP screen for unk1 and unk3.** In total of 200 μL reaction, 2 mM of each pNP substrate was incubated with 5 μg of either unk1 or unk3 in 50 mM HEPES pH7.0. The samples were incubated for 2 hrs at room temperature and absorbance was measured at 410 nm. Reaction mixture without the proteins were used as blanks. B) **Nelson Simogyi reducing sugar assay was used to test for the activity of unk1 and unk3 (n=2).** In total of 100 μL reaction, 5 μg of each of unk1 and unk3 were added to 0.5% (w/v) of each polysaccharides in 50 mM HEPES pH6.5 buffer containing 5mM MgCl$_2$. The samples were incubated at 37°C for 24 hrs. For the blank samples, water was added instead of the proteins. All reagents for the Nelson-Somogyi assay were prepared as previously described (Mcclavey 1999). After the incubation period, 75 μl of Reagent D was added and the samples were incubated at 100°C for 20 min. The samples were cooled to a room temperature and then 75 μl of Reagent E was added and mixed. The samples were spun at 1000xg for 5 min, and 200 μL of the reaction was used to read absorbance at 520 nm.

Unk1 and unk3, as well as FjoAcXE were tested for ability to bind to polysaccharides using affinity gel electrophoresis (Figure C4) but no binding on any of the substrates tested was observed. Function of unk1 and unk3 remain unknown.
Figure C4: Affinity gel electrophoresis. 5 µg of each protein was run on 7.5% (w/v) native polyacrylamide gel (25 mM Tris, 250 mM glycine buffer pH8.3) containing 0.01% of each substrate for 2 hrs at 90V. Gray dotted bar indicates migration distance of bovine serum albumin, which can be used as a reference to compare distance traveled by the tested proteins in relation to the no substrate control gel. MGX = 4-O-Methylglucuronoxylan; OSX = oat spelt xylan; CMC = carboxymethylcellulose; WAX = wheat arabinoxylan.
Appendix D: Predicted secondary structure of FjoAcXE using the PSIPRED protein sequence analysis workbench (McGuffin, Bryson, and Jones 2000)
Appendix E: Summary and sequences for the candidates selected from PULDB

<table>
<thead>
<tr>
<th>Name</th>
<th>Unk1</th>
<th>Unk2</th>
<th>Unk3</th>
<th>Unk4/FjoAcXE</th>
</tr>
</thead>
<tbody>
<tr>
<td>GenID</td>
<td>BACOVA_03430</td>
<td>BACOVA_03440</td>
<td>BACOVA_03447</td>
<td>Fjoh_3879</td>
</tr>
<tr>
<td>Organism</td>
<td>Bacteroides ovatus ATCC 8483</td>
<td>Flavobacterium johnsoniae UW101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SignalP</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Mature chain</td>
<td>23-344</td>
<td>21-155</td>
<td>19-139</td>
<td>22-414</td>
</tr>
<tr>
<td># of glycosylation sites</td>
<td>1</td>
<td>2</td>
<td>None</td>
<td>3</td>
</tr>
<tr>
<td>Amino acids</td>
<td>343</td>
<td>156</td>
<td>142</td>
<td>414</td>
</tr>
<tr>
<td>Size (signal seq cut + Histag on N-terminal) in kDa</td>
<td>38.38088</td>
<td>17.35749</td>
<td>16.19919</td>
<td>45.28108</td>
</tr>
<tr>
<td>PI</td>
<td>5.46</td>
<td>7.01</td>
<td>5.96</td>
<td>6.4</td>
</tr>
<tr>
<td>Predicted function</td>
<td>?Acylhydrolase</td>
<td>?</td>
<td>?CBM</td>
<td>Esterase</td>
</tr>
<tr>
<td>PUL</td>
<td>PUL73 and PUL114</td>
<td>PUL73 and PUL114</td>
<td>PUL73 and PUL114</td>
<td>PUL20</td>
</tr>
<tr>
<td>Pfam domain</td>
<td>PF0852</td>
<td>None</td>
<td>None</td>
<td>PF13472</td>
</tr>
<tr>
<td>Other domains</td>
<td>DUF1735</td>
<td>hypothetical protein</td>
<td>hypothetical protein</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Structure</td>
<td>YES (3N91)</td>
<td>--</td>
<td>In progress</td>
<td>In progress</td>
</tr>
<tr>
<td>Sequenced clone in DH5alpha</td>
<td>Clone1.3P</td>
<td>Clone 2P</td>
<td>Clone 3.3</td>
<td>Clone 4.1</td>
</tr>
<tr>
<td>Clone in BL21</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Expressed in E.coli?</td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

Original amino acid sequence:

- **Unk1**: MKKNLAIYGLVLILTTWTSCEDSNFDPDFTYQTVYFANQYQLRTILGGESEFDVNTLTDNQKVMIKAAPONGGTYRNNNVTIVKDESLCDNLNFKTDQJPLVMPASSYTLADRIAIPIQGMAGVEVQLTDFFADEKSIENYVPLLMTNVQGADSILQGKVPVENPVT
- **Unk2**: MKKYLVALVACLGLILSVAQDVKTQESVQCEANNLTVGQTLTSLTSGNPRTYGYSRGDSETQITYDHHAYQQMGXNTKNDL5YHIVHLEIEATDNNATNAICSLKIEYQTGDKTVKSMGNRVEYLYEGLEKSTIGCIDAGTKEVTYFGKKKVFQSAFDEKAGIAIDFPVTCEHLIDKVAIYIAKSVGKTVK
- **Unk3**: SITKSVSIMGKNNKVINDNNLVAK
- **Unk4/FjoAcXE**: MKKYLYFCLFAAANAAKEEPIQLFMRLNLLVFCLAAANAAKEEPIQLFMPRTGRFCGSETQQITYDHHAYQQMGXNTKNDL5YHIVHLEIEATDNNATNAICSLKIEYQTGDKTVKSMGNRVEYLYEGLEKSTIGCIDAGTKEVTYFGKKKVFQSAFDEKAGIAIDFPVTCEHLIDKVAIYIAKSVGKTVK
### Appendix F: Summary and sequences for the candidates selected from metagenomes

<table>
<thead>
<tr>
<th>Name</th>
<th>METAL</th>
<th>METAB</th>
<th>METAC</th>
<th>METAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original amino acid sequence</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**GeneID** | **METAL** | **METAB** | **METAC** | **METAD** |
---|---|---|---|---|
**SignalP (assuming gram negative)** | YES | YES | NO | YES |
**Mature chain** | 18.967 | 22.1393 | N.A | 25.634 |
**# of glycosylation sites** | 2 | 5 | 3 | 1 |
**AA** | 971 | 1193 | 1179 | 631 |
**Size (signal seq cut + Histag on N-terminal in iKDA)** | 112.0357 | 155.58602 | 136.49201 | 71.19818 |
**PI** | 5.9 | 5.8 | 6.12 | 9.23 |
**Predicted function** | GH115 | GH115 | GH115 | GH115 |
**PUL** | No | BD-PH PUL117 | BD-PH PUL97 | No |
**Organism** | Unknown | Unknown | Unknown | Unknown |
**Pham domain** | PF03648 (Glyco_hydro_67N); PF15979 (GH115) | PF15979 (GH115); PFO0331 (GH115) | None | |
**Sequenced clone in DHSalpha** | Clone 1.5 | Clone 2.3 | Clone 3.2 | Clone 4.2 |
**Clone in BL21** | YES | YES | YES | YES |
**Expressed in E.coli?** | YES | YES | NO | YES |
Appendix G: Polysaccharide structures.

**Arabinoglucomuronoxylan**

![Arabinoglucomuronoxylan structure](image)

**Arabinan**

![Arabinan structure](image)

**Arabinoxylan**

![Arabinoxylan structure](image)

**β-glucan**

![β-glucan structure](image)

**Carboxymethyl cellulose**

![Carboxymethyl cellulose structure](image)

**Galactan**

![Galactan structure](image)

**Galactomannan**

![Galactomannan structure](image)

**Glucuroarabinoxylan**

![Glucuroarabinoxylan structure](image)

**Glucomannan**

![Glucomannan structure](image)

**(Me)Glucoronoxylan**

![(Me)Glucoronoxylan structure](image)

**Starch**

![Starch structure](image)

**Xyloglucan**

![Xyloglucan structure](image)