STRUCTURE-FUNCTION ANALYSES OF PLANT GLYCAN-DEGRADING

ENZYMES

by

Nicholas McGregor

B.Sc., McMaster University, 2012

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Chemistry)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2019

© Nicholas McGregor, 2019

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

STRUCTURE-FUNCTION ANALYSES OF PLANT GLYCAN-DEGRADING ENZYMES

submitted by	Nicholas McGregor	in partial fulfillment of the requirements for
the degree of	Doctor of Philosophy	
in	Chemistry	
Examining Com	mittee:	
Harry Brumer		
Supervisor		
Lawrence McIn Supervisory Co	tosh mmittee Member	
Stephen Wither	8	
Supervisory Co	mmittee Member	
Lindsay Eltis		
University Exar	niner	
Shawn Mansfiel	ld	
University Exar	niner	

Additional Supervisory Committee Members:

David Chen Supervisory Committee Member

Abstract

Plant biomass is both the most abundant organic carbon source and the most abundant organic carbon sink on our planet. This carbon is stored primarily in the cell wall, where carbohydrates, proteins and polyphenols are interwoven to form complex purpose-built composite materials. Within plants, diverse polysaccharides are built up and broken down as part of their natural life cycle. Within the guts of multicellular organisms, a diverse and adaptable collection of bacteria anaerobically ferments complex plant polysaccharides. In this thesis, the structure and function of enzymes involved in these two processes are described.

The *xyloglucan endo-transglycosylase/hydrolase* (*XTH*) gene family encodes enzymes of central importance to plant cell wall remodelling. Investigations into the ancestry of the *XTH* family revealed a subfamily of *endo*-glucanases which share a common ancestor with the *XTHs*. Based on product analysis, kinetics, and X-ray crystallography these EG16s have been identified as a family of $\beta(1,4)$ -specific *endo*-glucanases with an uncommon mode of substrate recognition. Although the biological role(s) of EG16 orthologues remains to be fully resolved, the presented biochemical and tertiary structural characterisation provide insight into plant glycoside hydrolase evolution, and will continue to inform studies of plant cell walls.

Within the gut, *Prevotella* are an important genus of Gram-negative bacteria associated with carbohydrate-rich diets. Recent genome sequencing has shown that they possess many undescribed polysaccharide utilisation loci (PULs). A revisited broad-specificity cross-linking glycan-degrading *endo*-glucanase (PbGH5A) is associated with a PUL of unknown function within *Prevotella bryantii*, an obligate anaerobe originally isolated from the bovine gut. Based on X-ray crystallography, product identification, binding assays, and kinetic measurements, the structures and functions of a variety of proteins involved in the recognition and breakdown of complex β -mannans have been determined. These include a broad-specificity *endo*- β -glucanase, an *endo*- β -mannanse, two β -mannan-binding proteins, two β -mannan acetylesterases, a mannobiose-2-epimerase, and a mannosylglucose phosphorylase. Characterisation of the two β -mannan acetylesterases provides a basis for the expansion of CAZy family CE7 and the formation locus provides a genetic template for identifying systems which degrade complex galactomannans and glucomannans across the Bacteroidetes phylum.

Lay Summary

Plant biomass is both one of the most abundant carbon sources and carbon sinks on our planet. This carbon is stored primarily in the cell wall, where carbohydrates, proteins and polyphenols are interwoven to form complex purpose-built composite materials. Breaking these walls down requires disentangling large, complex, and diverse carbohydrates. This thesis takes a close look at some of the enzymes that perform this task in both plants and gut bacteria. The structure and function of an enzyme from a new family of plant cell wall-degrading is presented. Furthermore, the function and organisation of a genetic system responsible for the breakdown of a key part of the cell walls of many plants, including softwoods and legumes, is elucidated. Together, these studies add to our understanding of carbohydrate breakdown throughout the biosphere and provide insight into the behaviours of both plants, and the bacteria that help us degrade their cell walls.

Preface

Chapter 2: Crystallographic Insight into the Evolutionary Origins of Xyloglucan Endotransglycosylases and Endohydrolases. This chapter was written by the author (Nicholas McGregor). The manuscript was reviewed and revised by Dr. Harry Brumer (professor, PhD supervisor). This chapter was also reviewed by Dr. Filip van Petegem (collaborating professor). Intact protein mass spectra were collected by Victor Yin in the course of his undergraduate honours thesis. Crystal screening experiments were designed by Ching-Chieh Tung. All cloning, protein expression, mutagenesis, kinetic analysis, product analysis, phylogenetic analysis, enzyme crystallisation, diffraction data collection, structure solution, and substrate/ligand preparation were performed by the author. A version of this chapter has been published:

McGregor, N., Yin, V., Tung, C.-C., Van Petegem, F., and Brumer, H. (2017) Crystallographic Insight into the Evolutionary Origins of Xyloglucan Endotransglycosylases and

Endohydrolases. *Plant J* 89, 651–670.

Chapter 3: Structure-Function Analysis of a Mixed-Linkage β-Glucanase/ Xyloglucanase from the Key Ruminal Bacteroidetes *Prevotella bryantii* **B**₁**4.** This chapter was co-written by the author (Nicholas McGregor) with contributions related to enzyme structural analysis from Dr. Mariya Morar, a post-doctoral fellow and collaborator. The manuscript was reviewed and revised by Dr. Harry Brumer (professor, PhD supervisor) and Dr. Alexei Savchenko (collaborating professor). Cloning, mutagenesis, enzyme crystallisation, diffraction data collection, and structure determination were performed by Dr. Mariya Morar and Elena Evdokimova. Inhibitors were synthesised by Dr. Thomas Fenger. Phylogenetic analysis was performed by Dr. Nicolas Lenfant. Protein expression, kinetic analysis, product analysis, intact protein mass spectrometry, and substrate/ligand preparation was performed by the author. A version of this chapter has been published:

McGregor, N., Morar, M., Fenger, T.H., Stogios, P., Lenfant, N., Yin, V., Xu, X., Evdokimova, E., Cui, H., Henrissat, B., Savchenko, A., Brumer, H. (2016) Structure-Function Analysis of a Mixed-linkage β-Glucanase/Xyloglucanase from the Key Ruminal Bacteroidetes *Prevotella bryantii* B₁4. *J. Biol. Chem.* **291**, 1175–1197. **Chapter 4: Structural and Biochemical Analyses of β-Mannan Recognition Machinery from** *Prevotella bryantii* **B**₁**4.** This chapter was written by the author (Nicholas McGregor) with contributions from Dr. Peter Stogios related to structural characterisation. The manuscript was reviewed and revised by Dr. Harry Brumer (professor, PhD supervisor). Mary Wang and the author performed cloning and protein expression. Elena Evdokimova crystallised all of the proteins. Peter Stogios collected diffraction data and determined protein structures. All mutagenesis, kinetic analysis, product analysis, phylogenetic analysis, and substrate/ligand preparation was performed by the author. A version of this chapter will be submitted for publication.

Chapter 5: Characterisation of Two Mannan/Xylan Acetylesterases from *Prevotella bryantii* **B**₁**4 Reveals a New Esterase Family.** This chapter was written by the author (N. McGregor). The manuscript was reviewed and revised by Dr. Harry Brumer (professor, PhD supervisor). Tatiana Skarina crystallised the protein. Dr. Peter Stogios determined the protein structure. All cloning, protein expression, kinetic analysis, product analysis, phylogenetic analysis, and substrate/ligand preparation was performed by the author. A version of this chapter will be submitted for publication.

Table of Contents

Abstract	ii
Lay Summary	iv
Preface	v
List of Tables .	xiv
List of Figures	xv
List of Abbrev	iations xix
Acknowledgen	nents xxiii
Dedication	
Chapter 1: Int	troduction1
1.1	Complex Carbohydrate-Enzyme Interactions in the Biosphere1
1.1.1	The Plant Cell Wall1
1.1.1.1	Xyloglucans2
1.1.1.2	Mixed-Linkage Glucans
1.1.1.3	β-Mannans4
1.1.1.4	Xylans
1.1.1.5	Pectins
1.1.2	Enzyme Specificities in the Breakdown of Plant Polysaccharides
1.1.2.1	Glycoside Hydrolases
1.1.2.2	Polysaccharide Lyases
1.1.2.3	Glycosyltransferases
1.1.2.4	Carbohydrate Esterases
1.1.2.5	Auxiliary Activities
1.1.2.6	Carbohydrate-Binding Modules10
1.2	Carbohydrate-Active Enzyme Discovery
1.2.1	CAZymes in Plants 11
1.2.2	Fungal CAZymes
1.2.3	Bacterial and Archaeal CAZymes
1.2.4	Polysaccharide Utilisation Loci in Gut Bacteroidetes
1.2.4.1	Surface Glycan-Binding Proteins15
	vii

1.2.4.2	TonB-Dependent Transporters	16
1.2.4.3	Carbohydrate-Degrading Enzymes	16
1.2.4.4	Sensor/Regulators	17
1.3	Quantifying CAZyme Activities	18
1.3.1	Quantifying Glycoside Hydrolysis	18
1.3.1.1	Reducing End Quantitation	18
1.3.1.2	Enzyme-Linked Assays	19
1.3.1.3	HPLC Assays	19
1.3.1.4	Chromogenic Substrates for Glycoside Hydrolases	20
1.3.2	Quantifying Carbohydrate Esterase Activity	21
1.3.2.1	Chromogenic Substrates	21
1.3.2.2	Enzyme-Linked Assays	21
1.3.3	Measuring CAZyme-Substrate Interactions through Kinetic Analysis	22
1.3.3.1	Models of Enzyme Kinetics	22
1.3.3.2	Relating Kinetic Parameters to Molecular Interactions	25
1.4	Determining Enzyme Specificity through Product Analysis	26
1.4.1	Chemical Transformations for Carbohydrate Structure Determination	26
1.4.2	Chromatographic Analysis of Carbohydrate Mixtures	27
1.4.3	Identification of Carbohydrates using Mass Spectrometry	28
1.4.4	Nuclear Magnetic Resonance Spectroscopy of Carbohydrates	30
1.5	Relating Enzyme Function to Structure using X-ray Crystallography	31
1.5.1	Glycoside Hydrolase Active-Site Structure	32
1.5.2	Active Site Probes for Enzyme Active Site Labelling	35
1.6	Aims of this Thesis	35
Chapter 2: Cr	ystallographic Insight into the Evolutionary Origins of Xyloglucan End	lo-
Transglycosyla	ses and <i>Endo</i> -Hydrolases	37
2.1	Introduction	37
2.2	Materials and Methods	39
2.2.1	Bioinformatic Analysis	40
2.2.2	Cloning	40
2.2.3	Protein Expression and Purification	41

2.2.4	Carbohydrate Analysis	
2.2.5	Substrates	
2.2.6	Enzyme Activity Determination	
2.2.7	Co-crystallisation of VvEG16 Variants in Complex with Oligosacch	arides 44
2.2.8	Data Collection and Refinement	
2.3	Results	
2.3.1	Molecular Phylogeny of the EG16 Clade	
2.3.2	Recombinant Production of VvEG16 in <i>E.coli</i>	
2.3.3	VvEG16 is a bi-function MLG/XyG endo-glucanase	
2.3.4	Kinetic Subsite Mapping of the VvEG16 Active Site	50
2.3.5	Three-Dimensional Structure of VvEG16 Variants in Complex with	Matrix
Glycan Oli	gosaccharides	54
2.3.5.1	EG16 Tertiary Structure	54
2.3.5.2	Key Features of MLG Recognition by VvEG16	56
2.3.5.3	Key Features of XyG Recognition by VvEG16	58
2.3.5.4	Supporting GGGG and XXXG Complex Structures	61
2.3.5.5	EG16 Tertiary Structure vis-à-vis GH16 Licheninases and XTH	Gene
Products	5	61
2.4	Discussion	64
2.4.1	Plant EG16 Members Represent a Unique Class of Bi-Functional Ma	atrix
Glycan Hy	drolases	64
2.4.2	EG16 Members Represent Extant Transitional Enzymes Linking the	
Evolution	of Bacterial Licheninases and Plant XTG Gene Products	65
2.4.3	In Vivo Roles of EG16 Members	67
Chapter 3: Stu	ructure-Function Analysis of a Mixed-Linkage β-Glucanase/Xylogluc	anase
from the Key I	Ruminal Bacteroidetes <i>Prevotella bryantii</i> B ₁ 4	69
3.1	Introduction	69
3.2	Materials and Methods	
3.2.1	Analytical Methods	
3.2.1.1	HPAEC-PAD Carbohydrate Analysis	
3.2.1.2	Mass Spectrometry	

3.2.2	Substrates and Inhibitors72
3.2.2.1	Commercial Substrates72
3.2.2.2	Synthetic Substrates and Inhibitors73
3.2.2.3	Xyloglucan oligosaccharides (tXyGOs)73
3.2.2.4	Mixed-linkage glucan oligosaccharides (bMLGOs)74
3.2.3	Enzyme Cloning and Expression74
3.2.4	Enzyme Kinetics and Product Analysis75
3.2.4.1	Polysaccharide Hydrolysis75
3.2.4.2	Chromogenic Oligosaccharide Hydrolysis78
3.2.4.3	Native Oligosaccharide Hydrolysis79
3.2.4.4	Inhibition Kinetics
3.2.5	Enzyme Crystallisation
3.2.6	X-ray Crystal Structure Determination
3.2.7	Phylogenetic Analysis
3.3	Results
3.3.1	Polysaccharide Kinetics
3.3.2	Polysaccharide Hydrolysis Product Distributions
3.3.3	Chromogenic Substrate Kinetics
3.3.4	Native Oligosaccharide Kinetics
3.3.5	Inhibition and Covalent Labelling with an Active Site-Directed Inhibitor 92
3.3.6	Structural Characterisation of PbGH5A Variants in the apo Form and in
Complexes	s with Oligosaccharides
3.3.6.1	Overall Structure of PbGH5A
3.3.6.2	Non-Covalent Complexes of PbGH5A Variants with Branched and
Unbranc	bed Ligands
3.3.6.3	Active Site Affinity-Labelled Complex of PbGH5A 103
3.4	Discussion 106
3.4.1	PbGH5A is a Predominant Mixed-Linkage Endo-Glucanase, but also a
Competent	<i>Endo</i> -Xyloglucanase
3.4.2	The Active Site of PbGH5A Comprises Seven Subsites in Total 109

3.4.3	PbGH5A Exhibits Subtle Discrimination of β -Glucan Linkage	
Regiochem	istry in the Active Site	
3.4.4	Implications for Specificity Prediction in GH5 Subfamily 4	
Chapter 4: Str	ructural and Biochemical Analyses of β-Mannan Recognition Ma	chinery
from Prevotella	bryantii B 1 4 Error! Bookmark	not defined.
4.1	Introduction	
4.2	Materials and Methods	
4.2.1	Analytical Methods	
4.2.1.1	HPAEC-PAD Carbohydrate Analysis	
4.2.1.2	Mass Spectrometry	
4.2.2	Substrates and Ligands	
4.2.2.1	Commercial Substrates	
4.2.2.2	Acetylated Oligosaccharides	
4.2.3	Enzyme Cloning and Expression	
4.2.3.1	Mutagenesis	
4.2.4	Enzyme Kinetics and Product Analysis	
4.2.4.1	Polysaccharide Hydrolysis	
4.2.4.2	Native Oligosaccharide Hydrolysis	
4.2.4.3	Activities of PbEpiA and PbGH130A	
4.2.5	Carbohydrate Affinity Determination	
4.2.5.1	Carbohydrate Affinity Polyacrylamide Gel Electrophoresis (CA-PAGE)
4.2.5.2	Insoluble Mannan Pull-Down Assay	
4.2.5.3	Isothermal Titration Calorimetry (ITC)	
4.2.6	Enzyme Crystallisation and Structure Determination	
4.2.7	Bioinformatic Analysis	
4.3	Results	
4.3.1	β-Mannan Utilisation Locus Model Building	
4.3.2	Recombinant Protein Production and Purification	
4.3.3	Glycan Recognition by Surface Glycan-Binding Proteins	
4.3.4	Tertiary Structural Characterisation of PbSGBP-B	

4.3.5	Initial Substrate Cleavage by Endo-Glycanases	
4.3.6	Tertiary Structural Characterisation of PbGH26A-GH5A and PbGH	26A.135
4.3.7	Mannobiose Breakdown as the Ultimate Step in Saccharification	137
4.3.8	Identification of Putative β-Mannan Utilisation Loci	138
4.4	Discussion	139
4.4.1	Genetic Markers of β-Mannan Utilisation	139
4.4.2	PbSGBP-B and PbSGBP-C Bind β -Mannan Utilizing a Dual-Trypto	ophan
Platform		
4.4.3	PbSGBP-A has no Affinity for β -Glucans or β -Mannans	141
4.4.4	PbGH26A has a Long Active Site Tailored for Glucomannan Degra	dation 142
4.4.5	The Fusion of PbGH26A and PbGH5A Connects Complementary A	Activities
without Co	operativity	143
4.4.6	β-Mannan Oligosaccharides are Hydrolysed in the Periplasm	
4.4.7	PbEpiA and PbGH130A Generate α-Mannose-1-Phosphate in the C	ytoplasm
		1 4 4
Chapter 5: Ch	aracterisation of Two Mannan/Xylan Acetylesterases from <i>Prevotel</i>	144 la
Chapter 5: Ch bryantii B14 Re	aracterisation of Two Mannan/Xylan Acetylesterases from <i>Prevotel</i> veals a New Esterase Family	144 la 146
Chapter 5: Ch bryantii B14 Re 5.1	aracterisation of Two Mannan/Xylan Acetylesterases from <i>Prevotel</i> veals a New Esterase Family Introduction	144 146
Chapter 5: Ch <i>bryantii</i> B ₁ 4 Re 5.1 5.2	aracterisation of Two Mannan/Xylan Acetylesterases from Prevotela veals a New Esterase Family Introduction Materials and Methods	144 146 146 147
Chapter 5: Ch bryantii B14 Re 5.1 5.2 5.2.1	aracterisation of Two Mannan/Xylan Acetylesterases from Prevotela veals a New Esterase Family Introduction Materials and Methods Mass Spectrometry	144 146 146 147 147
Chapter 5: Ch bryantii B14 Re 5.1 5.2 5.2.1 5.2.1 5.2.2	aracterisation of Two Mannan/Xylan Acetylesterases from Prevotela veals a New Esterase Family Introduction Materials and Methods Mass Spectrometry Substrates and Ligands	144 146 146 147 147 147
Chapter 5: Ch bryantii B14 Re 5.1 5.2 5.2.1 5.2.2 5.2.2 5.2.2.1	aracterisation of Two Mannan/Xylan Acetylesterases from Prevotela veals a New Esterase Family Introduction Materials and Methods Mass Spectrometry Substrates and Ligands Commercial Substrates	144 146 146 147 147 147
Chapter 5: Ch bryantii B14 Re 5.1 5.2 5.2.1 5.2.2 5.2.2 5.2.2.1 5.2.2.1 5.2.2.2	aracterisation of Two Mannan/Xylan Acetylesterases from Prevotela veals a New Esterase Family Introduction Materials and Methods Mass Spectrometry Substrates and Ligands Commercial Substrates Oligosaccharide Mixtures	144
Chapter 5: Ch bryantii B14 Re 5.1 5.2 5.2.1 5.2.2 5.2.2 5.2.2.1 5.2.2.2 5.2.2.2 5.2.3	aracterisation of Two Mannan/Xylan Acetylesterases from Prevotela veals a New Esterase Family Introduction Materials and Methods Mass Spectrometry Substrates and Ligands Commercial Substrates Oligosaccharide Mixtures Enzyme Cloning and Production	144 <i>la</i> 146 147 147 147 148 148
Chapter 5: Ch bryantii B14 Re 5.1 5.2 5.2.1 5.2.2 5.2.2 5.2.2.1 5.2.2.2 5.2.2.2 5.2.3 5.2.4	aracterisation of Two Mannan/Xylan Acetylesterases from Prevotela veals a New Esterase Family Introduction Materials and Methods Mass Spectrometry Substrates and Ligands Commercial Substrates Oligosaccharide Mixtures Enzyme Cloning and Production Enzyme Kinetics and Product Analysis	144 Ja
Chapter 5: Ch bryantii B14 Re 5.1 5.2 5.2.1 5.2.2 5.2.2 5.2.2.1 5.2.2.2 5.2.2.1 5.2.2.2 5.2.3 5.2.4 5.2.5	aracterisation of Two Mannan/Xylan Acetylesterases from Prevotel. veals a New Esterase Family Introduction Materials and Methods Mass Spectrometry Substrates and Ligands Commercial Substrates Oligosaccharide Mixtures Enzyme Cloning and Production Enzyme Kinetics and Product Analysis Enzyme Crystallisation and X-ray Diffraction	<i>I</i> 44 <i>Ia</i> <i>I</i> 46 <i>I</i> 47 <i>I</i> 47 <i>I</i> 47 <i>I</i> 47 <i>I</i> 47 <i>I</i> 47 <i>I</i> 47 <i>I</i> 48 <i>I</i> 48 <i>I</i> 48 <i>I</i> 49 <i>I</i> 50
Chapter 5: Ch bryantii B14 Re 5.1 5.2 5.2.1 5.2.2 5.2.2 5.2.2 5.2.2 5.2.2 5.2.3 5.2.4 5.2.5 5.2.6	aracterisation of Two Mannan/Xylan Acetylesterases from Prevotela veals a New Esterase Family Introduction Materials and Methods Mass Spectrometry Substrates and Ligands Commercial Substrates Oligosaccharide Mixtures Enzyme Cloning and Production Enzyme Kinetics and Product Analysis Enzyme Crystallisation and X-ray Diffraction Bioinformatic Analysis	<i>a</i> <i>144</i> <i>146</i> <i>147</i> <i>147</i> <i>147</i> <i>147</i> <i>147</i> <i>147</i> <i>147</i> <i>148</i> <i>148</i> <i>148</i> <i>148</i> <i>149</i> <i>150</i>
Chapter 5: Ch bryantii B14 Re 5.1 5.2 5.2.1 5.2.2 5.2.2 5.2.2 5.2.2 5.2.3 5.2.4 5.2.5 5.2.6 5.3	aracterisation of Two Mannan/Xylan Acetylesterases from Prevotela veals a New Esterase Family Introduction Materials and Methods Mass Spectrometry Substrates and Ligands Commercial Substrates Oligosaccharide Mixtures Enzyme Cloning and Production Enzyme Kinetics and Product Analysis Enzyme Crystallisation and X-ray Diffraction Bioinformatic Analysis Results	<i>la</i> <i></i> 144 <i></i> 146 <i></i> 147 <i></i> 147 <i></i> 147 <i></i> 147 <i></i> 148 <i></i> 148 <i></i> 148 <i></i> 150 <i></i> 151
Chapter 5: Ch bryantii B14 Re 5.1 5.2 5.2.1 5.2.2 5.2.2 5.2.2 5.2.2 5.2.2 5.2.3 5.2.4 5.2.5 5.2.6 5.3 5.3.1	aracterisation of Two Mannan/Xylan Acetylesterases from Prevotela veals a New Esterase Family Introduction Materials and Methods Mass Spectrometry Substrates and Ligands Commercial Substrates Oligosaccharide Mixtures Enzyme Cloning and Production Enzyme Kinetics and Product Analysis Enzyme Kinetics and Product Analysis Bioinformatic Analysis Results Primary Structure Analysis and Phylogeny	<i>la</i> 144 <i>la</i> 146 147 147 147 147 148 148 148 149 150 151 152

5.3.3	Site Selectivity of PbCExA	
5.3.4	Tertiary Structural Analysis of PbCExA	
5.4	Discussion	
5.4.1	PbCE7A and PbCExA Display Complementary Acetylestera	se Specificities.
5.4.2	PbCE7A is Part of a Distinct Group within CE7	
5.4.3	PbCExA Utilizes an Novel Ser-His-Cys Catalytic Triad	
5.4.4	PbCExA is a Founding Member of a New CE Family	
5.5	Conclusions	
Chapter 6: C	General Conclusions	165
6.1	PbGH5A and the Discovery of the <i>P</i> . <i>bryantii</i> β -Mannan Utilisa	ation System 165
6.2	EG16s are of Emerging Interest in Fundamental Biology	
References		170
Appendices		212
Appendix A	A Supporting Information for Chapter 2: Crystallographic Insight in	to the
Evolutionar	ry Origins of Xyloglucan Endo-Transglycosylases and Endo-Hydro	lases 212
Appendix E	3 Supporting Information for Chapter 4: Structural and Biochemica	l Analyses of β-
Mannan Re	cognition Machinery from Prevotella bryantii B ₁ 4	
Appendix C	C Supporting Information for Chapter 5: Characterisation of Two M	Iannan/Xylan
Acetylester	ases from <i>Prevotella bryantii</i> B ₁ 4	
Appendix I	O Additional Publications Not Included in this Thesis	

List of Tables

Table 2-1: Apparent Michaelis-Menten kinetic constants for hydrolysis and transglycosylation	
reactions catalysed by wild-type VvEG16 and VvEG16(ΔV152)	51
Table 3-1: Kinetic parameters for the hydrolysis of various substrates by PbGH5A	84
Table 3-2: X-ray diffraction data and refinement statistics for PbGH5A crystal structures	96
Table 3-3: Dali search results for PbGH5A	97
Table 4-1: X-ray diffraction data collection and refinement statistics for PbGH26A and PbSGB	P-
B1	31
Table 4-2: Apparent Michaelis-Menten kinetics of various hydrolysis reactions catalysed by the	9
GH26 and GH5 domains of PBR_03681	33
Table A-5: Sequences used for generation of the GH16 phylogenetic tree	26
Table A-6: Primer sequences used for cloning and mutagenesis in Chapter 2	26
Table A-3: X-ray diffraction data statistics for VvEG16 crystal complexes. 2	27
Table A-1: EG16 expression constructs screened for protein production	28
Table A-2: ¹⁸ O-labelling of each product from cellooligosaccharide hydrolysis by	
VvEG16(ΔV152)	28
Table A-4: Privateer validation results for VvEG16 crystal complexes	29

Table B-1: Primers used for the cloning and mutagenesis of the P. bryantii MUL genes 241

List of Figures

Figure 1-1: Representative chemical structures of the major cross-linking glycans in plant	t cell
walls	
Figure 1-2: Retaining and inverting mechanisms of glycoside hydrolysis	
Figure 1-3: The serine hydrolase mechanism of carbohydrate esterase activity	10
Figure 1-4: The current model of the prototypical starch utilisation locus from <i>B</i> .	
thetaiotaomicron	15
Figure 1-5: Kinetic models of enzyme kinetics	
Figure 1-6: Glycoside hydrolase active site architectures	33
Figure 1-7: Wall-eyed stereo view of the interactions between human pancreatic amylase	and
acarbose	33
Figure 1-8: Michaelis complex of Pseudomonas cellulosa GH26 mannanase with 2,4-	
dinitrophenyl 2-deoxy-2-fluoro-β-mannotrioside	35
Figure 2-1: Protein sequence-based phylogeny of EG16 homologs within Glycoside Hydr	rolase
Family 16 (GH16)	46
Figure 2-2: VvEG16 limit-digest products	48
Figure 2-3: Site of MLG hydrolysis by VvEG16 compared to other known MLG-active et	ndo-
glucanases	49
Figure 2-4: Gel formation by VvEG16-catalysed hydrolysis of MLG	50
Figure 2-5: Structure of VvEG16(Δ V152,E89A) in complex with $\beta(1,3)/\beta(1,4)$ mixed-line	kage
gluco-oligosaccharides	57
Figure 2-6: Structure of VvEG16(Δ V152,E89A) in complex with xylogluco-oligosacchar	ides.60
Figure 2-7: Tertiary structural comparison of $VvEG16(\Delta V152, E89A)$ with representative	;
licheninase, XET, and XEH enzymes of GH16	63
Figure 3-1: Polysaccharide structures used in this study	70
Figure 3-2: pH-rate profiles for PbGH5A	77
Figure 3-3: Thermal stability of PbGH5A	
Figure 3-4: PbGH5A polysaccharide specificity	83
Figure 3-5: HPAEC-PAD chromatograms of the limit-digests of bMLG (A) and tXyG (B)
hydrolysed by PbGH5A	86

Figure 3-6: Michaelis-Menten plots for the hydrolysis of various chromogenic substrates b	у
PbGH5A	88
Figure 3-7: Michaelis-Menten plots for the hydrolysis of various model oligosaccharides b	ру
PbGH5A	90
Figure 3-8: HPAEC-PAD product analysis of the digestion of cellohexaose by PbGH5A	91
Figure 3-9: Mass spectrum of the products of cellopentaose degradation by PbGH5A in H	2 ¹⁸ O 91
Figure 3-10: Inhibition of PbGH5 with active-site affinity labels	93
Figure 3-11: Intact MS of PbGH5A and several mutants	94
Figure 3-12: Overall Structure of PbGH5A	95
Figure 3-13: Interactions in the PbGH5A active site	99
Figure 3-14: A schematic representation of the PbGH5A active site bound to two XXXG	
oligosaccharide units	100
Figure 3-15: Comparison of PbGH5A complexed structures	102
Figure 3-16: PbGH5A active site in complex with XXXG-NHCOCH ₂ Br	104
Figure 3-17: Comparison of GH5_4 structural homologs	107
Figure 4-1: The <i>Prevotella bryantii</i> MUL and two model β-mannan substrates	117
Figure 4-2: Cartoon model of the putative Prevotella bryantii MUL	126
Figure 4-3: Identification of polysaccharide ligands for surface-glycan binding proteins	127
Figure 4-4: Overall structure of PbSGBP-B	130
Figure 4-5: HPAEC-PAD analysis of β -mannans digested by PbGH26A-GH5A	133
Figure 4-6: Mass spectrometric analysis of the hydrolysis products of PbGH26A acting on	1
mannohexaose in H ₂ ¹⁸ O	135
Figure 4-7: Structure of PbGH26A-GH5A	136
Figure 5-1: Molecular phylogeny of PbCExA	153
Figure 5-2: Analysis of oligosaccharides deacetylated by PbCExA	156
Figure 5-3: Structural analysis of PbCExA	158

Figure A-1: Protein sequence alignment of EG16s with Licheninases and XTH gene products 213
Figure A-2: Intact MS and SDS-PAGE analyses of purified VvEG16 and its mutants 213
Figure A-3: Comparison of wild-type VvEG16 and VvEG16 (ΔV152) activities
Figure A-4: Dependence of VvEG16(Δ V152) activity on pH and temperature

Figure A-5: Identification of oligosaccharide series produced by action of VvEG16(Δ V152) on
bMLG
Figure A-6: Identification of the MLGO produced by the action of $VvEG16(\Delta V152)$ on bMLG
(Glc ₄ -MLGO) by HPAEC-PAD
Figure A-7: Identification of the MLGO produced by the action of VvEG16(Δ V152) on bMLG
(Glc ₄ -MLGO) by LC-MS
Figure A-8: Time-dependent MLG hydrolysis by VvEG16(ΔV152)
Figure A-9: Kinetic data and Michaelis-Menten curves used to determine the k_{cat} and K_M values
for the hydrolysis of various chromogenic substrates by VvEG16(ΔV152)
Figure A-10: Kinetic data and Michaelis-Menten curves used to determine the k _{cat} and K _M values
for the hydrolysis of various oligosaccharide substrates by VvEG16(ΔV152)
Figure A-11: Mass spectra of products produced by the hydrolysis of cello-oligosaccharides by
VvEG16(Δ V152) in H ₂ ¹⁸ O
Figure A-12: Homo- and hetero-transglycosylation of oligosaccharides by VvEG16(Δ V152). 222
Figure A-13: Structure of VvEG16(ΔV152,E89A) in complex with a cellooligosaccharide 223
Figure A-14: Structure of VvEG16(ΔV152,C22S,C188S) in complex with a xyloglucan
oligosaccharide
Figure A-15: Superimposition of the carbohydrate ligands bound within the active site of each
experimentally determined structure of VvEG16
Figure B-1: SDS-PAGE following inM pull-down
Figure B-2: ITC of PbSGBP-B and PbSGBP-C with oligosaccharides
Figure B-3: Alignment of BoMan26A with PbGH26B
Figure B-4: Fluorescent gel images of GFP-PbGH26B CA-PAGE
Figure B-5: Alignment of SusD homologues with PbSGBP-A
Figure B-6: CA-PAGE of SGBP-A fused to E. coli maltose binding protein
Figure B-7: PbSGBP-B mutants run on CA-PAGE gels
Figure B-8: PbSGBP-C mutants run on CA-PAGE gels
Figure B-9: A) pH-activity and B) temperature-activity plots for PbGH26A
Figure B-10: Michaelis-Menten plots showing the relationship between substrate concentration
and activity for PbGH26A
Figure B-11: LC-MS of oligosaccharides produced by the action of PbGH26A on inM

Figure B-12: HPAEC-PAD traces demonstrating the activities of PbEpiA and PbGH130A 239
Figure B-13: Michaelis-Menten plots showing the relationship between substrate concentration
and activity for PbGH26A
Figure C-1: HPAEC-PAD analysis of the XOs mixture
Figure C-2: Plots of change in E-value vs. hit number for NCBI RefSeq BLAST hits
Figure C-3: Phylogenetic tree of PbCExA homologues
Figure C-4: Phylogenetic tree of PbCE7A homologues
Figure C-5: Molecular phylogeny of PbCE7A
Figure C-6: Acetonitrile tolerance, pH optimum and thermal stability of PbCExA and PbCE7A
Figure C-7: Michaelis-Menten plots showing the relationship between substrate concentration
and initial hydrolytic rates for PbCExA and PbCE7A

List of Abbreviations

2-D NMR: Two-Dimensional Nuclear Magnetic Resonance AA: Auxiliary Activity Ac: Acetate BCA: Bicinchoninic Acid **BLAST: Basic Local Alignment Search Tool** bMLG: Barley Mixed-Linkage Glucan **BSA:** Bovine Serum Albumin Bu: Butyrate bX: Beechwood Xylan CA-PAGE: Carbohydrate Affinity Polyacrylamide Gel Electrophoresis CAZy: Carbohydrate Active Enzyme Database CAZymes: Carbohydrate-Active Enzymes CBM##: Carbohydrate-Binding Module Family ## CBM: Carbohydrate-Binding Module CE##: Carbohydrate Esterase Family ## CE: Carbohydrate Esterase cGM: Carob Galactomannan CID: Collision-Induced Dissociation CMC: Carboxymethylcellulose CNP: 2-Chloro-4-Nitrophenyl **COSY:** Correlation Spectroscopy ddH2O: Sterile Ultra-Pure Water diH2O: Deionised Water DMSO: Dimethylsulfoxide DNA: Deoxyribonucleic Acid DNP: 2,4-Dinitrophenyl DNSA: 3,5-Dinitrosalicylic Acid dTKP: De-Oiled Tamarind Kernel Powder **DTT:** Dithiothreitol EDTA: Ethylenediaminetetraacetic Acid

ESI: Electrospray Ionisation

FAD: Flavin Adenine Dinucleotide

G: $\beta(1,4)$ -Linked Glucose

G3: $\beta(1,3)$ -Linked Glucose

Gal: Galactose

GAX: Glucuronoarabinoxylan

GGM: Galactoglucomannan

gGM: Guar Galactomannan

GH##: Glycoside Hydrolase Family ##

GH: Glycoside Hydrolase

Glc: Glucose

GT: Glycosyltransferase

HEC: Hydroxyethylcellulose

HEPES: 4-(2-Hydroxyethyl)-1-Piperazineethanesulfonic Acid

HMBC: Heteronuclear Multiple Bond Correlation

HPAEC-PAD: High-Performance Anion Exchange Chromatography Coupled to Pulsed

Amperometric Detection

HPLC: High-Performance Liquid Chromatography

HTCS: Hybrid Two-Component System

ID: Sequence Identity

IMG: Integrated Microbial Genomics

inM: Insoluble Mannan

ITC: Isothermal Titration Calorimetry

JGI: Joint Genome Institute

kDa: Kilodalton

kGM: Konjac Glucomannan

kGMO: Konjac Glucomannan-Derived Oligosaccharide

L: $\beta(1,2)$ -Galactosyl- $\alpha(1,6)$ -Xylosyl- $\beta(1,4)$ -Glucose

LN₂: Liquid Nitrogen

LC-MS: Liquid Chromatography Coupled to Mass Spectrometry

m/z: Mass-to-Charge Ratio

M: $\beta(1,4)$ -Linked Mannose

MALDI-TOF: Matrix-Assisted Laser Desorption Ionisation Coupled to Time-of-Flight Mass

Spectrometry

Man: Mannose

MBP: Maltose-Binding Protein

MeCN: Acetonitrile

MES: 2-(*N*-Morpholino)Ethanesulfonic Acid

MLG: Mixed-Linkage Glucan

MLGOs: Mixed-Linkage Glucan-Derived Oligosaccharides

MOPS: 3-(N-Morpholino)Propanesulfonic Acid

MR: Molecular Replacement

MS/MS: Tandem Mass Spectrometry

MS: Mass Spectrometry

MSn: Tandem Mass Spectrometry with n Sequential Fragmentation Steps

MUL: Mannan-Utilisation Locus

MWCO: Molecular Weight Cut-Off

NAD+: Nicotinamide Adenine Dinucleotide

NADH: Reduced Nicotinamide Adenine Dinucleotide

NCBI: National Center for Biotechnology Information

NMR: Nuclear Magnetic Resonance

NOESY: Nuclear Overhauser Effect Spectroscopy

NTA: Nitrilotriacetic Acid

PASC: Phosphoric Acid-Swollen Cellulose

PDB ID: Alphanumeric Protein Structure Identifier within the Protein Data Bank

PDB: Protein Data Bank

PEG: Polyethylene Glycol

PL: Polysaccharide Lyase

PMAA: Partially Methylated Alditol Acetate

PNP: 4-Nitrophenyl

PUL: Polysaccharide Utilisation Locus

Q-TOF: Quadrupole Time-of-Flight

RMSD: Root-Mean-Square Deviation

SAD: Single Anomalous Dispersion

SDS-PAGE: Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

SEC: Size-Exclusion Chromatography

SeMet: Selenomethionine-Derivatised

SEMP: Symporter-Epimerase-Mannosylglucose Phosphorylase

SGBP: Surface Glycan-Binding Protein

SUMO: Small Ubiquitin-Like Modifier

TBDT: TonB-Dependent Transporter

TCEP: Tris(2-Carboxyethyl)Phosphine

TLC: Thin-Layer Chromatography

TOCSY: Total Correlation Spectroscopy

TOF: Time-of-Flight

TPR: Tetratricopeptide Repeat

TRIS: Tris(Hydroxymethyl)Aminomethane

tXyG: Tamarind Xyloglucan

UPLC: Ultra-Performance Liquid Chromatography

UV: Ultraviolet

wAX: Wheat Arabinoxylan

X: $\alpha(1,6)$ -Xylosyl- $\beta(1,4)$ -Glucose

XEH: Xyloglucan Endo-Hydrolase

XET: Xyloglucan Endo-Transglycosylase

XO: Xylan-Derived Oligosaccharide

XTH: Xyloglucan Endo-Transglycosylase/Hydrolase

XyG: Xyloglucan

XyGO: Xyloglucan-Derived Oligosaccharide

Xyl: Xylose

Acknowledgements

I must express my eternal gratitude to my family. Their love and support has made everything I do possible. I want to thank my parents, Debbie and Andrew, for their endless love and support, and for being patient, honest, dependable, and caring guides. I want to thank my brother, Alex, for sharing so many great stories about the world beyond the academy and for helping become who I am today. I want to thank my wife, Jessica, for good conversations and great company on our adventure, for always taking good care of me, for playing house with me, and for introducing me to the wonderful Risley family. I want to thank my parents-in-law, Karen and Charley, for welcoming me into their home and into their family.

I want to thank all of the great members of the Brumer lab who I have been fortunate to share my time at UBC with. There are too many excellent people to mention, but I want to highlight the undergraduate students whom had the privilege of supervising. The keen intellects and dedication of Vincent Chan, Mary Wang, and Victor Yin never ceased to impress me. I wish you the best of luck with all of your future endeavours. I also want to thank Shaheen Shojania for demonstrating what it means to be a great lab citizen every day. I have not met as capable of a cat herder.

I want to thank UBC and NSERC. The support that I received throughout my time in Vancouver has made it possible for me to work and study while thoroughly experiencing life on the West coast and enjoying many of the treasures that British Columbia has to offer. I want to specifically thank the Faculty of Graduate Studies at UBC for financial support through the Graduate Support Initiative and the Four-Year Fellowship program and for furthering my education through the Graduate Pathways to Success program and the Graduate Student Travel Fund.

I want to thank the Graduate Student Society of UBC Vancouver for welcoming me into their community. My time advocating for students with the GSS was truly inspirational. I want to thank the former general manager of the GSS, Mark Wellington, the former president of the GSS, Tobias Friedel, and the amazing executive team I had the pleasure of working with, Genevieve Cruz, Jennifer Deboer, Dante Mendoza, Mashid Ghafar Tehrani, and Xiaolei Deng, for teaching me about the strengths, benefits, and challenges of working with diverse, and often fluid, teams. I hope to one day find such a diverse collection of brilliant people to work with again.

xxiii

Finally, I want to thank my supervisor, Harry Brumer, and the rest of my supervisory committee for providing advice, direction, and support over the past five years. I have learned so much about academic life, intellectual pursuits, and what it means to be a good scientist from the mentorship provided by each of you.

Thank you all

To Debbie, Andrew, Jessica, Karen, and Charley Thank you for helping me explore my potential



Chapter 1: Introduction

1.1 Complex Carbohydrate-Enzyme Interactions in the Biosphere

Carbohydrates are one of the four fundamental classes of biological molecules alongside nucleic acids, peptides, and lipids. The fundamental biological importance of carbohydrates is reflected in their multiplicity of functions which include energy storage and transport, cell shape and tissue structure, and communication and signalling. Yet glycobiology is a field hindered by the myriad possible saccharide combinations and the phylogenetic diversity of enzymes dedicated to saccharide synthesis, modification, and breakdown. Continued exploration of this diversity is necessary to build our understanding of the potential that carbohydrates hold.

Plant carbohydrates play two invaluable roles in modern society; they provide structural materials and food. Structural polysaccharides are incorporated into the composite cell wall matrix that supports the growth and development of plants (Carpita and McCann, 2000). They are an essential component of important materials such as linen and timber. Storage polysaccharides are found in high concentrations in the walls of endosperm cells (Buckeridge, 2010). Crops with starch-containing endosperms, such as corn, wheat, and rice, are essential food sources. Endosperms rich in non-starch polysaccharides, such as that of guar, are important sources of soluble dietary fibre and thickeners.

The build-up and breakdown of carbohydrates are essential processes governing carbon cycling within the biosphere. Enzymes expressed during germination liberate monosaccharides from storage polysaccharides providing a source of energy to drive early plant growth. Structural or storage polysaccharides form the basis of most animal diets. Essential nutrients, such as phosphorus and sulfur, are returned to soil by the action of microbes breaking down dead or discarded plant materials. Cycles, such as leaf growth and decay, are critically dependent on the efficient breakdown of polysaccharides. To understand and harness these processes we must understand the diversity of polysaccharide materials synthesised by plants and the myriad enzymes produced by both plants and bacteria to break them down.

1.1.1 The Plant Cell Wall

Terrestrial plant cell walls make up one of the largest stores of organic carbon on Earth. The accessibility of their constituent polysaccharides to heterotrophic organisms is often limited by the complexity and chemical stability of the cell wall. Structurally, the plant cell wall can consist of many layers generally classified as belonging to either the primary cell wall, the secondary cell wall, or the middle lamella which connects adjacent cells (Carpita and McCann, 2000). The middle lamella is a pectin-rich outer layer which regulates cell-cell adhesion. The primary cell wall is a thin layer of elastic hydrogel which supports the plant cell during pressure-driven expansion. Insoluble cellulose fibrils are a significant component of the primary wall dry weight. These weakly-interacting fibrils are held together by cross-linking glycans (i.e. hemicelluloses) to form a robust, elastic network. Pectins and structural proteins make up the rest of the matrix of the primary wall. The major cross-linking glycans in the primary wall are glucuronoarabinoxylans (GAXs), xyloglucans (XyGs), and, in some species, mixed-linkage glucans (MLGs) (Figure 1-1) (Albersheim *et al.*, 2010). Deposited under the primary wall after cell expansion is complete, the often multi-layered secondary cell wall generally contains high levels of lignin, a phenolic polymer, and is much thicker and more rigid than the primary cell wall. Thus, it is much more structurally robust and resistant to degradation. The major cross-linking glycans in secondary cell walls are typically galactoglucomannans (GGMs) and GAXs.

1.1.1.1 Xyloglucans

XyGs are a major class of water-soluble cross-linking glycans found in the primary cell wall of all land plants (Nishitani, 1998). They have a strong affinity for cellulose (Hayashi *et al.*, 1994, Lopez *et al.*, 2010) and evidence has been found for an essential role tethering cellulose microfibrils together within the plant cell wall (Takeda *et al.*, 2002, Cosgrove, 2005, Kong *et al.*, 2015). XyG, like most cross-linking glycans, is synthesised in the Golgi apparatus (Scheller and Ulvskov, 2010). As shown in Figure 1-1, it is built from a core structure consisting of a poly $\beta(1,4)$ -glucan backbone with $\alpha(1,6)$ -xylose branches on – depending on the plant species – two or three of every four consecutive glucose residues (Vincken *et al.*, 1997). Fry *et al.* proposed the current systematic nomenclature for XyGs where G denotes an unsubstituted $\beta(1,4)$ -glucose moiety and X denotes an $\alpha(1,6)$ -xylosyl- $\beta(1,4)$ -glucose moiety (Fry *et al.*, 1993). $\beta(1,2)$ -galactosylation of xylose residues (denoted by the letter L) is the most common modification of the core xyloglucan structure, yet the diversity of possible side chains is exemplified by the recent expansion of the xyloglucan nomenclature to include 19 possible moieties (Tuomivaara *et al.*, 2015). Reported acetylation further complicates the description of XyG structures (Gille and Pauly, 2012). The most common source of XyG for laboratory studies is a simple storage XyG

consisting of G, X, and L moieties (York *et al.*, 1993) which can be readily isolated from the endosperm of tamarind (*Tamarindus indica*) seeds (Aspinall, 1969).



Figure 1-1: Representative chemical structures of the major cross-linking glycans in plant cell walls.

1.1.1.2 Mixed-Linkage Glucans

Mixed-linkage glucans are relatively uncommon in an evolutionary context. They have only been found only in commelinid monocots (e.g. cereals) and eusporangiate monilophytes (e.g. horsetails) (Xue and Fry, 2012). MLGs are nonetheless a significant source of dietary fibre in the human diet. They are found at high levels in cereal grains (e.g. wheat, barley, rice, corn), making up roughly 65-75% of the non-starch soluble glycan content of the barley endosperm (Fincher, 1975, Bacic and Stone, 1981, Ebringerová, 2005). The structure of MLGs is a series of $\beta(1,3)$ -linked cellooligosaccharides (short $\beta(1,4)$ -linked glucans) (Woodward and Fincher, 1982, Woodward *et al.*, 1983). The distribution of cellooligosaccharide moiety lengths varies between organisms, with Iceland moss producing MLG with >90% cellotriose moiety content, *Equisetum sylvaticum* producing MLG with >90% cellotetraose moiety content, and barley producing MLG with roughly 60% cellotriose moieties, 30% cellotetraose moieties and 10% cellopentaose and cellohexaose moieties (Wood *et al.*, 1994, Fry *et al.*, 2008b, Simmons *et al.*, 2013). Like XyG, MLG has been shown to bind cellulose tightly with evidence suggesting that affinity is independent of the average cellooligosaccharide moiety length (Kiemle *et al.*, 2014). Similar to cellulose, but unlike other cross-linking glycans, MLG appears to be synthesised at the cell membrane (Wilson *et al.*, 2015). For laboratory studies, the most common sources of watersoluble MLG are oat (*Avena sativa*) and barley (*Hordeum vulgare*) seed. Hot, mildly alkaline water extraction followed by amylase treatment and ethanol precipitation yields a material that is 75% MLG, having protein, lipid, and some insoluble material as the major contaminants (Redmond and Fielder, 2006).

1.1.1.3 β-Mannans

β-Mannans are found in all land plants and come in many forms, all containing $\beta(1,4)$ linked mannose residues in their backbone structure (Carpita and McCann, 2000). Unsubstituted β -mannan is rare, having only been identified once in the pseudobulb of an orchid (Wang *et al.*, 2006). Mannan isolated from ivory nut (*Phytelephas macrocarpa*), having $\alpha(1,6)$ -galactosyl residues on only 1-7% of mannose residues (Aspinall et al., 1953), is often used as a readily isolable model of unsubstituted mannan. Galactomannans, having between one and three $\alpha(1,6)$ galactosyl residues for every four residues of the mannan backbone, can be readily isolated from guar (Cyamopsis tetragonoloba) or carob (Ceratonia siliqua) seed, where they serve as a storage polysaccharide (Buckeridge, 2010). Glucomannans are fundamentally different from galactomannans. They contain short stretches of $\beta(1,4)$ -linked glucan interspersed among longer stretches of $\beta(1,4)$ -linked mannan (Katsuraya *et al.*, 2003). Water-soluble glucomannan can be isolated from konjac (Amorphophallus konjac) root, where is serves as a storage polysaccharide. Glucomannans are generally partially acetylated (Maekaji, 1978, Tenkanen et al., 1993, Moreira and Filho, 2008). For example, konjac glucomannan has roughly one acetylation for every 10-20 backbone hexose units (Kato and Matsuda, 1969). Shown in Figure 1-1, galactoglucomannans (GGMs) have the same structure as glucomannans with the addition of $\alpha(1,6)$ -galactose

4

substitutions (Buckeridge, 2010). GGMs are the major cross-linking glycan found in the cell walls of softwoods where they serve as structural polysaccharides (Ebringerová, 2005). GGMs can be extracted from softwoods in a crude form by high-pressure hot-water extraction and further purified for use in a variety of potential applications (Willför *et al.*, 2008).

1.1.1.4 Xylans

Xylans are found in all vascular plants (Xue and Fry, 2012). Having a core structure consisting of $\beta(1,4)$ -linked xylose residues, xylans can be modified at the 2 or 3 positions with glucuronic acid (which may be further methylated at the 4-position), arabinofuranose, or acetate (Figure 1-1) (Carpita and Gibeaut, 1993, Carpita and McCann, 2000). Furthermore, feruloylation is a common modification of the arabinofuranose residues of xylans found in grasses. The major sources of xylans are hardwoods and cereals (Ebringerová, 2005). Thus, xylans are an essential component of the dietary fibre intake of most animals. Wheat (*Triticum aestivum*) and beechwood (*Fagus sp.*) are both common sources of alkaline-extracted arabinoxylan for laboratory study. Acetylated xylan or glucuronoxylan can be isolated from the cell walls of mature dicot tissues (such as hardwood) by dimethylsulfoxide (DMSO) extraction (Rowley *et al.*, 2013) or prepared through the partial chemical acetylation of certain xylan extracts (Johnson *et al.*, 1988).

1.1.1.5 Pectins

Pectins are complex polysaccharides rich in galacturonic acids (Carpita and McCann, 2000). The most common pectin is homogalacturonan, a linear $\alpha(1,4)$ -linked polysaccharide which is methyl-esterified and partially acetylated. Some homogalacturonans are further modified with short neutral sugar branches (Schols *et al.*, 1995). Rhamnogalacturonan I is similar to homogalacturonan, but has a dimeric backbone repeating unit of $\alpha(1,4)$ -galacturonic acid- $\alpha(1,2)$ -rhamnose. Lastly, the structures of highly complex rhamnogalacturonan II samples are, in some cases, still being worked out (Ndeh *et al.*, 2017). Pectins are generally found at the highest levels in the primary cell wall and their major roles are cell adhesion and signalling, with pectinase activity controlling stomatal opening and respiration (Caffall and Mohnen, 2009). Pectins are readily isolated from a variety of fruits, such as apples, grapes, and oranges, by mild acid extraction and precipitation. They are commonly used as gelling agents in cosmetics and food.

1.1.2 Enzyme Specificities in the Breakdown of Plant Polysaccharides

A vast abundance of carbohydrate-active enzymes (CAZymes) exist in nature. The carbohydrate-active enzyme database (CAZy) was launched in 1999 to provide a resource that correlates protein amino acid sequence, structure, and function using the method outlined by Bernard Henrissat (Henrissat, 1991, Lombard et al., 2014). For enzymologists, the biggest value of CAZy classification lies in its ability to predict the function of uncharacterised proteins. However, this is limited by the low number of characterised representatives within each CAZy family, the narrow substrate scope tested for many enzymes that have been characterised, and the identification of multiple substrate specificities within each family. In practice, classification into a CAZy family predicts a loosely-bounded space of potential substrates that an enzyme is likely to recognise. For example, as of 2017, 147 of 9905 sequences in the GH2 family have been characterised. Among those 147 characterised sequences are reported glucuronidases, glucosaminidases, xylosidases, mannosidases, galactosidases, and one L-arabinofuranosidase (www.cazy.org/GH2). However, with the exception of the arabinofuranosidase, all of these enzymes act on β -linked D-hexopyranoses in an *exo* fashion. Thus, for any enzyme belonging to GH2, experiments testing for the recognition of a relatively small set of β -linked D-hexopyranose substrates can reasonably be expected to identify a competent substrate, giving direction to the enzyme characterisation effort. The application of heuristics in this way has had a significant positive impact on the efficiency of CAZyme discovery.

The CAZy database defines families of glycoside hydrolases (EC 3.2.1.x), glycosyl transferases (EC 2.4.x.y), polysaccharide lyases (EC 4.2.2.x), carbohydrate esterases (EC 3.5.1.x/3.1.1.y), auxiliary activities (EC 1.x.y.z), and carbohydrate-binding modules based on amino acid sequence similarity. Each CAZy family shares a common 3-dimensional protein fold, set of catalytic residues, and mechanism of action.

1.1.2.1 Glycoside Hydrolases

Glycoside hydrolases (GHs) catalyse the hydrolytic cleavage of glycosidic bonds. The two most common mechanisms of action were described by Daniel Koshland (Koshland, 1953) (Figure 1-2). The inverting mechanism of glycoside hydrolysis begins with a water molecule positioned adjacent to a catalytic aspartate or glutamate base in the active site. A carbohydrate substrate binds across the active site with its hemiacetal carbon positioned over this water molecule. The glycosidic oxygen is placed within hydrogen bonding distance of a catalytic

aspartic or glutamic acid. This arrangement enables the simultaneous protonation of the leaving group oxygen and deprotonation of the nucleophilic water molecule. Combined with conformational strain to facilitate nucleophilic attack (Ardèvol and Rovira, 2015), this finely-tuned molecular arrangement is able to accelerate the rate of hydrolysis by as much as 10^{18} -fold (assuming an uncatalysed hydrolysis rate for α -glucose of 2×10^{-15} s⁻¹ (Wolfenden *et al.*, 1998) and a turnover frequency as high as 1500 s^{-1} for an α -glucanase under similar conditions (Yang *et al.*, 2013)).

The retaining mechanism of glycoside hydrolysis is called a ping-pong mechanism due to the formation of a substituted enzyme intermediate state. It begins with the same molecular arrangement with the exception that, instead of a water molecule, the catalytic nucleophile is positioned next to the hemiacetal carbon for direct nucleophilic attack. This leads to the formation of a glycosyl-enzyme intermediate, which is then cleaved by a water molecule which enters the active site. The water molecule fills the position of the glycosidic oxygen following the departure of the leaving group. If, however, a carbohydrate replaces the leaving group, this mechanism gives rise to a second catalytic path called transglycosylation.

Most retaining GHs will perform some transglycosylation at a sufficiently high substrate concentration; some have evolved to preferentially catalyse transglycosylation. The glycoside hydrolase family 2 (GH2) galactosidase, LacZ, catalyses an intramolecular transglycosylation to convert the $\beta(1,4)$ linkage of lactose into the $\beta(1,6)$ linkage of allolactose, which induces expression of the well-known lactose-metabolizing system (Wheatley *et al.*, 2013). The GH13 glycogen debranching enzyme rearranges highly branched glycogen molecules to improve their accessibility to glycogen phosphorylase without sacrificing the energy stored in the glycosidic bond (Berg *et al.*, 2002). The GH16 xyloglucan transglycosylases continuously cleave and reform long xyloglucan chains within the plant cell wall (Rose *et al.*, 2002, Eklöf and Brumer, 2010). It is hypothesised that this activity facilitates cell wall expansion.



Figure 1-2: Retaining and inverting mechanisms of glycoside hydrolysis. A) The inverting mechanism of glycoside hydrolysis is shown as a single concerted step. B) The two-step retaining mechanism of glycoside hydrolysis is shown for a $\beta(1,4)$ -linked glucoside. The divergence of the hydrolysis and transglycosylation pathways is highlighted.

1.1.2.2 Polysaccharide Lyases

Polysaccharide lyases (PLs) are fundamentally different from glycoside hydrolases in that they cleave glycosidic bonds through an elimination mechanism instead of a substitution mechanism (Yip and Withers, 2006). PLs take advantage of the ability of 4-O substituted hexuronic acids to undergo base-catalysed elimination. This breaks glycosidic bonds, forming a new reducing end and a hexenuronic acid. PLs are particularly important in the study of pectin and alginate breakdown due the high content of hexuronic acids in those substrates (Aspinall and Cañas-Rodriguez, 1958, Edstrom and Phaff, 1964).

1.1.2.3 Glycosyltransferases

Glycosyltransferases (GTs) are responsible for the synthesis of new glycosidic bonds from substrates having higher energy glycosidic bonds. Leloir GTs catalyse the transfer of a donor monosaccharide from a sugar mono- or diphosphonucleotide to an acceptor sugar (Lairson *et al.*, 2008). Non-Leloir GTs catalyse the same reaction using other high-energy donors such as sugar pyrophosphates. As with GHs, GTs can follow retaining or inverting mechanisms, although their details are outside of the scope of this thesis; see (Lairson *et al.*, 2008) for a review of GT mechanisms. Highly specific donor and acceptor recognition ensure the formation of specific glycosidic bonds at specific positions (Zheng *et al.*, 2011).

1.1.2.4 Carbohydrate Esterases

Carbohydrate esterases (CEs) are an essential part of the carbohydrate saccharification process since modifications such as acetylation and feruloylation can interfere with the action of some GHs, preventing carbohydrate catabolism (Grohmann *et al.*, 1989). CEs catalyse the hydrolysis of O- or N-acyl groups from carbohydrates. Although several CE mechanisms have been described, CEs primary act through a lipase-like ping-pong mechanism involving a serine-histidine-aspartate catalytic triad (Biely, 2012) (Figure 1-3). The substrate is positioned in the active site with the oxygen of a serine residue adjacent to the central, electrophilic carbon of the acyl group. The proton of the serine residue is abstracted by the histidine group which may be polarised and oriented by an aspartate group. A carbohydrate esterase employing a novel cysteine-containing catalytic triad is described in Chapter 5 of this thesis.

1.1.2.5 Auxiliary Activities

Auxiliary activity enzymes are so-named for their ability to facilitate the degradation of carbohydrates by other enzymes (Levasseur *et al.*, 2013). All of the currently classified AA families consist of redox-active enzymes. These act on substrates including lignin and lignols, monosaccharides and oligosaccharides, alcohols, and soluble and insoluble polysaccharides. The mechanisms of many of these enzymes are still being elucidated, but most use an ionic copper cofactor or an organic cofactor such as flavin adenine dinucleotide (FAD). AA enzymes have

been shown to enhance the rate and completeness of the enzymatic deconstruction of complex cellulosic biomass (Hemsworth *et al.*, 2015).



Figure 1-3: The serine hydrolase mechanism of carbohydrate esterase activity. The deacetylation of β-glucose is shown in 2 key steps: cleavage of the carbohydrate ester to form an acyl-enzyme intermediate and hydrolysis of the acyl-enzyme intermediate by water. The common aspartate residue is not shown.

1.1.2.6 Carbohydrate-Binding Modules

Carbohydrate-binding modules (CBMs) are protein domains which display affinity, but not catalytic activity, towards specific carbohydrates (Gilkes *et al.*, 1988, Boraston *et al.*, 2004). CBMs are often found fused to GHs. They can potentiate catalysis through a targeting effect (helping the enzyme find its ideal site of attack) (Hervé *et al.*, 2010), a proximity effect (increasing the local concentration of enzyme around the substrate) (Gilbert *et al.*, 2013), or an adhesive effect (attaching cells to their desired substrate) (Montanier *et al.*, 2009a). CBMs have been organised into over 80 families. Roughly a third of these families have members which bind cellulose and roughly a fifth have members which bind chitin. Interestingly, roughly half of the known CBM families have members which recognise cross-linking glycans or pectins. The most common structure of a CBM is a simple β -sandwich fold with a single carbohydrate binding site which has a shape complementary to that of the carbohydrate ligand.

1.2 Carbohydrate-Active Enzyme Discovery

The advent of highly efficient DNA sequencing technologies has led to an explosion in the number of known protein-coding sequences; however, there has been comparably little parallel growth in the efficiency of enzyme discovery. Thus, the functions of the large majority of CAZymes remain unknown. This preponderance of uncharacterised sequences remains a valuable potential source of new enzymatic specificities.

The search for new enzymatic activities is generally accomplished through careful detective work. Clues that a novel enzymatic specificity may be expressed can include the utilisation of nutrient sources requiring activities which no known enzyme displays (Speciale et al., 2016), the co-expression of a known gene with one having no characterised homologues (Martens et al., 2009, Ndeh et al., 2017), or the identification of phylogenetic clades with no characterised representatives (Eklöf et al., 2013). Plants, bacteria and fungi have all historically been fruitful sources of CAZymes. However, polysaccharide utilisation loci (PULs), a recently discovered type of bacterial gene cluster devoted to the deconstruction of complex carbohydrates, have been a particularly valuable source of clues about enzyme function (Terrapon *et al.*, 2017). The characterisation of PULs devoted to the breakdown of α -mannans (Cuskin et al., 2015), β-mannans (Bågenholm et al., 2017), starch (Martens et al., 2009), xyloglucans (Larsbrink et al., 2014a), mixed-linkage glucans (Tamura et al., 2017), laminarins (Kabisch et al., 2014), chitins (Larsbrink et al., 2016), arabinans (Arnal et al., 2015), galactans (Hehemann et al., 2012), fructans (Sonnenburg et al., 2010), xylans (Dodd et al., 2010b, Rogowski et al., 2015), mucins (Crost et al., 2013), and pectins (Luis et al., 2017, Ndeh et al., 2017) have each uncovered the activities of diverse arrays of enzymes. The identification and characterisation of members of a PUL in Prevotella bryantii, a Gram-negative species of bacteria isolated from bovine rumen, is presented in Chapter 4 of this thesis.

1.2.1 CAZymes in Plants

In spite of their life-cycle being dependent on the synthesis of large quantities of polysaccharides by GTs, plant genomes encode equally large arrays of GHs. For example, the genome of *Arabidopsis thaliana*, a well-developed model plant species, encodes 360 GTs and 387 GHs among its 1224 putative CAZymes (Ekstrom *et al.*, 2014). Due to gene multiplicity and the presence of complex regulatory mechanisms, the functions of many of these genes have only been determinable using modern recombinant enzyme production technologies. Among the GHs commonly found in plant genomes are large numbers of GH1s, including the prototypical almond β -glucosidase (Armstrong *et al.*, 1908), which are known to break down specific glycosides to release secondary metabolites (Czjzek *et al.*, 2000, Ahn *et al.*, 2007), GH9s, which are primarily *endo*- β (1,4)-glucanases (Libertini *et al.*, 2004), GH16s, which are primarily xyloglucanases (Kaewthai *et al.*, 2013), GH17s, which are primarily *endo*- β (1,3)-glucanases (Varghese *et al.*, 1994), and GH28s, which are primarily pectinases (Markovič and Janeček,
2001). The structure and function of a representative of a new group of plant enzymes related to the GH16 xyloglucanases is presented in Chapter 2 of this thesis.

A recent example of plant CAZyme discovery efforts has uncovered a thermophilic enzyme in cactus tissue with potential industrial applications (Vikramathithan *et al.*, 2010). Plants have also been the source of many GHs harbouring surprising mutations in their catalytic machinery. For example, myrosinase, a GH1 glucosidase which hydrolyses glucosinolates as part of a defense mechanism employed by *Brassica* species, has evolved to function without the normal GH1 catalytic nucleophile (Burmeister *et al.*, 2000). In a very significant departure from enzymatic origins, a GH18-homologous gene without a catalytic amino acid residue participates in transmembrane signalling in *Nicotanio tabacum* (Kim *et al.*, 2000).

1.2.2 Fungal CAZymes

Fungi have been the source of a significant number of CAZymes used industrially today. Following observations of its ability to grow on canvas, *Trichoderma reesei* (a filamentous fungus from the Solomon Islands) was identified as the first source of cellulases (Reese, 1956, Bischof *et al.*, 2016). *T. reesei* secretes a variety of cellulose-degrading GHs including processive cellobiohydrolases, which cleave cellobiose from the end of cellulose chains, *endo*- β glucanases, which cleave cellulose chains randomly, *exo*- β -glucosidases, which cleave cellooligosaccharides into glucose, and AA9 lytic polysaccharide monooxygenases, which oxidatively cleave cellulose chains. Cross-linking glycan-degrading enzymes, such as xylanases and mannanases, have also been isolated from *T. reesei* (Wong and Saddler, 1992, Tenkanen *et al.*, 1997, Martinez *et al.*, 2008).

While *Trichoderma* species form the basis of modern industrial cellulose degradation, many other genera of fungi have proven to be valuable sources of CAZymes, including *Aspergillus, Fusarium, Neurospora, Penicillium, Phanerochaete, Schizophyllum,* and *Podospora* (van den Brink and de Vries, 2011). Enzymes identified in these fungi provide the ability to degrade virtually all of the known components of plant cell walls.

Fungal enzyme discovery has historically been laborious. It generally begins with a culture of an isolated fungus on a carbohydrate substrate ranging from complex lignocellulosic biomass (Couger *et al.*, 2015) to a purified plant cell wall component (Silva *et al.*, 2015). Enzymes are identified in the culture by assaying for desired activities, and the enzyme(s) responsible for the positive assay result are isolated by a tailor-made multi-step purification

protocol. Protein and DNA sequencing technologies are then used to identify the gene giving rise to the observed activity. Significant advances in modern transcriptomics, proteomics, and bioinformatics have facilitated an increase in throughput and completeness of the identification of fungal CAZymes (Wang *et al.*, 2011), yet these analyses remain complex and costly.

1.2.3 Bacterial and Archaeal CAZymes

As significant contributors to the breakdown of plant biomass throughout the biosphere, Gram-negative and Gram-positive soil bacteria have been excellent sources of CAZymes and highly efficient cellulosomal biomass-degrading systems (DeBoy *et al.*, 2008, Blouzard *et al.*, 2010).

Carbohydrate-degrading systems in many Gram-negative soil bacteria have been mined extensively for CAZymes. Several examples related to the work presented in this thesis include the *Bacillus licheniformis* GH16 mixed-linkage glucanase (Planas *et al.*, 1992, Hahn *et al.*, 1995). This highly efficient *endo*- β -glucanase is the closest characterised bacterial relative of the EG16s presented in Chapter 2 of this thesis (30% amino acid sequence ID, reciprocal best hit). The first identified GH5 *endo*-xyloglucanase, PpXG5, was isolated from *Paenibacillus pabuli*. This enzyme is a close relative (29% amino acid sequence ID, reciprocal best hit) of the GH5 endo- β -glucanase presented in Chapter 3 of this thesis. One of β -galactosidases used to prepare degalactosylated xyloglucan-derived substrates in Chapters 2 and 3 of this thesis, Bgl35A, was originally isolated from *Cellvibrio japonicus*, a cellulolytic bacterium isolated from Japanese soil (Larsbrink *et al.*, 2014b).

Gram-positive bacteria are remarkable for their extensive employment of cellulosomal architectures. These large non-covalently associated clusters rich in CAZymes assemble on the surfaces of bacteria such as *Ruminiclostridium cellulolyticum* (Desvaux, 2005) and *Clostridium thermocellum* (Hirano *et al.*, 2016). Many important non-cellulosomal CAZymes have also been isolated from Gram-positive bacteria such as the GH5 *endo*-β-glucanases and GH11 *endo*-β-xylanases originally isolated from *Ruminococcus flavefaciens* (Zhang *et al.*, 1994).

To serve the many commercial processes requiring high temperatures or other extreme conditions, many CAZymes have been identified in and isolated from extremophilic bacteria or archaea. Particularly, a thermophilic β -glucosidase, optimally active near 100°C, has been isolated from *Pyrococcus furiosus* (Kengen *et al.*, 1993) and a similarly thermophilic cellulase

has been isolated from *Thermotoga maritima* (Bronnenmeier *et al.*, 1995). Thermophilic xylanases have been isolated from species of *Geobacillus* bacteria (Bibi *et al.*, 2014, Bhalla *et al.*, 2015) and a dual-function GH9 *endo*-β-glucanase/GH5 *endo*-β-mannanase has been isolated from thermophilic *Caldicellulosiruptor bescii* bacteria.

1.2.4 Polysaccharide Utilisation Loci in Gut Bacteroidetes

The human genome encodes 97 glycoside hydrolases, of which only 8 have been confirmed to be involved in the digestion of food (Kaoutari *et al.*, 2013). These genes encode enzymes which break down sucrose, lactose, and starch. To digest more diverse carbohydrates, most animals rely on large populations of microorganisms in their gut which, together, encode tens of thousands of CAZymes (Hess *et al.*, 2011). Under the anaerobic conditions of the gut, these organisms derive energy from the fermentation of complex carbohydrates into fatty acids which are secreted, taken up by the host and metabolised aerobically (Filippo *et al.*, 2010). The microorganisms that perform this fermentation provide the diversity of metabolic function necessary to match the diversity of food sources available in the biosphere.

The major Gram-negative bacteria in the mammalian gut are Bacteroidetes. These bacteria are remarkably versatile carbohydrate-degraders. Individual species can possess a wide array of carbohydrate-degrading capacities (Martens *et al.*, 2011) and can occupy highly specific niches (Hehemann *et al.*, 2010). The genome of a typical Bacteroidetes species encodes 20-70 GHs or PLs (Kaoutari *et al.*, 2013). Thus, a typical gut symbiont species is able to degrade a wider variety of plant carbohydrates than a human. Gut microbiota provide the plasticity necessary to make use of changing food sources. Species abundance changes in response to diet (Wu *et al.*, 2011), genetic make-up changes with the seasons (Smits *et al.*, 2017), and new capacities are acquired from the environment when new food sources are exploited (Hehemann *et al.*, 2010). Transcriptomic work has shown that the genes which encode the enzymes responsible for the breakdown of a particular carbohydrate are co-localised in gene clusters which are highly upregulated in the presence of a specific substrate (Martens *et al.*, 2009). These gene clusters are generally referred to as polysaccharide utilisation loci, or PULs.

The archetypical PUL, the starch utilisation system (SUS) first identified in *Bacteroides thetaiotaomicron*, imbues *Bacteroides* species with the ability to ferment starch. It encodes four major elements: surface glycan-binding proteins (SusDEF), a TonB-dependent transporter (SusC), a complement of extracellular or periplasmic glycoside hydrolases (SusABG), and a

sensor/regulator (SusR) (D'Elia and Salyers, 1996, Reeves *et al.*, 1997, Shipman *et al.*, 2000) (Figure 1-4). Most functional PULs require versions of each of these elements.



Figure 1-4: The current model of the prototypical starch utilisation locus from *B. thetaiotaomicron*. Starch is displayed as a series of dark blue circles with lines drawn to indicate the presence of $\alpha(1,6)$ -linked branches. SusD, SusE, SusF, and SusG are all shown anchored to the surface of the outer membrane (OM) of the cell. SusC is shown as a transmembrane barrel with a periplasmic domain in contact with TonB. SusA and SusB are shown floating freely in the periplasm. SusR is shown as a transmembrane sensor/regulator. An additional passive transporter is shown embedded in the inner membrane (IM) to facilitate diffusion of glucose into the cytoplasm. Proteins possessing enzymatic activities are coloured dark blue. Figure adapted from (Koropatkin *et al.*, 2012).

1.2.4.1 Surface Glycan-Binding Proteins

Surface glycan-binding proteins (SGBPs) are responsible for adhering to specific carbohydrate substrates. They have variable sequences and numbers of binding sites, but often

share common domain organisations and folds. SusD, one of the first characterised SGBP, contains a single starch-binding site (Koropatkin *et al.*, 2008) which is dispensable for locus functionality (Cameron *et al.*, 2014), SusE contains two binding sites, and SusF contains three (Cameron *et al.*, 2012). SusD folds as a single mostly helical protein domain containing a tetratricopeptide repeat (TPR) motif often associated with protein-protein interactions (Blatch and Lässle, 1999, Bakolitsa *et al.*, 2010). SusE and SusF fold as three and four separate immunoglobulin-like domains, respectively. In contrast, non-homologous SusE-like proteins from the *Bacteroides ovatus* xyloglucan and xylan utilisation loci both have single carbohydrate-binding sites in a larger C-terminal domain following two non-carbohydrate-binding immunoglobulin-like domains (Rogowski *et al.*, 2015, Tauzin *et al.*, 2016). Overall, because of the high variability in sequence and lack of conserved carbohydrate-binding sites, the structures and functions of most SGBPs cannot yet be reliably predicted from phylogenetic analyses.

1.2.4.2 TonB-Dependent Transporters

TonB-dependent transporters (TBDTs) are active transporters embedded in the outer membrane which are known to transport vitamins, minerals, and carbohydrates into the periplasm (Noinaj *et al.*, 2010). They are actuated, using a proton motive force, by the TonB-ExbB-ExbD inner-membrane complex. Recently, it has been shown that a TBDT forms a complex with a SusD-like protein (Glenwright *et al.*, 2017). Separately, the presence of SusD, with or without carbohydrate-binding capacity, been shown to be essential for carbohydrate uptake in *Bacteroides* (Cameron *et al.*, 2014, Tauzin *et al.*, 2016).

TonB-dependent transporters and surface glycan-binding proteins are tightly associated in PULs. TBDTs and SusD-like SGBPs show significant sequence conservation between PULs targeting the same substrate. Thus, these two elements are commonly used as the genetic hallmark of the presence of a PUL (Terrapon *et al.*, 2017). Additional SGBPs (SusEF-like) are commonly found associated with SusCD-like pairs, but are not conserved between PULs.

1.2.4.3 Carbohydrate-Degrading Enzymes

The organisation of carbohydrate-degrading enzymes is highly dependent on the structure and complexity of the target substrate. The starch-utilisation system has one surface-exposed *endo*-amylase which cleaves large maltooligosaccharides from starch (Reeves *et al.*, 1997). This system also encodes two periplasmic *exo*- α -glucanases which are responsible for generating free glucose by cleaving the $\alpha(1,4)$ and $\alpha(1,6)$ linkages in the imported oligosaccharides. A similar organisation, with an *endo*-glycanase on the bacterial surface and a collection of *exo*-glycanases which generate monosaccharides in the periplasm, has been reported for almost all of the PULs listed at the top of section 1.2. Exceptionally, the recently reported *B. thetaiotaomicron* rhamnogalacturonan II utilisation locus appears to lack a surface-exposed *endo*-glycanase, yet encodes 21 GHs, 1 PL, and 3 CEs (Ndeh *et al.*, 2017), reflecting the small size and high complexity of the substrate.

1.2.4.4 Sensor/Regulators

susR encodes a transmembrane sensor/regulator which regulates gene transcription through maltooligosaccharide-dependent DNA-binding (D'Elia and Salyers, 1996, Martens *et al.*, 2009). Starch is thought to be sensed by *B. thetaiotaomicron* through low levels of background locus expression (Martens *et al.*, 2011). Providing similar regulatory behavior to SusR, hybrid two-component system (HTCS) sensor/regulators are common in glycan-degrading systems in B. *thetaiotaomicron* and *B. ovatus* (Sonnenburg *et al.*, 2006). The SusR and HTCS architectures are able to take advantage of the unique linkage patterns associated with different polysaccharide substrates by transmitting a signal from the periplasm, where complex oligosaccharides are actively being hydrolysed into monosaccharides, into the cytoplasm. This relay system creates a specific and responsive negative feedback loop.

Regulation by a cytosolic carbohydrate-binding transcription regulator has not been reported for any PUL characterised to date, but several carbohydrate-binding bacterial transcription regulators are known in Gram-negative bacteria. The *E. coli* AraC gene is the prototype for a large family of bacterial transcription regulators (Gallegos *et al.*, 1997). Similar to LacI, the well-known transcription regulator (Jacob and Monod, 1961) used in recombinant protein production (Studier and Moffatt, 1986), these regulators have a ligand-binding domain, and a DNA-binding domain which activates the transcription of some genes and inhibits the transcription of others when a ligand is bound (Greenblatt and Schleif, 1971, Rodgers and Schleif, 2009). AraC-homologous regulators have been implicated in a variety of carbohydrate-dependent processes including arabinose metabolism (Lee *et al.*, 1987), carbohydrate transport (Russell *et al.*, 1992), and biofilm synthesis (Rowe *et al.*, 2016). An AraC-like regulator is found in the PUL presented in Chapter 4 of this thesis.

1.3 Quantifying CAZyme Activities

The measurement of CAZyme activities is best thought of as an exercise in measuring *changes* in carbohydrate-containing materials. For example, lysozyme catalyses the hydrolysis of $\beta(1,4)$ linkages between N-acetylmuramic acid and N-acetyl-D-glucosamine (Manchenko, 2002). This can be measured based on, for example, the formation of new reducing ends or the lysis of bacterial cells. However, to understand the mechanism that underlies these observations, the specific bonds being broken must be identified. The identification of lysozyme's activity was dependent on knowledge of the existence of such a linkage within an isolable material, and the ability to distinguish a material containing such linkages from one which no longer contains them. This section describes tests, many of which are used throughout this thesis, which have been developed to detect and quantify chemical changes induced by CAZymes. Section 1.4 describes techniques for the chemical analyses of both carbohydrate substrates and products which enable us to pinpoint the bond recognised by an enzyme.

1.3.1 Quantifying Glycoside Hydrolysis

1.3.1.1 Reducing End Quantitation

Within the relatively narrow enzymological scope of measuring glycoside hydrolase activities, there are many observable changes induced by an enzyme which have been documented: an increase in the total concentration of aldehydes (reducing ends) in solution (Reese *et al.*, 1950, Jermyn, 1952), a loss of viscosity (Fuwa, 1954), reduced iodine-binding capacity, or the release of dye-labelled oligosaccharides from insoluble polysaccharides (Barnes and Blakeney, 1974). While not all polysaccharides bind to dyes or iodine, the release of reducing ends is always evidence of the hydrolysis of any polysaccharide. This action produces limited spectroscopic changes, with no changes in the readily accessible ultraviolet or visible spectroscopic ranges. However, free reducing ends have significantly different reactivity from glycosidic linkages. This difference can be exploited in a variety of ways. For example, reducing sugars will, when heated in alkaline medium, quantitatively reduce Cu(II) ions (Smogyi, 1952). This can be exploited to generate a visible colour change, based on the different spectroscopic properties of Cu(I) and Cu(II) complexes (Smith *et al.*, 1985). Similarly, when heated in an acidic medium, 3,5-dinitrosalicylic acid reacts with reducing ends to quantitatively generate 3-amino-5-nitrosalicylic acid which has a distinct visible spectroscopic signature (Miller, 1959).

Though not as commonly used, the *p*-hydroxybenzoic acid hydrazide (Lever, 1973) and 3methyl-2-benzothiazolinone hydrazone (Anthon and Barrett, 2002) methods can also be used to quantify reducing ends in solution. Overall, the measurement of reducing ends is a relatively crude technique prone to interference from other biological materials (Gusakov *et al.*, 2011), making more specific assays desirable.

1.3.1.2 Enzyme-Linked Assays

Enzymes are remarkably specific catalysts, thus enzyme-linked assays are generally better suited to complex mixtures or assays in which the starting material and product cannot be otherwise differentiated. Two common glucose-specific assays are the hexokinase assay (Scheer *et al.*, 1978) and the glucose oxidase assay (Huggett and Nixon, 1957). The hexokinase assay is based on the quantitative generation of UV-absorbing reduced nicotinamide adenine dinucleotide (NADH) by glucose-6-phosphate dehydrogenase following the generation of glucose-6phosphate from glucose and adenosine triphosphate (ATP) by hexokinase. The glucose oxidase assay is based on the oxidation of the reducing end of glucose, converting oxygen into hydrogen peroxide (Gibson *et al.*, 1964). Hydrogen peroxide can then be quantified through electrochemical detection (Cass *et al.*, 1984) or through reaction with chromogenic reagents such as Amplex Red (Zhou *et al.*, 1997). While these enzyme-linked assays are highly sensitive, inexpensive, and highly specific, they require enzymes which can recognise the molecule(s) of interest and generate a usable signal. Thus, a significant up-front investment in enzyme discovery and assay optimisation is needed for each analyte of interest.

1.3.1.3 HPLC Assays

Techniques such as mass spectrometry, HPLC, and NMR offer the analytical power to identify and quantify the products of virtually any enzymatic reaction. However, these are all "low-throughput" techniques: They offer little-to-no ability to perform measurements in parallel, require comparably long measuring times, and can require chemical derivatisation or specially prepared standards. Of the low-throughput methods, HPLC techniques are generally simplest and most robust. For a given enzyme-substrate combination, a general HPLC assay requires a column which is selective enough to separate substrate, products, and any interferents from one another and a detector which can quantify products. For carbohydrates, high-performance anion exchange coupled to pulsed amperometric detection (HPAEC-PAD, *vide infra*) offers the

19

sensitivity, selectivity, and robustness necessary for kinetic measurements of glycoside hydrolases (McGregor *et al.*, 2017a). Products of glycoside hydrolysis ranging from monosaccharides to complex oligosaccharides can be unambiguously quantified over time to determine reaction rates. In Chapter 2 of this thesis HPAEC-PAD was used to measure the hydrolytic kinetics of an *endo*- β -glucanase acting on a variety of oligosaccharides. The use of HPAEC-PAD enabled the simultaneous quantification of the products of different modes of substrate recognition. The application of this HPAEC-PAD kinetic method shown in Chapter 3 of this thesis enabled the quantitation of kinetic parameters at exceptionally low substrate concentrations.

1.3.1.4 Chromogenic Substrates for Glycoside Hydrolases

Chromogenic assays have been used successfully in the discovery of new glycoside hydrolases for several decades. Teather and Wood screened microorganisms for secreted cellulases on agar plates containing carboxymethylcellulose (CMC) stained with Congo red, a cellulose-binding dye (Teather and Wood, 1982). Colonies of microorganisms secreting cellulases were identified based on the formation of a "halo" with less staining due to the hydrolysis of CMC and loss of dye-binding. This assay, though used qualitatively by Teather and Wood, can be used to quantify relative enzyme activities based on the release of dye into solution. Similarly, azurine cross-linked polysaccharides are widely used to detect a variety of *endo*-glycosidase activities (Li *et al.*, 2011).

Reducing-end modified synthetic chromogenic oligosaccharides and monosaccharides offer the ability to quantify enzyme turn-over rates with a variety of substrates. These substrates are typically prepared through the glycosylation of a labile chromogenic leaving group. For the leaving group to be a useful reporter of hydrolysis, it must have different spectroscopic properties in its glycosylated and deglycosylated forms. Phenol derivatives, such as 4-nitrophenol (PNP), 2-chloro-4-nitrophenol (CNP), and 2,4-dinitrophenol (DNP), are commonly used in absorbance-based assays because they have low pKa values (7, 5.5, and 4 respectively) and are only visibly coloured in their anionic forms. This makes them good leaving groups and good reporters. Many glycosidases are optimally active below pH 7, so CNP and DNP derivatives are often also favoured because the development of colour can be monitored in real-time, accelerating data collection and improving data quality. Due to their sensitivity, scalability, and suitability for *in vivo* experiments, fluorescence-based assays are commonly employed using

derivatives of resorufin (Coleman *et al.*, 2007, Ibatullin *et al.*, 2009) or 4-methylumbelliferone (Rosenthal and Saifer, 1973), which both undergo significant changes in absorbance and fluorescent properties upon deprotonation.

The favourable properties of chromogenic oligosaccharides and monosaccharides have been used in a variety of ways for high-throughput enzyme discovery and enzyme characterisation. For example, DNP-cellobioside has been used to screen metagenomic clone libraries for expressed cellulase (or glucosidase) activity in a high-throughput multi-well platebased robotic assay (Mewis *et al.*, 2011). Resorufin-galactoside has been used in a microfluidic system to probe galactosidase activity in sub-nanolitre droplets (Sjostrom *et al.*, 2013). Libraries of chromogenic substrates have also been used in active-site mapping (Ariza *et al.*, 2011). This method, using CNP derivatives of cellooligosaccharides and xyloglucan-derived oligosaccharides to map part of the active site of an *endo*- β -glucanase, has been applied to the characterisation of enzymes in both Chapters 2 and 3 of this thesis.

1.3.2 Quantifying Carbohydrate Esterase Activity

1.3.2.1 Chromogenic Substrates

The measurement of CE activity, similar to the measurement of GH activity, relies on either synthetic chromogenic substrates, or natural oligosaccharide or polysaccharide substrates. The most common chromogenic substrate used in the detection of acetylesterase activity is 4nitrophenyl acetate (PNP-Ac) (Huggins and Lapides, 1947, Levine *et al.*, 2008). PNP-Ac is a highly labile substrate, making it generally suitable for the detection of even low levels of acetylesterase activity and suitable for the detection of esterase activity under even mildly alkaline conditions. Fluorimetric esterase assays can also be performed using fluorescein diacetate as a substrate (Roberts and Rosenkrantz, 1966). Due to the similarity between carbohydrate acetylesterases and lipases, acetylesterase specificity towards acetate is generally experimentally verified by comparison with a butyrate derivative (e.g. PNP-Butyrate). The longer carbon chain of PNP-Bu is generally favoured by lipases and disfavoured by acetylesterases (Kastle, 1906).

1.3.2.2 Enzyme-Linked Assays

The release of acetate from non-chromogenic substrates can be followed using an enzyme-linked assay. This method, first described by Rose *et al.* (Rose *et al.*, 1954), takes

advantage of three different enzymatic activities to convert NADH into NAD⁺, giving a reduction in the absorbance of 340 nm light. Acetate kinase performs the first step of the reaction, using added ATP to phosphorylate all of the free acetate in solution. Secondly, added phosphoenolpyruvate and pyruvate kinase convert the ADP thus generated back into ATP, releasing free pyruvate which is reduced into lactate using added lactate dehydrogenase. This process consumes NADH quantitatively, giving a measurable change in UV absorbance. This assay is applied to the measurement of acetate release from glucomannan and xylan in Chapter 5 of this thesis.

1.3.3 Measuring CAZyme-Substrate Interactions through Kinetic Analysis

In his writings on enzymes, John Haldane wrote "The key to a knowledge of enzymes is the study of reaction velocities" (Haldane, 1930). Through the measurement of reaction velocities, kinetic experiments provide valuable information about the catalytic efficiency and specificity of an enzyme, and about how it recognises its substrates. Kinetic experiments also determine the conditions (e.g. pH, temperature, cofactor concentration) under which an enzyme is optimally active. Measuring the parameters within a kinetic model for any given enzyme in otherwise controlled conditions generally requires a series of time-dependent quantitative experiments performed at varying substrate concentrations using a method such as one described in sections 1.3.1 and 1.3.2. A plot of observed reaction rate (normalised to enzyme concentration) vs. substrate concentration forms a curve such as those shown in (Figure 1-5). With these data in hand, it is possible, with the use of a well-considered kinetic model and an iterative non-linear fit, to determine key parameters governing interactions between enzyme and substrate. Comparisons of such parameters between enzymes acting on a particular substrate or between substrates being acted on by a particular enzyme support a general understanding of enzyme-substrate interactions of different types.

1.3.3.1 Models of Enzyme Kinetics

How the rate of an enzyme reaction is affected by changes in conditions such as substrate concentration, co-factor concentration, or inhibitor concentration must be considered through the lens of a kinetic model. The most common model used in the characterisation of enzymes is the Michaelis-Menten model (Michaelis and Menten, 1913). The Michaelis-Menten model is the derived from minimal chemical pathway assuming a single binding event and a single catalytic

event. The chemical pathway for the mechanism employed by retaining glycoside hydrolases is depicted in (Figure 1-5A). This model allows the description of the catalytic efficiency that an enzyme displays with a particular substrate based on both the substrate concentration at which half of the total enzyme population is bound to substrate (Equation 1-1) and the rate at which enzyme-substrate complexes generate product molecules (Equation 1-2).



Figure 1-5: Kinetic models of enzyme kinetics. A) A kinetic model for the retaining mechanism of glycoside hydrolysis is shown including the initial formation of the ES complex, the release of the aglycone

(P₁) and formation of the glycosyl-enzyme intermediate (EP₂), and the subsequent hydrolysis of this intermediate to release P₂. The binding of a second substrate molecule to the glycosyl-enzyme intermediate, leading to transglycosylation to form P₃ is also shown. B) The derived mathematical formula depending on only observed initial rates (v₀) and initial substrate concentrations ([S]₀) and an exemplar curve ($K_M = 0.5$ mM, $k_{cat} = 35$ min⁻¹) resulting from the fitting of measured rates across a range of substrate concentrations.

C) A derived mathematical formula including the equilibrium constant for the binding of the second

susbtrate (K_i) and an exemplar curve ($K_M = 0.5 \text{ mM}$, $k_{cat} = 200 \text{ min}^{-1}$, $K_i = 1 \text{ mM}$) showing a decrease in the observed rate of the enzyme-catalysed reaction as initial substrate concentration increases.

Equation 1-1

$$K_M = \frac{k_{-1} + k_2}{k_1} \cdot \frac{k_3}{k_2 + k_3}$$

Equation 1-2

$$k_{cat} = \frac{k_2 k_3}{k_2 + k_3}$$

This model is valid under the same assumptions as the Michaelis-Menten model: 1) that the vast majority of substrate molecules are not bound to enzyme molecules, 2) that the enzyme is at thermodynamic binding equilibrium with its substrate, and 3) that the substrate concentration does not change over time. The first assumption is generally true in both kinetic experiments and in cells because enzymes rarely have high substrate affinity and low catalytic turnover rates. The second assumption is harder to ensure, but analysis of the model shows that any violation of this assumption will only affect the apparent K_M value obtained. The last assumption is true so long as no more than a few percent of the total substrate pool is consumed during the measurement period.

Deviations from the curve shape in (Figure 1-5B) indicate that the simple Michaelis-Menten model is not valid. This can occur for a variety of reasons including cooperativity, allosteric interactions, or inhibition (Bardsley *et al.*, 1980). For example, substrate inhibition, observed in experiments involving the retaining *endo*- β -glucanases described in Chapters 2 and 3 of this thesis, is observed when, following the formation of the glycosyl-enzyme intermediate, a second substrate molecule binds in such a manner as to prevent hydrolysis from taking place (Kaiser, 1980). A kinetic model (Figure 1-5A) can be developed from which, under the additional assumption that the SEP₂ complex is not catalytically active ($k_{cat3} = 0$), a modified Michaelis-Menten model can be derived which fits with the observed behaviour. Although additional data points are required to obtain a meaningful fit from which kinetic parameters can be extracted, the same basic experimental design can be used. Solving for the K_M and k_{cat} requires also determining the dissociation constant ($K_i = k_{off}^*/k_{on}^*$) for the second binding event. This gives a plot similar to that shown in (Figure 1-5C). If $k_{cat3} \neq 0$ (e.g. transglycosylation takes place), then the curve observed Figure 1-5C simply tends towards the value of k_{cat3} .

1.3.3.2 Relating Kinetic Parameters to Molecular Interactions

Measurements of enzyme kinetics offer the information necessary to map the interactions between enzymes and their substrates (Fersht, 1974). By measuring the catalytic efficiency of an enzyme acting on a family of substrates, the differences in the change in free energy required for the reaction to take place (or change in transition state free energy, ΔG^{\ddagger}) between each substrate ($\Delta\Delta G^{\ddagger}$) can we determined. A carefully chosen set of structurally-related substrates will provide information about which parts of the substrate molecule the enzyme is most dependent on for catalysis.

Within the context of the Michaelis-Menten model, the ratio of k_{cat} to K_M (often called the 'specificity constant', but preferably called the 'performance constant' (Koshland, 2002)) is generally taken as a measure of enzyme catalytic efficiency (Eisenthal *et al.*, 2007). This convention enables meaningful comparisons of enzyme performance across a library of substrates. Changes in transition state free energy can be related to changes in performance constant through Equation 1-3 (Street *et al.*, 1989).

Equation 1-3

$$\Delta \Delta G^{\ddagger} = RT ln \frac{(k_{cat}/K_M)_{s2}}{(k_{cat}/K_M)_{s1}}$$

The choice of substrates to test depends on both the specificity of the enzyme and the desired level of molecular detail. High levels of detail, down to looking at the effects of individual functional groups, can be invaluable in enzyme engineering (Namchuk and Withers, 1995). Lower levels of detail, such as looking at monosaccharide arrangement or linkage types, are valuable for understanding the full range of substrates an *endo*-glycanase is able to degrade (Brayer *et al.*, 2000). This information is particularly informative in combination with the molecular structure of the enzyme under study.

1.4 Determining Enzyme Specificity through Product Analysis

The use of carbohydrate substrates for the discovery and characterisation of CAZymes requires an understanding of what the components of the polysaccharide are, how they are linked together, and how they may have been modified. There can exist a vast multitude of different carbohydrate structures. Hexoses, such as glucose, have 5 chiral centres in their pyranose form. Thus, they can be configured in up to 32 unique isomeric ways. Restricting possible disaccharides to those linked through at least one anomeric oxygen, there can be 9 unique linkages between any two different hexoses. Thus, over 4,000 unique disaccharides can be formed using only unmodified hexoses. The inclusion of furanose forms, pentoses, modifications such as oxidation, amination, acetylation and sulfation, and additional branching all amplify the potential complexity of carbohydrate materials found in nature. Fortunately, structural studies of carbohydrate complexity.

1.4.1 Chemical Transformations for Carbohydrate Structure Determination

The first characteristic usually determined when working with an unknown carbohydrate structure is its monosaccharide composition. Acid-catalysed hydrolysis of a carbohydrate into a mixture of monosaccharides greatly reduces sample complexity. Generally speaking, only one enantiomer of any given monosaccharide is typically found in nature and samples of most naturally occurring monosaccharides are available commercially. Monosaccharides can often be identified through a comparison of chromatographic behaviour to a library of monosaccharide standards (*vide infra*).

The second characteristic often determined when working with an unknown carbohydrate structure is the type and abundance of substitutions such as glycosidic linkages. This can be accomplished through a technique called methylation analysis (Pettolino *et al.*, 2012). Carbohydrate materials are treated with methyl iodide to methylate all free hydroxyl groups and then hydrolysed in acid. The stability of methyl ethers to acid gives a monosaccharide mixture with hydroxyl groups wherever glycosidic linkages (or other modifications) prevented methylation. Subsequent reduction and acetylation yields partially methylated alditol acetates (PMAAs). Gas chromatography is then used to identify and quantify the positions and degrees of methylation and acetylation on each type of monosaccharide in comparison to a mixture of

PMAAs prepared from a material with a known structure or from mixtures of PMAAs prepared by the partial methylation of pure monosaccharide standards.

1.4.2 Chromatographic Analysis of Carbohydrate Mixtures

Many enzyme-induced changes, and the subtleties of enzyme specificities, cannot be identified through monosaccharide and methylation analysis. Chromatographic techniques have been exploited to gain a much more detailed view of carbohydrate mixtures. This enables the identification of smaller and more specific changes. Thin-layer chromatography (TLC) has been a common carbohydrate analysis technique since the 1960s (Stahl and Kaltenbach, 1961). Monosaccharides or oligosaccharides are separated based on affinity for a stationary phase, such as silica or cellulose, by the capillary action-driven flow of an eluent, commonly an alcohol-water mixture, through a thin layer of the stationary phase. Following separation, the plate is dried and carbohydrates are visualised using a developing reagent such as anisidine phthalate or ceric ammonium molybdate in 10% sulfuric acid. The development of TLC into a mature analytical field with a wide variety of high-quality commercial plates, reagents, protocols, and equipment has solidified its place in carbohydrate analysis. However, TLC has largely been supplanted by faster and more powerful pressure-driven high-performance liquid chromatography (HPLC).

HPLC separations occur when a sample is applied to the end of a cylindrical column of chromatographic medium and an eluent is forced through the medium by a pump. While the separation mechanisms in HPLC are broadly the same as in TLC, HPLC offers several advantages including faster separations, higher resolution, a greater variety of chromatographic media, and a variety of on-line detectors. Carbohydrate HPLC separations are routinely performed using size-exclusion, normal phase, reverse-phase, or ion-exchange chromatographic media.

Arguably, the most powerful separation technique currently available for carbohydrate analysis is high-performance anion exchange chromatography coupled to pulsed amperometric detection (HPAEC-PAD). The on-line pulsed amperometric detection of carbohydrates was first described by Dennis Johnson and Scott Hughes (Hughes and Johnson, 1983). Alkaline-stable high-performance ion-exchange media were described by Dionex Corporation in the same year giving rise to HPAEC-PAD (Rocklin and Pohl, 1983). Carbohydrates are separated on the basis of their affinity for a modified synthetic cationic latex medium in a highly alkaline eluent. The alkalinity of the eluent is generally sufficient to deprotonate any carbohydrate (Feng *et al.*, 2013) so that it interacts with the cationic functionality of the medium, while the hydroxide concentration is generally sufficient to elute monosaccharides from the column. Acidic sugars and larger oligosaccharides have much higher affinities for the medium, so they must be eluted with "pusher" anions such as acetate or nitrate (Corradini *et al.*, 2012).

The outflow from the column passes by a small (1-2 mm diameter) circular gold electrode which is pulsed with an oxidizing potential (Hughes and Johnson, 1983). Current flow is measured and integrated over time to determine the total charge which was evolved during the pulse. At low carbohydrate loadings (often <10 nanomoles), this charge is directly proportional to the concentration of carbohydrate in the outflow from the column. Sensitivity can vary with both the concentration of sodium hydroxide and with the nature of the carbohydrate. Thus, quantitative work requires the use of a consistent separation gradient and the determination of a calibration curve using a purified standard of the analyte of interest. A drawback of HPAEC is that the use of sodium hydroxide in the eluent requires additional equipment to enable the use of non-electrochemical detectors, such as mass spectrometers, which can provide additional data about the molecule giving rise to an observed chromatographic peak.

1.4.3 Identification of Carbohydrates using Mass Spectrometry

Mass spectrometry is the separation and detection of ions separated in the gas phase based on their mass-to-charge ratio (m/z) (Hoffmann and Stroobant, 2007). The full breadth of techniques for generating and separating carbohydrate ions is beyond the scope of this thesis, and has been recently reviewed in (Kailemia *et al.*, 2014). Most mass spectrometry methods are very sensitive, requiring only nanograms of sample or less. Common quadrupole mass spectrometers are generally able to separate ions with m/z differences of at least one, allowing the quantification of molecular isotope ratios. High resolution mass spectrometry techniques, such as the electrospray ionisation (ESI) coupled to quadrupole-time-of-flight (Q-TOF) MS used throughout this thesis, often enable the determination of the atomic compositions of molecular ions due to the unique mass defect associated with each element. Within Q-TOF instruments, molecular ions can also be fragmented into smaller ions to aid in identifying unknown molecules. This is referred to as tandem mass spectrometry (MS/MS) because an ion must be isolated before fragmentation and a second round of separation. One of the most valuable aspects of mass spectrometry that it provides a rapid separation that is complimentary to most forms of chromatography.

Liquid chromatography coupled to mass spectrometry (LC-MS) is a very powerful analytical tool for characterizing glycoside hydrolases. Carbohydrates separated using an HPLC column are passed into a mass spectrometer which provides a wealth of information about each isolated molecule. The mass of a carbohydrate provide information about its composition, including the monosaccharide composition (e.g. pentoses, hexoses, deoxy sugars), modifications (e.g. acetylation, amination, methylation), and the total number of residues. Additional information can be obtained from tandem mass spectrometry (MS/MS). This technique is used to confirm the identity of an oligosaccharide in Chapter 2 of this thesis and it is used to identify the oligosaccharides which are selectively deacetylated by esterases presented in Chapter 5 of this thesis.

Performing informative MS/MS experiments on carbohydrate analytes can be quite challenging. Variation in the mode of fragmentation, adduct imparting charge on the ion, or type of carbohydrate can all have significant impacts on the information obtained (Mutenda and Matthiesen, 2007). A common mode of fragmentation for LC-MS/MS is collision-induced dissociation (CID). Ions are fragmented by electrically-driven acceleration in a chamber containing a small amount of chemically inert gas (e.g. argon, helium, nitrogen). Collisions with the gas convert kinetic energy into internal energy resulting in the cleavage of a chemical bond (Hoffmann and Stroobant, 2007). This cleavage event leaves a charge on one of the fragments allowing its mass to be determined. The identification of fragments can provide diagnostic information about the structure of the molecule.

Different adducts fragment in different ways (Mutenda and Matthiesen, 2007). Adducts with acidic cations (e.g. ammonium, proton) tend to fragment via an elimination mechanism to give a loss of 18 mass units (water) providing little useful information. Adducts with hard ions (e.g. Na⁺, K⁺) are more likely to undergo "cross-ring" fragmentation and fragmentation at random glycosidic bonds. This enables the "sequencing" of linear glycosides containing monosaccharide residues of different masses. Negative ions can be readily formed, particularly by the formation of chloride adducts (Vinueza *et al.*, 2013), and have been consistently found to fragment at glycosidic bonds at the reducing end of a carbohydrate. This enables the "sequencing" of branched oligosaccharides in MSⁿ experiments.

Mass spectrometry also offers the unique ability to perform stable isotope labelling experiments. Common isotopes used for labelling experiments with organic molecules include deuterium (²H), carbon-13 (¹³C), nitrogen-15 (¹⁵N), and oxygen-18 (¹⁸O). Isotope labelling experiments are necessary when there is a need to differentiate one part of a molecule, or one source of a molecule, from another without introducing any chemical change. Isotopes are incorporated through some form of synthetic control. One example is the use of $H_2^{18}O$ to label new reducing ends formed through enzymatic hydrolysis (Schagerlöf et al., 2009). This provides information about the orientation with which a substrate molecule binds to an enzyme active site when protein structural information is limited. Another example is the use of sodium borodeuteride to quantify aldehyde formation by a carbohydrate oxidase (Yalpani and Hall, 1982). Reduction of enzymatically-formed aldehydes back to alcohols with borodeuteride adds 1 Da of mass to the molecule, allowing quantification of the aldehyde based on an isotope ratio. Isotope labelling can act as a valuable complement to chemical techniques. Consecutive methylation, desulfation, and deuteromethylation has been used to distinguish between free hydroxyl groups, hydroxyl groups modified with glycosidic linkages, and sulfated hydroxyl groups in marine polysaccharides (Lei et al., 2009). In spite of the power of LC-MS, it is not always possible to unambiguously assign carbohydrate structures using mass spectrometry.

1.4.4 Nuclear Magnetic Resonance Spectroscopy of Carbohydrates

Nuclear magnetic resonance (NMR) is a technique used routinely by chemists to determine the structures of molecules. NMR spectrometers measure the response of an atomic nucleus to radio waves in the presence of a powerful homogeneous magnetic field (Levitt, 2008). The response of each atom within a molecule to a pattern of radio excitation is defined by its relationship to the atoms near it (in terms of chemical bonding or, less commonly, physical space). NMR typically requires milligrams of a purified sample to obtain high quality signal within a reasonable experimental timeframe. Hydrogen and carbon-13 are the major nuclei probed in carbohydrate NMR experiments.

The use of NMR in carbohydrate structure determination typically takes advantage of two-dimensional NMR (2-D NMR) techniques (Kapaev and Toukach, 2016). 2-D NMR techniques allow the determination of the electronic environment around atoms and the through-bond or through-space distances between atoms. Due to their relative strength, proton signals are typically used to determine carbohydrate structures. There are a handful of common experiments

for determining carbohydrate structure. 1D-¹H spectra provide coupling constants which enable the determination of relative stereochemistry (Karplus, 1963). 2-D correlation spectroscopy (COSY) detects signals which reveal protons on neighbouring carbon atoms (Levitt, 2008). Total correlation spectroscopy (TOCSY) provides information about how atoms are grouped into monosaccharide residues. Nuclear Overhauser effect spectroscopy (NOESY) detects signals with an intensity related to through-space proximity. This provides information about linkage position and monosaccharide sequence. Similar information can be obtained using heteronuclear multibond coupling (HMBC) spectroscopy, requiring relay through a ¹³C nucleus (Keeler, 2010). The use of a variety of complementary NMR techniques enables the determination of monosaccharide type, sequence, linkage position, and the position and type of any modifications (Kapaev and Toukach, 2016).

1.5 Relating Enzyme Function to Structure using X-ray Crystallography

Enzymes, under the right conditions, can form regularly ordered protein crystals which diffract X-rays. From these diffraction patterns, the arrangement of atoms in an repeating unit can be deduced so long as the phase of the diffracted photons can be experimentally determined based on the incorporation of certain heavy atoms (for anomalous dispersion techniques) or reasonably correctly guessed (as in molecular replacement) (Rupp, 2009). *a priori* knowledge of the expected atomic composition of the protein is generally essential to both designing diffraction experiments and determining the correct protein structure (Kleywegt, 2000). Furthermore, the information that can be derived from structural data depends heavily on the experimental setups used. The most common motivations for crystallizing an enzyme are to understand interactions with small molecules such as substrates and inhibitors, to understand enzyme domain organisation and homology, and to understand mechanistic elements of enzyme function.

While *apo*-enzyme structures are generally sufficient to determine domain organisation and perform structural comparisons, they do not provide information about interactions with substrates or inhibitors. Co-crystallisation, in which a ligand is added to the crystallisation solution, and crystal soaking, in which a ligand is introduced to a pre-formed *apo*-enzyme crystal, are the most common methods for obtaining complexes between enzymes and substrates, products, or inhibitors. These experiments are often performed on enzyme mutants, such as those specifically mutated to prevent substrate hydrolysis. It is not always possible to find a natural ligand which can be incorporated into enzyme crystals, so inhibitors are sometimes used to obtain information about substrate recognition and catalytic mechanisms (Notenboom *et al.*, 1998, Brayer *et al.*, 2000, Varrot *et al.*, 2003). This strategy is demonstrated in Chapter 3 of this thesis. The first crystal complex of a GH5 enzyme with all of its active site residues engaged in ligand binding was generated using a xyloglucan-derived alkyl bromide active site label.

1.5.1 Glycoside Hydrolase Active-Site Structure

The most interesting feature of any enzyme is its active site. The active sites of glycoside hydrolases generally form one of three overall architectures: a pocket, a cleft, or a tunnel (Davies and Henrissat, 1995) (Figure 1-6). Pocket active sites are associated with the breakdown of small oligosaccharides or with *exo*-hydrolase activity. Cleft active sites allow the recognition of a target segment in a longer polymeric material and are generally found in the structures of *endo*-hydrolases. Tunnel active sites are significantly less common; they are generally associated with the processive hydrolysis of insoluble substrates such as cellulose. Each of these active site architectures is built around a core set of catalytic machinery described in section 1.1.2.1. The catalytic machinery is heavily dependent on the recognition of specific substrates by a complement of additional amino acid residues positioned throughout the active-site.

Neutral carbohydrate substrates are generally recognised using a combination of wellplaced hydrogen-bonding and hydrophobic functionalities. The active site cleft of human amylase is an excellent example of the interactions typically observed between glycoside hydrolases and their substrates (Brayer *et al.*, 2000) (Figure 1-7). In a crystal structure determined in complex with acarbose, a maltopentaose-like amylase inhibitor, several glucosyl hydroxyl groups were found apparently hydrogen-bonded to residues which were positioned to form hydrogen bonds with ideal bond-lengths of 1.5-3.0 Å and ideal bond angles of 140-180° (Jeffrey and Saenger, 1994b). The enzyme also takes advantage of the hydrophobic faces of several sugar rings by placing hydrophobic or aromatic residues in face-to-face contact. This promotes binding through the displacement of structured water and maximisation of Van der Waals interactions (Breiten *et al.*, 2013). To better understand the importance of these different interactions in the active site, a quantitative view of enzyme-substrate interactions is needed.



Figure 1-6: Glycoside hydrolase active site architectures: A) pocket (PDB ID: 3GLY) B) cleft (PDB ID: 1TML) and C) tunnel (PDB ID: 3CBH). Catalytic residues are highlighted in red in each case. Taken from (Davies and Henrissat, 1995)



Figure 1-7: Wall-eyed stereo view of the interactions between human pancreatic amylase and acarbose (PDB ID: 1PPI). Acarbose is shown with salmon-coloured carbon atoms and amino acid residues

interacting with acarbose are shown with slate-coloured carbon atoms. Polar interactions, including hydrogen bonds, are shown as black dashes and apparent hydrophobic interactions are shown as orange dashes.

Alan Fersht was among the first enzymologists who sought to measure the strengths of enzyme-substrate interactions. Fersht and Wilkinson took advantage of one of the first sitedirected mutants generated by Michael Smith and Mark Zoller to show that the sulfhydryl group of C35 of the Bacillus stearothermophilus tyrosyl-tRNA synthetase contributes up to 1 kcal/mole to the stabilisation of the transition state (Wilkinson et al., 1983). Namchuk and Withers demonstrated an inverted approach, synthesizing substrates with or without specific hydroxyl groups, to show that the contributions to catalysis of interactions between the hydroxyl groups of glucose and the active site of Agrobacterium faecalis β -glucosidase were strongest at the 2- and 3-positions (up to 6 and 2 kcal/mole, respectively) and weaker at the 4- and 6-positions (up to 1 kcal/mole). The strengths of the interactions at the 4-, and 6-positions are commensurate with those predicted for a single hydrogen bond in an aqueous medium (Jeffrey and Saenger, 1994a). Hydrophobic interactions are far more challenging to estimate due to the complex nature of Van der Waals attraction and water structuring (Breiten et al., 2013). However, they have been successfully incorporated into modern models of protein folding to enhance enzyme stability (Kim et al., 2012) and into models of protein-ligand interactions to design better inhibitors (Simone *et al.*, 2013).

It is important to note that the binding energy between the enzyme active site and its substrate is generally more heavily invested into reducing transition state energy than into holding the substrate in place. As postulated by Linus Pauling (Pauling, 1948), enzyme active sites have been shown to have the highest affinity not for substrate or product, but for transition state analogues, such as acarbose (Mosi *et al.*, 1998, Schramm, 2011). For GHs, this investment of binding energy into catalysis manifests as a strained interaction with the substrate at the cleavage site. This has been observed in a crystallised *endo*- β -mannanase (Ducros *et al.*, 2002) (Figure 1-8) and in a crystallised *endo*- β -glucanase (Sulzenbacher *et al.*, 1996).



Figure 1-8: Michaelis complex of *Pseudomonas cellulosa* GH26 mannanase with 2,4-dinitrophenyl 2deoxy-2-fluoro-β-mannotrioside (PDB ID: 1GVY) showing the strained ¹S₅ conformation of the mannose residue adjacent to the catalytic nucleophile (E320).

1.5.2 Active Site Probes for Enzyme Active Site Labelling

Active site probes are a promising family of molecules which, in combination with modern protein and DNA sequencing technologies, can greatly simplify and accelerate the identification of enzymes responsible for detected activities. Active site probes are generally designed and synthesised with two parts: a carbohydrate or carbohydrate-mimic which is specifically recognised by the enzyme active site, and a reactive "warhead" which forms a covalent linkage to the enzyme. Additional accessory motifs, such as fluorophores or pull-down tags, can be further appended to the probe to facilitate purification or detection of labelled enzymes. A variety of active site probes have been discovered to date (Rempel and Withers, 2008). Of particular relevance to this thesis, N-bromoacetylglycosylamines and bromoketone C-glycosides have been used in the inhibition of β -glucosidases and xyloglucanases (Black *et al.*, 1993, Howard and Withers, 1998, Fenger and Brumer, 2015). These inhibitors have also been used to identify of active-site amino acid residues (Tull *et al.*, 1996, Brayer *et al.*, 2000) and they offer considerable advantages in enzyme crystallography as well, facilitating the formation of enzyme-substrate complexes *in crystallo* (Brayer *et al.*, 2000, McGregor *et al.*, 2016).

1.6 Aims of this Thesis

The central thesis of my work is that performing detailed structure-function analyses with a diverse library of substrates will enable the delineation of new functionally-related groups of *endo*-glucanases and improve our ability to predict glycoside hydrolase function. Furthermore, I hypothesize that applying more detailed enzyme characterisation methodologies will illuminate applications of enzymes under study. Many of the methodologies described above are applied to the characterisation of enzymes throughout this thesis.

My first aim was to define the key elements of structure and function which define a new group of recently identified glycoside hydrolases called the EG16s (Eklöf *et al.*, 2013). Following the identification of an unusual *XTH*-like gene from *Physcomitrella patens* (Yokoyama *et al.*, 2010), Eklöf *et al.* characterised a homologue from *Populus trichocarpa* and defined a new group of GH16 *endo*-glucanases in plants which are distinct from the classical *XTH* gene family. I aimed to determine the full range of substrates accepted by these new EG16s and to determine the first atomic-resolution structure of an EG16. I aimed to use these data to better define the EG16 family and its relationships with the plant XTHs and bacterial licheninases.

My second aim was to identify the molecular determinants of specificity in a recently crystallised *endo*-glucanase (PbGH5A) originally isolated from *Prevotella bryantii* B₁4 (Matsushita *et al.*, 1990). On the basis of its inclusion in GH5 subfamily 4, the GH5 subfamily containing all known GH5 xyloglucanases, it was hypothesised that PbGH5A is an *endo*-xyloglucanase. I aimed to determine the structures of PbGH5A in complex with a carbohydrate derived from its substrate and to map its active site on the basis of kinetic data measured with a library of structurally related oligosaccharide substrates. With these data I aim to refine our understanding of specificity in GH5 subfamily 4 and extend our ability to predict *endo*-xyloglucanase activity in GH5.

Finally, sequencing performed by the Russell and Wilson groups (Gardner *et al.*, 1997) showed that PbGH5A is fused to a second catalytic domain which was reported to have β -mannanase activity. Within the recently sequenced genome of *P. bryantii*, PbGH5A is found fused to a putative GH26 mannanase. Furthermore, this fusion appears to be embedded in a polysaccharide utilisation locus containing a variety of genes homologous to known glycoside hydrolases which are active on β -mannans. My third aim was to characterize the GH26 domain fused to PbGH5A and determine the role of this gene fusion in the putative *P. bryantii* β -mannan utilisation locus.

36

Chapter 2: Crystallographic Insight into the Evolutionary Origins of Xyloglucan *Endo*-Transglycosylases and *Endo*-Hydrolases

2.1 Introduction

Plant cell walls are complex barriers constructed from a wide variety of carbohydrate and non-carbohydrate polymers (Carpita and McCann, 2000, Albersheim *et al.*, 2010). The wall serves as a rigid structural support, yet has a dynamic morphology that enables plant cells to assume a vast array of structures. The strength and flexibility of plant cell walls comes from the carefully programmed assembly of a composite matrix of cellulose microfibrils, matrix glycans ("hemicelluloses"), pectins, structural proteins, and polyphenolics. This complexity necessitates a large complement of proteins and enzymes for assembling, remodelling, and recycling the diverse cell wall components. Despite significant advances, many of these enzymes remain uncharacterised or unknown (Mewalal *et al.*, 2014).

The matrix glycans, including mannans, xylans, mixed-linkage glucans (MLGs), and xyloglucans (XyGs), are biosynthesised in the Golgi apparatus by a diversity of glycosyltransferases (GTs). These glycans are then shuttled to the apoplast for assembly into the cell wall (Carpita and McCann, 2000, Scheible and Pauly, 2004, Albersheim *et al.*, 2010, Pauly *et al.*, 2013). Wall remodelling may occur subsequently through hydrolysis (leading to glycan degradation) and transglycosylation (leading to glycan rearrangement), catalysed by enzymes in diverse glycoside hydrolase (GH) families (Minic and Jouanin, 2006, Lopez-Casado *et al.*, 2008, Fincher, 2009, Eklöf and Brumer, 2010, Tyler *et al.*, 2010, Buchanan *et al.*, 2012, Sampedro *et al.*, 2012, Franková and Fry, 2013, Kaewthai *et al.*, 2013, Simmons *et al.*, 2015).

In this context, the xyloglucan *endo*-transglycosylases (XETs, EC 2.4.1.207), which catalyse the non-hydrolytic cleavage and re-ligation of XyG chains, are the archetype of apoplastic wall remodelling enzymes (Rose *et al.*, 2002, Eklöf and Brumer, 2010). XET activity has been implicated in the incorporation of nascent XyG into primary walls, the expansion of primary walls, the morphogenesis of secondary cell walls, and the function of reaction wood (Bourquin *et al.*, 2002, Scheller and Ulvskov, 2010, Gerttula *et al.*, 2015). XETs are encoded by the large and diverse *xyloglucan endo-transglycosylase/hydrolase* (*XTH*) gene subfamily of Glycoside Hydrolase Family 16 (GH16), members of which number 20-60 in all vascular plant

genomes sequenced thus far (Eklöf and Brumer, 2010, Goodstein *et al.*, 2012). As implied by their name, a select number of *XTH* genes encode predominantly hydrolytic xyloglucan *endo*-hydrolases (XEH, EC 3.2.1.151) (Baumann *et al.*, 2007, Kaewthai *et al.*, 2013).

Given their widespread distribution across monocots, dicots, and early vascular plants (Eklöf and Brumer, 2010), revealing the functional diversity of XTH gene products is fundamental to understanding plant cell wall development, and thus remains an extremely active research area (for examples, see Becnel et al., 2006, Lee et al., 2010, Ye et al., 2012, Hara et al., 2014, Han *et al.*, 2015 and references therein). In this context, the evolution of catalytic specificity in these enzymes is particularly intriguing. The majority of XTH gene products characterised thus far appear to encode strict transglycosylases (XETs, EC 2.4.1.207) with little to no hydrolytic (XEH, EC 3.2.1.151) activity (reviewed in Eklöf and Brumer, 2010; for recent examples, see Maris et al., 2011, Hara et al., 2014, Han et al., 2015). On the other hand, previous work has indicated that XEH activity is likely to have arisen from ancestral XETs through a specific amino acid loop insertion in a small clade of XTH gene products (Baumann et al., 2007). This fundamental structure-function analysis has enabled the further refinement of XTH phylogeny, which significantly facilitates functional prediction from sequence data (Eklöf and Brumer, 2010, Kaewthai et al., 2013, Hara et al., 2014). Continuing efforts to characterize XTH gene products are uncovering new catalytic function that can be readily mapped onto, and thus further enhance, the predictive power of this molecular phylogeny (Simmons et al., 2015).

A number of fundamental questions remain regarding the evolutionary origin of the *XTH* gene products in the broader context of GH16. GH16 is a large family of β -jelly-roll proteins whose individual activities include the hydrolysis and/or transglycosylation of a range of β -glucans and β -galactans (Lombard *et al.*, 2014). In early phylogenetic analyses of GH16, it was noted that the closest homologs of plant *XTH* gene products are bacterial licheninases ($\beta(1,3)/\beta(1,4)$ - β -D-glucan 4-glucanohydrolases, EC 3.2.1.73) (Barbeyron *et al.*, 1998, Michel *et al.*, 2001) (Figure 2-1A). This is particularly remarkable, in light of the significant structural and catalytic differences between both groups of enzymes; beyond the conserved active site residues and a common overall fold, primary and tertiary structural similarity is generally low due to major loop differences and the presence of the canonical *XTH* C-terminal extension (InterPro Domain IPR010713, "XET_C") (Johansson *et al.*, 2004). Both groups also exhibit widely different substrate specificities: the licheninases hydrolyse linear, undecorated MLG chains,

whereas XETs and XEHs act upon the heavily decorated xyloglucan backbone. These large differences obfuscate the evolutionary pathway in the diversification of extant licheninases and *XTH* gene products.

Aided by the significant recent expansion in the number of sequenced plant genomes (Goodstein *et al.*, 2012), we recently identified a small clade of homologs that represent a key phylogenetic link between the plant *XTH* gene products and the bacterial licheninases in GH16 (Eklöf *et al.*, 2013). These homologs lack several highly conserved features of XETs and XEHs, including disulfide-bridged C-terminal extensions, N-linked glycosylation sites, and cell-wall-targeting N-terminal signal peptides. Also in contrast with *XTH* gene products, this emergent clade is uniquely represented by a single member in those genomes in which they are found; various exemplar dicots (e.g. *Arabidopsis thaliana*) have none. Enzymological characterisation of a heterologously expressed representative from *Populus trichocarpa* (black cottonwood), *Pt*EG16, demonstrated that this enzyme is a broad specificity *endo*- β (1,4)-glucanase, strikingly capable of hydrolysing both of the major distinguishing cell wall glycans of dicots and commelinoid monocots, XyG and MLG, respectively (Eklöf *et al.*, 2013). However, the commonality of this catalytic ability across the clade, and especially its structural basis, remained experimentally unresolved.

We now report the recombinant production, enzymological characterisation, and X-ray crystallography of a new EG16 homolog from a *Vitis vinifera* (wine grape) expressed-sequence tag (EST) library (Peng *et al.*, 2007). Hydrolytic kinetics on a variety of polysaccharide, oligosaccharide, and synthetic substrates, together with MLG and XyG oligosaccharide crystal complexes, reveal the molecular basis for the substrate plasticity of EG16 active-sites. In turn, a refined phylogenetic analysis enabled by these new data provides unprecedented molecular insight into the evolution of GH16 *endo*-glucanases and *XTH* gene products in plants.

2.2 Materials and Methods

All chemicals and resins were obtained from Sigma-Aldrich unless otherwise specified. UV/Vis spectroscopy was run on a Cary 60 (Agilent) equipped with a single-cell temperature controller. Chromogenic substrate degradation was monitored on a Cary 300 Bio (Agilent) with a temperature-controlled 8-cell dual-beam sample holder. SDS-PAGE was run using Bio-Rad mini Protean TGX 4-20% gels. Gels were imaged in a Bio-Rad Gel Doc XR+ imager. Michaelis-Menten parameters (K_m , k_{cat}) were determined by non-linear fit using the Michaelis-Menten model with or without substrate inhibition in OriginPro 9.1. Sequence alignments were displayed using ESPript 3.0 (Robert and Gouet, 2014).

2.2.1 Bioinformatic Analysis

Amino acid sequences of GH16 enzymes with confirmed activities were extracted from the CAZy database, trimmed using ScanProsite (Castro *et al.*, 2006) and aligned using the Expresso method (Armougom *et al.*, 2006). The resulting alignments were manually refined using the alignment explorer in MEGA6 v6.06. A phylogenetic tree was derived from the resulting sequence alignment using the maximum likelihood method in MEGA6. The reliability of the tree was tested by bootstrap analysis using 100 resamplings of the data set. Five cellulases from GH7 were used as an outgroup to root the tree. Where apparent substrate-specific clades could be identified, branches were collapsed and displayed as triangles (see Table A-1 for accession codes).

EG16 protein sequences were identified through BLAST searches of the NCBI GenBank and JGI Phytozome databases. Bioinformatic analyses were performed on all 33 of the EG16 protein sequences found to date (accession numbers and species of origin can be found in Figure 2-1). Protein sequence alignments were performed using the MUSCLE algorithm in MEGA6 v6.06 (Tamura *et al.*, 2013) with the UPGMB clustering method. The resulting alignment was manually refined using the alignment explorer in MEGA6 on the basis of the secondary structure of VvEG16 (this work). A phylogenetic tree was derived from the resulting sequence alignment using the maximum likelihood method in MEGA6. Two licheninases, from *B. subtilis* and *B. licheniformis* as well as two *XTHs*, TmNXG1 and PttXET16-34, were used as outgroups after removal of all signal peptides and the *XTH* C-terminal extensions. The reliability of the tree was tested by bootstrap analysis using 100 resamplings of the data set.

2.2.2 Cloning

10G Hi-Control and BL21(DE3) Hi-Control *E.coli* were obtained from an Expresso T7 cloning kit (Lucigen). pET24a vector was purchased from EMD Millipore. T5 exonuclease, Q5 High-Fidelity DNA Polymerase and Phusion High-Fidelity DNA Polymerase were purchased from New England Biolabs (Ipswich, MA). Taq DNA ligase was purchased from MCLAB (San Francisco, CA). Nucleotides were purchased from Amresco. Oligonucleotide primers were purchased from Integrated DNA Technologies (IDT, San Diego, CA) (Table A-2). PCR

40

reactions were run in a Bio-Rad S1000 Thermal Cycler. Genes of interest were purchased from IDT as GBlocks and amplified by PCR using Q5 DNA polymerase (New England Biolabs). All cloning was done using Gibson assembly (Gibson *et al.*, 2009) and primers for Gibson assembly were designed using the NEBuilder web tool (New England Biolabs). Site-directed mutagenesis was performed using the QuikChange method (Stratagene).

2.2.3 **Protein Expression and Purification**

BL21(DE3) Hi-Control bearing VvEG16 (NCBI RefSeq: XP_002273975.1), VvEG16(Δ V152) or VvEG16(Δ V152/E89A) expression plasmids were grown at 37^oC in Studier media (1x YT supplemented with 25 mM Na₂HPO₄, 25 mM KH₂PO₄, 50 mM NH₄Cl, 2 mM MgSO₄, 5 mM Na₂SO₄, 0.05% glucose, and 0.5% glycerol) to an OD₆₀₀ of 1.8-2.0 prior to induction with 0.2 mM IPTG overnight (16-18 h) at 16^oC. The cells were collected by centrifugation and resuspended in buffer A (300 mM NaCl, 20 mM imidazole, 20 mM NaPi, pH 7.5) supplemented with 1 mM EDTA (to prevent proteolysis). The cells were lysed by a single pass through a French press. The lysate was clarified by centrifugation before protein purification using a HisTrap FF crude (GE Life Sciences) column. After elution with a 10 CV linear gradient from 20-500 mM imidazole, the protein was desalted into SEC buffer (20 mM MOPS, 1 mM EDTA, pH 7.5), supplemented with 1 mM DTT and concentrated to 5 mg/mL using a 10 kDa Centricon (EMD Millipore) maintained at 4^oC. The protein was then cleaved by TEV protease (1 mg TEV protease/50 mg VvEG16, 16h, 4^oC), run over a freshly-charged HisTrap column, concentrated down to 10 mg/mL and purified over an XK 16/100 column (GE Life Sciences) packed with Superdex 75 (GE Life Sciences) run with SEC buffer at 1 mL/min. The pure protein was then concentrated down to 20-50 mg/mL and flash-frozen in LN₂. All purified proteins were analysed for sequence correctness and post-translational modification by intact mass spectrometry using a Waters nanoACQUITY UPLC coupled to a Xevo G2-S QTof as previously described (Sundqvist et al., 2007).

Wild-type PpXG5 (Gloster *et al.*, 2007) was generated by site-directed mutagenesis from a clone of PpXG5 (E323G) in pET21a with a 6x-C-terminal polyhistidine tag (Spadiut *et al.*, 2011). This was desalted into 20 mM NH₄HCO₃ pH 7.5 prior to concentration to ~10 mg/mL using a 30 kDa Centricon. Sucrose was then added (9 mg/mg PpXG5) and the solution was frozen in LN₂ and lyophilised. CjBgl35A and PtEG16 were expressed as described (Eklöf *et al.*, 2013, Larsbrink *et al.*, 2014b). TEV protease was prepared as described (Tropea *et al.*, 2009) using BL21(DE3) Hi-Control cells bearing the pRARE2 plasmid for expression.

2.2.4 Carbohydrate Analysis

HPAEC-PAD was performed on a Dionex ICS-5000 system equipped with an AS-AP auto-sampler with a temperature-controlled sample tray run in a sequential injection configuration using Chromeleon 7 control software. The injection volume was 10 µL unless otherwise specified. A 3x250 mm Dionex CarboPac PA200 column and a 3x50mm guard column were used for all HPAEC separations. Separations were run using gradients A-C previously specified (Larsbrink *et al.*, 2011).

MALDI-TOF analysis of oligosaccharides was performed on a Bruker Autoflex system (Bruker Daltonics) operated in reflectron mode. Oligosaccharide samples (0.1-10 mg/mL) were mixed 1:1 with 10 mg/mL 2,5-dihydroxybenzoic acid in 1:1 H₂O:MeOH directly on a Bruker MTP 384 ground steel MALDI plate and allowed to dry under ambient conditions.

Liquid chromatography-mass spectrometry was performed on a Waters Xevo Q-TOF with a nanoACQUITY UPLC system. Samples were separated on a 0.32x150 mm Hypercarb KAPPA column packed with 3 µm porous graphitised carbon particles run at 8 µL/min. Buffer A was 25 mM formic acid in 95:5 H₂O:MeCN, raised to pH 5 with ammonium hydroxide, buffer B was 25 mM formic acid in 5:95 H₂O:MeCN with equivalent ammonium hydroxide added. Mass spectrometry was run in positive mode, optimised for resolution, with a 2 second scan time. MS/MS was performed by CID using 15 V collision energy. Following the injection of a 1 µL of sample, the gradient was run at 30°C: 0-5 min, 90% A, 10% B; 5-25 min, linear gradient to 30% B; 25-25.5 min, linear gradient back to 10% B; 25.5-30 min, equilibration with 10% B.

2.2.5 Substrates

Tamarind xyloglucan (tXyG), barley β -glucan (bMLG), konjac glucomannan (kGM), carboxymethylcellulose (CMC), wheat arabinoxylan, mixed-linkage glucan oligosaccharides (MLGOs) and cellooligosaccharides (G₃₋₆) were purchased from Megazyme International. Hydroxyethyl cellulose (HEC) was purchased from Fluka. Guar gum (gGM) was purchased from West Point Naturals, a local store in Vancouver, Canada.

tXyGOs were prepared essentially as described in (Eklöf *et al.*, 2012) using 5 U of His₆-PpXG5 per gram of tXyG in 10 mM NH₄OAc, pH 5.5 incubated at 30^oC, instead of *T. reesei* cellulase. XXXG and XXXGXXXG were subsequently produced by incubation with AnBgl35A as described (McGregor *et al.*, 2016). [β -D-Glcp-(1 \rightarrow 3)- β -D-Glcp-(1 \rightarrow 4)- β -D-Glc

4-nitrophenyl β -glucoside (G-PNP) and 4-nitrophenyl β -cellobioside (GG-PNP) were purchased from Sigma-Aldrich. 4-nitrophenyl β -cellotrioside (GGG-PNP), 2,4-dinitrophenyl β cellotrioside (GGG-DNP) and 2,4-dinitrophenyl β -cellobioside (GG-DNP) were kind gifts from the Withers lab (UBC). 2-chloro-4-nitrophenyl β -XXXG (XXXG-CNP) was prepared in-house as previously described (Ibatullin *et al.*, 2008). 2-chloro-4-nitrophenyl β -cellobioside (GG-CNP) and 2-chloro-4-nitrophenyl β -cellotrioside (GGG-CNP) were purchased from Megazyme International.

2.2.6 Enzyme Activity Determination

The bicinchoninic acid-copper (BCA) assay was performed as previously described (Arnal *et al.*, 2017). The pH-activity optimum of VvEG16(Δ V152) was determined by incubating 1 mg/mL tXyG with 5 μ g/mL VvEG16(Δ V152) for 15 minutes at room temperature in 50 mM buffer containing 1 mM EDTA. The buffers used were: sodium citrate (pH 3.75-5.5), sodium phosphate (pH 6.1-8.3) and glycine-HCl (pH 8.6-9.9). The temperature optimum was determined in 20 mM citrate, pH 6.0 with 1 μ g/mL VvEG16(Δ V152) and 1 mg/mL tXyG incubated for 15 minutes. Chromogenic substrate hydrolysis was quantified essentially as described (McGregor *et al.*, 2016) using 20 mM sodium citrate, pH 6, incubated at 30°C. Incubation times and enzyme concentration were minimised for GGG-PNP due to cleavage of GGG-PNP into GG and G-PNP. To monitor polysaccharide hydrolysis by HPAEC-PAD, polysaccharides (tXyG or bMLG, 0.1 mg/mL final) were mixed with an appropriate amount of enzyme in a 1 mL reaction containing 20 mM sodium citrate, pH 6.0, and incubated at 30°C. 100 µL samples were diluted into 400 µL of 0.1 M Na₂CO₃ and run on HPAEC-PAD using gradient C. XXXGXXXG, MLGO and cellooligosaccharide kinetics were determined as previously described (McGregor et al., 2016). The regiospecificity of cellooligosaccharide hydrolysis was determined using the $H_2^{18}O$ labelling method previously described (McGregor *et al.*, 2016).

2.2.7 Co-crystallisation of VvEG16 Variants in Complex with Oligosaccharides

Crystals were grown using the sitting-drop method at 4^oC; 10 mg/mL

VvEG16(Δ V152/E89A) in SEC buffer supplemented with 5 mM DTT was co-crystallised with 2 mM cellotetraose in 0.8 M NaH₂PO₄ + 1.2 M K₂HPO₄ yielding large hexagonal prismic needle clusters. The same enzyme stock was co-crystallised with 5 mM G3GGG3GGG (potentially contaminated by ~0.5 mM of putative GG3GGG3GGG according to HPAEC-PAD) in 0.1 M NaCl, 0.1 M HEPES pH 7.5, 1.6 M (NH₄)₂SO₄ to yield small rectangular prisms with rounded ends. Co-crystallisation with 10 mg/mL tXyGO₂ in 20% PEG 6000 in 100 mM pH 6.0 MES buffer yielded small tetragonal bipyramidal crystals. 10 mg/mL VvEG16(Δ V152/C22S/C188S) in SEC buffer was co-crystallised with 5 mM XXXG in 30% PEG 6000 in 0.1 M Bicine at pH 9.0 and 4°C yielding large hexagonal prismic needle clusters. All crystals were cryoprotected with 25% (v/v) glycerol in well solution and flash-frozen in LN₂.

2.2.8 Data Collection and Refinement

Diffraction experiments were performed at the Canadian Light Source (Saskatoon) on beamline 08ID-1 run with MxDC for the XXXG, cellotetraose and GG3GGG3GGG crystals. Diffraction experiments were performed at the Stanford Synchrotron Radiation Lightsource (Menlo Park) on beamline 7-1 run with Blu-Ice (McPhillips et al., 2002) for the tXyGO₂ complex. Datasets were processed using XDS (Kabsch, 2010). For initial phasing, a search model was generated from PttXET16-34 in complex with a tXyGO (PDBID: 1UMZ) using Chainsaw (Stein, 2008). The initial structure (VvEG16(Δ V152/E89A)-tXyGO₂) was phased by molecular replacement using phaser-2.5.0 (McCoy et al., 2007) yielding poor initial phases. This initial model was improved via auto-building in ARP/wARP 7.4 (Langer et al., 2008) in CCP4 (Winn et al., 2011), which improved the phases, followed by manual adjustment using Coot (Emsley and Cowtan, 2004) and refinement with Refmac5 (Murshudov et al., 1997, Pannu et al., 1998, Vagin et al., 2004). Final refinement was completed with simulated annealing using Phenix.refine (Adams et al., 2010, Chen et al., 2010, Afonine et al., 2012). Subsequent structures were solved using VvEG16(Δ V152/E89A)-tXyGO₂ as the search model and refined using Phenix.refine. The cellotetraose complex was refined with anisotropic B-factors for all non-hydrogen atoms and the GG3GGG3GGG complex was refined with anisotropic B-Factors for all non-hydrogen and nonwater atoms. Models were validated using the *Molprobity* and *Coot* validation tools along with the PDB ADIT server and Privateer (Agirre et al., 2015). No residues were found in the

disallowed region of the Ramachandran plot for the tXyGO₂ or GG3GGG3GGG complexes. However, H41 of the GGGG complex was identified as a Ramachandran outlier by the wwPDB X-Ray validation server (Table A-3) due to a minor alternate conformation observed for the backbone of D40 which could not be likewise modelled for the backbone of H41. All structure figures were prepared using PyMOL (Schrödinger).

2.3 Results

2.3.1 Molecular Phylogeny of the EG16 Clade

Thirty-three EG16 sequences were identified by BLAST queries of both the NCBI GenBank and JGI Phytozome databases using *Pt*EG16 (Eklöf *et al.*, 2013) as a query sequence. Protein sequence alignment revealed high similarity among EG16 homologs, and clear differences with bacterial licheninases and plant XETs and XEHs of GH16 (Figure A-1). GH16 enzymes hydrolyse glycosidic bonds using the canonical double-displacement, anomeric configuration-retaining mechanism, which employs both a catalytic nucleophile and a general acid/base residue (Koshland, 1953, Planas, 2000, Eklöf and Brumer, 2010). The consensus active-site motif, DEID(F/I)EFLG, contains the key catalytic glutamate/glutamic acid residues, as well as a catalytic "helper" aspartate residue that is thought to electrostatically modulate catalysis (underlined). This motif is strictly conserved among bacterial licheninases and all plant GH16 members. A phylogenetic tree constructed from the alignment shown in Figure A-1 and rooted with two licheninase sequences reveals a division between monocot and dicot EG16s with moderate support (bootstrap value 82%), in addition to outlying *Physcomitrella patens* and Selaginella moellendorffii sequences (Figure 2-1B). Inspection of the alignment indicates that the monocot/dicot division likely arises from sequence differences in β -strand 6 and loops 6 and 7, as well as in the Glu/Asp/Ala-rich C-terminal tails (Figure A-1). Notably, these variable Ctermini bear no homology to the distinctive C-terminal extensions of XETs and XEHs (InterPro Domain IPR010713, "XET C") (Johansson et al., 2004, Baumann et al., 2007). Unlike the licheninases and the XTH gene products in GH16, EG16 members are rich in cysteine residues throughout the main β -jelly-roll domain, containing between 3 (e.g., BdEG16, HvEG16) and 12 (e.g., PeEG16) per protein. Positional conservation of these cysteine residues is generally poor, although there are four homologous cysteine residues in dicot EG16s, and one in monocot EG16s (Figure 2-1). In contrast, the four cysteine residues in the C-terminal domain of XTH gene products (InterPro Domain IPR010713, "XET_C"; which is absent in EG16 homologs) are widely conserved and participate in the formation of 2 structural disulfide bonds (Johansson *et al.*, 2004, Baumann *et al.*, 2007).



Figure 2-1: Protein sequence-based phylogeny of EG16 homologs within Glycoside Hydrolase Family 16 (GH16).A) Overall phylogeny of GH16 encompassing all specificities identified to date. Each collapsed branch represents 3-5 sequences. The tree is rooted with 5 GH7 cellulases; GH7 and GH16 together form Clan GH-B. A dashed line separates enzymes containing the EXDXXE "beta-bulge" active-site motif from those with the regular beta-strand EXDXE active-site motif. B) Phylogenetic tree of EG16 homologs identified in GenBank and Phytozome, based on the protein sequence alignment shown in Figure A-1.
Abbreviated protein names are derived from the genus and species of origin, and the common name (where available) is given along with the accession code. The tree is rooted using two Bacillus licheninases and additionally includes a xyloglucan *endo*-transglycosylase (PttXET16A, (Johansson et al., 2004)) and a xyloglucan *endo*-hydrolase (TmNXG1, (Baumann et al., 2007)) as outliers with known tertiary structures.
Box diagrams, to scale, indicate key protein sequence features: Black, signal peptides; red, active site EXDXE motif; light purple, licheninase loop extension; dark purple, XEH loop extension; blue, C-terminal XET/XEH extension (XET_C); yellow, conserved cysteine residues in the monocot or dicot EG16 clades; brackets, crystallographically observed disulfide bonds; fork, conserved XET N-glycosylation site. Bootstrap values from 100 maximum likelihood resamplings are shown next to each branch of both trees.

2.3.2 Recombinant Production of VvEG16 in E.coli

From this phylogeny, four genes encoding EG16 targets, representing bryophyte (Physcomitrella patens EG16), lycophyte (Selaginella moellendorffii EG16), monocot (Brachypodium distachyon EG16), and dicot sequences (Vitis vinifera EG16), were selected for heterologous expression in E.coli in native form and in tandem with solubilizing fusion polypeptides (Table A-1). Of these, only native VvEG16 and an N-terminal SUMO conjugate could be produced in a stable, soluble form in an initial screen. The expression and purification of native VvEG16, a fortuitous mutant VvEG16(Δ V152) (having a single amino acid deletion of Val152), as well as the corresponding site-directed catalytic-nucleophile mutants (VvEG16(E89A) and VvEG16(Δ V152/E89A)) and surface-cysteine deletion mutants (VvEG16(C22S/C188S) and VvEG16(ΔV152/C22S/C188S)) under optimised conditions gave exceptional yields (~100 mg/L of E. coli culture) of electrophoretically pure protein following TEV protease cleavage of the His₆ affinity tag (Figure A-2). All recombinant VvEG16 variants were significantly more stable than PtEG16, which was prone to cysteine oxidation and aggregation upon storage (Eklöf et al., 2013). VvEG16(Δ V152), in particular, was amenable to crystallisation (vide infra). In light of this observation, the majority of the subsequent biochemical analyses were performed on this variant.

2.3.3 VvEG16 is a bi-function MLG/XyG endo-glucanase

Recombinant VvEG16(Δ V152) hydrolysed barley MLG (bMLG) with the highest apparent k_{cat} value and lowest apparent K_M value of any polysaccharide tested, at the pH optimum of 6.0 and 30^oC. (Table 2-1, Figure A-3A, Figure A-4A,B). Under these conditions, tamarind seed XyG (tXyG) was hydrolysed less efficiently, evident from a >5-fold higher apparent K_M value and a 2-fold lower k_{cat} value. The recombinant wild-type enzyme had nearly identical kinetics on both substrates, thus providing further justification for our focus on the crystallisable VvEG16(Δ V152) variant (Table 2-1, Figure A-3). Very poor activity of VvEG16(Δ V152) was observed on konjac glucomannan (kGM), carboxymethylcellulose (CMC), and hydroxyethylcellulose (HEC), all of which contain β (1,4)-glycosyl backbone residues, while no activity was detected with wheat arabinoxylan, crystalline cellulose, laminarin, or guar galactomannan. Under reducing conditions (dithiothreitol-containing buffer),
VvEG16(Δ V152) is thermally unstable above 50^oC (Figure A-4C), which is near the thermal stress limit of *V. vinifera* (Greer and Weston, 2010).



Figure 2-2: VvEG16 limit-digest products. A) HPAEC-PAD chromatogram of XyGOs produced by the action of VvEG16(ΔV152) on tXyG. B) MALDI-TOF analysis of the mixture shown in A). C) HPAEC-PAD chromatogram of oligosaccharides produced by the action of VvEG16(ΔV152) on bMLG. D) MALDI-TOF analysis of the oligosaccharide mix shown in C).

HPLC analysis of the limit-digest of tXyG by VvEG16(Δ V152) revealed the production of the canonical mixture of tXyGOs arising from hydrolysis at the anomeric position (C-1) of the unbranched backbone glucosyl residue (Figure 2-2A,B; oligosaccharide nomenclature is according to (Tuomivaara *et al.*, 2015)). However, prior to the completion of xyloglucan hydrolysis, a notable increase in the proportion of the Xyl₃Glc₄ heptasaccharide XXXG was observed relative to the galactosylated congeners XLXG, XXLG and XLLG (XXXG:XLXG:XLLG:XLLG observed: 2.2:1:3.9:5.6, versus 1.4:1:3:5.4 (York *et al.*, 1990)). This result corroborates the kinetic analyses, and collectively suggests that extended polysaccharide sidechains are disfavored in the active-site.

Digestion of bMLG using a similar amount of VvEG16(Δ V152) resulted in the production of an oligosaccharide series corresponding to 3n+1 glucose residues (where *n* is an integer ≥ 1), along with less abundant 3n+2 and 3n+3 series (Figure A-5). Exhaustive digestion of bMLG using a higher concentration of the enzyme resulted in three products: glucose, cellobiose and the mixed-linkage tetrasaccharide β -D-Glcp-(1,3)- β -D-Glcp-(1,4)- β -D-Glcp-(1,4)-(G3GGG), based on comparison with elution times of standard samples in HPLC (Figure 2-2C,D, Figure A-6A). Due to impurities in commercial G3GGG, the identity of this product was further confirmed by HPLC analysis following partial digestion of the purified oligosaccharide with Agrobacterium sp. β -glucosidase (Wakarchuk et al., 1986) (Figure A-6B) and by independent LC-MS/MS analysis (Figure A-7). Notably, the production of G3GGG indicates that VvEG16 selectively hydrolyses the $\beta(1,4)$ glycosidic bond to a $\beta(1,3)$ -Glc residue. This further implies that the dominant oligosaccharide series observed in partial digests (Figure A-5) is comprised of repeats of $\beta(1,3)$ -linked cellotriose with a $\beta(1,3)$ -Glc residue at the non-reducing end. To the best of our knowledge, this strict linkage specificity is unique not only among GH16 members, but also among all known endo-glucanases with activity toward MLG (Zverlov and Velikodvorskaya, 1990, Malet et al., 1993, Henriksson et al., 1995) (Figure 2-3).



Figure 2-3: Site of MLG hydrolysis by VvEG16 compared to other known MLG-active *endo*glucanases. The primary bond targeted by each enzyme is indicated with a solid black arrow and any known secondary activity is indicated with a dashed black arrow and an approximate relative activity.

Remarkably, we observed that hydrolysis of bMLG and oat MLG at elevated concentrations (>10 g/L) by VvEG16 resulted in the formation of a gel (Figure 2-4). The structural changes in the polysaccharide that gave rise to this gelation were initially investigated by following the hydrolysis of bMLG by VvEG16 using HPAEC-PAD analysis and licheninase digestion (Figure A-8). Consistent with the product analysis above, a complex mixture of barley MLG oligosaccharides (bMLGOs) is formed after 5 minutes of digestion and the solution remains fluid. After heat-inactivation of VvEG16, complete digestion with a bacterial licheninase revealed a depletion of stretches of $\beta(1,4)$ -linked glucosyl residues longer than Glc4 (Figure A-8B) in this sample. Extending the VvEG16 incubation time to 60 minutes resulted in gel formation. This gel could be dissolved by heating to 65°C, and subsequent HPLC analysis revealed that the oligosaccharide mixture had become simpler (Figure A-8C). Licheninase digestion of this material indicated a selective depletion of $\beta(1,4)$ -linked glucan motifs longer than Glc₃ (Figure A-8D).



20 g/L oat MLG gel

Figure 2-4: Gel formation by VvEG16-catalysed hydrolysis of MLG. A) Gel formation with purified barley MLG. B) Gel formation with partially purified oat MLG.

2.3.4 Kinetic Subsite Mapping of the VvEG16 Active Site

The active-sites of GHs can be delineated into positive subsites, which extend toward the reducing-end of oligosaccharide and polysaccharide substrates from the point of cleavage, and negative subsites extending toward the non-reducing-end (Davies *et al.*, 1997). To better understand the substrate and product specificity of VvEG16, enzyme kinetics on a series of pure, native oligosaccharides and synthetic chromogenic oligosaccharides were quantified to map the contributions of individual active-site subsites to catalysis (Table 2-1).

Substrate	Enzyme	$k_{cat, app}$ (min ⁻¹)	<i>K_{m, app}</i> (mM)	$k_{cat, app}/K_{m, app}$ (M ⁻¹ s ⁻¹)	Assay
bMLG	VvEG16(ΔV152)	143±7	0.34±0.03*	7.00*	BCA
	VvEG16	163±5	$0.36 \pm 0.02*$	7.55*	BCA
tXyG	VvEG16(ΔV152)	68±8	$1.8 \pm 0.4*$	0.63*	BCA
	VvEG16	70±6	2.1±0.4	0.56*	BCA
HEC	VvEG16(ΔV152)	10.4 ± 0.8	1.1±0.2*	0.16*	BCA
СМС	VvEG16(ΔV152)	17 ± 2.2	2.8±0.5*	0.10*	BCA
kGM	VvEG16(ΔV152)	ND	ND	$0.040\pm5*$	BCA
Guar Galactomannan	VvEG16(ΔV152)	ND	ND	ND	BCA
Wheat Arabinoxylan	VvEG16(ΔV152)	ND	ND	ND	BCA
Laminarin	VvEG16(ΔV152)	ND	ND	ND	BCA
GG-PNP	VvEG16(ΔV152)	0.023±0.002	5.9±0.6	0.065	PNP
GG-CNP	VvEG16(ΔV152)	0.45 ± 0.02	3.5 ± 0.4	2.1	CNP
GG-DNP	VvEG16(ΔV152)	20±1	2.0 ± 0.2	170	DNP
GGG-PNP	VvEG16(ΔV152)	0.15 ± 0.006	2.8 ± 0.2	0.9	PNP
GGG-CNP	VvEG16(ΔV152)	1.03 ± 0.04	1.1 ± 0.1	16	CNP
	VvEG16	0.98 ± 0.03	0.8 ± 0.1	20	CNP
GGG-DNP	VvEG16(ΔV152)	95±2	0.80 ± 0.06	2000	DNP
XXXG-CNP	VvEG16	12.0 ± 0.5	1.1±0.3	180	CNP
	$VvEG16(\Delta V152)$	13.9±0.6	1.2 ± 0.1	190	CNP
GG → 2xG	VvEG16(ΔV152)	ND	ND	ND	HPAEC-PAD
GGG → GG+G	VvEG16(ΔV152)	1.44 ± 0.02	0.43 ± 0.02	56	HPAEC-PAD
GGGG → 2xGG	VvEG16(ΔV152)	66±1	0.094 ± 0.005	1.2×10^4	HPAEC-PAD
GGGG → GGG+G	VvEG16(ΔV152)	70±10	0.5 ± 0.2	2300	HPAEC-PAD
$2xGGGGG \rightarrow GG+GGGGGGG$	VvEG16(ΔV152)	98±6	1.3±0.2	1250	HPAEC-PAD
GGGGGG→GGG+GG	VvEG16(ΔV152)	62±2	0.26 ± 0.04	3500	HPAEC-PAD
G3GGG	VvEG16(ΔV152)	ND	ND	ND	HPAEC-PAD
GG3GG→2xGG	VvEG16(ΔV152)	ND	ND	0.3 ± 0.05	HPAEC-PAD
GGG3G → GGG+G	VvEG16(ΔV152)	ND	ND	28±5	HPAEC-PAD
GGG3G → GG+G3G	VvEG16(ΔV152)	36±3	1.0 ± 0.1	600±20	HPAEC-PAD
G3GGG3GGG→ G3GGG+GGG	VvEG16(ΔV152)	ND	ND	25±1	HPAEC-PAD
XXXGXXXG→2xXXXG	VvEG16	14±1	0.055 ± 0.007	4200	HPAEC-PAD
XXXGXXXG → 2xXXXG	TmNXG1 (Baumann <i>et al.</i> , 2007)	9.32±0.16	0.076±0.007	2000	HPAEC-PAD

 Table 2-1: Apparent Michaelis-Menten kinetic constants for hydrolysis and transglycosylation reactions

 catalysed by wild-type VvEG16 and VvEG16(ΔV152).

ND = not detected

In the first instance, the contributions of the negative subsites to catalysis were elucidated using substituted phenyl β -glycosides as chromogenic substrates. In the first instance, the contributions of the negative subsites to catalysis were elucidated using substituted phenyl β -glycosides as chromogenic substrates (Figure A-9). 4-Nitrophenyl β -glucoside (G-PNP) was not

hydrolysed by VvEG16(Δ V152), indicating that the presence of single Glc residue capable of binding in subsite -1 is not sufficient for catalysis. In contrast, 4-nitrophenyl and 2-chloro-4nitrophenyl β -cellobiosides were competent substrates, with k_{cat}/K_M values inversely dependent on leaving group p K_a (Table 2-1), as expected (Planas, 2000). The corresponding cellotriosyl congeners were consistently hydrolysed with ca. 10-fold greater k_{cat}/K_M values, thereby indicating that a third negative subsite contributes approximately -6 kJ/mol ($\Delta\Delta G^{\ddagger}$) to catalysis. Analysis of the contribution of a potential -4 subsite using GGGG-CNP was not possible due to internal cleavage of the oligosaccharide to yield primarily GG and GG-CNP. A *ca.* 10-fold greater k_{cat}/K_M value was observed for XXXG-CNP hydrolysis over GGG-CNP that likely reflect specific interactions with internal xylosyl branches observed crystallographically (*vide infra*).

To extend this analysis to the positive subsites, initial-rate kinetics of the hydrolysis of a series of cello-oligosaccharides (Glc₂-Glc₆) were quantified by HPLC (Figure A-10A-D). Hydrolysis of cellobiose could not be detected under any conditions, including at high enzyme and substrate concentrations. This recapitulates the observation with G-PNP; binding in subsite - 1, with or without a contribution of Glc binding in +1, is insufficient for catalysis. Hydrolysis of cellotriose (GGG) to cellobiose and glucose (GG + G) by VvEG16(Δ V152) was poor ($k_{cat}/K_m = 56 \text{ M}^{-1}\text{s}^{-1}$). In contrast, the hydrolysis of cellotetraose (GGGG) via either of two modes, yielding two molecules of cellobiose (2 x GG), or yielding cellotriose plus glucose (GGG + G), was considerably more efficient, with 200- and 40-fold greater k_{cat}/K_M values, respectively. Cellopentaose was hydrolysed through a single mode, producing only cellotriose and cellobiose, with a k_{cat}/K_M value comparable to those of cellotetraose as a substrate.

Additional product analysis by mass spectrometry from assays performed in ¹⁸O-enriched water indicated that for asymmetrical cleavage, cellotriose yielded GG(¹⁸O)+G though a -2 \rightarrow +1 subsite binding mode, cellotetraose yielded GGG(¹⁸O)+G though a -3 \rightarrow +1 binding mode, and cellopentaose yielded GGG(¹⁸O)+GG though primarily a -3 \rightarrow +2 binding mode (Figure A-11, Table A-5). The kinetically favored hydrolysis of cellotetraose via the -2 \rightarrow +2 subsite binding mode over the -3 \rightarrow +1 indicates a slightly stronger contribution of the +2 versus the -3 subsite, which is not fully realised in the hydrolysis of cellopentaose in either the -3 \rightarrow +2 or -2 \rightarrow +3 binding modes.

In light of the clear specificity of VvEG16 for bMLG, we determined the hydrolysis kinetics of VvEG16(Δ V152) acting on various model mixed-linkage glucan oligosaccharides, to

understand the effects of $\beta(1,3)$ linkages on catalysis (Figure A-10E-G, Table 2-1). Notably, VvEG16(Δ V152) did not hydrolyse G3GGG to any detectable degree, indicating that the enzyme discriminates against $Glc\beta(1,3)$ residues in the -2 and -3 subsites (cf. GGGG, Table 2-1). This observation also explains the accumulation of G3GGG as a major product in the limit-digest of the bMLG polysaccharide (vide supra). GG3GG was hydrolysed only to two molecules of GG, with a low specificity constant ($k_{cat}/K_m = 0.3 \text{ M}^{-1}\text{s}^{-1}$) that is 40000- and 7600-fold lower than the - $2 \rightarrow +2$ and $-3 \rightarrow +1$ hydrolysis modes of GGGG, respectively. This indicates a strong preference for the hydrolysis of $\beta(1,4)$ - versus $\beta(1,3)$ -glucan linkages by VvEG16. Correspondingly, GGG3G was preferentially hydrolysed to GG and G3G via a $-2 \rightarrow +2$ binding mode. Interestingly, the hydrolysis of the $\beta(1,3)$ linkage in GGG3G to give GGG and G via a $-3 \rightarrow +1$ binding mode can occur, although this is 20-fold less favored than $-2 \rightarrow +2$ cleavage. The $-3 \rightarrow +1$ hydrolysis mode of GGG3G had a ca. 200-fold lower specificity constant than the analogous reaction with the all- $\beta(1,4)$ -linked GGGG, again demonstrating the limited capacity of VvEG16(Δ V152) to cleave β (1,3)-linkages. Thus, the heptasaccharide G3GGG3GGG was specifically hydrolysed at the $\beta(1,3)$ bond with a nearly identical specificity constant ($k_{cat}/K_m =$ 25 M⁻¹s⁻¹) to the hydrolysis of GGG3G into cellotriose and glucose ($k_{cat}/K_m = 28 \text{ M}^{-1}\text{s}^{-1}$, Table 2-1). The lack of hydrolysis of this substrate to G3GG+G3GGG, G3GGG3G+GG, or G3GGG3GG+G via the favored $\beta(1,4)$ bond hydrolysis again reflects a strong bias against $\beta(1,3)$ -linkages between the negative subsites. Regarding positive subsite interactions, the hydrolysis of GGG3G into GG and G3G had a ~20-fold lower specificity constant than the analogous hydrolysis of GGGG into two GG molecules (Table 2-1), which suggests that $\beta(1,3)$ linkages are slightly disfavoured between subsites +1 and +2. Analogous to the cellooligosaccharide data, the composite analysis of MLGO hydrolysis demonstrates clear evidence for +1 and +2 subsite binding in VvEG16, but no indication of a kinetically relevant +3 subsite.

The hydrolysis of the XyG tetradecasaccharide XXXGXXXG at the internal, unbranched Glc residue exhibited a 20-fold higher specificity constant than XXXG-CNP, thus indicating a significant contribution of positive subsite binding to catalysis (Table 2-1). The k_{cat}/K_m value for XXXGXXXG hydrolysis was also similar to those of cellotetraose and cellopentaose, which further highlights the accommodation of xylosyl branches in the VvEG16 active-site. Notably, the kinetic constants for XXXGXXXG hydrolysis by VvEG16(Δ V152) were similar to those of the archetypal XEH from nasturtium (*Tropaeolum majus*), TmNXG1 (Baumann *et al.*, 2007).

In light of the defining ability of XETs to perform substrate transglycosylation, we also examined the capacity of VvEG16 to catalyse this reaction using well-defined oligosaccharide substrates. Indeed, HPLC analysis under the initial-rate conditions used to measure GGGG hydrolysis indicated that transglycosylation also occurs at a significant rate to produce cellohexaose (GGGGGG) and GG (Figure A-12A, Table 2-1). However, similar analysis with XXXGXXXG at elevated substrate concentrations revealed no detectable formation of (XXXG)₃. The potential of VvEG16 to catalyse hetero-transglycosylation was also tested (Hrmova *et al.*, 2007, Fry *et al.*, 2008a). Using GGGG as a glycosyl donor substrate in the presence of an excess of XXXG as a potential acceptor, no transglycosylation products were observed by HPLC. On the other hand, with XXXGXXXG as a donor substrate in the presence of an excess of cellobiose as an acceptor, a small peak at a retention time between that of XXXG and XXXGGXXXG slowly increased over time (Figure A-12B). This peak may be due to XXXGGG, however, the small amount formed precluded structural characterisation. Taken together, the data suggest that significant transglycosylation by VvEG16 is only observed with linear β -glucans at elevated concentrations (>1 mM acceptor).

2.3.5 Three-Dimensional Structure of VvEG16 Variants in Complex with Matrix Glycan Oligosaccharides

To illuminate the structural basis for the unique catalytic specificity of VvEG16, we performed X-ray crystallography of variants of the enzyme in complex with representative oligosaccharide substrates.

2.3.5.1 EG16 Tertiary Structure

Despite extensive efforts, wild-type VvEG16 and the catalytically inactive variant VvEG16(E89A) resisted crystallisation independently and in the presence of oligosaccharide substrates and products. Likewise, we were unable to crystallize VvEG16(Δ V152/E89A) in the *apo* form. In contrast, high-quality crystals of VvEG16(Δ V152/E89A) were obtained in the presence of the linear MLG octasaccharide GG3GGG3GGG (β -D-Glc*p*-(1,4)- β -D-Glc*p*-(1,3)- β -D-Glc*p*-(1,4)- β -D-Glc*p*-(1,4)- β -D-Glc*p*-(1,4)- β -D-Glc*p*) and the branched xyloglucan tetradecasaccharide XXXGXXXG (1.65 Å and 1.79 Å resolution, respectively; Table A-3). Strikingly, both of these extended oligosaccharides acted to template crystallisation by spanning two protein molecules in the asymmetric unit that do not have direct

protein-protein contacts with one another. (Figure 2-5A and Figure 2-6A, *vide infra*). Two additional oligosaccharide complexes were also obtained: VvEG16(Δ V152/E89A) in complex with cellotetraose (GGGG; Figure A-13) and VvEG16(Δ V152/C22S/C188S) in complex with the heptasaccharide XXXG (Figure A-14; 0.97 Å and 1.59 Å resolution, respectively; Table A-3). The VvEG16(Δ V152/E89A):GGGG structure appears to be the highest resolution structure determined for a GH16 member to-date (Vasur *et al.*, 2009, Hehemann *et al.*, 2012, Labourel *et al.*, 2014).

Overall, VvEG16 has a β -jelly roll fold that is typical of GH16 members (Lombard *et al.*, 2014) and comprises 16 β-strands and 17 loops (Figure 2-7A, cf. Figure A-1). Superposition of all four complexes reveals that there are no major differences in protein conformation: Chain A of the bMLGO and tXyGO complexes each superpose onto the GGGG complex with all-atom RMSD values for the protein residues of 0.63 and 0.87 Å, respectively and superpose onto the XXXG complex with *all-atom* RMSD values of 0.70 and 0.87 Å, respectively. In the VvEG16(Δ V152/E89A) complexes, the general acid/base catalytic residue, E93, the mutated catalytic nucleophile, E89A, and the catalytic "helper" residue, D91, are found co-linearly on strand β 8, with their sidechains directed into the active-site cleft. This strand delineates the positive enzyme subsites, which extend toward the reducing-end of bound saccharide substrates, from the negative subsites, which extend in the opposite direction (Figure 2-7A). Extending from the rigid core of β-strands, high relative B-factors, indicative of conformational flexibility, were observed in all complexes for loops 5 and 12, residues 154-159 of loop 14, and residues 177-181 of loop 15. Notably, the $\Delta V152$ mutation is located at the beginning of Loop 14 on the convex side of the β -jelly-roll and thus is distant from the active-site cleft. This structural observation explains the insignificant effect of this mutation on catalysis (Table 2-1). Furthermore, our ability to crystallize the VvEG16(Δ V152) variants, but not the VvEG16(E89A) and wild-type variants, may be due to decreased disorder and more favorable packing of the shortened loop. Of the six cysteine residues found in VvEG16, C22 and C188 are surface-exposed, and C26, C64, C124 and C194 are buried. C64 and C124, which are conserved among all dicot EG16s (Figure A-1), appear to be well-positioned to form a disulfide bond, yet only partial occupancy was observed in the high-resolution (0.97 Å) cellotetraose complex, possibly due to the reducing conditions used during crystallisation.

2.3.5.2 Key Features of MLG Recognition by VvEG16

The 1.7 Å resolution complex of two VvEG16(Δ V152/E89A) molecules bridged by their mutual recognition of a single GG3GGG3GGG molecule provides a comprehensive view of the active-site interactions with the preferred substrate of the wild-type enzyme (Figure 2-5). The asymmetric unit contains two polypeptide chains, two oligosaccharides and a monosaccharide. Chain A was modelled from G1 to V207 (except missing the complete H9 residue) and Chain B was modelled from E10 to V207. Electron density consistent with the presence of the α -anomer of G3GGG in the negative subsites and the non-reducing-terminus of GG3GGG3GGG in the positive subsites of Chain A was also observed (Figure 2-5B,C). Likewise, electron density was observed for the α -anomer of the reducing terminus of GG3GGG3GGG in the negative subsites of Chain B. A single β -glucose was also modelled into the +1 subsite of Chain B. Strikingly, the two polypeptide chains are completely separated by solvent water molecules and the bridging oligosaccharide. To our knowledge, such crystal packing is unique among GHs, and is reminiscent of the mutual recognition of a single DNA molecule by multiple DNA-binding proteins (Shi *et al.*, 1998, Fujii, 1999, Murphy Iv, 1999).

Strongly supported by kinetic data for cello-oligosaccharide and MLGO hydrolysis (Table 2-1), the VvEG16(Δ V152/E89A):MLGO complex allows the clear definition of 5 glucosyl-binding subsites (-3 \rightarrow +2) and highlights a potential, weakly interacting -4 subsite. At the catalytic centre of Chain A, the reducing-end glucose in subsite -1 interacts with strands β 7 and β 8 and loop 8, forming hydrogen bonds with Q87, E82, and W171, as well as an aromatic stacking interaction with Y77 (Figure 2-5B,C). Toward the negative subsites, the glucosyl residue in subsite -2 interacts with strands β 2, β 5, β 7 and β 14, forming hydrogen bonds with R46, S169, Y77 and Y21, as well as an aromatic stacking interaction with Y171. The glucosyl residue in subsite -3 forms a hydrogen bond with R46 of strand β 5 and an aromatic stacking interaction with Y21 of strand β 2. The β (1,3)-linked glucosyl residue in the -4 position of chain A forms an apparent hydrogen bond with G43 of loop 5 which is not observed in any other complex.



Figure 2-5: Structure of VvEG16(ΔV152,E89A) in complex with β(1,3)/β(1,4) mixed-linkage gluco-oligosaccharides: A) The asymmetric unit of VvEG16(ΔV152,E89A) contains two protein molecules. The surface representation is shown in white and the secondary structure cartoon is coloured according to B-factors. The two protein molecules are bridged by a single molecule modelled as GG3GGG3GGG, in purple. A second oligosaccharide, modelled as G3GGG, in green, is observed in the negative subsites of one protein

molecule. The active site residues EXDXE are shown in black and the 2mFo-DFc map, contoured at 1.5 Å, is shown for both ligands and active site residues. B) Active-site detail shown in wall-eyed stereo, with the ligands coloured as in panel A. Sidechains which interact with the ligands are shown in white with all oxygen atoms red and all nitrogen atoms blue. The 2mFo-DFc map is contoured at a 1.5σ level. Individual glucose residues are labelled according to the subsite they occupy. Amino acid numbering begins from the N-terminal glycine remaining after TEV-protease digestion. Putative hydrogen bonding interactions are indicated with black dotted lines and hydrophobic interactions are indicated with orange dotted lines C) Schematic representation of the interactions between VvEG16(ΔV152,E89A) chain A and G3GGG in the negative subsites and GG in the positive subsites. Putative hydrogen bonding interactions are shown as dotted lines and hydrophobic interactions are shown as curved surfaces. Amino acid residues are numbered as in panel B.

The crystallographic observation of binding a linear arrangement of $\beta(1,4)$ glucosyl residues in subsites -3 to -1, and tolerance of a $\beta(1,3)$ kink in subsite -4, succinctly rationalizes the kinetic data on MLGO hydrolysis, in which rigorous exclusion of Glc $\beta(1,3)$ residues from subsites -3 and -2 is observed (Table 2-1). Notably, no hydrogen bonding interaction is observed with the unsubstituted 6-OH of the glucosyl residue in the -2 subsite. This, along with the significant twist of the glucan chain (>90^o between -1 and -3) is more typical of *endo*-xyloglucanases than *endo*-glucanases that act on unbranched substrates, such as cellulases or licheninases (Gloster *et al.*, 2007, Mark *et al.*, 2009, McGregor *et al.*, 2016).

In the positive subsites of Chain A of the VvEG16(Δ V152/E89A):MLGO complex, the glucosyl residue in subsite +1 interacts with strands β 8 and β 9 and loops 10 and 15, forming hydrogen bonds with E93, E115, Q103, N105 and Y107, and an aromatic stacking interaction with W181. Despite the kinetic importance of +2 subsite binding for $\beta(1,3)$ - and $\beta(1,4)$ -linked oligosaccharides (Table 2-1), the only apparent interaction in this position is a hydrophobic interaction with I176 of loop 15. At the same time, significant binding plasticity in the positive subsites allows the accommodation of both linkages; positioning of Glc $\beta(1,3)$ in subsite +1 is required to realize the essentially exclusive mode of cleavage of MLG (Figure 2-3). All protein-carbohydrate interactions through the negative and positive subsites discussed above were essentially recapitulated in Chain B.

2.3.5.3 Key Features of XyG Recognition by VvEG16

To understand the consequences of polysaccharide branching on substrate recognition in light of the significant activity of VvEG16 on tXyG and XXXGXXXG (Table 2-1), we

crystallised VvEG16(Δ V152/E89A) in the presence of a mixture of variably galactosylated, Glc₈-based tXyGOs. Similar to the bMLGO complex, the resulting 1.8 Å resolution structure has an asymmetric unit containing 2 protein molecules bridged only by their mutual recognition of an extended oligosaccharide spanning the positive subsites of Chain A and the negative subsites of Chain B. This bridging ligand was modelled as the tetradecasaccharide XXXGXXXG, since no significant electron density from terminal galactosyl branches was observed (Figure 2-6). Chain A additionally contained an apparent XXXG moiety in the negative subsites, while the positive subsites of Chain B were not occupied by a carbohydrate.

The backbone glucosyl residues of the tXyGO ligands superpose perfectly with those of the bMLGO complex across the -3 to +1 subsites (Figure A-15). Analogous to the VvEG16(Δ V152/E89A):MLGO complex, the glucosyl residue of XXXG in Chain A at potential subsite -4, and especially its pendant xylosyl residue, have no apparent interaction with the protein (Figure 2-6B,C). The -3' xylosyl residue participates in a hydrogen bonding interaction with R46, which displaces the amino acid sidechain from the interaction observed with 6-OH of the -3 Glc in the MLGO complex. Interestingly, this interaction also forces the -3' xylose into an unusual $^{0}S_{2}$ skew-boat conformation in Chain A, while a $^{2}S_{0}$ skew-boat conformation, Table A-6, as verified by Privateer (Agirre *et al.*, 2015)). This suggests a significantly strained interaction and may partially account for the lesser activity of VvEG16 on tXyG versus bMLG (Table 2-1). The 4-OH of the -2' Xyl participates in a strong hydrogen bonding interaction with E82, which remains properly positioned to form another hydrogen bond with the 3-OH of the -1 glucose.

In the positive subsites of Chain A, Glc in subsite +1 is tightly anchored, yet there is only a single potential (~3.4 Å) hydrogen bonding interaction between Q87 and the 4-OH of the +1' Xyl. In light of the comparatively limited number of positive subsites and corresponding proteincarbohydrate interactions, this region does not appear to be a major discriminating factor for polysaccharide specificity.



Figure 2-6: Structure of VvEG16(ΔV152,E89A) in complex with xylogluco-oligosaccharides: A) The asymmetric unit of VvEG16(ΔV152,E89A) contains two protein molecules. The surface representation is shown in white and the secondary structure cartoon is coloured according to B-factors. The two protein molecules are bridged by a single molecule modelled as XXXGXXXG, in purple. A second oligosaccharide, modelled as XXXG, in green, is observed in the negative subsites of one protein molecule. B) Active-site detail

shown in wall-eyed stereo; colouring is as in Figure 2-5B. C) Schematic representation of active-site interactions, analogous to Figure 2-5.

2.3.5.4 Supporting GGGG and XXXG Complex Structures

The 0.97 Å resolution structure of VvEG16(Δ V152/E89A) in complex with cellotetraose bound in the negative subsites and glucose in the +1 subsite (Figure A-13) was determined with an *R*/*R*_{free} ratio of 0.136/0.150 (Table A-3). These relatively high individual values can be partially attributed to density in the positive subsites which could not be modelled with confidence, due to the partial occupancy of many species, including glucose, water and possibly glycerol. In contrast, cellotetraose (GGGG) was clearly modelled in the negative subsites, and superposed perfectly with the glucan backbones in the -3 through -1 subsites of the VvEG16(Δ V152/E89A):MLGO and VvEG16(Δ V152/E89A):XyGO complexes (Figure A-15). In contrast to the well-defined electron density in these subsites, the weighted 2*m*F₀-*D*F_c difference density for the non-reducing-terminal glucosyl residue could only be observed when the cut-off was lowered to 1.5 σ , due to a high relative B-factor (21.7 Å², compared to 6.5-8.4 Å² for Glc residues in the -3 to -1 positions). As in the bMLGO and tXyG complexes, this suggests that the putative -4 subsite may only be weakly interacting or absent.

The surface-cysteine variant VvEG16(Δ V152/C22S/C188S) was originally produced in an unsuccessful attempt to eliminate over-labelling by a XXXG-*N*-bromoacetylglycosylamine inhibitor (Fenger and Brumer, 2015). Fortuitous crystallisation of this variant in complex with the heptasaccharide XXXG in the negative subsites revealed identical interactions to those observed in the VvEG16(Δ V152/E89A):XyGO complex (Figure A-14, Table A-3). Here again, strict superposition of glucosyl residues in subsites -3 to -1 was observed (Figure A-15), whereas interactions in a potential weakly interacting -4 subsite were not apparent.

2.3.5.5 EG16 Tertiary Structure vis-à-vis GH16 Licheninases and XTH Gene Products

The superposition of VvEG16(Δ V152/E89A) with representatives from the bacterial licheninase and plant XET/XEH clades reveals how major sequence insertions and deletions along the evolutionary trajectories of extant GH16 enzymes are manifested in their tertiary structures (Figure 2-7 *cf.* Figure 2-1). A key defining feature in the phylogeny of these enzymes is the presence of a regular β -strand (β 8) bearing the catalytic residues. This motif is distinct

from other members of GH16, including the closely related laminarinases (β (1,3)-glucanases, EC 3.2.1.39), in which an additional residue in this strand produces a " β -bulge" (Michel *et al.*, 2001) (Figure 2-1).

The superposition of the high-resolution VvEG16(Δ V152/E89A):GGGG complex with the *Paenibacillus macerans* licheninase:GGG3G product complex (PDB ID 1U0A) (Gaiser *et al.*, 2006) reveals a striking structural homology and almost perfect alignment of amino acids in the positive subsites (Figure 2-7B). Although the glucosyl residue in the -1 subsite is similarly positioned by the same hydrogen bonding and aromatic stacking interactions in both VvEG16 and the *P. macerans* licheninase, the distal negative subsites in these enzymes are highly divergent. This is the result of significantly shorter Loops 3 and 5 in EG16 members than in both licheninases (Gaiser *et al.*, 2006) and laminarinases (Fibriansah *et al.*, 2007) (Figure 2-7 cf. Figure A-1). These loop differences dramatically alter the trajectory of unbranched β -glucans across the concave surface of the β -jelly-roll fold (Figure 2-7B). In particular, the recognition of a $\beta(1,3)$ linkage between subsites -2 and -1 is a key defining feature of licheninase specificity, which arises directly from steric restriction by the extended Loops 3 and 5 in these enzymes (Planas, 2000). In VvEG16, the more open active-site not only enables binding of all- $\beta(1,4)$ linked glucan chains in subsites -3 to -1, but also curiously disfavors binding of Glc $\beta(1,3)$

Not least, this relief of steric constriction is also central to the accommodation of highly branched xyloglucan chains in the VvEG16 active-site cleft (Figure 2-7C). As such, the glucan backbone of the XXXG moiety in the negative subsites of the VvEG16(Δ V152/E89A):tXyGO complex closely superposes with the glucan backbone of XLLG in the TmNXG1(Δ YNIIG) complex (PDB ID 2VH9, (Mark *et al.*, 2009)). Despite VvEG16 and TmNXG1 displaying similar hydrolysis kinetics toward XyG and XXXGXXXG substrates, there is limited similarity in the positions of xylosyl branches in the corresponding complexes. This suggests either significant flexibility in substrate recognition or the evolution of distinct binding modes in these broad, negative-subsite clefts.

62



Figure 2-7: Tertiary structural comparison of VvEG16(ΔV152,E89A) with representative licheninase, XET, and XEH enzymes of GH16. A) Cartoon representation of VvEG16(ΔV152,E89A) with active site residues (A89, D91, E93) shown as black sticks (based on PDB ID: 5DZE). Loops and β-strands are labelled in order from the N-terminus to C-terminus (cf. Figure A-1) and the directionality of active-site

subsites indicated. B) Superposition of the VvEG16(ΔV152,E89A):GGGG complex (dark blue, PDB ID: 5DZE) with *P. macerans* licheninase:GGG3G complex (green, PDB ID: 1U0A). The ligands are shown as sticks in colours corresponding to the protein structures. The inset highlights the similarity of the two enzymes in the positive subsites, in contrast to the major loop differences observed in the negative subsites (loop 3 corresponds to the "licheninase loop"). C) Wall-eyed stereo view of the superposition of the VvEG16(ΔV152,E89A):XyGO complex (dark blue, PDB ID: 5DZG) with TmNXG1:XLLG negative-subsite complex (dark pink, PDB ID 2UWA, XLLG from PDB ID: 2VH9), and PttXET16-34:XLG positive-subsite complex (light pink, PDB ID: 1UMZ).

A clear distinction of plant XTH gene products in GH16 is the presence of a large Cterminal extension (InterPro Domain IPR010713, "XET C"), as well as a lengthening of Loop 9, which significantly increase the surface area in the positive subsite region and directly enable specific recognition of branched xyloglucan substrates (Johansson et al., 2004, Mark et al., 2009, Mark et al., 2011). Superposition of the VvEG16(Δ V152/E89A):tXyGO complex with the Populus tremula x tremuloides XET16-34:XLG positive-subsite complex (Johansson et al., 2004) reveals notable tertiary structural differences in this region of the active-site (Figure 2-7C). Firstly, a lesser number of positive subsites in VvEG16 is clearly evident, due to the lack of the "XET C" domain. Furthermore, VvEG16 and PttXET16-34 exhibit distinct differences in the trajectory of the glucan backbone through the positive subsites, which arise from alternate conformations of Loop 14. Yet despite these differences, VvEG16 demonstrates a much higher degree of overall tertiary structural similarity with archetypal XET and XEH (excluding the Cterminal "XET C" extension), in spite of lower sequence similarity, than with licheninases (Figure 2-7B,C, Figure A-1). Thus, while VvEG16 exhibits overall a more open, XET/XEH-like active-site cleft, which enables recognition and hydrolysis of both MLG and XyG, residuespecific variations and the addition of the C-terminal extension appear to be key contributors to the evolution of the strict XyG-specificity observed for the majority of characterised XTH subfamily members.

2.4 Discussion

2.4.1 Plant EG16 Members Represent a Unique Class of Bi-Functional Matrix Glycan Hydrolases

In light of the widespread distribution of MLG and XyG across the plant kingdom, there is a clear and pressing need to achieve a comprehensive understanding of the enzymology of the

biosynthesis and biodegradation of these key matrix glycans (Fincher, 2009, Burton *et al.*, 2010, Attia *et al.*, 2016, Pauly and Keegstra, 2016). Endogenous *endo*-hydrolysis of MLG, the key initial step in glucose mobilisation, was previously thought to be catalysed exclusively by plant Glycoside Hydrolase Family 17 (GH17) members (Fincher, 2009). On the other hand, endogenous *endo*-hydrolysis of the XyG backbone was only known among a small subset of GH16 members, the XEHs, which are encoded by Group III-A *XTH* genes (Baumann *et al.*, 2007, Eklöf and Brumer, 2010).

Building upon our previous work on a *Populus trichocarpa* homolog (Eklöf *et al.*, 2013), we demonstrated here that members of the EG16 clade within GH16, which exclusively comprises plant homologs (Figure 2-1), uniquely catalyse the *endo*-hydrolysis of both MLG and XyG with comparable efficiency. Importantly, the kinetic analysis of VvEG16 presented here represents the first quantitative demonstration that predominant mixed-linkage *endo*-glucanase activity is found among plant enzymes that are phylogenetically and structurally distinct from canonical GH17 members. Moreover, these EG16 members have notably broader substrate specificity than both bacterial licheninases (Planas, 2000) and plant XEHs (Eklöf and Brumer, 2010) of GH16.

2.4.2 EG16 Members Represent Extant Transitional Enzymes Linking the Evolution of Bacterial Licheninases and Plant XTG Gene Products

Molecular phylogeny in combination with detailed structural enzymology of VvEG16 has allowed us to elucidate key protein features that give rise to the bi-functionality of EG16 members. Moreover, this composite analysis allows us to propose that EG16 members are extant "transitional enzymes" (by analogy with *transitional fossils* in paleontology) in the evolution of GH16 members. Specifically, the experimentally determined crystal complexes of VvEG16 directly reveal key, stepwise structural changes that give rise to the functional diversification of extant bacterial licheninases, plant EG16s, and plant XETs and XEHs.

Our global phylogenetic analysis of GH16 β -jellyroll domains across species and activities demonstrates a clear delineation of licheninases, EG16 members, and *XTH* gene products into distinct, well-supported clades (Figure 2-1A). This phylogeny essentially recapitulates that first presented by Barbeyron, Michel, *et al.* (Barbeyron *et al.*, 1998, Michel *et al.*, 2001), but with the key insertion of the EG16 clade intermediate between the licheninases and *XTH* gene products. In consideration of the early evolutionary origin of bacteria and key

sequence features present in all bacterial GH16 enzymes which have been lost in EG16 enzymes (such as the extended loop 3), it is most likely that a licheninase-like protein served as the ancestor to both extant bacterial licheninases and plant EG16 enzymes. In addition to the common β -jellyroll fold, composed of 13 core β -strands and a regular β 8-strand bearing the catalytic EXDXE motif, this ancestral protein also would have included an extended version of Loop 3 (referred to here as the "Licheninase Loop"), as found in extant bacterial mixed-linkage *endo*-glucanases (Figure A-1).

Subsequent truncation of Loop 3 in the divergence of EG16 members (Figure 2-1B) resulted in the widening of the active-site cleft, thereby allowing these enzymes to accept both the defining licheninase substrate, MLG, as well as highly branched XyG chains (Figure 2-7). It is presently unclear in which kingdom this mutation may have occurred; just as there are no EG16-like (short Loop 3) proteins among known bacterial sequences, there are no licheninase-like (long Loop 3) sequences among known plant sequences. However, EG16 members are found in the genomes of early-diverging plants, currently represented by the mosses *Physcomitrella patens* and *Selaginella moellendorffii* (Figure 2-1B), which highlights the ancient evolution of this class of enzymes.

In turn, we propose that the broadened active-site cleft of an ancestral EG16 member poised this protein scaffold for further evolution into extant xyloglucan *endo*-transglycosylases (Group I, II, and III-B *XTH* gene products (Eklöf and Brumer, 2010)) by addition of the distinguishing "XET_C" C-terminal domain extension (Figure 2-1B). This is the most parsimonious explanation in light of the existence of EG16 members in plants, vis-à-vis a conceivable tandem event directly giving rise to a XET-like homolog via the simultaneous truncation of Loop 3 and gain of the XET_C extension in an ancestral licheninase. Notably, no proteins are currently known that comprise both a long, licheninase-like Loop 3 and a XET_C extension, which supports the proposed evolution of XETs from an EG16-like (short Loop 3) ancestor. Collectively, the truncated Loop 3 and XET_C extension, together with various activesite point mutations, has resulted in the high specificity of XETs for XyG over MLG. In this context, it is interesting to note that a recently discovered *Equisetum* β-glucan:XyG heterotransglycosylase does not represent an evolutionary intermediate between licheninases and XETs, but appears to represent a later functional divergence in Group I/II *XTH* gene products (Simmons *et al.*, 2015). Finally, we have previously provided phylogenetic, biochemical, and tertiary structural evidence that the comparatively fewer number of predominant XEHs (Group III-A *XTH* gene products) evolved subsequently from strict XETs by the introduction of a short, 5 amino-acid loop insert on the active-site cleft, immediately preceding strand β 10 (Figure A-1, "XEH loop"; Figure 2-1B) (Baumann *et al.*, 2007). These XyG hydrolases have apparently evolved to meet needs for seed-storage XyG mobilisation and fruit ripening in select species (reviewed in (Baumann *et al.*, 2007)), and also appear to be active in vegetative tissues of Arabidopsis (Kaewthai *et al.*, 2013).

Thus, the structural enzymology of VvEG16 vis-à-vis other GH16 members provides compelling evidence for the following evolutionary trajectory: ancestral licheninase—ancestral EG16—ancestral XET—ancestral XEH. Through evolution, the ancestral bacterial licheninaseencoding genes have expanded greatly across species (Lombard *et al.*, 2014), while ancestral plant *XTH* genes have also seen massive expansion to comprise 20-50 members in individual bryophytes, lycophytes, and angiosperms (reviewed in (Eklöf and Brumer, 2010, Eklöf *et al.*, 2013)). Strikingly, EG16 members appear to be restricted to one member in each plant genome (Figure 1-1B), with many prominent dicots, including the model species *Arabidopsis thaliana*, lacking EG16 representatives altogether (Eklöf *et al.*, 2013). Indeed, EG16 members have been largely overlooked, or in some cases have been mistakenly classified as *XTH* gene products (Geisler-Lee *et al.*, 2006, Yokoyama *et al.*, 2010, Eklöf *et al.*, 2013), for this reason. The now clear delineation of EG16 will significantly enable future bioinformatics analysis, as well as consideration of functional data in an updated phylogenetic context (Hatfield and Nevins, 1987).

2.4.3 In Vivo Roles of EG16 Members

Despite their limited distribution, transcriptional analysis suggests that EG16 members are not simply vestigial, but are likely to play functional roles in plants. In particular, VvEG16 was first identified in mRNA extracted from each of root, leaf and flower of *V. vinifera* (Peng *et al.*, 2007). Also in retrospect, EG16 members have been found in EST libraries from flowers, leaves, and roots of a variety of plants at different developmental stages, including citrus trees and cotton (Arpat *et al.*, 2004, Forment *et al.*, 2005). In the absence of systematic studies, it is unclear under which conditions EG16-encoding genes might be specifically upregulated. Curiously, however, all EG16 members identified thus far lack a trafficking signal peptide, lack N-glycosylation sites, and contain numerous unpaired cysteine residues. These characteristics argue against apoplastic role for EG16 members (unless secretion occurs via a non-classical pathway (Rose and Lee, 2010)) This is in sharp contrast to other plant XyG- and MLG-active enzymes, namely the *XTH* gene products of GH16 (Eklöf and Brumer, 2010) and the mixed-linkage β -glucanases of GH17 (Fincher, 2009). As such, we tentatively suggest that EG16 members may not be not involved in wall polysaccharide remodelling in the classical sense, but may function in housekeeping or modifying roles within the confines of the cell membrane, or after cell death. Certainly, the unique MLG cleavage specificity of VvEG16, which leads to polysaccharide gelation *in vitro*, is worthy of consideration in light of potential wall-modifying roles and may also have biotechnological application.

Although the biological role(s) of the EG16 members presently remain enigmatic, the present enzyme structure-function analysis sets the stage for future genome mining, transcriptomic analysis, cellular localisation, and forward- and reverse-genetic analyses across plant species. It is especially intriguing that mosses, grasses, and trees all encode EG16 members. Elucidating to what extent EG16 function *in vivo* is conserved or specialised among species with such intrinsically diverse cell wall compositions and growth habits now remains to be resolved.

Chapter 3: Structure-Function Analysis of a Mixed-Linkage β-Glucanase/Xyloglucanase from the Key Ruminal Bacteroidetes *Prevotella bryantii* B₁4

3.1 Introduction

The chemical and structural complexity of plant cell walls pose a challenge to organisms, from bacteria to humans, in extracting energy from biomass via polysaccharide saccharification and further metabolism. A diverse collection of amorphous polysaccharides ("hemicelluloses" and "pectins"), structural (glyco)proteins and polyphenolics ("lignin") associate with paracrystalline cellulose microfibrils within the plant cell wall to form a composite framework that is both strong and dynamic (Albersheim et al., 2010). Among the many matrix glycans in land plants, the family of xyloglucans and the mixed-linkage glucans predominate in varying ratios, depending on the plant lineage and tissue type (Fry, 1989, Hayashi, 1989, Fry et al., 2008b, Xue and Fry, 2012). Mixed-linkage glucans have a general structure composed of short stretches of $\beta(1,4)$ -linked glucosyl residues (typically 3-5 residues) which are linked together by $\beta(1,3)$ linkages (Figure 3-1B) (Burton and Fincher, 2009). In contrast, xyloglucans are composed of a linear backbone of $\beta(1,4)$ -linked glucosyl residues decorated with a regular pattern of $\alpha(1,6)$ linked xylosyl residues, which are further extended with galactosyl, fucosyl, and/or arabinosyl residues (Figure 3-1A) (Tuomivaara *et al.*, 2015). The $\beta(1,3)$ "kinks" of mixed-linkage glucan and the complex branches of xyloglucan appear to serve a similar function of inducing structural disorder, thereby endowing these polysaccharides with significant water solubility and hydrogellation properties, while at the same time maintaining affinity to cellulose.

The vast diversity of glycoside hydrolases (GHs) directed toward plant cell walls is a testament to the importance, and the challenge, of biomass degradation in the biosphere. Indeed, hundreds of thousands of GHs have been annotated in over 130 structurally related families in the Carbohydrate-Active Enzymes (CAZy) classification, the majority of which are directed to plant polysaccharides (Davies *et al.*, 2005, Davies and Sinnott, 2008, Lombard *et al.*, 2014). Moreover, considerable divergent evolution has occurred within individual GH families giving rise to substrate specificity differences among members. Mapping functional diversity in such

polyspecific families has been enabled by further division into phylogenetic subfamilies in some cases (Stam *et al.*, 2006, Garron and Cygler, 2010, Lombard *et al.*, 2010, Aspeborg *et al.*, 2012).



Figure 3-1: Polysaccharide structures used in this study. A) Structure of tamarind xyloglucan depicting the repeating Glc₄ oligosaccharide moiety and variable galactosylation (a, b = 0 or 1). The primary site of PbGH5A attack (and canonical cleavage site of XyG) is marked with a solid arrow and the secondary site is marked with a dashed arrow. B) Structure of barley $\beta(1,3)/\beta(1,4)$ -mixed-linkage glucan (depicting GGG3GG). The primary and secondary sites of PbGH5A attack are indicated as for tXyG.

Glycoside hydrolase family 5 (GH5) is a key example of family diversity, with members demonstrating over 20 known specificities. GH5 members are united by a canonical doubledisplacement, anomeric configuration-retaining mechanism of hydrolysis, which involves two key catalytic carboxylic acid sidechains present on a conserved $(\beta/\alpha)_8$ protein fold (Barras *et al.*, 1992). The recent division of GH5 into subfamilies has shown that many of these activities cluster into phylogenetic clades (Aspeborg et al., 2012). Among these, GH5 subfamily 4 (GH5_4) constitutes one of the largest, which generally encompasses *endo*- $\beta(1,4)$ -glucanases, including cellulases (EC 3.2.1.4), mixed-linkage endo- $\beta(1,3)/\beta(1,4)$ -glucanases (EC 3.2.1.73), and highly-specific endo-xyloglucanases (EC 3.2.1.151) evolved for the saccharification of plant biomass. GH5_4 endo-xyloglucanases (Gloster et al., 2007, Larsbrink et al., 2014a) are particularly distinguished by their ability to accommodate and harness the numerous extended $\beta(1,6)$ -xylosyl branches in diverse xyloglucans (Eklöf *et al.*, 2012). Unfortunately, the current level of functional characterisation of GH5_4, which includes the observation that most of the characterised members have not been tested consistently on the same panel of substrates (e.g. including xyloglucan) (Faure et al., 1989, Hamamoto et al., 1992), means that clear delineation of polysaccharide specificity in this subfamily is not straightforward. This presents a significant

difficulty for *in silico* analysis of (meta)genomes for functional prediction, as well as for the selection and application of specific enzymes for industrial biomass utilisation.

To address this issue, we present here the characterisation of a GH5_4 member, PbGH5A, from the symbiotic gut bacterium *Prevotella bryantii* B₁4 involved in dietary polysaccharide breakdown (Avguštin et al., 1997, Flint and Bayer, 2008, Purushe et al., 2010). Locus PBR_0368 of the Prevotella bryantii B₁4 genome encodes a bi-modular gene product composed of a predicted N-terminal, Signal Peptidase I-cleavable signal peptide, followed by a GH26 module and a C-terminal GH5 module (PbGH5A) (Purushe et al., 2010). Early efforts to clone PBR_0368 and characterize its product (equivalent to GenBank AAC97596, also known as ORF4, CMCase, or Cel5A) revealed general endo-glucanase activity via assay on carboxymethylcellulose (Matsushita et al., 1990, Gardner et al., 1997). However, detailed specificity data are currently lacking, especially in light of the identified bimodularity of this protein and diverse specificities found within GH26 and GH5 (Cantarel et al., 2009). Notably, PBR_0368 is located in a predicted Polysaccharide Utilisation Locus encoding hallmark SusDand SusC-like proteins and at least two other GHs whose collective function is currently unknown (Martens et al., 2014). In the present study, kinetic analyses on a range of natural and artificial substrates, together with tertiary structures of enzyme variants in complex with oligosaccharides and an active-site affinity label, yielded molecular-level insight into interactions along the entire active-site cleft responsible for the specificity of recombinant PbGH5A for mixed-linkage glucan over xyloglucan.

3.2 Materials and Methods

3.2.1 Analytical Methods

3.2.1.1 HPAEC-PAD Carbohydrate Analysis

High-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) was performed on a Dionex ICS-5000 system equipped with an AS-AP autosampler with a temperature-controlled sample tray run in a sequential injection configuration using Chromeleon 7 control software. The injection volume was 10 μ L unless otherwise specified. A 3x250 mm Dionex CarboPac PA200 column with a 3x50 mm guard column was used for all HPAEC separations. Solvent A was 18.2 M Ω H₂O, solvent B was 1.0 M NaOH prepared from a carbonate-free 50-52% stock and solvent C was 1.000 M NaOAc prepared from anhydrous BioUltra-grade solid (Sigma). The gradients used were as follows: Gradient A: 0-5 min, 10 % B, 2 % C; 5-12 min 10 % B, 2-30 % C linear gradient; 12-12.1 min 50 % B, 50 % C; 12.1-13 min return to initial conditions (exponential profile 9); 13-17 min, initial conditions. Gradient B: Gradient A with 0 % C initially. Gradient C: Gradient A with 6 % C and a 12 minute linear gradient. Gradient D: Gradient A with 3.5 % C initially.

3.2.1.2 Mass Spectrometry

Intact protein masses were determined on a Waters Xevo Q-TOF with a nanoACQUITY UPLC system, according to the method described by (Sundqvist *et al.*, 2007). MALDI-TOF analysis was performed on a Bruker Autoflex MALDI-TOF equipped with a Bruker Smartbeam-II 355 nm laser system. Samples dissolved in 18.2 M Ω water (0.1-10 mg/mL) were mixed 1:1 with 10 mg/mL 2,5-dihydroxybenzoic acid (DHB) dissolved in 1:1 H₂O:MeOH. The samples were left to dry under ambient conditions. MALDI spectra were acquired in positive reflectron mode, averaging 500 laser shots per spectrum. External calibration was performed using the standard mix of XXXG, XXLG/XLXG and XLLG obtained from canonical enzymatic hydrolysis of tamarind tXyG.

3.2.2 Substrates and Inhibitors

Oligosaccharides and their derivatives are abbreviated using the general shorthand for xyloglucan oligosaccharides, in which G represents Glc*p*, X represents [Xyl*p*(α 1-6)]Glc*p*, and L represents [Galp(β 1-2)Xyl*p*(α 1-6)]Glc*p*, with β (1,4) linkages between backbone glucosyl units as the default (Tuomivaara *et al.*). In mixed-linkage gluco-oligosaccharides, β (1,3)-linked glucosyl residues are denoted as G3 (e.g. G3G is laminaribiose and GG is cellobiose).

3.2.2.1 Commercial Substrates

High purity (>94%) mixed-linkage glucan (beta-glucan (barley; high viscosity)) (bMLG), carboxymethylcellulose (CMC), konjac glucomannan (kGM), carob galactomannan (cGM), tamarind xyloglucan (tXyG), wheat arabinoxylan (wAX), beechwood xylan (bX), cellooligosaccharides (Glc₃-Glc₆), mixed-linkage-glucan oligosaccharides (Glc₄), mannohexaose (MMMMM) and 2-chloro-4-nitrophenyl β-cellotrioside (GGG-CNP) were purchased from Megazyme International (Ireland) and used for all activity measurements and HPAEC-PAD experiments. Hydroxyethylcellulose (HEC) was purchased from Fluka (Sigma-Aldrich). 4-Nitrophenyl β-glucoside (Glc-PNP) and 4-nitrophenyl β-cellobioside (GG-PNP) were purchased from Sigma-Aldrich. 2-Chloro-4-nitrophenyl β -cellobioside (GG-CNP) was purchased from Carbosynth (UK). Phosphoric acid-swollen cellulose (PASC) (Walseth, 1952) was prepared according to (Zhang *et al.*, 2006).

3.2.2.2 Synthetic Substrates and Inhibitors

4-nitrophenyl β -cellotrioside (GGG-PNP) was a kind gift from Prof. S. Withers (UBC). The 2-chloro-4-nitrophenyl glycoside of XXXG (XXXG-CNP) (Ibatullin *et al.*, 2008) and xyloglucan-derived inhibitors (Fenger and Brumer, 2015) were synthesised as previously described.

3.2.2.3 Xyloglucan oligosaccharides (tXyGOs)

The tetradecasaccharide XXXGXXXG was prepared by partial digestion of xyloglucan from de-oiled tamarind kernel powder (dTKP, Premcem Gums) with His₆-PpXG5 (Gloster et al., 2007) followed by degalactosylation with AnBgl35A (Megazyme). Briefly, 100 grams of dTKP were slowly added to 1 L of 10 mM NH₄OAc, pH 5.5 containing 500 U (~2.5 mg) of PpXG5 (where 1 unit (U) is defined as the amount of enzyme which releases 1 µmol of glucoseequivalent reducing ends per minute). The reaction was stirred at 50 °C until a smooth, tan opaque suspension formed (~30 min). The reaction was sampled regularly. The samples were filtered, run over Dowex 1X2 Cl and analysed using HPAEC-PAD (gradient C) until the population of Glc₈-tXyGOs was maximal (~4 h). The pH was raised to 8 (using 1 M NH₄OH) to stop the reaction and the solution was centrifuged for 15 min at 4000 g. The translucent yellow supernatant was then decolourised by passage through Dowex 1X2 Cl and passed through a HisTrap FF crude column (GE Healthcare Life Sciences) to fully remove His₆-PpXG5. The pH was then returned to 5.5 with 1 M AcOH and 1400 U of AnBgl35A was added and stirred at 30 °C overnight. The degalactosylated tXyGOs were lyophilised for storage. 500 mg of this was then dissolved in 5 mL of diH₂O, 0.45-µm filtered and purified using a 90 cm P6 Bio-Gel (Bio-Rad) column (XK 26/100, GE Life Sciences) run at 6 cm/h at room temperature. Fractions were monitored by HPAEC-PAD (Gradient C) and homogeneous fractions of XXXG, XXXGXXXG and XXXGXXXGXXXG were pooled and lyophilised to give white foam (final yield: 200 mg XXXG, 55 mg XXXGXXXG, 30 mg XXXGXXXGXXXG).

3.2.2.4 Mixed-linkage glucan oligosaccharides (bMLGOs)

 $Glc-\beta(1,3)-Glc-\beta(1,4)-Glc-\beta(1,4)-Glc-\beta(1,3)-Glc-\beta(1,4)-Glc-\beta(1,4)-Glc (G3GGG3GGG)$ was prepared by the digestion of oat β -glucan (B-CAN, Garuda International) with *Vitis vinifera* family 16 endo-glucanase (VvEG16) (expressed and purified in-house, manuscript in preparation) to give a mixture of oligosaccharides with the formula G3GGG(3GGG)_n. 10 g of B-CAN was initially swollen in 500 mL of H₂O at 25 °C for 15 min. The B-CAN was then collected by centrifugation at 1000 g for 2 min and the supernatant was discarded. The material was washed in this manner 3 times to extract glucose, unidentified oligosaccharides, fines, and coloured material. The swollen particles were then resuspended in 500 mL of 10 mM NH₄OAc, pH 5.5 and heated to 80 °C. The solution was stirred until dissolved (~15 min) and allowed to cool to 37 °C. 50 U (~10 mg) of VvEG16 was then added, and the reaction was stirred at 37 °C overnight. 30 min into the digestion, the now significantly less viscous solution was centrifuged at 4000 g for 5 min to remove a small amount of insoluble matter. The reaction completion was confirmed based on the oligosaccharide distribution observed by HPAEC-PAD (Gradient A) and the opaque tan solution was centrifuged at 4000 g for 5 min at room temperature to separate insoluble bMLG from soluble bMLG. The now clear, faintly yellow solution was then adjusted to pH 8 using 1 M NH₄OH and decolourised by running through a 5 g plug of Dowex 1X2 Cl. The product was then precipitated from the clear, colourless solution by the addition of 1 L of acetone. After cooling to -20 °C in the freezer, a well-flocculated white precipitate was collected by centrifugation (in a HDPE bottle) at 1000 g for 2 min. The product was dried under vacuum for several hours to give 1.52 g of a white powder. 500 mg of this was then dissolved in 5 mL of diH₂O and purified using a 90 cm P2 Bio-Gel (Bio-Rad) column (XK 26/100, GE Life Sciences) run at 6 cm/h at room temperature. Fractions were monitored by HPAEC-PAD (Gradient A), and homogeneous fractions of G3GGG and G3GGG3GGG were pooled and lyophilised (final yield: 41 mg G3GGG, 62 mg G3GGG3GGG).

3.2.3 Enzyme Cloning and Expression

The PbGH5A gene fragment of PBR_0368 corresponding to amino acid residues 425 to 776 was received from the Joint Genome Institute (http://jgi.doe.gov) in a pET101 plasmid and subcloned into a cloning vector p15Tv-LIC (Eschenfeldt *et al.*, 2009) providing an N-terminal His₆-tagged fusion with a tobacco etch virus (TEV) protease cleavage site between the tag and the enzyme. PbGH5A was expressed in *E. coli* BL21(DE3) grown in auto induction media

(Studier, 2005) for 3 h at 37 °C and continued overnight growth at 18 °C. Cells were harvested via centrifugation at 5000 *g*. The resulting pellet was resuspended in a binding buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5 mM imidazole, and 14% glycerol (v/v)), lysed via sonication, and cell debris removed via centrifugation at 30000 *g* for 30 min. Cleared lysate was loaded onto a 5 mL Ni-NTA column (QIAGEN) pre-equilibrated with the binding buffer, and the column was washed with the binding buffer containing 30 mM imidazole. Bound proteins were eluted using the binding buffer with 250 mM imidazole. The His₆ tag was removed by cleavage with TEV protease, expressed and purified in-house per (Waugh, 2011), overnight at 4 °C while in dialysis buffer 0.5 M NaCl, 10 mM HEPES pH 7.5, and 0.5 mM tris[2-carboxyethyl]phosphine (TCEP), followed by binding to Ni-NTA resin and capture of the flow-through. Fractions containing the protein of interest were identified by SDS-polyacrylamide gel electrophoresis (SDS-PAGE).

3.2.4 Enzyme Kinetics and Product Analysis

3.2.4.1 Polysaccharide Hydrolysis

Polysaccharide hydrolysis was quantified using either the BCA (Doner and Irwin, 1992) or DNSA (Miller, 1959) assay. For BCA assays, reactions were prepared to a final volume of 100 μ L and heated to the incubation temperature for 0, 15 and 30 min before being quenched by the addition of fresh BCA reagent (100 μ L). A glucose series (10-500 μ M) was run with each assay. Colour was developed by heating to 80 °C for 10 min before reading the absorbance at 563 nm. For DNSA assays, 100 μ L of the reaction was quenched by adding 100 μ L of DNSA reagent. The reaction was then heated to 95 °C for 10 min to develop colour, cooled to room temperature and centrifuged at 1000 g for 1 min. The absorbance was read at 540 nm.

The pH optimum of the enzyme was initially determined using the BCA assay to quantify reducing ends over 15 minutes of incubation of 0.05 nM of native enzyme with 1 mg/mL bMLG at 37°C using 50 mM MES (pH 5.3-7.9), MOPS (pH 7.4-8.5), TRIS (pH 7.2-8.8), acetate (pH 3.75-5.5), citrate (pH 3.5-6.5), phosphate (pH 6.1-7.9) and glycine (pH 8.4-9.4) buffers. However, using the polysaccharide substrate, different kinetic pKa values were observed for different buffers (Figure 3-2B). The pH optimum of the enzyme was further determined using citrate (pH 3-6), acetate (pH 3.75-5.5), MES (pH 5.3-7.9), MOPS (pH 7.4-8.5) and TRIS (pH 7.2-8.4) buffers (50 mM) (Figure 3-2A) with a chromogenic oligosaccharide substrate, giving

two consistent kinetic pKa values. The native enzyme (0.33 nM) was incubated for 30 min with GGG-CNP (0.5 mM) at 37 °C, and then free CNP was determined by diluting 5:1 into 100 mM Na₂CO₃ and measurement of A₄₀₅. Rates in TRIS buffer were barely detectable at all pH values indicating that TRIS is strongly inhibitory.

The temperature optimum was determined in 50 mM pH 5.5 sodium citrate buffer using 1 mg/mL bMLG as substrate (Figure 3-3A) with 0.02 nM enzyme. The reaction was mixed at 4 °C and incubated at a temperature ranging from 30 to 55 °C for 30 min before reducing ends were quantified using the BCA assay. The specific activity of PbGH5A was standardised with 1 mg/mL bMLG substrate at 37 °C in 10 mM pH 5.5 sodium citrate buffer. 1 unit (U) was defined as the amount of enzyme which releases 1 μ mol of glucose-equivalent reducing ends per minute. The thermal stability of PbGH5 was determined by incubating the enzyme (1 μ g/mL in 20 mM pH 5.5 citrate) at temperatures ranging from 30-74 °C. At regular time intervals, samples were taken, diluted into room-temperature citrate buffer (pH 5.5) and assayed using 200 μ M XXXG-CNP.

To determine limit-digestion products, PbGH5A (10 μ g) was added to 1 mL of 0.1 mg/mL substrate in 50 mM NaOAc pH 5.5 and incubated for 4 hours at 37 °C. 10 μ L of the reaction was then analysed by HPAEC-PAD directly using gradient A.



Figure 3-2: pH-rate profiles for PbGH5A. A) Activity of PbGH5A on 0.5 mM GGG-CNP in various buffers across a range of pH at 37°C. Error bars are the standard deviation of three single-point replicates. The values (excluding the TRIS series) were fit to determine the pH optimum (pH 5) and apparent kinetic pKa values (3.5 and 6.5). B) Activity of PbGH5A on 1 mg/mL bMLG in various buffers across a range of pH at 37°C. Each point is the average of two replicates.



Figure 3-3: Thermal stability of PbGH5A. A) Activity of PbGH5A on 1 mg/mL bMLG in pH 5.5 NH4OAc as a function of temperature. The reaction was incubated for 15 minutes to minimize instability effects. Each point is the average of two single-point replicates after subtraction of a boiled-enzyme control. B) Long-term thermal stability of PbGH5A in pH 5 citrate buffer across a range of temperatures over 4 hours. After incubation at temperature, the enzyme was diluted and assayed using 0.2 mM XXXG-CNP.

3.2.4.2 Chromogenic Oligosaccharide Hydrolysis

4-nitrophenol substrate kinetics were determined by mixing enzyme (20-1000 nM), buffer (50 mM pH 5.5 citrate) and substrate (0.1-25 mM) to a final volume of 200 μ L. At 5 minute intervals, 60 μ L of the reaction was diluted into 540 μ L of 50 mM Na₂CO₃ and A₄₀₅ was measured on a Cary 60 UV/Vis spectrometer with a 1 cm path length quartz cuvette. An extinction coefficient of 18.2 mM⁻¹·cm⁻¹ was used to quantify 4-nitrophenol release. 1 U was defined as the amount of enzyme which releases 1 μ mol of 4-nitrophenol per minute. 2-chloro-4-nitrophenol (CNP) substrate kinetics were determined by preheating 180 μ L of 1.11x substrate stock (to give 0.02-10 mM final) and adding 20 μ L of 10x enzyme stock to give 0.01-100 nM final in 20 mM NaOAc pH 5.5. The change in absorbance at 405 nm was followed continuously over 10 min at 37 °C in 200 μ L quartz cuvettes using a Cary 300 UV/Vis Spectrometer with an 8-cell sample changer and thermostat. The extinction coefficient for CNP was determined to be 10.7 mM⁻¹cm⁻¹ in the buffer used. For the XXXG-CNP substrate the assay was optimised for the conditions compatible for residual activity measurement of inhibitor evaluation. The hydrolysis was monitored in 50 mM citrate buffer at pH 5.5, absorbance was measured at 405 nm, and the extinction coefficient for CNP was determined to be 11.2 mM⁻¹ cm⁻¹ in the buffer used. Specific activity measurements for wild-type enzyme and the three mutants, E280A, S119A, H112A were determined using GGG-CNP at 500 μ M and XXXG-CNP at 200 μ M in 50 mM citrate buffer at pH 5.5.

3.2.4.3 Native Oligosaccharide Hydrolysis

HPLC kinetics were determined by mixing a 10x enzyme in buffer stock (to give 0.02-10 nM and 20 mM pH 5.5 citrate final enzyme and buffer concentration) with a 1.11x substrate stock (to give 0.005-1 mM final substrate concentration) preheated to 37 °C. For example, 10 μ L of 0.2 nM PbGH5 in 200 mM sodium citrate buffer (pH 5.5) was added to 90 μ L of 1.11 mM GGGGGG in ddH₂O preheated to 37 °C. The reaction was then injected 4 times (10 μ L each) at regular time intervals, and the change in peak area over time was quantified. Gradient A was used for monitoring cellooligosaccharide and mixed-linkage glucan oligosaccharide degradation, gradient D was used for monitoring XXXGXXXG degradation. An 8-point linear calibration series from 0.4-100 μ M was run for each product quantified. Rates were fit to the Michaelis-Menten model (Michaelis and Menten, 1913, Johnson and Goody, 2011) using OriginPro graphing software (Origin Lab).

To determine the regiospecificity of cellopentaose hydrolysis, ¹⁸O incorporation from ¹⁸O-water was determined by mass spectrometry (Schagerlöf *et al.*, 2009). 1 μ L of PbGH5A (0.10 μ g/mL in 20 mM NH₄OAc, pH 5.5) and 1 μ L of 0.5 M NH₄OAc, pH 5.5 containing 5 mM NaOAc (to control adduct formation) were added to 22 μ L of 97% ¹⁸O-water (Cambridge Isotope Laboratories) and mixed thoroughly by reciprocal pipetting. To this was then added 1 μ L of 10 mM cellopentaose. The reaction was then mixed thoroughly again to give an estimated final ¹⁸O concentration of 85%. The reaction was drawn into a 50 μ L gas-tight Hamilton syringe

(Hamilton, model 1705) and infused into a Waters Xevo QTof at 2 μ L/min using a syringe pump (Harvard Apparatus 11 Plus). The degree of isotopic labelling was quantified as the area ratio of [M+Na]⁺(¹⁶O-1) to [M+Na]⁺(¹⁸O-1).

3.2.4.4 Inhibition Kinetics

Inhibition kinetic parameters were determined at 37 °C using 0.038 μ M PbGH5A in 25 mM citrate buffer at pH 5.5, 37 °C containing 1% bovine serum albumin (BSA), incubation with various putative inhibitor concentrations (0.1 mM – 3.5 mM) 1/5 K_i to 5 K_i. 10 μ L of enzyme-inhibitor solution was added to 190 μ L of 0.13 mM XXXG-CNP in 5 mM sodium citrate buffer, pH 5.5, and the reaction was monitored at 405 nm over 1-2 min in 1 cm quartz cuvette maintained at 37°C. The inhibition data were fit according to the Kitz-Wilson model (Kitz and Wilson, 1962) and apparent inactivation rate constants (k_{app}) were determined by fitting the residual activity data to equation 3-1. K_i and k_i values were determined by fitting plots of inactivation rate constants versus putative inhibitor concentrations to equation 3-2 by nonlinear regression using OriginPro graphing software.

Equation 3-1 $v = v_0 e^{-k_{app}t}$

Equation 3-2
$$k_{app} = \frac{k_i[I]}{K_i + [I]}$$

3.2.5 Enzyme Crystallisation

Wild-type PbGH5A enzyme was crystallised at room temperature using the hanging drop method, with 1.8 μ L of protein solution at 28 mg/mL mixed with 1.8 μ L of reservoir solution (0.1 M sodium cacodylate pH 6.3 to 7.1, 0.2 M calcium acetate, 25% PEG8K). The PbGH5A-E280A mutant was crystallised in the same crystallisation solution as above and using serial dilution seeding of wild-type crystals as starting material. The PbGH5A:XXXG and PbGH5A-E280A:GGGG complexes were obtained by soaking *apo*-enzyme crystals in reservoir solution supplemented with 20 mM XXXG or 10 mM GGGG for 3.5 h and 2 h, respectively. Prior to data collection, crystals were cryoprotected with Paratone-N oil and flash frozen in liquid nitrogen.

The PbGH5A:XXXG-NHCOCH₂Br and PbGH5A-E280A:XXXGXXXG complexes were obtained through co-crystallisation with the tag-less protein at a final concentration of 28 mg/mL. For the PbGH5A:XXXG-NHCOCH₂Br complex 6.6 mM EDTA was added to the protein and the mixture was incubated at 4 °C overnight. The inhibitor was then added to a final concentration of 8 mM, and the mixture was further incubated at 37 °C for 3 h. For the PbGH5A-E280A:XXXGXXXG complex the protein solution was incubated with 2.4 mM ligand at 37 °C for 3 h. The crystals were grown at room temperature using the sitting drop method, with 0.5 µL of complex solution mixed with 0.5 µL of reservoir solution: 0.2 M calcium chloride, 20% (w/v) PEG3350 for the XXXGXXXG complex. All complex, and 0.2 M magnesium acetate, 20% (w/v) PEG3350 for the XXXGXXXG complex. All complex crystals were cryoprotected by Paratone-N oil and flash frozen in liquid nitrogen.

3.2.6 X-ray Crystal Structure Determination

Diffraction data at 100 K were collected using a Rigaku Micromax-007 HF rotating copper anode source with a Rigaku R-AXIS IV image plate detector (for the apo-enzyme and PbGH5A:XXXG complexes), or using a Rigaku Micromax-007 HF rotating copper anode source with a Rigaku Saturn A200 CCD (at the Structure Genomics Consortium, for the PbGH5A-E280A:GGGG, PbGH5A:XXXG-NHAcBr and PbGH5A-E280A:XXXGXXXG complexes). All X-ray data were reduced with HKL-3000 (Minor et al., 2006). First, the apo-enzyme structure was determined by molecular replacement using a model generated by Phyre2 (Kelley and Sternberg, 2009) of the PbGH5A sequence with the Paenibacillus pabuli GH5 structure as a template (PDB ID:2JEQ, (Gloster et al., 2007)) and Phenix.phaser (Adams et al., 2010). This was followed by automated model building using *Phenix.autobuild*. The PbGH5A ligand complex structures were determined by molecular replacement using the *apo*-enzyme structure as the search model and *Phenix.phaser* to obtain phasing information. All refinement was performed with Phenix.refine with manual editing in Coot (Emsley and Cowtan, 2004). During refinement B-factors were defined as anisotropic for all non-hydrogen atoms and TLS parameterisation was utilised. The final atomic model of the structures included chain A residues 7-352 and chain B residues 6-352. Average B-factor and bond angle/length root mean square deviation (RMSD) values were calculated using Phenix.b_factor_statistics. All geometry was verified using the Phenix Molprobity and Coot validation tools plus the wwPDB Deposition server. The data collection and refinement statistics are listed in

Table 3-2.

3.2.7 Phylogenetic Analysis

Sequences from Glycoside Hydrolase Family 5 (Aspeborg *et al.*, 2012) from bacteria were obtained from the CAZy Database (URL: http://www.cazy.org/) (Lombard *et al.*, 2014) and aligned with MAFFT (Katoh and Standley, 2014). The distances between sequences were calculated using FastTree (Price *et al.*, 2010) based on the multiple sequence alignment. The resulting tree was displayed with Dendroscope (Huson and Scornavacca, 2012).

3.3 Results

3.3.1 Polysaccharide Kinetics

In light of the diverse specificities observed in GH5, we tested recombinant PbGH5A for activity on a library of linear β -glycan polysaccharides, including hydroxyethylcellulose (HEC), carboxymethylcellulose (CMC), phosphoric acid-swollen cellulose (PASC), xyloglucan (tXyG), glucomannan (kGM) and mixed-linkage glucan (bMLG). As anticipated from its membership in GH5 subfamily 4, significant activity toward tamarind tXyG ($k_{cat} = 6800 \text{ min}^{-1}$, $K_M = 1.1 \text{ mM}$; Table 3-1) was observed at the pH optimum of 4.5-5.5 (Figure 3-2A) and at 37 °C. However, our kinetic analysis revealed that PbGH5A is significantly more selective for barley mixed-linkage glucan, with a k_{cat} of almost 3.5x10⁴ min⁻¹ and K_M of 0.12 mg/mL (Figure 3-4, Table 3-1).



Figure 3-4: PbGH5A polysaccharide specificity (A) Michaelis-Menten plots of PbGH5A acting on various β-glucan substrates. Activities were measured as rates of reducing-end production using the BCA assay. (B) Graphical expansion of data for tXyG, kGM, CMC, HEC, PASC, bX and wAX from panel A.

PbGH5A was poorly active on the synthetic, soluble cellulose mimics CMC and HEC, perhaps reflecting detrimental interactions of the pendant groups in the active site cleft. Interestingly, PASC was a worse substrate for PbGH5A than CMC (Figure 3-4). Low activity was also measured for the hydrolysis of kGM, which suggested some tolerance of $\beta(1,4)$ -linked mannosyl residues in the polysaccharide backbone. While very poor activity was observed for PbGH5A acting on xylans, no activity was observed with either galactomannan polysaccharide or mannohexaose, confirming the glucan-specificity of the enzyme.
Sh streets	k _{cat}	К_{М, арр}	$k_{\rm cat}/K_{\rm M}$	A	
Substrate	(min ⁻¹)	(mM)	(M ⁻¹ s ⁻¹)	Assay	
bMLG	$3.49{\pm}0.05{x}10^{4}$	0.122±0.007*	4.8x10 ³ *	BCA	
tXyG	$6.8 \pm 0.4 \times 10^3$	$1.1\pm0.2*$	103*	BCA	
kGM	$3.2\pm0.2x10^{3}$	0.90±0.17*	59*	BCA	
cGM	ND	ND	ND	BCA	
CMC	1.83±0.14x103	2.3±0.4*	13*	BCA	
HEC	$2.0\pm0.2x10^{3}$	0.62±0.13*	54*	BCA	
PASC	$2.0\pm0.2x10^{3}$	10.8±1.3*	3.1*	BCA	
wAX	ND	ND	0.23 ± 0.04	BCA	
bX	ND	ND	0.23±0.03	BCA	
G-PNP	ND	ND	ND	PNP	
GG-PNP	ND	ND	4.4 ± 0.2	PNP	
GG-CNP	460±30	11.4±1.3	670	CNP	
GGG-PNP	360±25	9.2±1.2	650	PNP	
GGG-CNP	$9.1\pm0.3x10^{3}$	1.5 + 0.1	1.01×10^5	CNP	
GGGG-CNP	$8.1\pm0.4x10^3$	0.7 ± 0.1	1.94×10^{5}	CNP	
XXXG-CNP	$9.7 \pm 0.8 \times 10^3$	0.064 ± 0.013	2.53×10^{6}	CNP	
G3GGG3GGG → G3GGG+GGC	61.97±0.13x10 ⁴	0.570 ± 0.040	5.8x10 ⁵	HPAEC-PAD	
G3GGG	ND	ND	ND	HPAEC-PAD	
GG3GG → GG + GG	ND	ND	1.46±0.2x103	HPAEC-PAD	
GGG3G→GGG+Glc	ND	ND	3.13±0.4x10 ³	HPAEC-PAD	
GGG3G → GG+G3G	ND	ND	600±200	HPAEC-PAD	
GGGG → GG+GG	106±4	0.77 ± 0.05	2.3x10 ³	HPAEC-PAD	
GGGG → GGG+G	ND	ND	300±50	HPAEC-PAD	
GGGGGG → GGG+GG	$1.21 \pm 0.02 \times 10^4$	0.753±0.026	2.68x10 ⁵	HPAEC-PAD	
GGGGGGG → GGGGG+GG	$9.3 \pm 1.5 \times 10^3$	0.79±0.22	1.96x10 ⁵	HPAEC-PAD	
GGGGGGG → GGG+GGG	$6.3 \pm 1.1 \times 10^3$	0.84±0.26	1.25x10 ⁵	HPAEC-PAD	
XXXGXXXG→2xXXXG	422±10	0.037 ± 0.009	1.9x10 ⁵	HPAEC-PAD	
MMMMMM	ND	ND	ND	HPAEC-PAD	

Table 3-1: Kinetic parameters for the hydrolysis of various substrates by PbGH5A.

The pH-rate profile of PbGH5A was affected by the substrate used for its determination (Figure 3-2). The pH-rate profile of PbGH5 with XXXG-CNP gave a pH optimum of 5 with kinetic pKa values of 3.5 and 6.5; however, the pH-rate profile with MLG demonstrated different kinetic pKa values depending on the buffer used. The activity-temperature profile of PbGH5A on bMLG substrate indicates that the enzyme has limited activity enhancement above 37 °C (Figure

3-3A), hence, kinetic measurements were routinely performed at 37 °C in pH 5.5 buffer. The enzyme is stable below 45°C (Figure 3-3B), but exhibits rapid ($t_{1/2} = 15$ minutes) permanent inactivation at elevated temperatures.

3.3.2 Polysaccharide Hydrolysis Product Distributions

Analysis of the limit-digestion products was subsequently performed to determine the cleavage specificity of PbGH5A. HPAEC-PAD analysis of the initial digest of bMLG (Figure 3-5A) contained three peaks with short retention times, corresponding to primarily cellotriose and cellotetraose with a small amount of cellobiose. When allowed to run significantly longer, the limit-digestion of bMLG gave glucose, cellobiose and cellotriose (data not shown). Interestingly, a small number of peaks with longer retention times were also generated, but not further degraded in longer incubations. The major late-eluting peak was determined to be G3GGG based on retention time and standard addition; the other peaks were not identified (see Experimental Procedures, Substrates and Inhibitors section, for oligosaccharide nomenclature).

The presence of more than the canonical four peaks corresponding to XXXG, XLXG, XXLG and XLLG (ratio ~13:9:28:50 (York *et al.*, 1990, Eklöf *et al.*, 2012)) in the limit-digest of tamarind tXyG (Figure 3-5B) indicates that the enzyme is able to cut at sites other than the unbranched glucosyl residues. Indeed, MALDI-MS analysis of the digest revealed the presence of fragments with masses corresponding to XLLGX and XLG/LXG, which confirmed an alternate cleavage mode in which xylosyl-branched glucosyl units bind in the -1 and +1 subsites (active-site nomenclature according to (Davies *et al.*, 1997)).



Figure 3-5: HPAEC-PAD chromatograms of the limit-digests of bMLG (A) and tXyG (B) hydrolysed by PbGH5A. Identifiable hydrolysis products are labelled. *XLLGX and XLG/LXG have been identified by mass spectrometry, but have not been unambiguously assigned as the chromatographic peaks specified.

3.3.3 Chromogenic Substrate Kinetics

To map the negative enzyme subsites and determine their specific contributions to catalysis, we employed a series of initial-rate kinetic experiments measuring the release of the aglycone from the 2-chloro-4-nitrophenyl (CNP) and 4-nitrophenyl (*p*-nitrophenyl, PNP) β -glycosides of glucose (G), cellobiose (GG), cellotriose (GGG), cellotetraose (GGGG), and the xyloglucan heptasaccharide XXXG. Hydrolysis of G-CNP was undetectable, and only weak activity ($k_{cat}/K_M = 670 \text{ M}^{-1}\text{s}^{-1}$) was observed with GG-CNP (Figure 3-6, Table 3-1). GGG-CNP

was a significantly better substrate ($k_{cat}/K_{M} = 1.01 \times 10^{5} M^{-1} s^{-1}$), thus indicating a significant contribution to catalysis due to binding of the additional Glc residue in a -3 subsite; a similar trend was observed for the PNP congeners (Table 3-1). The specificity constant for GGGG-CNP hydrolysis ($k_{cat}/K_{M} = 1.94 \times 10^{5} M^{-1} s^{-1}$) was only 2-fold higher than that of GGG-CNP, suggesting little to no contribution from a -4 subsite. In keeping with the poorer leaving-group ability of the aglycone (Kempton and Withers, 1992, Ibatullin *et al.*, 2008), GG-PNP and GGG-PNP were hydrolysed significantly more slowly than the CNP congeners. Comparison of the kinetic constants for XXXG-CNP (Xyl₃Glc₄-CNP, composed of a GGGG backbone) with GGGG-CNP revealed a similar k_{cat} value, but a significantly (10-fold) lower K_{M} value, yielding a corresponding increase in specificity constant ($k_{cat}/K_{M} = 2.53 \times 10^{6} M^{-1} s^{-1}$, Table 3-1) for the branched substrate. However, the observation of significant substrate inhibition and deviation from classical Michaelis-Menten kinetics with XXXG-CNP (Figure 3-6) suggests that caution is warranted in interpreting the apparent positive effects of xylosyl branches in the negative subsites.



Figure 3-6: Michaelis-Menten plots for the hydrolysis of various chromogenic substrates by PbGH5A. Activities are measured as CNP or PNP release rates measured by monitoring A₄₀₅ over time. Error bars represent the fitting error of each measured rate. (A-F) Michaelis-Menten plots for the hydrolysis of cellobiose-CNP, cellotriose-CNP, cellotetraose-CNP, XXXG-CNP, cellobiose-PNP and cellotriose-PNP by PbGH5A, respectively.

3.3.4 Native Oligosaccharide Kinetics

To gain insight into the contribution of the positive subsites to substrate binding and catalysis, we determined the initial-rate kinetics of PbGH5A on cello-oligosaccharides, mixedlinkage $\beta(1,3)/\beta(1,4)$ -glucan oligosaccharides, and xyloglucan oligosaccharides, using an HPLCbased assay. No activity was observed with laminaribiose (G3G), cellobiose (GG) or cellotriose (GGG), suggesting that PbGH5A requires the occupancy of at least 4 subsites for initiation of the glycosidic bond cleavage. Indeed, cellotetraose (GGGG) was readily hydrolysed through two modes, one yielding two molecules of cellobiose (2xGG), and one yielding glucose (G) plus cellotriose (GGG); Michaelis-Menten analysis revealed that the symmetric cleavage mode was favored by a seven-fold greater k_{cat}/K_{M} value (Figure 3-7, Table 3-1). Notably, cellohexaose (GGGGGG) was degraded to GG, GGG and GGGG with similar kinetic constants to cellopentaose (GGGGG), which was exclusively converted to cellotriose (GGG) and cellobiose (GG), with a $k_{\text{cat}}/K_{\text{M}}$ value 130-fold higher than that for symmetric cleavage of cellotetraose (Figure 3-7 & Figure 3-8, Table 3-1). Exclusive isotopic labelling of the product cellotriose (GGG) in H₂¹⁸O revealed that recognition across the $-3 \rightarrow +2$ subsites was responsible for this cleavage mode (Figure 3-9). Specifically, the M+2 peak of cellobiose did not increase in relative intensity above the natural abundance, while the intensity of the M+2 peak of cellotriose indicated 74% ¹⁸O labelling (theoretical, 85%).



Figure 3-7: Michaelis-Menten plots for the hydrolysis of various model oligosaccharides by PbGH5A.
Activities are measured as the rate of increase in product peak integration using HPAEC-PAD. Error bars represent the fitting error of each measured rate. (A-C) Michaelis-Menten plots for the hydrolysis of cellooligosaccharides (GGGG, GGGGG, and GGGGGG) by PbGH5A. (D-F) Michaelis-Menten plots for the hydrolysis of mixed-linkage glucan oligosaccharides (GG3GG, GGG3G, and G3GGG3GGG) by PbGH5A.
(G) Michaelis-Menten plot for the hydrolysis of XXXGXXXG by PbGH5A.



Figure 3-8: HPAEC-PAD product analysis of the digestion of cellohexaose by PbGH5A.



Figure 3-9: Mass spectrum of the products of cellopentaose degradation by PbGH5A in H₂¹⁸O. A) Isotopic distribution of produced cellobiose. No increase in the M+2 peak indicates no ¹⁸O incorporation. B) Isotopic distribution of produced cellotriose showing significant increase in the M+2 peak intensity.

Turning our attention to mixed-linkage $\beta(1,3)/\beta(1,4)$ -glucan oligosaccharides, we observed that G3GGG was not hydrolysed, which suggested that $\beta(1,3)$ bonds are not tolerated between the first three negative subsites. In contrast, GG3GG was a competent substrate, yielding cellobiose as the only product (Figure 3-7, Table 3-1). This recapitulated the rejection of

 $\beta(1,3)$ bonds between the negative subsites, and furthermore highlighted the importance of -2 subsite binding. The specificity constant of GG3GG degradation is only ca. 1.5-fold lower than that of cellotetraose (Table 3-1), which indicated a lack of selectivity for $\beta(1,3)$ or $\beta(1,4)$ bonds in the cleavage site. Interestingly, GGG3G was hydrolysed via two modes, in which the production of cellotriose (GGG) plus glucose, via binding in the $-3 \rightarrow +1$ subsites and cleavage of the $\beta(1,3)$ linkage, was favored by a factor of 5 in k_{cat}/K_M over the production of cellobiose (GG) plus laminaribiose (G3G), via cleavage of the $\beta(1,4)$ linkage and $-2 \rightarrow +2$ binding subsites (Figure 3-7, Table 3-1). The extended mixed-linkage heptasaccharide G3GGG3GGG was hydrolysed most rapidly of all the substrates tested, with a k_{cat}/K_M exceeding that of cellopentaose or cellohexaose by 2-fold, to give only cellotriose (GGG) and G3GGG as products (Figure 3-7, Table 3-1).

3.3.5 Inhibition and Covalent Labelling with an Active Site-Directed Inhibitor

We have previously introduced *N*-bromoacetylglycosylamines and bromoketone *C*-glycoside derivatives of xyloglucan oligosaccharides as specific active-site affinity labels for *endo*-xyloglucanases (Figure 3-10) (Fenger and Brumer, 2015). Incubation of PbGH5A with the *N*-bromoacetylglycosylamine derivative of XXXG (XXXG-NHCOCH₂Br, **1**) led to rapid, timeand concentration-dependent inactivation (Figure 3-10), with a dissociation constant, K_i , 0.63 ± 0.03 mM and an irreversible inactivation constant, k_i , of 0.0364 ± 0.0006 min⁻¹ ($k_i/K_i = 0.06$ mM⁻¹·min⁻¹).

Intact protein MS after a 3 h incubation of PbGH5A with 1 at 1.4 mM and 37 °C revealed exclusive, single-labelling of the enzyme (Figure 3-11). The bromoketone *C*-glycoside isostere (XXXG-CH₂COCH₂Br, **2**) was a less potent, but nonetheless effective, inhibitor of PbGH5A, with a 3-fold lower k_i value and 2.5-fold lower K_i ($K_i = 0.27 \pm 0.07$ mM, $k_i = 0.0113 \pm 0.0008$ min⁻¹, $k_i/K_i = 0.04$ mM⁻¹·min⁻¹). Intact protein MS of PbGH5A under conditions similar to those giving essentially complete inactivation (7.3 µM PbGH5A, 1.4 mM inhibitor **2**, 3 h incubation at 37 °C) indicated near-complete labelling of the enzyme, also at 1:1 stoichiometry (Figure 3-11).



Figure 3-10: Inhibition of PbGH5 with active-site affinity labels. (A,B) Inhibition of PbGH5A with 1.
(A) Plot of reaction velocity versus time at different inhibitor concentrations. (B) Plot of pseudo first-order rate constants versus Inhibitor concentration. (C,D) Inhibition of PbGH5A with 2. (C) Plot of reaction velocity versus time at different inhibitor concentrations. (D) Plot of pseudo first-order rate constants versus Inhibitor concentrations. (D) Plot of pseudo first-order rate constants versus Inhibitor concentrations. (D) Plot of pseudo first-order rate constants versus Inhibitor concentration. (C,D) Inhibiting error of the exponential decay function. (E) Chemical structure of inhibitor 1. (F) Chemical structure of inhibitor 2.



Figure 3-11: Intact MS of PbGH5A and several mutants. A) Intact MS of wild-type PbGH5A at 7.7 μM, expected mass: 41649.0, found 41650.6. B) MS of PbGH5A at 7.7 μM and inhibitor 1 at 1.4 mM, 3h incubation 37°C. The peak at 42753.3 corresponds to the mono-labelled enzyme adduct (expected mass: 42751.7, found 42753.3). C) MS of PbGH5A at 7.4 μM and inhibitor 2 at 1.4 mM, 3h incubation at 37°C. The peak at 42751.2 corresponds to the mono-labelled enzyme (expected mass: 42751.2, found 42751.2). D) MS of the PbGH5A mutant E280A at 6.4 μM, 3h incubation at 37°C with 1.4 mM Inhibitor 1 XXXG-NHCOCH2Br. The peak at 42693.3 corresponds to mono-labelled protein. (Expected mass: 42693.7, found 42693.3)

3.3.6 Structural Characterisation of PbGH5A Variants in the *apo* Form and in Complexes with Oligosaccharides

To provide molecular-level insight into substrate recognition by PbGH5A, we determined the crystal structure of this protein to 1.65 Å resolution (PDB ID 3VDH). We also obtained high-resolution (1.6 - 1.9 Å) structures of enzyme variants with four different ligands (

Table 3-2). The complex structures of the catalytically inactive PbGH5A(E280A) sitedirected mutant with the tetradecasaccharide substrate XXXGXXXG (PDB ID 5D9M), and that of the wild-type enzyme with the covalent inhibitor XXXG-NHCOCH₂Br (**1**, PDB ID 5D9P) contained clear electron density corresponding to ligand molecules which spanned the length of the active-site cleft for both complexes (Figure 3-12). Together, these complexes provide the most complete view of enzyme-substrate interactions across the entire active site of a GH5 member to date. In the complex structures between the wild-type enzyme and heptasaccharide XXXG (PDB ID 5D9N), and between the E280A variant and the linear glucan cellotetraose GGGG (PDB ID 5D9O), the respective ligands occupied the positive subsites of the PbGH5A active site, thereby providing a unique opportunity to directly compare binding for branched and unbranched ligands in the GH5_4 subfamily.



Figure 3-12: Overall structure of PbGH5A. (A) PbGH5A in complex with inhibitor 1. Secondary structure of PbGH5 is shown in cartoon representation and colour-coded with strands in green, helices in

 blue, and loops in yellow. Two ligand molecules, XXXG-NHCOCH₂Br, are shown in ball-and-stick, in grey. The active site positive and negative subsites are indicated. (B) Asymmetric unit for the PbGH5A-E280A:XXXGXXXG complex. Secondary structure of PbGH5 is shown in cartoon representation and colour-coded blue and pink for monomer A, and green and pink for monomer B. Two ligand molecules, are shown in ball-and-stick, in grey and black.

v	PhCH5A and	DLCH5A E280A.	PhCH5A.		РЬСН5А	
enzyme		XXXGXXXG complex	XXXG complex	cellotetraose complex	XXXG- NHCOCH2Br complex	
DDB Codo	21/10/1	5D0M	5D0N	5000	5000	
r DD Coue	SVDH		JUSIN	5090	5091	
Space group	P2 ₁	$P2_12_12_1$	$P2_12_12_1$	P2 ₁	P1	
a. b. c. Å	57.91.82.25.	77.43.81.68.	75.56.84.11.	49.17.85.06.	48.19.49.02.	
в°	74.56	108.49	138.63	74.96	85.97	
r,	109.15	100115	100100	101.41	76.39, 89.98, 66.80	
Resolution. Å	50.0 - 1.60	36.5 - 1.90	25.0 - 1.86	15.0 - 1.63	31.9 - 1.80	
\mathbf{R}_{sym}^{a}	0.039 (0.443) ^b	0.048 (0.50)	0.051 (0.437)	0.060 (0.526)	0.046 (0.309)	
$I/\sigma(I)$	38.95 (4.67)	23.26 (3.88)	41.71 (5.98)	21.86 (2.09)	11.47 (2.13)	
Completeness, %	84.0 (61.6)	0.(61.6) 99.2 (98.3) 99.9 (99.7) 97.8		97.8 (95.0)	92.8 (86.5)	
Redundancy	3.4 (4.0)	7.5 (7.2)	6.2 (6.0)	4.0 (4.0)	2.0 (1.7)	
5	~ /	REFIN	EMENT			
No. of reflections:						
working, test	79400, 1951	54601, 1994	74720, 2000	85156, 4269	60096, 1850	
R-factor/free R-	15.2/19.9	17.3/20.4	13.9/16.8	15.3/18.0	16.3/20.4	
factor ^c (29.4/31.6)		(22.9/30.0) (19.4/22.2)		(22.7/24.5)	(25.9/35.0)	
No. of refined atom	s, molecules:					
Protein	5493	5499	5542	5510	5503	
Ligand	and N/A		144	90	300	
Solvent	939	728 962		987	750	
B -factors						
Protein	24.4	26.1	25.2	11.8	25.5	
Ligand	N/A	26.7	49.1	28.0	30.6	
Solvent	28.5	34.5	37.6	24.7	34.6	
RMSD						
Bond lengths, Å	0.010	0.005	0.015	0.007	0.008	
Bond angles, °	1.222	0.915	1.370	1.118	1.189	
Ramachandran Pl	ot:					
Allowed, %	89.4	96.3	95.9	96.5	96.6	
Additionally	dditionally 10.1 3.4		3.7	3.1	4.0	
allowed, %						
Disallowed, %	0.4^{d}	0.3	0.4	0.4	0.4	

Table 3-2: X-ray	diffraction data	and refinement	statistics for	PbGH5A cr	vstal structures
Table J-2. II-Iay	unnachon uata	and remember	statistics for	1 DOHOM CI	your ou ucui co

 ${}^{a}R_{sym} = \sum_{h}\sum_{i} |I_{i}(h) - \langle I(h) \rangle / \sum_{h}\sum_{i}I_{i}(h)$, where $I_{i}(h)$ and $\langle I(h) \rangle$ are the *i*th and mean measurement of the intensity of reflection *h*.

^bFigures in parentheses throughout the table indicate the values for the outer shells of the data.

 ${}^{c}R = \Sigma |F_{p}{}^{obs} - F_{p}{}^{calc}|/\Sigma F_{p}{}^{obs}$, where $F_{p}{}^{obs}$ and $F_{p}{}^{calc}$ are the observed and calculated structure factor amplitudes, respectively. ^dResidues in the disallowed region are: A116 (which comprises the active site cleft) and/or T272

3.3.6.1 Overall Structure of PbGH5A

The PbGH5A *apo* enzyme structure was determined by molecular replacement using the structure of *Paenibacillus pabuli* GH5 (PDB ID: 2JEQ) as a search model. The asymmetric unit contained two polypeptide chains corresponding to PbGH5A residues 7 to 352. According to analytical gel filtration analysis (data not shown) the PbGH5A protein predominantly exists as a monomer in solution, suggesting that intermolecular contacts observed in the crystal structure are most probably a result of crystal packing. The overall structure of PbGH5A is a (β/α)₈-barrel fold typical of the GH5 family (Figure 3-12A). Structural comparison using the Dali server (Holm and Rosenström, 2010) identified other characterised GH5 subfamily 4 enzymes as the closest structural homologues of PbGH5A. The best match was the structure of endoglucanase A from *Piromyces rhizinflata* (PDB 3AYS) (Tseng *et al.*, 2011) which superimposed with PbGH5A with an RMSD value of 1.7 Å over 323 of 357 C α positions (Table 3-3).

PDB ID	Organism	Protein name	Z score	RMSD (Å)	% ID	Activ Poc Area (Å ²) V	e Site ket Volume (Å ³)	Activity Spectrum	Ref.
3VDH	Prevotella bryantii	PbGH5A	-	-	-	600	1020	bMLG, tXyG, kGM, CMC, HEC	this work
3AYS	Piromyces rhizinflata	PrGH5	43	1.7	34	600	990	Avicel, MLG, CMC, lichenin	(Sato <i>et al.</i> , 2001)
2JEQ	Paenibacillus pabuli	PpXG5	43	1.8	33	880	1620	XyG	(Gloster <i>et al.</i> , 2007)
4W87	uncultured bacterium	XEG5A	42	1.8	34	500	740	XyG, lichenin, CMC	(dos Santos <i>et al.</i> , 2015)
3ZMR	Bacteroidetes ovatus	BoGH5	42	2.0	30	830	1630	XyG	(Larsbrink <i>et al.</i> , 2014a)
4V2X	Bacillus halodurans	BhGH5	42	1.9	28	866	537	XyG, MLG	(Venditto <i>et al.</i> , 2015)
1EDG	Clostridium cellulolyticum	Cel5A	41	1.8	34	617	943	cellulose	(Ducros <i>et al.</i> , 1995)
4W8B	uncultured bacterium	XEG5B	40	2.1	33	680	1140	XyG	(dos Santos <i>et al.</i> , 2015)

 Table 3-3: Dali search results for PbGH5A

The active site in the homologous enzymes is located in a large solvent accessible cavity formed by loop regions at the top of the barrel. As inferred from comparison of primary and tertiary structure of GH5 members, E280 and E162 are the catalytic active-site nucleophile and general acid/base, respectively, in PbGH5A (Withers *et al.*, 1992, Miao *et al.*, 1994, Yuan *et al.*, 1994). Accordingly, the mutation of E280 to alanine resulted in a >18000-fold reduction in activity (below the limit of detection) compared with the wild type on both chromogenic substrates XXXG-CNP and GGG-CNP. Hence, the catalytically inactive PbGH5A(E280A) variant was used for co-crystallisation with tetradeca- and hepta-saccharide substrates.

3.3.6.2 Non-Covalent Complexes of PbGH5A Variants with Branched and Unbranched Ligands

The PbGH5A(E280A) variant in complex with tetradecasaccharide substrate exhibited unambiguous, well-ordered electron density corresponding to a single XXXGXXXG molecule spanning the active sites of both monomers found in the asymmetric unit (Figure 3-12B); one half of the substrate occupied the positive subsites of one monomer (monomer A), while the other half localised to the negative subsites of the second monomer (monomer B). In addition, the negative subsites of monomer A featured density corresponding to the XXXG moiety of the second substrate molecule, the rest of which was apparently disordered in the solvent channel. The positive subsites of monomer B did not contain any additional electron density.

The conformation of the XXXGXXXG substrate molecule in the negative subsites of both PbGH5A(E280A) monomers was well-defined and virtually identical. Detailed analysis of interactions between PbGH5A(E280A) monomers and the substrate molecule showed that the subsite -1 of the enzyme forms by far the most direct interactions with glucosyl moiety of the substrate (Figure 3-13A). More specifically, this subsite was occupied by the α -anomer of the glucose unit, with the C1-hydroxyl forming a hydrogen bond with the sidechain of Y240 (Figure 3-14). This glucose unit is also positioned by a stacking interaction with W324, and hydrogen bonds between C2-, C3-, and C6-hydroxyls and the sidechains of N161, H112, and D288, respectively. It is interesting to note that the C1-carbon atom itself is located 3.1 Å away from the C4-hydroxyl of the xylosyl-branched glucosyl unit bound in the +1 subsite, implying that little movement would be necessary to bring the two ligands within the distance required for formation of an intact β -1,4 bond in the reverse reaction. These observations suggested that

PbGH5A(E280A):XXXGXXXG complex structure represents a good proxy for wild-type enzyme:product interactions, despite having been generated from a catalytically inactive enzyme variant.



Figure 3-13: Interactions in the PbGH5A active site. (A). Stereo view of the negative subsites of XXXGXXXG complex. The ligand in the active site is in grey, in ball-and-stick, while PbGH5 side chains, involved in binding are in yellow; water molecules are shown as red spheres. Fo-Fc electron density (3σ level) is contoured in green around the ligand. The ligand in adjacent positive subsites is shown in line representation in black. (B). Stereo view of the positive subsites of XXXGXXXG complex. The ligand is in black ball-and-stick. The ligand shown in grey lines is bound in the negative subsites of the same monomer. The rest as in panel A. (C). Stereo view of the positive subsites of XXXG complex. The same representation as

for panel A. (D) Stereo view of the positive subsites of cellotetraose complex. The same representation as for panel A.

The -2 subsite is formed by the sidechains of N28 and D288, which hydrogen bond to the C2- and C3-hydroxyls of the corresponding glucosyl residue, respectively (Figure 3-14). Binding of the -2' xylosyl unit is water-mediated, with the exception of a hydrogen bond between the C4-hydroxyl and the backbone of A117. In subsite -3, the main interaction is stacking of the glucosyl residue against W48, while in subsite -4 the stacking is against F47. These structural observations (see also the inhibitor complex below) are in line with the kinetic analysis presented above, which likewise suggest the existence of a total of four negative subsites in PbGH5A.



Figure 3-14: A schematic representation of the PbGH5A active site bound to two XXXG oligosaccharide units, based on the structure of the PbGH5A in complex with the XXXGXXXG tetradecasaccharide. Key hydrogen bonds between the enzyme and the ligands are represented by dotted lines, while key stacking interactions are represented by curved lines and the residues involved in grey. Interactions with the solvent were omitted for visual clarity.

As mentioned above, the PbGH5A(E280A):XXXGXXXG complex features one substrate molecule extending from the positive subsites of one monomer into the negative subsites of another monomer in the asymmetric unit. In the positive subsites the substrate

binding maps from the +1 subsite, in which the glucosyl moiety is closely stacked against W170 (Figure 3-13B). The glucosyl moiety forms a hydrogen bond with the sidechain of E162 via the C4-hydroxyl, while the other hydroxyls are solvent-coordinated. The xylosyl unit in the +1 subsite occupies the space between the active site cleft and the glucan backbone and forms an intricate network of hydrogen bonds with the protein, making use of all the available hydroxyls of the sugar moiety. The glucosyl unit in the +2 subsite stacks against W243, and its C2-hydroxyl forms hydrogen bonds with the sidechain of D171. The xylosyl in the +2 position is bound away from the active site cleft and completely solvent exposed, as are the ligand units in the +3 and +4 positions.

In the wild-type PbGH5A:XXXG heptasaccharide complex, clear electron density corresponding to all seven monosaccharide residues bound in the positive subsites was observed (Figure 3-13C). While the general orientation of the ligand molecule in this complex structure is similar to that observed in the PbGH5A(E280A)-XXXGXXXG complex, it nevertheless has a distinct conformation in which the glucan backbone of the substrate is rotated ~15° away from enzyme catalytic centre (Figure 3-15B). This alternative conformation of the ligand results in a more parallel binding along the bottom of the active site cleft, and is observed for both PbGH5A monomers found in the asymmetric unit. Due to this shift, a unique set of interactions is formed between the enzyme and the ligand as compared to the tetradecasaccharide complex structure (Figure 3-15B). The interactions conserved between the two structures include W170 and W243 stacking with the ligand in the +1 and +2 subsites, respectively, and the hydrogen bond between K214 and the C4-hydroxyl of the +1' xylosyl residue. The latter moiety forms four unique hydrogen bond interactions not observed in the complex with XXXGXXXG, demonstrating that while different, the recognition of XXXG is as elaborate as for the tetradecasaccharide. The xylosyl units for the rest of the ligand are completely solvent exposed. The presence of available hydrogen bonding partners in the PbGH5A active site, yet lack of conserved interactions with the xylose moieties in the two complexes, implies that ligand binding is dominated by overall accommodation of the xyloglucan polymer rather than specific recognition of the pendant groups.



Figure 3-15: Comparison of PbGH5A complexed structures. (A) Overall shape of the active site pocket and conformation of ligand binding within. Surface representation of PbGH5 complexes with ligands from the four complexes shown superimposed. Enzyme regions in direct contact with the ligand are coloured dark blue for the XXXGXXXG complex, additional interactions are in grey for XXXG complex, and in pink for cellotetraose complex. Ligands are in ball-and-stick; in black (XXXGXXXG), in pink (XXXG-NHCOCH2Br), in orange (XXXG), in light blue (cellotetraose). (B) Pairwise comparison: binding of branched ligands. Stereo view of XXXGXXXG and XXXG ligand superposition is shown in ball-and-stick, ligands colour-coded as in (A). PbGH5A residues shown for orientation are in yellow and key differences in binding indicated by grey lines. (C) Pairwise comparison: binding of branched vs. unbranched ligand. Stereo view of XXXG and GGGG ligand superposition is shown in ball-and-stick, ligands colour-coded as in (A). PbGH5A residues shown for orientation are in yellow, and key differences in binding indicated by grey lines.

As observed for the complexes with branched ligands, the PbGH5A(E280A)cellotetraose (GGGG) complex contained the ligand bound in the positive subsites, although clear electron density was present only for glucosyl residues +1 to +3 (Figure 3-13D). In contrast to the xylogluco-oligosaccharide-bound structures, the binding conformation of the cellotetraose ligand is dramatically different. In particular, the glucosyl residue in the +1 subsite in the cellotetraose complex is flipped ~180° about the C1-C4 axis vis-à-vis the XXXG complex (Figure 3-15C). In this orientation, the glucan backbone is pushed deeper into the core of the protein and results in all four glucosyl moieties of cellotetraose forming direct interactions with the protein. In the +2 subsite, the C6-hydroxyl of the glucosyl moiety occupies the space equivalent to the +1'-xylosyl of the branched ligands and forms a hydrogen bond with the sidechain of D241 and the main chain amide of Y240. In the +3 subsite, the glucosyl moiety forms direct interactions with the protein not seen in the other complexes via hydrogen bonding interactions of the C3-hydroxyl with the sidechain of D241 and the main chain amide of W243. The limited electron density of the glucosyl unit in the +4 position also suggests hydrogen bonding with the protein. The more extensive hydrogen bonding network of the unbranched glucan vis-à-vis the xylose-branched congeners parallels the observed catalytic preference of PbGH5A for bMLG, while the flip in the binding conformation of this ligand provides structural rationalisation for the lack of discrimination in hydrolysis of $\beta(1,3)$ versus $\beta(1,4)$ bonds (as discussed in greater detail below).

3.3.6.3 Active Site Affinity-Labelled Complex of PbGH5A

In keeping with the anticipated reactivity of the electrophilic affinity label XXXG-NHCOCH₂Br, the XXXG-NHCOCH₂- moiety was observed in the negative subsites, covalently bound to the sidechain oxygen of the catalytic acid-base, E162, via displacement of the bromide nucleofuge (Figure 3-16A). The specific labelling of E162 was consistent with the observation by MS of a single protein-inhibitor covalent complex of both the wild type and the E280A catalytic nucleophile mutant in solution (Figure 3-11). Likewise, labelling of the general-acid base residue in a cellulase by a homologous *N*-bromoacetyl cellobiosylamine has previously been observed (Tull *et al.*, 1996). As with the PbGH5A(E280A)-XXXGXXXG complex structure, clear electron density for the entire ligand indicates well-ordered binding. The position of the oligosaccharide moiety of the label in the negative subsites superimposes remarkably well with that of XXXGXXXG in the corresponding complex structure (Figure 3-15A). This observation confirmed that these inhibitors retain their full ability to interact with the negative subsites of PbGH5A and are thus accurate substrate mimics. The key difference between these structures results from accommodation of the inhibitor's "handle" in the -1 subsite. Here, the strictly conserved H328 orients the amide moiety through a hydrogen bond, while the active site nucleophile, E280, forms a hydrogen bond with the C2-hydroxyl of the glucose moiety (Figure 3-16A).



Figure 3-16: PbGH5A active site in complex with XXXG-NHCOCH₂Br. (A) Negative subsites. The ligand in the active site is in grey, in ball-and-stick, while PbGH5A side chains, involved in binding are in yellow; water molecules are shown as red spheres. Fo-Fc electron density (3.5σ level) is contoured in green around the ligand. (B) Positive subsites. The ligand is in black, the residues from adjacent symmetry-related monomer are in cyan. Other colouring is as in panel (A).

An unexpected second inhibitor molecule occupied the positive subsites of the enzyme *in crystallo*. This inhibitor molecule was covalently linked to M62 of a neighbouring enzyme molecule within the crystal packing, suggesting that in this orientation the terminal group of the

inhibitor was solvent exposed and suitably poised to react with the nucleophilic thioether. Multiple labelling of PbGH5A in the presence of this inhibitor in solution was not observed in solution by MS, indicating that this was a fortuitous event, prompted by the particular crystal packing. Notably, this result follows a previous observation made by Black et al., who suggested that non-specific labelling could occur at solvent-exposed methionine residues (Black *et al.*, 1993).

This covalent pinning via the *N*-acetyl moiety resulted in well-ordered binding for all four glucosyl units of the second inhibitor moiety (Figure 3-16B). The oligosaccharide portion is oriented similarly to this in the PbGH5A:XXXG and PbGH5A(E280A):XXXGXXXG complexes and is likewise distinct from the PbGH5A(E280A):cellotetraose complex. Subtle differences in the conformation of XXXG-NHCOCH₂- in the positive subsites vis-à-vis XXXG and XXXGXXXG again suggests that the recognition of the XXXG motif is plastic. Here, the hydrogen-bonding network represents a mixture of that observed for these other branched complexes: The xylosyl residue in the +1' subsite retains the hydrogen bond with K214, and forms an additional hydrogen bond with the sidechain of D241 (Figure 3-16B). Similar to the XXXG complex, the +1' xylosyl residue of the inhibitor forms a hydrogen bond with the backbone of A212, yet similar to the XXXGXXXG complex, the same moiety forms a hydrogen bond to a hydroxyl of the +2 glucosyl residue. The xylosyl unit in the +3' position is hydrogen bonded to the main chain of W243 and the sidechain of D241, reminiscent of glucose binding in the +3 subsite of the cellotetraose complex. The remainder of the interactions are entirely water mediated.

In summary, the comparative analysis of PbGH5A-ligand complex structures revealed a striking variation in ligand-protein interactions within the positive, but not the negative subsites. This, however, does not translate into conformational changes in the protein structure itself. The enzyme backbones in all four complex structures superimposed with an average RMSD of 0.3 Å for the protein C α atoms. The binding of the various oligosaccharides in approximately the same location, but with a drastically different orientation therefore points to the great versatility of PbGH5A active site with respect to accommodation of different ligands within the positive subsites (Figure 3-15A).

3.4 Discussion

The combination of detailed kinetic analysis together with new insight brought by crystallographic complexes of PbGH5A provides a unique opportunity to explore key enzyme-substrate interactions which define substrate specificity within GH5_4, and to further elucidate the roles of this subfamily in glucan catabolism.

3.4.1 PbGH5A is a Predominant Mixed-Linkage *Endo*-Glucanase, but also a Competent *Endo*-Xyloglucanase

Polysaccharide kinetics reveal that although PbGH5A is a competent *endo*-xyloglucanase (EC 3.2.1.151), with an activity ($k_{cat} = 6800 \text{ min}^{-1}$, $K_{M} = 1.1 \text{ mM}$) similar to that of the highly specific bacterial GH5_4 *endo*-xyloglucanases from *Paenibacillus pabuli* (PpXG5, $v_0/[E]_t = 8700 \text{ min}^{-1}$ at 0.5 mg/mL substrate) (Gloster *et al.*, 2007) and *Bacteroides ovatus* (BoGH5, $k_{cat} = 2.61 \times 10^4 \text{ min}^{-1}$, $K_M = 0.82 \text{ mM}$) (Larsbrink *et al.*, 2014a), it is a superior bMLG hydrolase (EC 3.2.1.73), with $k_{cat} = 3.5 \times 10^4 \text{ min}^{-1}$ and $K_M = 0.12 \text{ mg/mL}$. PbGH5A thus has significant catalytic flexibility, having an ability to tolerate branching xylosylation on β -glucan chains. This activity profile clearly distinguishes PbGH5A from the strict *endo*-xyloglucanases of GH5_4 (Gloster *et al.*, 2007, Larsbrink *et al.*, 2014a) and provides further direct evidence of the polyspecificity of this subfamily GH5_4, which also includes a number of characterised carboxymethylcellulases and mixed-linkage *endo*-glucanases (Berger *et al.*, 1989, Foong *et al.*, 1991, Palackal *et al.*, 2007).

Hence, we compared PbGH5A to other well characterised GH5_4 enzymes: *Paenibacillus pabuli* XG5 (PpXG5, PDB ID 2JEQ) (Gloster *et al.*, 2007), *Bacteroides ovatus* (BoGH5, PDB ID 3ZMR) (Larsbrink *et al.*, 2014a), *Bacillus halodurans* GH5 (BhGH5, PDB ID 4V2X) (Venditto *et al.*, 2015), Xeg5A (PDB ID 4W88) and Xeg5B (PDB ID 4W8B) (dos Santos *et al.*, 2015). These were chosen as they have been subjected to detailed structure-function characterisation and have been specifically tested for both mixed-linkage *endo*-glucanase and *endo*-xyloglucanase activities. Of these, BhGH5 and Xeg5A, like PbGH5, accept both bMLG and xyloglucan substrates, while PpXG5, BoGH5 and Xeg5B are highly specific to xyloglucan.

Examination of the overall shape of the active site of these enzymes reveals a substantially shallower and narrower cleft of the predominant mixed-linkage *endo*- β -glucanases versus the predominant *endo*-xyloglucanases (Figure 3-17A). Quantitation of this difference using the CASTp server indicated that both the active site surface area and volume are greater by

approximately one-third for the former enzymes (Table 3-3, (Dundas *et al.*, 2006)). The contrast is particularly dramatic at the catalytic centre (the region surrounding subsites -1 and +1), where PbGH5, BhGH5 and Xeg5A possess a shallow cleft with a constriction in the middle, while BoGH5, PpXG5 and Xeg5B are more open with extra space visible both at the top and the bottom of the catalytic site.



Figure 3-17: Comparison of GH5_4 structural homologs. (A) Overall shape of the active site pocket. The active sites are shown as a semi-transparent blue surface representation for six structures: three MLG active enzymes PbGH5, Xeg5A (PDB ID: 4W88), *Bh*GH5 (PDB ID: 4V2X); and three XyG-specific enzymes, Xeg5B (PDB ID: 4W8B), *Pp*XG5 (PDB ID: 2JEQ), and *Bo*GH5 (PDB ID: 3ZMR). Ligands, if present, are shown in cyan ball-and-stick representation. High-lighted in red are the two catalytic glutamate residues present in all of the compared structures. High-lighted in green are two regions that contribute the most to the differences in the active site shape between the compared structures: the narrowing at the top of the -1

subsite in MLG-active enzymes (absent in XyG-specific enzymes), and the presence of a bulky aromatic residue making up the binding platform for the -2' xylose in XyG (absent in MLG-active enzymes). For emphasis a white circle contours the binding surface available at the -1 and -2 subsites of the XyG-specific enzymes and points to the lack thereof for the MLG-active enzymes. (B). Superposition of the MLG-active enzymes from panel A. For clarity, only the secondary structure of PbGH5 is shown. The loops making up the active site are shown in cyan (top four loops) and blue (bottom three loops). In ball-and-stick are the residues responsible for the unique shape of the active site: top acidic residues narrowing the -1 subsite and the bottom His residue forming the -2' subsite. PbGH5 residues are in green, Xeg5A in grey, and BhGH5 in

wheat. For general orientation, the XXXGXXXG ligand in PbGH5A structure is shown in line representation. (C) Comparison between PbGH5 and XyG-specific enzymes. The representation is the same as in panel B. Distinct residues in the -1 and -2'-subsites are in the following colour-code: PbGH5 – green, *Bo*GH5 – orange, Xeg5B – violet, *Pp*GH5 – pink.

Seven loop regions combine to form this distinct shape of the PbGH5 active site: four at the top (residues 27-48; 238-262; 280-295; 324-339), and three at the bottom (113-121, 152-165, and 210-214) (Figure 3-17). There is a great variability in the overall conformation of these loops compared to the other discussed GH5_4 enzymes, however, key features are conserved. The conserved regions encompass functionally equivalent residues participating in key protein-ligand interactions, which are generally found at the base of the loops and include catalytic residues: E162 and E280; stacking residues: W48, Y240, W243 and W324, and hydrogen bonding partners: H112 and N161.

The constriction at the top of the catalytic centre in PbGH5 is mainly formed by the loop residues 280-295, of which D288 forms a direct interaction with the glucosyl moiety bound in the -1 subsite. This feature, which is conserved in the bMLG-specific enzymes (Figure 3-16B), is absent in the xyloglucanases (Figure 3-16C). At the bottom of the active centre, the shallow pocket of PbGH5 is formed, in part, by a well-conserved histidine residue, H113. Here, the additional depth of the xyloglucan-specific enzymes is due to the presence of a bulky aromatic residue found in the -2 subsite and responsible for stacking of the -2' xylosyl (Figure 3-17A,C). The distinct conservation in this region has been observed previously and reported as a potential signature motif (Gloster *et al.*, 2007). The active site pocket widens beyond the -1 and +1 subsites, yet while the branching residues of xyloglucan saccharides can be accommodated by the protein here, there are no obvious pockets that appear to be specifically tailored for this purpose. In the negative subsites, it is the glucan backbone that is intricately bound via stacking

and hydrogen-bonding interactions, while the majority of the xylosyl moieties are solvent exposed. This is distinct from the specific *endo*-xyloglucanases of GH5_4, in which aromatic residues have been identified to provide binding platforms for -2' and -3' xylosyl residues (Gloster *et al.*, 2007) (Figure 3-17B,C). Likewise in the positive subsites, PbGH5 appears to accommodate, rather than specifically harness, branched oligosaccharide in an open cleft. It is particularly striking that the glucan backbones of cellotetraose (GGGG) and its triple-xylosylated congener XXXG are bound with different trajectories through the positive subsite region (Figure 3-15), which again implies significant flexibility in substrate binding.

In this context, it is notable that PbGH5A hydrolyses xyloglucans at non-canonical backbone cleavage sites. Of the GH5 endo-xyloglucanases characterised to-date, all cleave the dicot xyloglucan polysaccharide (exemplified by Tamarindus indica xyloglucan) at the unbranched backbone glucosyl unit (Figure 3-1) to generate oligosaccharides based on a Glc₄ backbone (Gloster et al., 2007, Larsbrink et al., 2014a, dos Santos et al., 2015). This cleavage pattern is also typical for GH7, GH9, GH12, GH16 members, with known exceptions of certain GH44 and GH74 members (Desmet et al., 2007, Gloster et al., 2007, Zhou et al., 2007, Gilbert et al., 2008, Ariza et al., 2011, Eklöf et al., 2012, Eklöf et al., 2013, Ravachol et al., 2014). Although the heptasaccharide XXXG was not hydrolysed in the presence of high enzyme concentrations (0.1 mg/mL of PbGH5A), the limit-digestion products of tamarind xyloglucan hydrolysis contained oligosaccharides consistent with cleavage via binding "X" ([Xylα(1,6)]Glc-) units at subsite -1 (Figure 3-15). Initial-rate kinetic analysis of the hydrolysis of the tetradecasaccharide XXXGXXXG revealed that cleavage of this substrate at the internal, unbranched glucosyl residue predominated, although slow ($k_{cat} = 422 \text{ min}^{-1}$, $K_M = 32 \mu M$, Table 3-1). Yet, analysis of the limit-digest (data not shown) showed alternative cleavage modes resulting in the formation of XXG and XXXGX. Taken together, the data indicate that glucan chain branching is generally not well-tolerated at the cleavage site due to constriction at subsites -1/+1 (Figure 3-17A), although an overall lack of specificity for xyloglucan motifs allows variable substrate positioning in the active-site cleft.

3.4.2 The Active Site of PbGH5A Comprises Seven Subsites in Total

Mapping the PbGH5A active site using chromogenic and native substrates, together with crystallographic analysis of enzyme-oligosaccharide complexes, suggests the presence of seven well-defined subsites – four negative subsites and three positive subsites – in an open active site

cleft. Indeed, the highest activity was observed for a mixed-linkage heptasaccharide, G3GGG3GGG (closely followed by cellopentaose and cellohexaose), whereas unbranched tetrasaccharides represent the smallest competent naturally occurring substrates for PbGH5 (Table 3-1). The mode of hydrolysis of the minimal substrate cellotetraose defined the smallest subset of subsites utilised for activity on linear glucans, with the $-2 \rightarrow +2$ binding/hydrolysis mode significantly favored over $-3 \rightarrow +1$. When two positive subsites are occupied, the importance of the -3 subsite contribution is emphasised by the 130-fold increase in k_{cat}/K_M value of $-3 \rightarrow +2$ the binding/hydrolysis mode for cellopentaose vs. the $-2 \rightarrow +2$ mode (Table 3-1). An essentially identical increase in k_{cat}/K_M values for release of the aglycones from GGG-CNP vs. GG-CNP and GGG-PNP vs. GG-PNP was observed. Collectively, these data indicate that binding in the -3 subsite of PbGH5A contributes a $\Delta\Delta G$ of -12 kJ/mol to catalysis.

Comparison of the $-3 \rightarrow +1$ binding/hydrolysis mode for cellotetraose with the $-3 \rightarrow +2$ binding/hydrolysis mode for cellopentaose reveals that binding in the +2 subsite contributes -17 kJ/mol to catalysis. As such, interactions in the +2 subsite (stacking with W243 and hydrogen bonding with D171) make a significantly greater contribution to catalysis than the interactions in the -3 subsite (stacking with W48).

Moving beyond the five core subsites spanning $-3 \rightarrow +2$, crystallographic complexes provide compelling evidence for an additional negative subsite which may explain the slight kinetic enhancement observed for the catalysis of GGGGGG \rightarrow GG+GGGG over GGGGGG \rightarrow GGG + GGG (Figure 3-15, Table 3-1). Specifically, the non-covalent PbGH5A(E280A):XXXGXXXG structure (Figure 3-13A) and the affinity-labelled PbGH5A:XXXG-NHCOCH₂- structure (Figure 3-12A) reveal a glucose-phenylalanine stacking interaction constituting subsite -4 (Figure 3-13A). On the other hand, differential binding of XXXG and GGGG ligands in the positive subsites makes clear definition of an additional positive subsite, +3, difficult. The apparent length of the active-site cleft (Figure 3-15) and wellordered electron density for the +3 glucose residue in all ligands (Figure 3-13B-D) implies that a binding surface may exist, although the breadth of the cleft at this point is not sufficient to restrict the backbones of all ligands to lie on the same trajectory. The absence of a +4 subsite is less ambiguous (Figure 3-15). Unfortunately, a lack of a sufficient diversity of higher oligosaccharide substrates precludes detailed kinetic dissection of these more distal subsites; yet the observation that the heptasaccharide G3GGG3GGG is cleaved exclusively through a $-4 \rightarrow +3$ binding/hydrolysis mode (Table 3-1) strongly supports the definition of seven subsites (Figure 3-14).

3.4.3 PbGH5A Exhibits Subtle Discrimination of β-Glucan Linkage Regiochemistry in the Active Site

Kinetic analysis of the hydrolysis of mixed-linkage oligoglucosides revealed that PbGH5 was essentially equally competent at hydrolysing $\beta(1,3)$ and $\beta(1,4)$ linkages at the catalytic centre, but demonstrated differential preference for these linkages in the positive and negative subsites: The tetrasaccharides GGGG (cellotetraose), GG3GG, and GGG3G are all hydrolysed with similar k_{cat}/K_m values for the $-2 \rightarrow +2$ binding/hydrolysis mode, yet G3GGG was not cleaved by PbGH5 (Table 3-1).

The k_{cat}/K_m value of GGGG is only 1.5-fold greater than that of GG3GG (-2 \rightarrow +2 binding/hydrolysis mode). This effective lack of linkage specificity can be rationalised in light of the oligosaccharide orientation in the positive subsites of the GGGG and XXXG heptasaccharide ligand complexes. The dramatic difference in the binding mode of these two ligands results in the close superposition of the C3-hydroxyl of cellotetraose and the C4-hydroxyl of XXXG due to a 180^o rotation of the glucan backbone (Figure 3-14C). Assuming that these structures reflect both the EP (enzyme:product) and corresponding ES (Michaelis) complexes, both C3- and C4-hydroxyl moieties can be suitably positioned as nucleofuges at the catalytic centre. Further, the apparent breadth of the active-site toward the positive subsites readily accommodates the different binding orientations distal to the catalytic centre required for longer β -glucan substrates (Figure 3-14). Here, a substantial number of ordered water molecules are present (not shown), which can potentially be variably displaced during the binding of alternative substrates.

Further reflecting an ambivalence to linkage regiochemistry at the catalytic centre, both GGGG and GGG3G were cleaved via the $-3 \rightarrow +1$ binding/hydrolysis mode to produce cellotriose (GGG) and glucose (G). However, the loss of +2 subsite binding and gain of -3 subsite binding had a large negative effect on the k_{cat}/K_M value of GGGG (7.7-fold lower than that of the $-2 \rightarrow +2$ mode). In contrast, the k_{cat}/K_M value for GGG3G hydrolysed via the $-3 \rightarrow +1$ mode is increased 5.2-fold vs. the $-2 \rightarrow +2$ mode. The results highlight the delicate balance between the contributions of subsite binding and glycosidic bond specificity to catalysis. Although it is difficult to fully disentangle these competing effects given the available kinetic and structural data, it is clear that +2 subsite binding is particularly important for catalysis of all- $\beta(1,4)$ linked

substrates: based on k_{cat}/K_M values (Table 3-1), cellopentaose is hydrolysed nearly 900-fold better in the $-3 \rightarrow +2$ mode (the exclusive hydrolysis mode) than cellotetraose is hydrolysed in the $-3 \rightarrow +1$ mode (which, again, is 7.7-fold poorer than in the $-2 \rightarrow +2$ mode).

The observation that GGG3G is efficiently hydrolysed to cellobiose (GG) and laminaribiose (G3G) indicates that $\beta(1,3)$ glucosidic bonds are tolerated between the +1 and +2 subsites. Comparison with the k_{cat}/K_M value for the $-2 \rightarrow +2$ mode of hydrolysis of cellotetraose (GGGG), indicates that $\beta(1,3)$ bonds are slightly disfavoured in this position by a factor of 4 (Table 3-1), although this equates to less than 2 kJ/mol of lost transition-state stabilisation. The recognition of $\beta(1,3)$ linkages between subsites +1 and +2 is likely to be responsible for the generation of G3GGG in the limit-digest of barley bMLG (Figure 3-5). The complexes of PbGH5 with GGGG and XXXG in the positive subsites suggests that the presence of a $\beta(1,3)$ linkage between the +1 and +2 subsites would necessarily cause the saccharide chain to adopt a different conformation, possibly disrupting the +2 hydrogen bonding interaction with D17, but stacking with W243 in subsite +2 would be anticipated to remain, due to the plasticity of this interaction (Figure 3-15).

Turning to the negative subsites, binding in subsite -2 is essential for catalysis; no substrates, including Glc-PNP, were hydrolysed to release glucose via $-1 \rightarrow +n$ modes (Table 3-1). Notably, kinetic analyses revealed that $\beta(1,3)$ linkages are not tolerated between three of four negative subsites. In particular, G3GGG is not hydrolysed through possible $-1 \rightarrow +3, -2 \rightarrow +2,$ nor $-3 \rightarrow +1$ modes (Table 3-1). The lack of $-2 \rightarrow +2$ and $-3 \rightarrow +1$ activity vis-à-vis the three other mixed-linkage tetrasaccharides provides clear evidence that $\beta(1,3)$ linkages are not accepted between subsites -2 and -1, as well as -3 and -2. Furthermore, GG3GG is not hydrolysed via the - $3 \rightarrow +1$ mode, unlike GGGG and GGG3G, which also indicates intolerance of $\beta(1,3)$ linkages between subsites -2 and -1. Similarly, the heptasaccharide G3GGG3GGG is only cleaved at the internal $\beta(1,3)$ glycosidic bond. The two $\beta(1,3)$ linkages prevent productive binding and cleavage at the four possible $\beta(1,3)$ glycosidic bonds, while the non-reducing-end $\beta(1,3)$ linkage is tolerated in subsite -4. The inability of PbGH5A to accept $\beta(1,3)$ bonds in the negative subsites is partially substantiated by the structures of complexes with xyloglucan oligosaccharides bound in these subsites. As discussed above, the xylosyl residues of these XXXG-based ligands are mostly solvent exposed, such that the observed binding of the backbone (Figure 3-3A and Figure 3-13A) might be anticipated to closely approximate that of the unbranched cellotetraosyl unit (GGGG).

In the -1 subsite, the enzyme forms intimate contacts with each ligand, with the C1-hydroxyl hydrogen bonding to the catalytic acid/base E162, and the C3-hydroxyl directly interacting with conserved residues H112 and H113. As such, accommodating a $\beta(1,3)$ link to the -2 subsite would break this interaction and require a major change in substrate orientation, likely altering the position of the scissile bond relative to the catalytic centre. Beyond the -2 subsite, the active-site cleft widens significantly, such that there are no obvious steric factors that would hinder substrate binding in this region. Binding of $\beta(1,3)$ -linked glucose across subsites -3 and -2 may be disfavored because the resulting kink in the glucan backbone could disrupt key stacking interactions with W48 and F47, which are the main contributors to the well-ordered ligand binding seen in the -3 and -4 subsites, respectively. Regardless, the presence of a $\beta(1,3)$ linkage between subsites -4 and -3 would appear to be structurally accommodated, as underscored by the superior kinetics of G3GGG3GGG (Table 3-1)

3.4.4 Implications for Specificity Prediction in GH5 Subfamily 4

Subfamily 4 is one of the largest GH5 subfamilies, which resulted from the merger of the previous cellulase subfamilies A3 and A4 (Aspeborg *et al.*, 2012). To explore the possibility of delineating the known "cellulase", mixed-linkage *endo*-glucanase, and *endo*-xyloglucanase activities within specific clades, we performed a new phylogenetic analysis of GH5_4 using all sequences in the public CAZy Database. Bootstrap analysis revealed several well-defined clades, however, *endo*-glucanase and *endo*-xyloglucanase activities were not absolutely segregated.

A lack of systematic enzymological data further hampers efforts to delineate specificities by phylogeny. While a generally low coverage of biochemical characterisation is a ubiquitous problem for all GH families, a further significant issue arises from the use of carboxymethylcellulose (CMC) as a proxy to measure "cellulase" activity. As the present reanalysis of PbGH5A activity shows (Table 3-1), the original use of CMC as a substrate to characterize this enzyme was misleading (Gardner *et al.*, 1997); in fact, the amorphous, phosphoric acid-swollen cellulose is an even poorer substrate for PbGH5A. Analogously, it is therefore unclear how many of the 56 GH5_4 members currently assigned as cellulases or *endo*- $\beta(1,4)$ -glucanases, often solely on the basis of activity toward this unnatural, anionic, polysaccharide derivative, have been incorrectly annotated. When assaying new GH5_4 members, a wider panel of soluble polysaccharide substrates must be tested, and more detailed re-evaluation of currently characterised members is certainly warranted. More broadly, it could be argued that CMC should be abandoned as a substrate altogether.

Regardless, a growing body of data suggests that GH5_4 members are more likely to be active on the amorphous cross-linking glycans of the composite plant cell wall, rather than on the para-crystalline cellulose component. Testing this hypothesis will require further characterisation of this large and historically significant subfamily via structure-function analyses that are at the same time systematic and deep. As our work here shows, such endeavours are likely to be fruitful in uncovering unanticipated specificities, thereby increasing the library of biocatalysts for potential applications.

Chapter 4: A Complex *Prevotella* Beta-Mannan Utilisation Locus in the Rumen and Human

4.1 Introduction

Gut microbiota support the metabolism of host organisms by providing the genetic diversity and plasticity necessary to make use of the various nutrient sources present within the biosphere (Cantarel *et al.*, 2012, Flint *et al.*, 2012). The recognition and saccharification of polysaccharides prior to fermentation into fatty acids is a key function of the gut microbiota. (Cummings and Macfarlane, 1997, Scott *et al.*, 2008). Among the dominant bacterial phyla responsible for this activity in humans are Gram-negative Bacteroidetes (Qin *et al.*, 2010). Within the human gut, *Prevotella* and *Bacteroides* are the major Bacteroidetes genera. A detailed survey of the relationship between food intake and the abundance of different Bacteroidetes species revealed a positive correlation between carbohydrate-rich diets and the abundance of *Prevotella* have been further associated with improved blood glucose regulation in individuals fed a dietary-fibre supplement (Kovatcheva-Datchary *et al.*, 2015).

In spite of their apparent importance, *Prevotella* in the mammalian gut remain significantly under-examined. *Prevotella* have proven challenging to manipulate *in vivo* (Accetto and Avguštin, 2007, Accetto and Avguštin, 2011) and, to date, only five species (*P. bryantii*, *P. ruminicola*, *P. oris*, *P. copri*, and *P. albensis*) have been isolated from the gut (Accetto and Avguštin, 2015). Most *Prevotella* isolates have been derived from oral or urogenital samples and have been assessed for pathogenicity (Brook, 2002, Charalampakis *et al.*, 2013). Given the ability to culture *P. bryantii* in the lab, it is a good model species in which to develop our understanding of carbohydrate catabolism by *Prevotella* in the gut.

Collectively, non-starch polysaccharides (i.e. dietary fibre) such as β -glucans, xylans, pectins, and β -mannans, are a formidably complex target for enzymatic degradation by gut bacteria. The wide range of monosaccharide moieties, linkage types, and branching patterns identified in these families of substrates necessitate the co-expression and assembly of complex molecular systems to sense, associate with, degrade, import, saccharify, and ferment these polysaccharides (Larsbrink *et al.*, 2014a, Martens *et al.*, 2014, Ndeh *et al.*, 2017). Within gut microbiomes, elements of such systems are often found genetically co-localised in

polysaccharide utilisation loci (PULs) (Martens *et al.*, 2009, Accetto and Avguštin, 2011). Though it was originally investigated as a source of cellulases, a variety of non-starch polysaccharide-degrading genes have since been identified within *P. bryantii* B₁4 (Matsushita *et al.*, 1990, Dodd *et al.*, 2010a, Yoshida *et al.*, 2011, Accetto and Avguštin, 2015).

Sequencing of the *P. bryantii* B₁4 genome revealed an impressive array of genes encoding carbohydrate active enzymes (Purushe *et al.*, 2010). Following the CAZy classification system, the *P. bryantii* B₁4 genome encodes at least 10 carbohydrate binding modules (CBMs), 19 carbohydrate esterases (CEs), 107 glycoside hydrolases (GHs), 53 glycosyl transferases (GTs), and 14 polysaccharide lyases (PLs). This carbohydrate-active enzyme (CAZyme) content is higher than the average gut Bacteroidetes species (Kaoutari *et al.*, 2013). Furthermore, the genome encodes at least 12 glycan-targeted SusC-SusD homologue pairs, the hallmarks of PULs (Accetto and Avguštin, 2015). This combination of a TonB-dependent transporter (TBDT, SusC homologue) and surface glycan binding protein (SGBP-A, SusD homologue) is responsible for the selective import of oligosaccharides produced by glycanases anchored to the cell surface (Shipman *et al.*, 2000, Neugebauer *et al.*, 2005). Only the xylan-degrading locus from *P. bryantii* has been characterised to date (Dodd *et al.*, 2010b).

Recently, we described the structure and function of a mixed-linkage glucanase (PbGH5A) from *P. bryantii* B₁4 (McGregor *et al.*, 2016). This enzyme was originally identified by Gardner *et al.* as a carboxymethylcellulase (Gardner *et al.*, 1997). Within the genome of *P. bryantii* B₁4, PbGH5A is found fused with a putative glycoside hydrolase family 26 (GH26) β -mannanase. Furthermore, the fused enzyme (PbGH26A-GH5A) is co-localised with a SusC-SusD homologue pair indicating the presence of a PUL (Figure 4-1). The fusion of PbGH5A with a putative *endo*- β -mannanase led us to hypothesize that glucomannans are the target substrate for this locus. Glucomannans are complex polysaccharides found in high levels in secondary plant cell walls (Ebringerová, 2005). The simplest glucomannan is a storage polysaccharide found in konjac root (Kato and Matsuda, 1969). It consists of stretches of, on average, two $\beta(1,4)$ -linked glucose resides interspersed among stretches of, on average, three $\beta(1,4)$ -linked mannose residues (Katsuraya *et al.*, 2003). Galactoglucomannans are further substituted with additional $\alpha(1,6)$ -linked galactose residues. Taken together, the sensing, import, and degradation of materials containing any or all of these structural elements requires significant molecular machinery.



Figure 4-1: The *Prevotella bryantii* MUL and two model β-mannan substrates. A) Core structural motifs of the konjac glucomannan and carob galactomannan (also known as locust bean gum) polysaccharides are shown. For clarity, glucose is shown in red, mannose in blue, galactose in green, and

acetate in pink. B) The genetic organisation of the *Prevotella bryantii* MUL is shown including manual annotation. Loci containing homologous SusC-SusD homologue pairs are also shown. Predicted transporters are shown in pink, predicted glycan binding proteins are shown in purple, predicted regulators are shown in yellow, predicted glycoside hydrolases are shown in blue, predicted epimerases are shown in cyan, predicted phosphorylases are shown in green and predicted esterases are shown in orange. Gene of unknown function are shown in black, while genes with predicted functions not related to carbohydrate-degradation are shown in white. Homologous SusC-SusD pairs (reciprocal best hits) are connected by a light pink band and the level of amino acid sequence identity to the *P. bryantii* genes is given as a percentage. Homology between "SEMP" cluster genes is also displayed in this way.

To provide insight into the target, organisation, and function of this PUL, we have expressed and characterised several components of this locus including the GH26 domain of PBR_0368 (PbGH26A), the surface glycan-binding proteins (SGBPs) associated with the colocalised SusC-SusD homologue pair, a 2-epimerase, and a GH130 phosphorylase. Molecular recognition has been explored using a variety of model polysaccharides and oligosaccharides. Xray crystallographic structures are presented for PbGH26A, the PbGH26A-GH5A fusion, and PbSGBP-B allowing for molecular rationalisation of their activities. We further explore the distribution of similar PULs throughout the Bacteroidetes phylum based on the identification of homologous SusC-SusD pairs and other, more tightly conserved, identifiable elements.

4.2 Materials and Methods

All buffers and reagents were purchased from Sigma Aldrich (St. Louis, MO) unless otherwise stated.

4.2.1 Analytical Methods

4.2.1.1 HPAEC-PAD Carbohydrate Analysis

High-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) was performed as described in (McGregor *et al.*, 2017b). The gradients used were as follows: Gradient A: 2% 1 M NaOH isocratic. Gradient B: 6% 1 M NaOH isocratic. Gradient C: 0–4 min - 10% 1 M NaOH, 0% 1 M NaOAc, 4–12 min - 10% 1 M NaOH, 0–30% 1 M NaOAc linear gradient; 12–12.1 min 50% 1 M NaOH; 50% 1 M NaOAc; 12.1–13 min return to initial conditions (exponential profile 9); 13–17 min, initial conditions.

4.2.1.2 Mass Spectrometry

Intact protein masses were determined on a Waters Xevo Q-TOF with a nanoACQUITY UPLC system, according to the method described by Sundqvist *et al.* (Sundqvist *et al.*, 2007). Carbohydrate LC-MS was performed using a 0.32 mm Hypercarb Kappa column as described in (McGregor *et al.*, 2017b). The column was eluted at 8 μ L/min and 30°C. β -mannooligosaccharides were separated using the following gradient: 0.0–5.0 min, 100% A (5% MeCN, 90% ddH₂O, 5% 200 mM NH₄COO, lowered to pH 5 with formic acid), 0% B (95% MeCN, 5% 200 mM NH₄COO pH 5); 5.0–15.0 min, linear gradient to 25% B; 15.0–15.1 min, linear gradient back to 0% B; 15.1–20.0 min, equilibration with 100% A.

4.2.2 Substrates and Ligands

Oligosaccharides and their derivatives are abbreviated using a general shorthand in which M represents $\beta(1,4)$ -linked D-mannopyranose and G represents $\beta(1,4)$ -linked D-glucopyranose. Phosphoric acid swollen cellulose (PASC) was prepared according to (Zhang *et al.*, 2006).

4.2.2.1 Commercial Substrates

High purity (>94%) mixed-linkage glucan (beta-glucan (barley; high viscosity)) (MLG), carboxymethylcellulose (CMC), konjac glucomannan (kGM), carob galactomannan (cGM), tamarind xyloglucan (tXyG), wheat arabinoxylan (wAX), beechwood xylan (bX), insoluble mannan (inM, product P-MANCB), mannobiose (MM), mannotriose (MMM), mannotetraose (MMMM), mannopentaose (MMMMM), and mannohexaose (MMMMMM) were purchased from Megazyme International (Ireland) and used for all activity measurements and HPAEC-PAD experiments. Acetylated xylan was purchased from Cambridge Glycosciences (Cambridge, UK).

4.2.2.2 Acetylated Oligosaccharides

Acetylated glucomannan oligosaccharides (kGMOs) were prepared from konjac glucomannan using following. 50 mg of kGM was wetted with 200 μ L of 95% ethanol prior to the addition of 9 mL of distilled water. The solution was boiled for 2 minutes to remove ethanol and dissolve the kGM. The clear, colourless solution was buffered with 1 mL of 100 mM NH4OAc reduced to pH 6.0 with acetic acid, diluted to a final volume of 10 mL, and cooled to 37°C. PbGH26A-GH5A was added to a final concentration of 20 μ g/mL and the reaction was incubated at 37°C for 2 hours. Enzyme was denatured by heating to 80°C for 5 minutes and the product was lyophilised to give 49.5 mg of white foam.
Acetylated xylan oligosaccharides (XOs) were prepared from acetylated xylan using CjXyn10A (Xylanase 10A, NZYtech, Lisbon). 2.5 mg of acetylated xylan was dissolved in 0.5 mL of 10 mM NH₄OAc reduced to pH 6.5 with acetic acid. Enzyme was added to a final concentration of 1.5 μ g/mL and incubated, denatured, and lyophilised as above.

4.2.3 Enzyme Cloning and Expression

PBR_0369 (PbGH26B), the N-terminal gene fragment of PBR_0368 (residues 24 to 424, PbGH26A), full-length PBR_0368 (residues 24-776, PbGH26A-GH5A), PBR_0364 (PbSGBP-A), PBR_0365 (PbSGBP-B), PBR_0366 (PbSGBP-C), PBR_0359 (PbEpiA), PBR_0357 (PbGH130A), PBR_0351 (PbGH5B), PBR_0352 (PbGH3A, i.e. CdxA), PBR_0353 (PbGH36A), PBR 0360 (PbTFAraC) and PBR 0356 (PbGH26C) were amplified from P. bryantii B14 genomic DNA obtained from DSMZ (https://www.dsmz.de/) with nucleotides encoding signal peptides, predicted by SignalP (Petersen et al., 2011), removed. Primers used for amplification can be found in Table B-1. Notably, a BLAST search for homologues of PBR_0369 suggested that the start site may have been mis-called in the original annotation, missing 15 out of the 18 codons encoding a signal peptide; the primers for both the N-terminus called by Purushe et al. and the N-terminus without the signal peptide are given. Gene fragments were cloned into the p15Tv-L vector (Addgene plasmid # 26093) using Gibson assembly (Eschenfeldt et al., 2009, Gibson et al., 2009) to add an N-terminal His6-tagged fusion with a tobacco etch virus (TEV) protease cleavage site between the tag and the enzyme. A greenfluorescent protein (GFP) tag was cloned in between the His6-tag and TEV protein cleavage site to solubilize PbGH26B. Proteins were expressed in E. coli BL21(DE3) Hi-Control as described in McGregor *et al.* (2017b) with the following exceptions: size-exclusion chromatography was performed using an XK 16/100 column (GE Life Sciences) packed with Superdex 200 (GE Life Sciences) run with SEC buffer at 1 mL/min, DTT was not added to any buffer, and TEV protease treatment was performed at 22°C to improve cleavage efficiency. Protein molecular weights were estimated from SEC elution volumes. The void volume was determined using blue dextran (2,000 kDa) and the column was calibrated using the Gel Filtration Standard (Bio-Rad).

4.2.3.1 Mutagenesis

Mutagenesis was performed using the Q5 mutagenesis kit (New England Biolabs, https://www.neb.ca/). Primers used for amplification can be found in Table B-1. Mutants were

confirmed by Sanger sequencing (Genewiz, https://www.genewiz.com/) and expressed, purified, and assayed in the same manner as the wild-type proteins.

4.2.4 Enzyme Kinetics and Product Analysis

4.2.4.1 Polysaccharide Hydrolysis

Polysaccharide hydrolysis was quantified using the BCA assay (Arnal *et al.*, 2017). The pH optimum of the enzyme was initially determined using the BCA assay to quantify reducing ends released over 40 minutes of incubation of 1 nM enzyme with 1 mg/mL cGM at 37°C using 50 mM citrate (pH 2.0-6.0), 50 mM phosphate (pH 6.0-8.0) or 50 mM glycine (pH 8.5-10.0) buffers. The temperature optimum was determined in 50 mM pH 5.5 sodium citrate buffer using 1 mg/mL cGM as substrate with 1 nM enzyme. The reaction was mixed at 4 °C and incubated at temperatures ranging from 25 to 60 °C for 40 min before reducing ends were quantified using the BCA assay.

Limit-digest products were identified by treatment of 1 mg/mL polysaccharide samples with 2 μ M PbGH26A overnight in 50 mM NaOAc pH 5.5. Samples were diluted 10-fold and 10 μ L of the reaction was then analysed by HPAEC-PAD. The cGM digest was separated using gradient C, the kGM digest was separated using gradient A, and the inM digest was separated using gradient B. Products were identified by comparison to a series of β -mannooligosaccharide standards. Prior to digestion, 10 mg of inM was solubilised with 1 mL of 2 M NaOH. Immediately following neutralisation with 5 M acetic acid (to pH ~5) and dilution to 10 mL with water, 2 μ M PbGH26A was added and incubated overnight at 37°C. No precipitation was observed following neutralisation or enzyme addition.

4.2.4.2 Native Oligosaccharide Hydrolysis

HPLC-based kinetic assays were run as described in McGregor *et al.* (2017a) using β mannooligosaccharide substrates ranging in length from mannotriose to mannohexaose. Product and substrate were separated using gradient B. An 8-point linear calibration series from 0.4-100 μ M was run for each product quantified. The Michaelis-Menten equation was fit to observed rates using *OriginPro* graphing software (Origin Lab).

To determine the bond specificity of mannohexaose hydrolysis, ¹⁸O incorporation from $H_2^{18}O$ was determined by mass spectrometry (Schagerlöf *et al.*, 2009). The method was

performed as per McGregor *et al.* (2016) using 1 μ L of 1 mg/mL PbGH26A and 1 μ L of 10 mM mannohexaose in 18 μ L of H₂¹⁸O.

4.2.4.3 Activities of PbEpiA and PbGH130A

The ability of PbEpiA to epimerize mannose, glucose, galactose, mannobiose, cellobiose, lactose, mannotriose, and cellotriose was assessed by incubating enzyme (1 μ g/mL final) with carbohydrate (0.1 mg/mL final) in 500 μ L of pH 7.5, 50 mM NaPi buffer for 1 hour at room temperature. Following enzyme heat denaturation as above, each sample was analysed in comparison to its starting material using gradient B.

PbGH130A (1 μ g/mL) was added to each solution which had been treated with PbEpiA and incubated for 1 hour at room temperature. Following enzyme heat denaturation, each sample was analysed using gradient C in comparison to glucose, mannose, galactose, and α -mannose-1-phosphate standards.

4.2.5 Carbohydrate Affinity Determination

4.2.5.1 Carbohydrate Affinity Polyacrylamide Gel Electrophoresis (CA-PAGE)CA-

PAGE was run as described in Moraïs *et al.* (2012). Briefly, 10 cm x 8 cm x 1 mm 10-well native polyacrylamide gels were cast supplemented with polysaccharide to a final concentration of 0.5 mg/mL. 5 μ g of protein was loaded into each well and the gel was run at 100 V for 3 h followed by fixation and staining with Coomassie R-250, and imaging using a Gel Doc XR+ imager (Bio-Rad).

4.2.5.2 Insoluble Mannan Pull-Down Assay

To qualitatively assess the ability of the SGBPs to recognise an insoluble substrate, 50 µg of each SGBP was mixed with 5 mg of inM to a final volume of 100 µL in ITC buffer. The solution was incubated at room temperature for 30 minutes. The mixture was centrifuged for 5 mins at 4000 g and the supernatant was collected. The inM was washed with 100 µL of ITC buffer and then suspended in another 100 µL of ITC buffer. Thirty-three microlitres of 4x SDS-PAGE loading buffer was added to each supernatant and inM suspension. The inM suspension was incubated for 5 minutes at room temperature and centrifuged as above. This inM extract and the original supernatant were then both heated to 95°C for 5 minutes prior to SDS-PAGE analysis.

4.2.5.3 Isothermal Titration Calorimetry (ITC)

ITC experiments were performed on an ITC-200 instrument (GE Healthcare). All protein samples were dialyzed for at least 24 h into ITC buffer (20 mM Na-P_i, pH 7.0) using a 12-14 kDa molecular weight cutoff (MWCO) membrane (Spectrum Laboratories). Ligands were prepared in the same buffer. Konjac glucomannan oligosaccharides (kGMOs) and carob galactomannan oligosaccharides (cGMOs) were prepared by exhaustive PbGH26A-PbGH5A digestion of 5 mg/mL polysaccharide in ITC buffer at 37°C followed by removal of the enzyme by filtration through a 30 kDa MWCO filter. Titrations were performed at 25°C with 100 μ M (PbSGBP-B) or 40 μ M (PbSGBP-C) protein and 5 mg/mL (for PbSGBP-B) or 2 mg/mL (for PbSGBP-C) polysaccharide titrations. The binding isotherms were analysed using a single-site binding model with the Microcal-modified version of *Origin 7.0* (OriginLab).

4.2.6 Enzyme Crystallisation and Structure Determination

For crystallisation, PbGH26A-PbGH5A and PbGH26A were purified as native proteins, and PbSGBP-B was purified as selenomethionine-derivatised (SeMet) protein using the standard M9 high-yield growth procedure according to the manufacturer's instructions (Shanghai Medicilon). PbGH26A-PbGH5A, PbGH26A and SeMet PbSGBP-B were crystallised at room temperature using the sitting-drop vapor diffusion method with 0.6 µL of protein solution mixed with 0.6 µL of reservoir solution. Protein concentrations and reservoir solutions were as follows: PbGH26A-GH5A (20 mg/mL), 0.1 M Tris pH 8.5, 1.6 M ammonium sulfate, 12% glycerol; PbGH26A (15 mg/mL), 2 M ammonium sulfate, 2% (w/v) PEG200; PbSGBP-B (15 mg/mL), 50 mM MES pH 6.5, 0.2 M ammonium sulfate, 30 % PEG 5K MME, 1% (w/v) MPD. Crystals were cryoprotected with Paratone-N oil or 2% glycerol plus Paratone-N before flash-freezing in liquid nitrogen prior to data collection. X-ray diffraction data was collected at -100 °C at the Structural Biology Center, Advanced Photon Source, beamline 19-ID for PbGH26A-GH5A, on a home source Rigaku 007-HF rotating anode with R-AXIS IV++ detector for PbGH26A, and at the Life Sciences Collaborative Access Team, Argonne National Laboratory, Beamline 21-ID-F for PbSGBP-B. Diffraction data was processed by HKL3000 or XDS (Kabsch, 2010) + Aimless (Winn et al., 2011). The structure of PbGH26A-PbGH5A was solved by Molecular Replacement (MR) using the structure of PbGH5A (PDB 3VDH) and a GH26 (PDB 3WDQ) modified with CHAINSAW (Stein, 2008). The structure of PbGH26A was solved by MR using the isolated

GH26 domain from the tandem PbGH26A-GH5A structure. The structure of PbSGBP-B was solved by the single anomalous dispersion (SAD) method using *Phenix.autosol* (Adams *et al.*, 2010). Model building was completed using either *ARP/wARP* (Langer *et al.*, 2008) or *Phenix.autobuild*, following by refinement using *Phenix.refine* (Afonine *et al.*, 2012) and *Coot* (Emsley and Cowtan, 2004). For all structures, *B*-factors were refined as isotropic and TLS parameterisation was included. Average B-factor and bond angle/length RMSD values were calculated using *Phenix*. All geometry was verified using the *Phenix*, *Coot*, and wwPDB validation tools. All structures were deposited in the Protein Databank (accession codes listed in Table 4-1).

4.2.7 Bioinformatic Analysis

A search of 20 genomes (found in Table B-2) from across the Bacteroidetes phylum for homologues of genes within the *P. bryantii* MUL was performed using the BLAST algorithm with an E-value cut-off of 10^{-40} . The local genetic neighbourhoods (±10 genes) of homologues identified by the BLAST algorithm were manually screened for putative β -mannan-degrading enzymes. Loci containing two or more predicted β -mannan-degrading glycoside hydrolases and a SusC-SusD homologue pair were considered to be putative MULs. To identify conserved features between MULs, genes from the same family which were conserved across most of the identified MULs were aligned using *ClustalW* and percent identities were determined on the basis of these alignments.

The subcellular localisation of gene products within the *P. bryantii* MUL was predicted using *PSORTb* (Yu *et al.*, 2010) and *LipoP* (Juncker *et al.*, 2003). Attachment to the membrane through a lipid anchor is proposed for any protein with a cysteine residue following a predicted SpII signal peptide cleavage site (Paetzel *et al.*, 2002).

4.3 Results

4.3.1 β-Mannan Utilisation Locus Model Building

To construct a model of β -mannan degradation by gene products from the mannan utilisation locus (MUL) from *P. bryantii* (Figure 4-2), the localisations of its various components were predicted. Signal peptides (21-26 amino acid) were identified for all of the *P. bryantii* MUL genes except PbEpiA, PbGH130A, and PbTFAraC, suggesting that these three are cytoplasmic. *PSORTb* predicts that PBR_0358 (a putative sodium-carbohydrate symporter) is embedded in the

inner membrane and that PBR_0363 (a putative TonB-dependent transporter) is embedded in the outer membrane, commensurate with the localisation of similar transporters involved in nutrient acquisition (Pourcher *et al.*, 1990, Franco and Wilson, 1999). In line with previously described PULs, *LipoP* predicts that PbSGBP-A, PbSGBP-B, PbSGBP-C, and PbGH26B are anchored to the outer membrane. It further predicts that PbGH36A, PbGH5A, and PbGH3A are free-floating in the periplasm, while PbGH26C is anchored to the inner membrane. Surprisingly, both *PSORTb* and *LipoP* fail to make confident predictions for the localisation of PbGH26A-GH5A. However, Gardner *et al.* (1995) showed that a antibodies raised against PbGH5A caused agglutination of *P. bryantii* cells, indicating that PbGH5A is associated with the outer membrane of *P. bryantii in vivo*.

4.3.2 Recombinant Protein Production and Purification

The recombinant production of several proteins cloned from this locus into *E. coli* yielded high levels of soluble protein, however, the production of PBR_0351 (PbGH5B), PBR_0360 (PbTFAraC), PBR_0353 (PbGH36A), PBR_0356 (PbGH26C), and PBR_0369 (PbGH26B) consistently resulted in the formation of inclusion bodies. Protein production from an N-terminal GFP fusion with PbGH26B yielded 0.4 mg/L of soluble, monomeric protein. For those that expressed well, yields ranged from 10 mg/L (PbSGBP-C) up to 50 mg/L (PbGH26A). Following cleavage with TEV protease, all proteins had masses matching those expected from their sequences suggesting no proteolysis or post-translational modification had occurred. All enzymes eluted from SEC at a volume consistent with monomeric protein. PbSGBP-C eluted at a volume consistent with stable dimerisation in solution. PbSGBP-A and PbSGBP-B eluted at a volume consistent with monomeric protein.

4.3.3 Glycan Recognition by Surface Glycan-Binding Proteins

The first step in carbohydrate acquisition using a polysaccharide utilisation locus pathway is the capture of a polysaccharide by surface glycan-binding proteins. PbSGBP-A, PbSGBP-B, and PbSGBP-C were assayed qualitatively for glycan affinity using CA-PAGE and a pull-down assay. Gels were run containing each of starch, bMLG, tXyG, bX, PASC, kGM, and cGM. In comparison to a gel run without polysaccharide present, PbSGBP-A showed no retardation in the presence of any polysaccharide (Figure 4-3). PbSGBP-B and PbSGBP-C both showed complete retardation in the presence of both kGM and cGM. The migrations of both



PbSGBP-B and PbSGBP-C were also retarded slightly in the presence of tXyG. The pull-down assay showed no evidence of affinity between inM and any of the SGBPs (Figure B-1)

Figure 4-2: Cartoon model of the putative *Prevotella bryantii* mannan utilization system. Complex galactomannans and glucomannans (shown using the graphical nomenclature from (Varki *et al.*, 2015) bind to PbSGBP-B and PbSGBP-C (Box A). These large glycans are degraded by *endo*-hydrolases (Box B). Short and medium-length oligosaccharide fragments are imported into the periplasm by the action of a TonB-Dependent Transporter. PbGH36A, PbGH5B, PbCE7A, PbCExA, and PbGH3A then saccharify these oligosaccharides. PbGH26C releases mannobiose from periplasmic oligosaccharides. We propose that mannobiose is actively transported into the cytoplasm by a sodium-carbohydrate symporter. Once in the



cytoplasm, PbEpiA and PbGH130 work sequentially to generate mannose-1-phosphate and glucose (Box C).

Figure 4-3: Identification of polysaccharide ligands for surface-glycan binding proteins. A) images of Coomassie-stained CA-PAGE gels containing the specified polysaccharides. PBR_0364, PBR_0365, and PBR_0366 are PbSGBP-A, PbSGBP-B, and PbSGBP-C respectively. B) Optimised isothermal titration

calorimetry (ITC) for PbSGBP-B and PbSGBP-C with cGM and kGM. The top graph in each pair shows the uncorrected heat evolved during titration, whereas the bottom graph shows the integrated heats after correction.

ITC was performed to quantify polysaccharide and oligosaccharide affinities. Titration of PbSGBP-B with cGM and kGM revealed clearly tighter binding for cGM ($K_d = 1.2\pm0.1 \mu$ M, Δ H = -10.5±0.2 kcal/mol, Δ S = -8 cal/mol/K, assuming 2.8 kDa of polysaccharide per binding site) compared to kGM ($K_d = 15\pm3 \mu$ M, Δ H = -12.5±0.8 kcal/mol, Δ S = -8 cal/mol/K, assuming 2.1 kDa of polysaccharide per binding site) (Figure 4-3). cGM ($K_d = 2\pm0.2 \mu$ M, Δ H = -10.0±0.2 kcal/mol, Δ S = -7 cal/mol/K, assuming 2.8 kDa of polysaccharide per binding site) also bound more tightly to PbSGBP-C than kGM ($K_d = 6\pm2 \mu$ M, Δ H = -16.0±1 kcal/mol, Δ S = -28 cal/mol/K, assuming 4.4 kDa of polysaccharide per binding site).

The length of the β -mannan-binding platform was investigated by titration with β mannooligosaccharides of different length and composition. Titrations with short kGM oligosaccharides produced by the action of PbGH26A-PbGH5A evolved no heat (Figure B-2A,B). Titrations with longer oligosaccharides from cGM showed some heat release from which fitting parameters could not reliably be extracted (Figure B-2C,D). While mannotriose, and mannotetraose gave no measurable heat release, PbSGBP-B and PbSGBP-C bound mannopentaose with low affinities ($K_d < 1 \text{ mM}$), and mannohexaose with affinities comparable to kGM, having dissociation constants of 46 μ M and 19 μ M, respectively (Figure B-2E,F).

An alignment of PbGH26B with its closest structurally characterized homologue, BoMan26A (Bågenholm *et al.*, 2017), revealed that PbGH26B lacks residues at the same position in the same sequence position as the catalytic glutamate residues (Figure B-3). The monomeric GFP-tagged PbGH26B which we obtained did not generate new reducing ends when incubated with cGM and kGM. CA-PAGE run with bMLG, cGM, and kGM showed no retardation of migration (Figure B-4).

4.3.4 Tertiary Structural Characterisation of PbSGBP-B

Attempts to crystallise PbSGBP-A and PbSGBP-C did not yield any protein crystals. To gain some insight into the tertiary structure of SGBP-A, an alignment of 4 crystallised SusD-like proteins with PbSGBP-A using T-Coffee Expresso (Armougom *et al.*, 2006) was prepared (Figure B-5). The alignment shows that PbSGBP-A shares the TPR-motif-containing secondary

structure of other SusD homologues, sharing TPR motifs 1 (helices 1 and 5, following the numbering in Figure B-5), 2 (helices 6 and 8), and 3 (helices 9 and 10) (Blatch and Lässle, 1999, Bakolitsa *et al.*, 2010). Furthermore, W82 and W306, the two conserved residues involved in ligand recognition in the SusD homologues from the *B. ovatus* xyloglucan utilisation locus (Tauzin *et al.*, 2016) and the *B. thetaiotaomicron* starch utilisation locus (Koropatkin *et al.*, 2008), are conserved in PbSGBP-A. However, none of the other 5 residues involved in ligand binding in the starch- or xyloglucan-binding SGBPs are conserved or shared with PbSGBP-A. Since the protein is stable in solution and a fusion protein prepared with the *E. coli* maltose-binding protein to modulate the folding behaviour of the protein showed no affinity for cGM, yet strong affinity for starch, demonstrating that the maltose-binding protein domain was well-folded (Figure B-6), we believe that this lack of affinity is not due to a failure to form the normal SusD-like fold.

Selenomethionine-derivatised PbSGBP-B was crystallised and its structure was solved by SAD phasing; crystallographic statistics are provided in Table 4-1. Residues 33-459 were resolved (the N-terminal signal peptide was not present). The protein contains three main subdomains: tandem 6- and 7-stranded β -sandwich Ig-like folds (residues 33-130 and residues 131-218), and a C-terminal 8-stranded β -sandwich fold (residues 219-459) (Figure 4-4A). The two Ig-like folds are highly similar in structure, and are found in an elongated arrangement with minimal interface between them *in crystallo*. The C-terminal domain is positioned to pack against the second domain, such that the full-length protein adopts a structure resembling a hammer. The protein is similar in structure (though not sequence) to BoSGBP-B from the *Bacteroides ovatus* xyloglucan utilisation locus (PDB: 5E7G (Tauzin *et al.*, 2016), RMSD 4.0 Å over 446 matching C α atoms, 18% amino acid sequence identity), and the xylan binding protein Bacova_04391 from the *B. ovatus* xylan-utilisation locus (PDB: 3ORJ, RMSD 3.8 Å over 429 matching C α atoms, 20% amino acid sequence identity); these three proteins share similar arrangement of their Ig-like folds, although PbSGBP-B and Bacova_04391 lack the third Ig-like fold domain found in BoSGBP-B (Figure 4-4B).

The loops between strands of the β -sandwich fold of the carbohydrate binding domains were not conserved. In spite of our failure to crystallize PbSGBP-B with mannooligosaccharides, the structural similarity between these SGBP structures allowed for prediction of the location of the mannooligosaccharide binding site based on the position of XyG bound to BoSGBP-B. The equivalent region in PbSGBP-B contained three prominently surface-exposed tryptophan residues (W242, W274 and W287) and nearby polar residues that may mediate interactions with mannooligosaccharides (H243, N276, N344, N360) (Figure 4-4C). We were able to model a bound glycerol molecule from the mother liquor, which is sometimes indicative of a carbohydrate binding site, into residual electron density nearby the W274 residue. Together, this platform is flat and spans approximately 26 Å and thus may form a recognition platform for a mannohexaose group (measuring 26 Å from O-1 of reducing end to O-4 of the non-reducing end when fully extended).



Figure 4-4: Overall structure of PbSGBP-B. Three domains are separately coloured. B) Comparison of structures of PbSGBP-B (blue), BoSGBP-B + XyG complex (green, PDB 5E7G) and Bacova_04391 (pink, PDB 3ORJ). The inset is a zoomed view of the C-terminal mannooligosaccharide-binding domains, rotated approximately 90° from view of overall structures. W242, W274 and W287 from PbSGBP-B and XyG from BoSGBP-B complex are shown in sticks. Loops from PbSGBP-B that impinge on the binding position of XyG from BoSGBP-B complex are show as thicker trace. C) Surface representation of PbSGBP-B, same view as B inset, showing putative mannan binding site and key tryptophan residues. The bound glycerol molecule is shown in sticks.

PbSGBP-B (W242A), PbSGBP-B (W274A), and PbSGBP-B (W287A) were constructed to determine the importance of the predicted binding platform identified in the crystal structure of PbSGBP-B. CA-PAGE of the W274A and W242A mutants of PbSGBP-B in the presence of kGM, cGM, and tXyG showed no retardation of migration, confirming that W274 and W242 are integral to polysaccharide recognition (Figure B-7). The W287A mutant of PbSGBP-B showed retardation in the presence of tXyG, cGM and kGM that was comparable to the wild type. Thus, W287 does not appear to be part of the β -mannan-recognition platform.

Though a crystal structure of PbSGBP-C could not be determined and no structurally characterised sequence homologues exist, a set of tryptophan mutants was also prepared for PbSGBP-C (PbSGBP-C (W524A), PbSGBP-C (W533A), PbSGBP-C (W541A), PbSGBP-C (W546A), PbSGBP-C (W568A), and PbSGBP-C (W579A)). CA-PAGE analysis of the PbSGBP-C mutants revealed a similar tryptophan-dependency in binding. PbSGBP-C (W533A) and PbSGBP-C (W579A) showed no affinity for β -mannans while PbSGBP-C (W541A) appears to have a partial loss of affinity (Figure B-8). Notably, the weak affinity of these SGBPs for tXyG is disrupted by the mutations affecting β -mannan recognition.

Structure	PbGH26A-GH5A	PbGH26A	PbSGBP-B
Data collection			
Space group	$P2_1$	P3 ₂ 2	P41212
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	54.48, 96.75, 138.1	73.58, 73.58, 105.1	82.90, 82,90, 152.47
α, β, γ (°)	90, 90, 90	90, 90, 120	90, 90, 90
Resolution, Å	40.0 - 2.10	30.0 - 1.72	43.33 - 2.19
\mathbf{R}_{merge}^{a}	0.110 (0.675)	0.082 (1.142)	0.082 (1.575)
R _{pim}	0.069 (0.444)	0.035 (0.492)	0.032 (0.601)
$CC_{1/2}^{c}$	0.583	0.607	0.762
$I / \sigma(I)$	9.89	2.15	20.2 (2.2)
Completeness, %	98.2 (96.50	99.9 (100)	99.7 (99.7)
Redundancy	3.2 (2.9)	6.4 (6.3)	14.3 (14.8)
Refinement			
Resolution, Å	19.81 - 2.10	27.24 - 1.72	43.3 - 2.19
No. of unique reflections:	77317, 2013	35526, 1782	28041, 1377
working, test			
<i>R</i> -factor/free <i>R</i> -factor ^d	14.4/19.4 (22.1/31.5)	15.2/19.6 (26.9/30.9)	20.4/24.1 (31.2/33.8)
No. of refined atoms, molecule	es		
Protein	10887	2715	3314
Solvent	63	68	31
Water	1780	577	176
B-factors			

Table 4-1: X-ray diffraction data collection and refinement statistics for PbGH26A and PbSGBP-B

Structure	PbGH26A-GH5A	PbGH26A	PbSGBP-B	
Protein	31.1	27.0	61.5	
Solvent	86.1	69.6	92.3	
Water	45.6	42.5	58.8	
r.m.s.d.				
Bond lengths, Å	0.004	0.010	0.004	
Bond angles, °	0.680	0.990	0.647	

 ${}^{a}R_{sym} = \Sigma_{h}\Sigma_{i}|I_{i}(h) - \langle I(h)\rangle/\Sigma_{h}\Sigma_{i}I_{i}(h)$, where $I_{i}(h)$ and $\langle I(h)\rangle$ are the *i*th and mean measurement of the intensity of reflection *h*.

^bFigures in parentheses indicate the values for the outer shells of the data.

^cValue refers to the outer shells of the data.

 ${}^{d}R = \Sigma |F_{p}{}^{obs} - F_{p}{}^{calc}| / \Sigma F_{p}{}^{obs}$, where $F_{p}{}^{obs}$ and $F_{p}{}^{calc}$ are the observed and calculated structure factor amplitudes, respectively.

4.3.5 Initial Substrate Cleavage by *Endo*-Glycanases

An initial screen for reducing end formation in the presence of PbGH26A with corn starch, barley mixed-linkage glucan (bMLG), laminarin, carob galactomannan (cGM), konjac glucomannan (kGM), tamarind xyloglucan (tXyG), beechwood xylan (bX), and wheat arabinoxylan (wAX) revealed hydrolytic activity on only kGM and cGM. Kinetic analysis at the pH and temperature optimum of the enzyme (Figure B-9) revealed preference for kGM over cGM (Table 4-2, Figure B-10).

To investigate potential cooperative effects between the two domains of the PbGH26A-GH5A fusion, full length PbGH26A-GH5A was assayed for activity against both cGM and kGM (Table 4-2) for comparison to PbGH26A and PbGH5A alone. We observed no significant kinetic enhancement for the full-length protein acting on cGM and an additive effect on kGM hydrolysis. Thus, the fusion of these domains shows no synergistic effect on enzyme performance when acting on soluble substrates *in vitro*.

β-Mannan hydrolysis product identification was confounded by the complexity of the oligosaccharide mixture produced by the digestion of cGM and kGM by PbGH26A alone (Figure 4-5B). inM was used as a model substrate to determine the backbone length of oligosaccharides produced by the action of PbGH26A. Overnight digestion of inM with 0.1 mg/mL PbGH26A gave a mixture of Man₁₋₆ oligosaccharides (Figure 4-5A). The masses of these oligosaccharides were confirmed by LC-MS (Figure B-11).

Substrate	Enzyme	k _{cat, app} (min ⁻¹)	K _{m,app} (mM)	$k_{cat, app}/K_{m,app}$ (M ⁻¹ s ⁻¹)	Assay
	PbGH5A ¹	3200±200	0.90±0.17*	60±12*	BCA
kGM	PbGH26A	2100±400	$0.40 \pm 0.05 *$	90±20*	BCA
	PbGH26A-GH5A	3400±300	$0.40 \pm 0.05 *$	140±20*	BCA
	PbGH5A ¹	ND	ND	ND	BCA
cGM	PbGH26A	3000±300	0.67±0.08*	80±12*	BCA
	PbGH26A-GH5A	3800±300	0.66±0.09*	100±15*	BCA
MMMMMM→ MMMMM+M	PbGH26A	2.1±0.2	0.9±0.1	39±4	HPAEC-PAD
MMMMMM→ MMMM+MM	PbGH26A	-	-	5.2±0.5	HPAEC-PAD
MMMMMM→ 2xMMM	PbGH26A	-	-	9.5±0.8	HPAEC-PAD
MMMMM→MMM M+M	PbGH26A	-	-	1.4±0.1	HPAEC-PAD
MMMMM→MMM+ MM	PbGH26A	-	-	3.3±0.2	HPAEC-PAD
MMMM→MMM+M	IPbGH26A	-	-	0.15 ± 0.01	HPAEC-PAD
MMMM→2xMM	PbGH26A	-	-	0.54 ± 0.03	HPAEC-PAD
MMM→MM+M	PbGH26A	-	-	0.019 ± 0.004	HPAEC-PAD

Table 4-2: Apparent Michaelis-Menten kinetics of various hydrolysis reactions catalysed by the GH26 and GH5 domains of PBR_0368.

¹Data from (McGregor *et al.*, 2016); ND = Not Detected; *Concentration expressed as g/L



Figure 4-5: HPAEC-PAD analysis of β-mannans digested by PbGH26A-GH5A. A) The digestion of inM is shown with six of eight peaks annotated based on comparison of elution times with βmannooligosaccharide standards. B) The digestion of carob galactomannan yielded a complex mixture of

oligosaccharides which could not be reliably assigned. C) The digestion of konjac glucomannan yields relatively simple mixture of small oligosaccharides including those annotated using retention time standards.

The extent of oligosaccharide binding within the active site cleft of PbGH26A was investigated using HPAEC-PAD to measure product release kinetics. Experiments with β -1,4linked β -mannooligosaccharide substrates ranging from mannobiose to mannohexaose revealed a large kinetic enhancement as substrate length increased (Table 4-2). Mannotriose, interacting with three subsites, is a very poor substrate for PbGH26A ($k_{cat}/K_M = 0.019 \pm 0.004 \text{ M}^{-1}\text{s}^{-1}$). The addition of a single mannose residue in the +2 subsite results in a 28-fold increase in k_{cat}/K_{M} and the hydrolysis of mannotetraose into mannotriose and mannose proceeds with an 8-fold greater $k_{\text{cat}}/K_{\text{M}}$ value. Thus, the +2 subsite contributes up to 8.6 kJ/mol to transition state stabilisation while the -3 subsite only contributes up to 5.4 kJ/mol (estimated following the method outlined in Wilkinson et al. (1983)). Mannopentaose hydrolysis into mannotriose and mannobiose occurs with a k_{cat}/K_{M} value 22-fold higher than the hydrolysis of mannotetraose into mannotriose and mannose, suggesting a contribution from the +2 subsite of up to 8.0 kJ/mol, roughly in line with the comparison of mannotriose hydrolysis and mannotetraose hydrolysis into mannobiose. The hydrolysis of mannopentaose into mannobiose and mannotriose occurred with a k_{cat}/K_M value only 6-fold higher than that for the hydrolysis of mannotetraose into mannobiose, giving a value of up to 4.6 kJ/mol for the contribution from the -3 subsite. The hydrolysis of mannopentaose into mannotetraose and mannose occurred with a k_{cat}/K_{M} value 10-fold higher than the hydrolysis of mannotetraose into mannotriose and mannose, giving a contribution of 5.9 kJ/mol from a -4 subsite.

Interestingly, the hydrolysis of mannohexaose favoured the release of mannose and mannopentaose. To confirm the point of cleavage of this reaction, hydrolysis was performed in $H_2^{18}O$. This confirmed that mannose is exclusively released from the reducing end giving ¹⁸O-labelled mannopentaose (Figure 4-6) and showed that mannotetraose is formed in the negative subsites in roughly 75% of catalytic encounters. The k_{cat}/K_M value for the formation of mannopentaose is roughly 30-fold higher than that of mannopentaose hydrolysis into mannotetraose and mannose, and roughly 8-fold higher than that of mannohexaose hydrolysis into mannotetraose and mannobiose.

134



Figure 4-6: Mass spectrometric analysis of the hydrolysis products of PbGH26A acting on mannohexaose in H2¹⁸O. The mass spectra shown are for A) mannobiose B) mannotriose C) mannotetraose D) mannopentaose and E) mannohexaose. For each spectrum, the expected mass of the base peak is given, and the degree of isotopic enrichment, corrected for added H2¹⁶O from the enzyme and substrate solutions, is calculated based on peak integral ratios.

4.3.6 Tertiary Structural Characterisation of PbGH26A-GH5A and PbGH26A

To obtain insight into the molecular specificity of the GH26 domain of PbGH26A-GH5A, we attempted the crystallisation of the full-length protein and the GH26 domain in isolation, and in complex with carbohydrate substrates. We successfully grew crystals of both the full-length GH26A-GH5A and the isolated GH26 domain as *apo* proteins, but were not successful in obtaining ligand-bound complexes. The crystal structures were solved by molecular replacement using the isolated PbGH5A structure (McGregor *et al.*, 2016, PDB 3VDH) and a

modified GH26 domain from (Tsukagoshi *et al.*, 2014); crystallographic statistics are listed in Table 4-1).



Figure 4-7: Structure of PbGH26A-GH5A. A) Overall two-domain architecture of PbGH26A-GH5A. Domains are separately coloured, active site clefts are coloured in light pink. XXXG as bound to GH5 domain

(from PDB 5d9n) is shown as sticks. B) Comparison of structures of PbGH26A (red), protist GH26 (blue, from PDB 3wdr), *P. anserina* GH26 (green, from PDB 3zm8), *B. subtilis* GH26 (brown, from PDB 2qha). The catalytic residues, E260 and E371, from PbGH26A and the glucomannan-derived oligosaccharide as bound to *P. anserina* GH26 are shown in sticks. The -1 to -5 subsites of *P. anserina* GH26 are also labelled.

The structure of PbGH26A-GH5A revealed a "back-to-back" arrangement of the GH26 and GH5 domains, which placed the active site clefts facing in opposite directions (Figure 4-7A). The domains are linked by a short 5-residue linker and are associated in an interface burying 680 $Å^2$. The opposing arrangement of active sites and relatively low contact area between domains is consistent with the observed absence of any kinetic enhancement of activity of the full-length protein relative to the isolated GH26 domain (Table 4-2).

The structure of the isolated PbGH26A did not show any significant conformational differences as compared to this domain in the context of the full-length PbGH26A-GH5A protein, except for a rearrangement of the active site loop that may contribute to the -4 and -5 subsites. While structural similarity searches with PbGH26A identify similarity with many GH26 domains, in general, its active site residues in the region comprising the catalytic centre and the nearby subsites (-1 and +1), including the catalytic acid/base E260 and nucleophile E371 are conserved, while loops surrounding the active site cleft are not conserved. PbGH26A is most closely related in sequence and structure with two eukaryotic GH26 members: a GH26 isolated from a symbiotic protist of termites (PDB 3WDR) (Tsukagoshi *et al.*, 2014) and one from *Podospora anserina* (PDB 3ZM8) (Couturier *et al.*, 2013); it is also similar to the bacterial GH26 BCman from *Bacillus subtilis* (PDB 2QHA) (Yan *et al.*, 2008) (Figure 4-7B). Comparison of these crystal structures shows most variation in active site loops of PbGH26A, over the presumed -4 and -5 subsites (as resolved in the crystal structure of the protist GH26).

4.3.7 Mannobiose Breakdown as the Ultimate Step in Saccharification

Due to the homology between PbEpiA and the mannobiose-2-epimerase described by Senoura *et al.* (2011), we performed assays for a variety of 2-epimerase activities. No activity was observed with glucose, mannose, or galactose. However, following treatment with PbEpiA, individual samples of mannobiose, cellobiose, and lactose contained mannosylglucose, glucosylmannose, and galactosylmannose, respectively, in thermodynamic equilibrium with the starting disaccharide (Figure B-12A-C). This resulted from the epimerisation of stereochemistry at the 2-position of the reducing monosaccharide residue. Mannotriose and cellotriose also underwent 2-epimerisation in the presences of PbEpiA.

PbGH130A and phosphate buffer were added to samples of mannobiose, cellobiose, and lactose with and without the addition of PbEpiA. As observed by Senoura *et al.* (2011), the selective degradation of the mannosylglucose peak in the sample of mannobiose incubated with PbEpiA only occurred in the presence of excess phosphate, confirming the roles of PbEpiA and PbGH130A proposed in Box C of Figure 4-2. This resulted in the formation of glucose and α -mannose-1-phosphate (Figure B-12D).

4.3.8 Identification of Putative β-Mannan Utilisation Loci

Homologous MULs were clearly identifiable within several *Prevotella* species based on SusC and SusD homology using a 30% sequence identity cut-off (Figure 4-1). None of the identified *Prevotella* MULs contain identical GH content. Every identified locus encodes at least one GH26 enzyme, yet, with the exception of the putative mannobiosidase (PbGH26C), they are not homologous between loci. GH5 β -mannanase and β -mannosidase homologues are also common, but not conserved, features. Out of the MULs identified based on SusC-SusD homology, GH36 is unique to the *P. bryantii* MUL. Based on previously observed activities in those families, α -galactosidase activity is likely supplied by the GH97 gene in *B. uniformis* and by the GH27 genes in *P. oris* and *P. salivae*. β -Glucosidase activity was only predicted to be present in the *P. bryantii* and *P. salivae* MULs, and *endo*-glucanase activity was only predicted in the *P. bryantii*, *P. paludivivens*, and *B. uniformis* MULs.

To better predict the existence of MULs across the Bacteroidetes phylum, we made a direct comparison between the genes found in the recently described *B. ovatus* MUL (Bågenholm *et al.*, 2017) and the *P. bryantii* MUL. Strikingly, the most conserved full-length genes were PbGH130A (74% ID), PbEpiA (56% ID), PbGH26C (42% ID), and PbGH36A (65% ID). In contrast, the markers typically used to identify homologous PULs, the SusC homologue (29% ID) and the SusD homologue (24% ID) were not significantly conserved. The conserved mannobiose-2-epimerase, mannosylglucose phosphorylase, and sodium-carbohydrate symporter gene cluster, first identified by Senoura et al., is hereafter referred to as the Symporter-Epimerase-Mannosylglucose Phosphorylase (or SEMP) cluster. To attempt the extension of MUL prediction using this motif, we chose 20 genomes (Table B-2) of non-starch polysaccharide-degrading Bacteroidetes within the JGI IMG and searched for the SEMP cluster

in association with SusC-SusD pairs. Searching with the SEMP cluster, and a sufficiently strict homology cut-off, facilitated the identification of putative MULs with no apparent false-positives in our test set. Subsequent examination of the local genetic neighbourhood consistently revealed homologues of known β -mannan-degrading glycoside hydrolases (Figure 4-1B). We were able to identify a putative MUL adjacent to a SEMP cluster in 12 of the 20 organisms chosen.

4.4 Discussion

It is becoming increasingly clear that *Prevotella* play an important role in a healthy gut. Their predominance in the guts of organisms consuming elevated quantities of dietary fibre suggests that they are particularly well-adapted to carbohydrate-rich diets (Gorvitovskaia *et al.*, 2016). This contrasts with *Bacteroides* species which are associated with carbohydrate-limited western diets. Given the centrality of carbohydrate availability to this population difference, it is essential to develop a stronger understanding of carbohydrate-degrading systems in *Prevotella*.

4.4.1 Genetic Markers of β-Mannan Utilisation

The first β -mannan-utilizing locus characterised within a *Prevotella* species appears to be far more versatile than the first β -mannan-utilisation locus identified in *Bacteroides* (Bågenholm *et al.*, 2017). The *P. bryantii* MUL encodes seven glycoside hydrolases in contrast to the three glycoside hydrolases identified in the recently characterised *B. ovatus* MUL. Furthermore, the activities of the genes found in the *P. bryantii* MUL are more diverse, encoding β -glucosidase, α -galactosidase, β -mannosidase, β -mannobiosidase, β -mannanase, and β -glucanase activities. Thus, the PbMUL endows *Prevotella* with the capacity to breakdown every known glycosidic linkage identified in complex galactomannans and glucomannans. This suggests that *P. bryantii* is able to fully utilize all known β -mannans.

The PbMUL consists of three apparent functional cassettes: PBR_0363-PBR_0369 are the outer-membrane machinery responsible for adhering to β -mannans, and hydrolysing $\beta(1,4)$ glucosidic linkages and $\beta(1,4)$ mannosidic linkages to generate oligosaccharides, and importing these oligosaccharides into the periplasm. PBR_0351-PBR_0354 are periplasmic enzymes capable of completely saccharifying β -mannan oligosaccharides. PBR_0355-PBR_0360 appear to make up a versatile β -mannan-specific regulatory system (*vide infra*).

The identification of putative MULs based on our understanding of the PbMUL presents a challenge due to the poor conservation of the SusC-SusD pair between known MULs and due

to the polyspecificity of CAZy families GH3, GH5, GH26, and GH36. To address this, we identified homologues of PbMUL genes within other bacterial genomes and examined their local genetic neighbourhoods for other known β -mannan-degrading genes. This allowed us to assess the predictive power of each conserved element. Elements of the "SEMP" motif, the set of the most conserved genes between the *B. ovatus* MUL and the *P. bryantii* MUL, were found to be strong indicators of putative MUL presence. These genes were used to identify putative MULs in 12 out of a set of 20 Bacteroidetes genomes, including an unusual putative MUL from an alkaliphilic carbohydrate-fermenter, *Alkaliflexus imshenetskii*, encoding only two GH26 enzymes and at least 4 proteins with unknown functions (Table B-2). Thus, while SusC and SusD homology are good metrics for identifying phylogenetically related PULs, our analysis suggests that other conserved elements can be superior predictors of PUL specificity.

4.4.2 PbSGBP-B and PbSGBP-C Bind β-Mannan Utilizing a Dual-Tryptophan Platform

CA-PAGE showed that PbSGBP-B and PbSGBP-C bind tightly to both glucomannan and galactomannan. Follow-up ITC measurements revealed that both proteins recognise galactomannan more tightly than glucomannan. Given the apparent specificity of PbGH26A-PbGH5A towards glucomannans, this was surprising. However, we also noted that there was a moderate retardation of both proteins in CA-PAGE gels containing tamarind xyloglucan. Based on our mutational study it is clear that this recognition is occurring in the β -mannan-binding site. However, the only monosaccharides shared between these polysaccharides are branching galactose residues. Galactomannans have galactose extending from the 6-position, while xyloglucans have galactose extending from the 2-position of a xylose added to the 6-position of the glucan backbone. Given the lack of affinity for cellulose or mixed-linkage glucan, we believe that the observed interaction is likely related to the presence of these galactose residues.

The length of the β -mannan-binding site was also investigated by ITC with β mannooligosaccharides. Mannopentaose and oligosaccharide mixtures derived from the action of PbGH26A-PbGH5A on kGM and cGM did not demonstrate meaningful affinity for the SGBPs. However, mannohexaose is a good ligand for both proteins, binding with affinities similar to those between the SGBPs and kGM. This suggests that mannohexaose saturates the backbone binding sites available on the surface of the SGBPs. Thus, we estimate the length of the β mannan-recognition site to be roughly equal to the length of mannohexaose. The crystal structure of PbSGBP-B revealed three roughly coplanar tryptophan residues (W242, W274, and W287) forming an apparent binding platform on the surface of the C-terminal domain of the protein. CA-PAGE with these W242A, W274A, and W287A mutants confirmed the essential nature of W242 and W274 to polysaccharide recognition and showed that W287 is dispensable. Thus, we propose that the ligand binding site is actually centred on W274 and extends into a cleft which is rich in hydrophilic residues. This gives a binding site measuring 25 Å from W242 to N271, comparable to the length of mannohexaose in an extended conformation (~26 Å).

Though we were unable to obtain a crystal structure of PbSGBP-C, we hypothesised the existence of a similar binding platform. The protein only contains 6 tryptophan residues (W524, W533, W541, W546, W568, and W579), all of which are found in the C-terminal domain. CA-PAGE analysis of mutants lacking each of these tryptophan residues identified W533 and W579 as essential for polysaccharide recognition. The W541A mutation only caused a partial loss of affinity, suggesting that this residue may also form part of the binding site. These results demonstrate that substituting tryptophan residues near the C-terminus of SusE-like and SusF-like proteins to alanine is an effective way, in the absence of structural data, to abrogate binding without disrupting other potential protein functions.

4.4.3 PbSGBP-A and PbGH26B have no Affinity for β-Glucans or β-Mannans

The lack of affinity between PbSGBP-A and β -mannan polysaccharides is surprising since previous SusD homologues have been shown to recognise the polysaccharide targeted by their locus (Shipman *et al.*, 2000, Larsbrink *et al.*, 2014a, Cuskin *et al.*, 2015). Recent work has shown that TBDT-SGBP-A pairs form a complex on the cell surface, independent of glycanbinding, which is indispensable for the function of a PUL *in vivo* (Cameron *et al.*, 2014, Tauzin *et al.*, 2016, Glenwright *et al.*, 2017). Thus, the observed lack of affinity for glycans suggests that this protein may not be a true "surface-glycan binding protein," but rather functioning only as an accessory protein for the TonB-dependent transporter (Glenwright *et al.*, 2017). However, in light of the observed SusD homology and the finding that the carbohydrate-binding site is found at its interface with SusC, we consider it likely that the carbohydrate-binding site of PbSGBP-A may not be properly formed without its cognate SusC.

On the basis of our alignment of PbGH26B with BoMan26A, we hypothesized that PbGH26B was an inactive GH26 member that may be acting as a β -mannan-binding protein

alongside PbSGBP-B and PbSGBP-C. The lack of β -mannan-specific affinity or activity observed for GFP-tagged PbGH26B refutes this hypothesis, suggesting that this protein is acting in some other, as yet unknown, capacity.

4.4.4 PbGH26A has a Long Active Site Tailored for Glucomannan Degradation

The inability of PbGH26A to cleave mannobiose demonstrates that the enzyme requires at least three subsites within its active-site cleft to be occupied by mannose residues for observable catalysis to take place. Together, the kinetic enhancements observed with β mannooligosaccharides demonstrate that the enzyme has strong substrate-binding subsites spanning the $-2 \rightarrow +2$ positions. -3 and -4 subsites which recognise the substrate relatively weakly also appear to be present in the active site. The addition of a sixth mannose residue revealed interactions at the -5 position which contribute to substrate affinity so much that the $K_{\rm M}$ drops by at least an order of magnitude (Figure B-13D). Thus, PbGH26A has a long active site cleft similar to that described in Tsukagoshi *et al.* (2014).

The existence of the +2 subsite combined with the proclivity for PbGH26A to hydrolyse mannohexaose into mannopentaose, incorporating ¹⁸O into the resulting oligosaccharide, demonstrates that this enzyme has an active site cleft spanning at least seven mannose residues. Although our kinetic evidence suggests that this is not likely to be the full length of the active site since the enzyme performance with both kGM and cGM is ~1000-fold higher than with mannohexaose. This may result from the contributions from the $-5 \rightarrow +2$ subsites with one additional strong negative subsite at the -6 position would give the observed polysaccharide hydrolysis performance. Alternatively, the observed kinetic enhancements with kGM could also be a consequence of specific recognition of glucose in the -3, or -4 subsites, while the kinetic enhancement observed with cGM may be attributed to recognition of $\alpha(1,6)$ -galactose branches. Assuming substrate binding along the trajectory of the glucomannan oligosaccharide shown in Figure 4-7B, the active site cleft spans from W165 to E354, a length of ~24 Å in the negative subsites and ~ 12 Å in the positive subsites. These lengths are comparable to those of the extended conformations of mannohexaose (~26 Å) in the negative subsites and mannobiose (~10 Å) in the positive subsites, providing further support for the $-6 \rightarrow +2$ subsite map over the specific recognition of backbone glucose residues or $\alpha(1,6)$ -galactose branches. In harness, the functional and structural data presented here suggest that PbGH26A has evolved to specifically recognise two clusters of mannose residues within the semi-regular glucomannan structure.

4.4.5 The Fusion of PbGH26A and PbGH5A Connects Complementary Activities without Cooperativity

The fusion of PbGH26A and PbGH5A observed in PBR_0368 is a unique feature of the PbMUL. Our initial hypothesis that this was favoured due to enhanced hydrolytic activity towards glucomannans was refuted by the lack of apparent cooperativity between the domains in the breakdown of konjac glucomannan. This lack of cooperativity was explained by the crystal structure of PbGH26A-PbGH5A which revealed a "back-to-back" arrangement of the two glycoside hydrolase domains. Given this arrangement, it is unlikely that a single polysaccharide molecule would interact with both domains simultaneously.

However, the activities of PbGH5A and PbGH26A are clearly complementary. The natural structure of glucomannan consists of short stretches of $\beta(1,4)$ -linked mannan and $\beta(1,4)$ -linked glucan. Thus, the actions of both enzymes are necessary for the polysaccharide to be broken down into short oligosaccharides for release from the SGBPs and import by the TBDT. We further propose that the cleavage of internal glucosidic linkages prevents the inhibition of the PbGH26A domain by its own relatively long oligosaccharide products. Thus, while this fusion does not enhance the hydrolytic performance of either enzyme, it may enhance the rate and completeness of oligosaccharide formation from polysaccharide molecules.

4.4.6 β-Mannan Oligosaccharides are Hydrolysed in the Periplasm

The processing of imported β -mannan oligosaccharides into monosaccharide in the periplasm is a complex step-wise process dependent on the source of the material. α -Galactosidase and β -mannosidase activities are sufficient to breakdown oligosaccharides derived from storage galactomannans, but structural glucomannans are significantly more complex, requiring broad specificity β -glucosidase, β -mannosidase, and α -galactosidase activities.

The β -glucosidase from the *P. bryantii* MUL has been previously characterised; Dodd *et al.* showed that it is the only GH3 glucosidase (CdxA) found in the genome of *P. bryantii* B₁4 (Dodd *et al.*, 2010a). The putative α -galactosidase (PbGH36A) and putative β -mannosidases (PbGH5B and PbGH26C) did not express in an active form in our hands. Fortunately, a close homologue of each has been characterised. PbGH36A shares 66% amino acid sequence similarity with AgaA (PDB ID: 4FNP), a thermostable *exo*- α -galactosidase from *Geobacillus stearothermophilus* (Merceron *et al.*, 2012) and PbGH5B shares 69% amino acid sequence similarity with CmMan5A (PDB ID: 1UUQ), an *exo*- β -mannosidase from *Cellvibrio mixtus*

(Dias *et al.*, 2004). Furthermore, PbGH26C shares 73% amino acid sequence similarity with a recently characterised β -mannobiosidase, BoMan26A (PDB ID: 4ZXO), from *B. ovatus* (Bågenholm *et al.*, 2017). Thus, the periplasmic genes from the *P. bryantii* MUL very likely possess all of the activities necessary to saccharify complex galactomannan and glucomannan oligosaccharides.

4.4.7 PbEpiA and PbGH130A Generate α-Mannose-1-Phosphate in the Cytoplasm

Following the action of periplasmic PbGH26C, we propose that the released mannobiose is imported into the cytoplasm by PBR_0358, a predicted sodium-carbohydrate symporter. Close homologues of the mannobiose-2-epimerase and mannosylglucose phosphorylase described by Senoura *et al.* (2011), having no signal peptides, then convert cytosolic mannobiose into glucose and α -mannose-1-phosphate (Figure B-12).

In contrast to the transmembrane maltooligosaccharide-dependent DNA-binding SusR (D'Elia and Salyers, 1996, Martens *et al.*, 2009) and the hybrid two-component system sensor/regulators which have been identified in several glycan-degrading systems in *B. thetaiotaomicron* and *B. ovatus* (Sonnenburg *et al.*, 2006), the PbMUL encodes PbTFAraC, an AraC-like transcription factor. Regulation by a cytosolic carbohydrate-binding transcription regulator has not been reported for any PUL characterised to date. The presence of PbTFAraC in the cytosol suggests a specific carbohydrate metabolite found in the cytosol is the primary signal regulating transcription of the PbMUL.

Though we were not able to recombinantly produce the PbTFAraC transcription factor to test this experimentally, we propose two possible ligands for PbTFAraC: mannobiose (MM) or α -mannose-1-phosphate (M1P). As AraC-like transcription repressors lose affinity for DNA upon ligand binding, either of these could form a negative feedback molecular logic system which initiates transcription of the *P. bryantii* MUL if mannobiose is drawn into the cytoplasm. It can be reasonably expected that if M1P is the ligand of PbTFAraC the signal would dissipate through the actions of phosphomannomutase (EC 5.4.2.8, PBR_2840, forming mannose-6-phosphate) and phosphomannose isomerase (EC 5.3.1.8, PBR_1766, forming fructose-6-phosphate). If MM is the ligand, then the signal would dissipate through the combined action of PbEpiA and PbGH130A.

With these as the last steps in β -mannan breakdown, key elements of the model of β mannan degradation in *P. bryantii* presented in Figure 4-2, ranging from the initial recognition of β -mannans to entry into glycolysis, have been validated experimentally. The unexpected complexity and apparent versatility of the PbMUL highlights the potential of *P. bryantii* as a model gut species. Though additional work will be required to identify the ligand of the MULassociated AraC-like transcription factor, the work presented here contributes to an understanding of the mechanisms and genetics underlying the acquisition of β -mannans by gut symbionts. This evolving understanding will be invaluable in the design of tools to manipulate the gut microbial community to promote host health. Going forward, enzymes from diverse putative MULs from across the Bacteroidetes phylum should be characterised to experimentally test the SEMP cluster as a marker of β -mannan utilisation. This continued mining of MULs for novel genes should uncover new β -mannan-degrading enzymes with valuable new activities and specificities.

Chapter 5: Characterisation of Two Mannan/Xylan Acetylesterases from *Prevotella bryantii* B₁4 Reveals a New Esterase Family

5.1 Introduction

Xylans and glucomannans, abundant in the secondary cell walls of plants, are known to be acetylated at positions 2, 3, or 6 of monosaccharide residues found in the polysaccharide main chain (Albersheim *et al.*, 2010). Acetyl groups modify the physicochemical properties of polysaccharides, enhancing the solubility of storage glucomannan found in konjac corm and facilitating the incorporation of xylans into the cell wall of *Arabidopsis thaliana* (Davé and McCarthy, 1997, Yuan *et al.*, 2016). However, acetylation is also known to hinder many hydrolytic enzymes which saccharify these polysaccharides, necessitating the production of carbohydrate esterases (CEs) for their catabolism (Selig *et al.*, 2009, Levisson *et al.*, 2012). This necessity has given rise to a diverse collection of CEs known to contribute to the degradation of acetylated polysaccharides (Nakamura *et al.*, 2017).

CEs are currently classified into 15 families within the CAZy database (Lombard *et al.*, 2014). Each family is defined by amino acid sequence similarities. This gives rise to families which share similar functions, common catalytic machinery, and a common protein fold. Acetylxylan esterase or acetylmannan esterase activities are among the most commonly identified in CEs, having been identified in members of CE families 1-7, and 16 (Biely, 2012). CE families 2, 3, and 6 belong to the serine-glycine-asparagine-histidine (SGNH) hydrolase superfamily, a broad class of serine hydrolases sharing a unique hydrogen bond network stabilizing their catalytic centre. CE families 1, 5, and 7 belong to the α/β hydrolase superfamily, a diverse collection of serine hydrolases sharing a common Ser-His-Asp/Glu catalytic triad and a common fold consisting a core β -sheet sandwiched between α -helices (Hotelier *et al.*, 2004).

We recently reported the identification and characterisation of key elements from a β mannan utilisation locus found in the genome of *P. bryantii* B₁4 (see Chapter 4). This locus encodes two serine hydrolases: PBR_0354 (NCBI GenBank: WP_006282900.1, hereafter PbCExA) is an SGNH hydrolase annotated as a sialate *O*-acetylesterase, while PBR_0355 (NCBI GenBank: WP_006282901.1, hereafter PbCE7A) is a CE7-like α/β hydrolase annotated as a cephalosporin C deacetylase. Considering the presence of these genes within the context of a β -mannan utilisation locus and the acetylated nature of plant cell wall glucomannans, it is highly unlikely that they display these activities. We hypothesize that PbCExA and PbCE7A are acetylmannan esterases responsible for the deacetylation of (galacto)glucomannan oligosaccharides formed in the course of β -mannan catabolism by *P. bryantii* B₁4. Here we report the cloning, recombinant production and functional characterisation of the full-length PbCExA and PbCE7A esterases in the context of glucomannan catabolism. We further report the structural characterisation of PbCExA, which revealed a novel active site architecture. The structural and phylogenetic analyses presented here suggest that PbCExA is the founding member of a new family of carbohydrate esterases.

5.2 Materials and Methods

All buffers and reagents were purchased from Sigma Aldrich unless otherwise stated.

5.2.1 Mass Spectrometry

Intact protein masses were determined on a Waters Xevo Q-TOF with a nanoACQUITY UPLC system, according to the method described by Sundqvist *et al.* (2007). Carbohydrate LC-MS(/MS) was performed using a 0.32 mm Hypercarb Kappa column as described in McGregor *et al.* (2017b). The column was eluted at 8 μ L/min at 30°C. The separation gradient used was 0.0–5.0 min, 100% A (5% MeCN, 10 mM NH₄COO, 0.1 mM NaOAc, pH 5), 0% B (95% MeCN, 10 mM NH₄COO, 0.1 mM NaOAc, pH 5), 0% B (95% MeCN, 10 mM NH₄COO, 0.1 mM NaOAc, pH 5); 5.0–40.0 min, linear gradient to 30% B; 40.0–40.1 min, linear gradient back to 0% B; 40.1–45.0 min, equilibration with 100% A.

5.2.2 Substrates and Ligands

Oligosaccharides and their derivatives are abbreviated using a general shorthand in which M represents $\beta(1,4)$ -linked D-mannopyranose, X represents $\beta(1,4)$ -linked D-xylopyranose, G represents $\beta(1,4)$ -linked D-glucopyranose, and Ac denotes acetylation on the preceding residue.

5.2.2.1 Commercial Substrates

High purity (>94%) high viscosity konjac glucomannan (kGM) was purchased from Megazyme International (Ireland). Naturally acetylated xylan was purchased from Cambridge Glycosciences (Cambridge, UK).

5.2.2.2 Oligosaccharide Mixtures

Acetylated glucomannan oligosaccharides (kGMOs) were prepared from konjac glucomannan using PbGH26A-GH5A (see Chapter 4). 50 mg of kGM was wetted with 200 μ L of 95% ethanol prior to the addition of 9 mL of distilled water. The solution was boiled for 2 minutes to remove ethanol and dissolve the kGM. The clear, colourless solution was buffered with 1 mL of 100 mM NH₄OAc reduced to pH 6.0 with acetic acid, diluted to a final volume of 10 mL, and cooled to 37°C. PbGH26A-GH5A was added to a final concentration of 20 μ g/mL and the reaction was incubated at 37°C for 2 hours. The enzyme was denatured by heating to 80°C for 5 minutes and the product was lyophilised.

Acetylated xylan oligosaccharides (XOs) were prepared from acetylated xylan using CjXyn10A (Xylanase 10A, NZYtech, Lisbon) (Charnock *et al.*, 1997). 2.5 mg of acetylated xylan was dissolved in 0.5 mL of 10 mM NH₄OAc reduced to pH 6.5 with acetic acid. Enzyme was added to a final concentration of 1.5 μ g/mL and incubated, denatured, and lyophilised as above. The oligosaccharides formed were analysed by HPAEC-PAD using gradient C from section 4.2.1.1. Xylooligosaccharides ranging from Xyl₁-Xyl₂₀ were observed in the mixture with the major peaks being xylotriose and xylopentaose, together representing 45% of the total peak area (Figure C-1). ¹⁸O-labelled XOs were prepared in the same manner, but using 0.1 mg of acetylated xylan in 50 μ L of H₂¹⁸O with 5 mM buffer. ¹⁸O-labelled oligosaccharides were used immediately without lyophilisation.

5.2.3 Enzyme Cloning and Production

PBR_0354 (PbCExA) and PBR_0355 (PbCE7A) were amplified from *P. bryantii* B₁4 genomic DNA obtained from DSMZ (https://www.dsmz.de/) with nucleotides corresponding to signal peptides, predicted by *SignalP* (Petersen *et al.*, 2011), excluded. Primers used for amplification are listed in Table C-1. Notably, the start site for PbCE7A appeared to be miscalled in the JGI Integrated Microbial Genomes database. The gene was cloned using primers which included codons encoding seven additional N-terminal amino acids, which we found to be essential for folding on the basis of recombinant expression results. PbCE7A and PbCExA were cloned into pMCSG53 (Eschenfeldt *et al.*, 2013) using Gibson assembly (Gibson *et al.*, 2009). The enzymes were expressed in *E. coli* BL21(DE3) Hi-Control as described in Chapter 4.

5.2.4 Enzyme Kinetics and Product Analysis

Initial studies of potential activities of PbCExA were performed with 4-nitrophenyl derivatives of glucose, cellobiose, mannose, acetate, and butyrate. 1 mM substrate was incubated with 0.1 mg/mL enzyme in PBS at 30°C. Changes in A₄₀₅ were assessed relative to a buffer blank after 1 hour. Kinetic analysis was performed using stocks of 4-nitrophenyl acetate (PNP-Ac) and 4-nitrophenyl butyrate (PNP-Bu) prepared at a concentration of 0.100 M in acetonitrile. A dilution series from 100 mM to 1.25 mM was prepared in acetonitrile from each of these stocks. Kinetic measurements were performed at 37°C in PBS supplemented with 0.5% Triton X detergent to maintain substrate solubility. For each measurement, 50 µL of 10x buffer was added to 390 µL of water, followed by 10 µL of substrate solution, and finally 50 µL of enzyme solution (20 µg/mL for PNP-Ac, and 300 µg/mL for PNP-Bu). A_{405} was monitored continuously for 5 minutes. Absorbance was converted into rate using an experimentally determined extinction coefficient of 12,500 M⁻¹ cm⁻¹.

The release of acetate from carbohydrates and cephalothin was detected using the Megazyme acetate kit (Acetate Kinase Manual Format, Megazyme International, Bray, Ireland) modified for 665 μ L reaction volumes and using 100 μ L of sample and 425 μ L of water to enhance the sensitivity of the assay. For standard activity determination, 2.5 g/L of glucomannan-based substrate, 0.5 mg/mL of xylan-based substrate, or 0.5 mM cephalothin (to give approximately equal total acetate content) was incubated with 3 μ g/mL enzyme for 30 minutes at 37°C in PBS. For limit-digest preparation, 1 mg/mL kGMOs or XOs were incubated with 100 μ g/mL PbCE7A or PbCExA for 16 hours at 37°C in PBS supplemented with 3 mM NaN₃. Released acetate was determined relative to a buffer blank in triplicate following 5 minutes of incubation at 80°C to denature the enzyme. The total hydrolysable acetate content of the kGMOs and XOs was determined by treatment of 2 mg/mL oligosaccharides with 0.1 M NaOH for 1 hour at 37°C. Free acetate was quantified following neutralisation with citric acid.

The pH optima of the enzymes were determined by measuring specific activity with kGMOs in a series of buffers ranging from pH 2-10, using citrate from pH 2.0-5.5, phosphate from pH 6.0-8.0 and glycine from pH 8.5-10.0. The thermal stability of the enzyme was determined by measuring specific activity with kGMOs following 1 hour of incubation in PBS at temperatures ranging from 25-70°C. Acetonitrile tolerance was measured using PNP-Ac as the substrate. A final substrate concentration of 1 mM was maintained across all measurements,

allowing as little as 1% acetonitrile to be tested. Supplemental acetonitrile was added to the buffer immediately prior to enzyme and substrate addition.

Oligosaccharides deacetylated by PbCExA and PbCE7A were identified by LC-MS/MS. Following separation using the gradient described above, limit-digest preparations of kGMOs and XOs were analysed by collision-induced dissociation MS/MS performed on the sodiated adducts of acetylated oligosaccharides using 40 V of collision energy. A limit-digest preparation of ¹⁸O-XOs was analysed with fragmentation of the M+2 peak to determine the site of acetylation in oligosaccharides selectively deacetylated by PbCExA.

5.2.5 Enzyme Crystallisation and X-ray Diffraction

PbCExA was expressed and purified in E. coli as selenomethionine-derivatised protein using our standard purification protocol as described previously (Quaile et al., 2018). Crystals were grown at 23°C using the hanging-drop vapor diffusion method by mixing 2 μ L of 50 mg/mL protein with 2 µL of reservoir solution containing 50 mM MES pH 6.5, 0.2 M sodium tartrate and 20% (w/v) PEG3350. Crystals were cryo-protected with reservoir solution supplemented with 1% ethylene glycol. X-ray diffraction data was collected at 100 K at beamline 21ID-F at LS-CAT, Advanced Photon Source at the wavelength of 0.97872 Å (selenium absorption peak). Diffraction data was reduced with XDS (Kabsch, 2010) and Aimless (Winn et al., 2011). Phenix.autosol (Adams et al., 2010) identified 20 of the 27 selenomethionine sites and model building was completed using Phenix.autobuild, Phenix.refine and *Coot* (Emsley *et al.*, 2010). The final structure includes three copies of PbCExA with all residues resolved plus one additional residue from the cleaved N-terminal TEV protease cleavage site. All geometries were verified with *Phenix.refine* and the wwPDB validation server. Structure coordinates were deposited to the Protein Databank. Structural orthologues were identified using the Dali lite server (Holm and Rosenström, 2010) and oligomerisation interactions identified by PDBePISA (Krissinel and Henrick, 2007).

5.2.6 Bioinformatic Analysis

The subcellular localisations of PbCE7A and PbCExA were predicted *LipoP* (Juncker *et al.*, 2003). Neither enzyme was predicted to be attached to the membrane through a lipid anchor due to their lack of a cysteine residue following a predicted signal peptide cleavage site (Paetzel *et al.*, 2002).

Homologues of PbCExA and PbCE7A were identified through BLAST searches of the NCBI RefSeq database. Bioinformatic analysis was performed on PbCExA and PbCE7A homologues selected on the basis of E-value cut-offs chosen using a plot of change in E-value vs. hit number. Protein sequence alignments were performed using the MUSCLE algorithm in MEGA6 v6.06 (Tamura *et al.*, 2013) with the UPGMB clustering method. Unrooted phylogenetic trees were derived from the resulting sequence alignments using the maximum likelihood method in MEGA6. The reliability of each tree was tested by bootstrap analysis using 100 resamplings of the data set.

Amino acid sequences of NanS (Rangarajan *et al.*, 2011) and characterised CE6 and CE7 enzymes with confirmed activities or 3-dimensional structures were extracted from the CAZy database and trimmed using *ScanProsite* (Castro *et al.*, 2006). PbCExA, NanS, and characterised CE6 enzymes were aligned using the Expresso method (Armougom *et al.*, 2006). PbCE7A and characterised CE7 enzymes were aligned in the same manner. The resulting alignments were manually refined using the alignment explorer in MEGA6 v6.06. An unrooted phylogenetic tree was derived from resulting PbCExA alignment using the maximum likelihood method in MEGA6. The reliability of the tree was tested by bootstrap analysis using 100 resamplings of the data set. Structurally characterised CE2 and CE3 enzymes were included as outgroups.

5.3 Results

Building on our recent identification of a β-mannan utilisation locus (MUL) from *P*. *bryantii* B₁4, we have cloned and recombinantly produced two putative esterases, PbCE7A and PbCExA. Predicted signal peptides suggest that these esterases are free-floating in the periplasm, co-localised with a variety of *exo*-glycosidases responsible for the saccharification of (galacto)glucomannan oligosaccharides generated by PbGH26A-GH5A, the vanguard *endo*-mannanase/*endo*-glucanase. The initial annotation of the *P. bryantii* B₁4 genome performed by Purushe *et al.* (2010) called PbCE7A a cephalosporin C deacetylase and did not make a prediction for the function of PbCExA. PbCExA has since been annotated as a sialate *O*-acetylesterase on the basis of similarity to NanS. Considering the genomic localisation of PbCE7A and PbCE7A and PbCExA.

5.3.1 **Primary Structure Analysis and Phylogeny**

1904 sequences, for none of which, to our knowledge, have the structure or function been experimentally determined, were extracted from the NCBI RefSeq database using the catalytic domain (E106-G374) of PbCExA as a BLAST search sequence with an E-value cut-off of 10⁻²⁰. The only activity predicted for the returned sequences was sialate *O*-acetylesterase. A plot of change in E-value, taken as a measure of evolutionary distance, vs. hit number (smoothed with a 5 sequence moving average) shows two major peaks at hits number 61 and 130 (Figure C-2A). A tree constructed from the first 130 BLAST results (sharing at least 40% sequence identity to PbCExA) contains 2 clades of esterases (Figure C-3). Clade I contains almost exclusively sequences identified in *Prevotella* species and clade II contains sequences primarily from *Bacteroides* and *Alistipes* species.

Analogues of the four motifs characteristic of SGNH hydrolases (I: GDSL, II: NxxxxG, III: GxN, IV: DxxH (Mølgaard *et al.*, 2000)) are present in the structure-guided alignment of PbCExA (Figure 5-1A). Motif I, containing the catalytic serine residue, is conserved between PbCExA, NanS, and CE6 esterases as GQSN. Motifs II and IV are not conserved with the exception of the active site glycine and catalytic histidine, respectively. Motif III is replaced by QGE in NanS and among CE6 esterases, but uniquely, QGC in PbCExA. Tree construction on the basis of this alignment successfully segregates CE2 and CE3 away from CE6, but does not segregate PbCExA and NanS away from the rest of CE6 (Figure 5-1B).

517 sequences, primarily of predicted acetyl-xylan esterases, were extracted from the RefSeq database as above using the catalytic domain (C117-E427) of PbCE7A as the search sequence. Of the sequences extracted, a single sequence, TmAcE (GenBank: AAD35171.1) from *Thermotoga maritima*, has been experimentally shown to be active on cephalosporin C and acetylated monosaccharides (Levisson *et al.*, 2012). A plot constructed as above for these hits gives three peaks at hits number 78, 246, and 370 (Figure C-2B), suggesting the presence of three clades. Notably, TmAcE is hit number 451 indicating significant evolutionary distance from PbCE7A. Enzymes most closely related to PbCE7A (hits 1-78) were consistently identified in the genomes of *Prevotella* and *Bacteroides* species (Figure C-4). Hits 79-246 were also primarily identified in *Prevotella* and *Bacteroides* species, suggesting that they likely share a distinct function. Hits 245-370 were identified in a diverse collection of Bacteroidetes species, including many marine bacteria.



Figure 5-1: Molecular phylogeny of PbCExA. A) Structure-guided alignment of PbCExA, NanS (PDB ID: 3PT5), two structurally characterised CE6 esterases (PDB ID: 1ZMB, 2AEA), six CE6 acetylxylan esterases, two CE2 acetylxylan esterases (PDB ID: 2WAO, 4XVH), and two CE3 acetylxylan esterases (PDB ID: 2VPT, 5B5L). The four motifs associated with SGNH hydrolases are labelled and the catalytic serine and histidine residues are marked with arrows. Secondary structure (mapped from PbCExA) is shown above each row of sequence data. Panel A was prepared using *ESPript 3.0* (Robert and Gouet, 2014). B) Unrooted maximum-

likelihood phylogenetic tree constructed on the basis of the alignment in panel A. Bootstrap values (out of 100) are provided next to each node.

An alignment of structurally or functionally characterised CE7 enzymes shows that PbCE7A shares the catalytic machinery and secondary structure features of CE7. S295, found in the conserved GxSQGG motif, H406, found in the semi-conserved GHE motif, and D377, found in the semi-conserved GLxD motif (Vincent *et al.*, 2003) make up the catalytic triad of CE7 (Figure C-5)).

5.3.2 PbCExA and PbCE7A Activity Optima and Substrate Specificities

To identify the substrates of PbCExA and PbCE7A we tested for activity with a variety of compounds selected in light of their genomic co-localisation with β -mannan-degrading enzymes, including kGM, kGMOs, and 4-nitrophenyl derivatives of acetate, butyrate, α -galactose, β -glucose, and β -mannose. We also tested PbCE7A for activity with cephalothin, a synthetic analogue of cephalosporin C. We did not test PbCExA for *N*-acetyl-9-*O*-acetylneuraminic acid esterase activity due to poor homology of NanS (15% ID, *vide infra*) to PbCExA and a lack of appropriate substrate.

Incubating PbCExA and PbCE7A with PNP-Ac led to the formation of a deep yellow colour, indicating hydrolysis of the substrate. Parallel incubations with 4-nitrophenyl carbohydrate derivatives of α -galactose, β -glucose, and β -mannose led to no measurable colour formation. Kinetic experiments were performed with PbCE7A and PbCExA using PNP-Ac and PNP-Bu following investigations of the pH and temperature optima, and acetonitrile tolerances of each enzyme (Figure C-6). PbCE7A clearly demonstrated specificity towards PNP-Ac ($k_{cat} = 26 \pm 1 \text{ s}^{-1}$, $K_M = 0.39 \pm 0.02 \text{ mM}$, $k_{cat}/K_M = 67000 \text{ M}^{-1}\text{s}^{-1}$) over PNP-Bu ($k_{cat} = 0.30 \pm 0.01 \text{ s}^{-1}$, $K_M = 0.44 \pm 0.04 \text{ mM}$, $k_{cat}/K_M = 680 \text{ M}^{-1}\text{s}^{-1}$), demonstrating that PbCE7A disfavours longer acyl chains (Figure C-7). Similarly, PbCExA displayed a 150-fold preference for PNP-Ac ($k_{cat} = 45 \pm 2 \text{ s}^{-1}$, $K_M = 0.8 \pm 0.1 \text{ mM}$, $k_{cat}/K_M = 56000 \text{ M}^{-1}\text{s}^{-1}$) over PNP-Bu ($K_M > 10 \text{ mM}$, $k_{cat}/K_M = 370 \text{ M}^{-1}\text{s}^{-1}$) (Figure C-7).

Overnight incubation of PbCE7A with cephalothin yielded no measurable free acetate, indicating a lack of cephalosporin C deacetylase activity and revealing the misannotation of this enzyme. Similar incubations with kGM yielded free acetate equal to the total acetate released following treatment of the substrate with 0.1 M NaOH. Thus, PbCE7A is able to fully

deacetylate kGM. Due to prediction of *LipoP* that PbCE7A and PbCExA are free-floating in the periplasm, we measured their specific activities with kGMOs as well. PbCE7A was 4-fold more active with kGMOs (0.14 ± 0.01 U/mg) than with kGM (0.03 ± 0.003 U/mg). Incubations of PbCExA with kGM yielded no detectable free acetate. However, incubations with kGMOs showed significant acetate release, with a specific activity of 7.5 ± 0.5 U/mg. Similar experiments with xylan and XOs revealed that PbCE7A is also able to fully deacetylate acetylated xylan substrates, having specific activities of 0.18 ± 0.02 U/mg on acetylated xylan and 0.55 ± 0.03 U/mg. Similar experiments with a specific activities of 0.18 ± 0.02 U/mg on acetylated xylan and 0.55 ± 0.03 U/mg on XOs, and that PbCExA is also able to deacetylate XOs with a specific activity of 5.0 ± 0.4 U/mg.

5.3.3 Site Selectivity of PbCExA

Overnight incubation with PbCExA resulted in the release of only 55% of the total acetate content present in kGMOs and 18% of the total acetate content present in XOs. LC-MS analysis of acetylated kGMOs and XOs before and after treatment with PbCExA revealed clear specificity towards some acetylated oligosaccharides over others (Figure 5-2A). Due to the relative simplicity of the XOs mixture and the apparent lack of kinetic preference for kGMOs over XOs, XOs were labelled on their reducing ends with ¹⁸O for MS/MS structural analysis.

Of two ¹⁸O-labelled acetylated xylotriose species present in the XOs mixture, only one was degraded by PbCExA. MS/MS spectra were recorded for the M+2 peak of the sodiated adducts of each acetylated xylotriose species revealed that PbCExA is able to deacetylate xylotriose acetylated at the 3-position of the second xylose residue, but is not able to deacetylate xylotriose acetylated at the 2-position of the non-reducing-terminal xylose residue (Figure 5-2B).


Figure 5-2: Analysis of oligosaccharides deacetylated by PbCExA. A) LC-MS traces of acetylated trisaccharides present in the kGMO (left) or XO (right) mixtures before and after treatment with PbCExA. B) Annotated MS/MS spectra of two ¹⁸O-labelled acetylated xylotriose molecules found in the XO mixture.

5.3.4 Tertiary Structural Analysis of PbCExA

To clarify the functional significance of the sequence variations observed in the primary structures of PbCE7A and PbCExA we attempted to crystallise both enzymes in the apoenzyme and substrate-bound states. We were not able to obtain diffraction patterns from any PbCE7A crystals. Soaking and co-crystallisation experiments also yielded no complexes. However, we were successful in solving the crystal structure of PbCExA in the apoenzyme form using a

selenomethionine-derivatised crystal and the single anomalous dispersion diffraction method (X-ray crystallographic statistics can be found in Table C-2).

The asymmetric unit of the PbCExA crystal contained three copies of the protein, each of identical overall structure, arranged in a triangular configuration (each chain contacts two other chains with ~1000 Å³ buried surface area). The structure contains a central catalytic domain (between residues 106-379) clearly adopting the SGNH esterase fold, and N- and C-terminal all- β domains (Figure 5-3A). The N-terminal domain, which is connected to the catalytic domain via a long shared β -sheet, adopts a fibronectin(III)-like fold with structural similarity (RMSD 2.54 Å over 72 matching C α atoms to PDB ID: 3PE9) to a region of *Clostridium thermocellum* cellobiohydrolase A that imparts thermostability (Brunecky *et al.*, 2012). The C-terminal domain of PbCExA shows structural similarity with a CBM21 domain found in *Rhizopus orzyae* glucoamylase (RMSD 3.47 Å over 78 matching C α atoms to PDB ID: 4BFN).

The catalytic Ser114-His353 pair of PbCExA of PbCExA was localised to an extended cleft bounded by three prominent arms (Figure 5-3A). Surprisingly, the imidazole ring of His353 is found 3.5 Å away from the sulfur atom of Cys126, suggesting the presence of a hydrogen bond forming a Ser-His-Cys catalytic triad which has not been observed previously (Bergen *et al.*, 2016). The four amino acid sequence motifs defining SGNH esterases are localised to the active site: motif I (Gly111, Gln112, Ser113 and Asn114); motif II (Gly204); motif III (Gln264, Gly265, Cys266, Ser267, Asn268); and motif IV (His353) (Figure 5-3B). Notably, PbCExA lacks the "N" from SGNH motif III.

A close analysis of the active site revealed extended electron density features on each Ser114, which we modelled as an acetyl group as in the active site of a *Bacillus pumilus* CE7 (PDB ID: 3FYU) (Figure 5-3B). The presence of the *O*-acetylserine residue enables the identification of key features comprising the catalytic apparatus of this enzyme: the acetyl group oxygen formed hydrogen bonds with Gly204 from motif II. This is a typical SGNH interaction, forming an oxyanion hole to stabilize the tetrahedral intermediate. This oxygen is also positioned to form a hydrogen bond with the primary amide group of Gln103. This latter interaction replaces the "N" component of the oxyanion hole in canonical SGNH hydrolases, occupying the same spatial location.

157



Figure 5-3: Structural analysis of PbCExA. A) Overall structure of PbCExA in solvent-exposed surface representation; domains, active site and N- and C-termini are labelled. B) Catalytic domain of PbCExA. Zoom shows key active site residues coloured by four motifs defining SGNH esterases plus two additionally

conserved hydrophobic residues. Electron density for *O*-acetylserine-114 is F₀-F_c density before modeling sidechain, contoured at 1.0 σ. C) Comparison of PbCExA with other CE enzymes At4g34215 CE6 (PDB ID: 2APJ), *C. acetobutylicum* CE6 (PDB ID: 1ZMB), and *E. coli* NanS (PDB ID: 3PT5). The zoom shows active site homology. Bolded residue labels indicate identical residues across all four enzymes. The PMSF modification on the catalytic serine of *C. acetobutylicum* CE6 is hidden for clarity.

Given this divergent catalytic architecture, we were interested in identifying structural orthologues of PbCExA that may share this active site composition. A structure similarity search performed using the catalytic domain of PbCExA identified both of the crystallised CE6 enzymes as close orthologues – At4g34215 from *Arabidopsis thaliana* (PDB ID: 2APJ, RMSD 2.2 Å over 199 matching C α atoms, and 24% sequence identity) and CE6 from *Clostridium acetobutylicum* (PDB ID: 1ZMB, RMSD 2.3 Å over 199 matching C α atoms, and 20% sequence identity) – as well as *E. coli* NanS, a *N*-acetyl-9-*O*-acetylneuraminic acid esterase, as a structural orthologue (PDB ID: 3PT5, RMSD 2.8 Å over 203 matching C α atoms, and 15% sequence identity). Figure 5-3C shows that the higher structural dissimilarity of NanS is related to the extended loops surrounding the active site. However, the active site architectures of these relatively divergent enzymes are nearly identical, sharing the glutamine-containing active site architecture described above.

Our structure similarity search also identified CE2 enzymes (a CE2 from *Corynascus thermopiles*, PDB ID: 4XVH, RMSD 2.9 Å over 176 matching Cα atoms, and 11% sequence identity; and a CE2 from *Ruminiclostridium thermocellum*, PDB ID: 2WAO, RMSD 2.5 Å over 165 matching Cα atoms, and 14% sequence identity) and a CE3 enzyme (from *Ruminiclostridium thermocellum*, PDB ID: 2VPT, RMSD 2.5 Å over 163 matching Cα atoms, and 8% sequence identity). While these enzymes share the SGNH fold with PbCExA, they conserve the traditional asparagine-containing SGNH active site architecture.

5.4 Discussion

In light of the abundance of sequenced carbohydrate esterase genes and the widespread distribution of acetylated polysaccharides, there is a pressing need to develop a comprehensive understanding of the role esterases play in the breakdown of plant matrix glycans (Gille and Pauly, 2012). Enzyme structure, function, and phylogeny are essential components of this

understanding. Building on our characterisation of the β -mannan utilisation locus from *P*. *bryantii*, we have determined the functions of two carbohydrate esterases.

5.4.1 PbCE7A and PbCExA Display Complementary Acetylesterase Specificities

Based on similarity to NanS, PbCExA was annotated as a sialate *O*-acetylesterase in the NCBI GenBank and PbCE7A was annotated as a cephalosporin C deacetylase, an antibiotic resistance gene. This is surprising considering that two structurally and functionally characterised CE7 enzymes, BpAXE from *Bacillus pumilus* and CAH from *Bacillus subtilis*, exhibit both acetylxylan esterase and cephalosporin deacetylase activities within the context of xylan-degrading operons (Vincent *et al.*, 2003, Montoro-García *et al.*, 2011). However, the closest functionally characterised homologue of PbCE7A, TmAcE from *Thermotoga maritima*, was shown to be active on cephalosporins and acetylated monosaccharides, but not acetylated xylan (Levisson *et al.*, 2012). The greater similarity between PbCE7A to TmAcE led us to the hypothesize that PbCE7A would be similarly polyspecific.

Explorations of the potential activities of PbCExA and PbCE7A in the context of βmannan degradation revealed clear glucomannan acetylesterase activity. Unlike other characterised CE7 esterases (Vincent *et al.*, 2003, Singh and Manoj, 2016), PbCE7A is able to fully deacetylate polysaccharide substrates. Considering the known diversity of acetylation sites in kGM and xylan, this demonstrates that PbCE7A is able to deacetylate equatorial 2- and 3positions of xylose residues in xylan, as well as the axial 2-position and primary 6-position of mannose residues in glucomannan. Interestingly, PbCE7A is a 10-fold better esterase than PbCExA when acting on PNP-Ac. Although it is not as effective against the oligosaccharide substrates which would be found in the periplasm during glucomannan catabolism, being 50-fold less active towards kGMOs than PbCExA. This suggests that PbCE7A is not as dependent on carbohydrate recognition for catalysis and, as a corollary, is not as able to take advantage of specific carbohydrate interactions to catalyse oligosaccharide deacetylation.

The closest homologues of PbCExA which have been functionally characterised and classified within CAZy are members of CE6. This family only contains characterised members which have been reported to display acetylxylan esterase or acetylmannan esterase activities. Among the first demonstrated acetylxylan esterase activities in what is now CE6 was BnaA from *Neocallimastix patriciarum*, an anaerobic ruminal fungus (Dalrymple *et al.*, 1997). BnaA

deacetylated partially chemically acetylated birch wood xylan. This activity was further identified in Axe6A and Axe6B from *Fibrobacter succinogenes* (McDermid *et al.*, 1990).

In contrast to these reported activities, PbCExA was not able to accommodate long polysaccharide substrates, but released acetate from partially acetylated *endo*-glycanase-degraded oligosaccharides. The comparable specific activities of PbCExA towards xylan- or glucomannan-derived oligosaccharide substrates demonstrate its ability to accommodate pentoses or hexoses with variable stereochemistry at the 2-position. Follow-up investigations into the specificity of PbCExA using ¹⁸O-labelled oligosaccharides yielded insight into the selectivity associated with the incomplete deacetylation of kGMOs and XOs by PbCExA. The enzyme completely deacetylated xylotriose acetylated at the 3-position of the middle xylose residue without hydrolysing any of the xylotriose acetylated at the 2-position of the non-reducing terminal xylose residue. It is not clear from these data whether the enzyme is selective for a specific residue within an oligosaccharide or a specific position of acetylation. However, based on its inability to deacetylate kGM, its lack of selectivity between kGMOs and XOs, and its open pocket active site structure, we believe that PbCExA is likely selective for the 3-position of monosaccharide residues adjacent to the reducing or non-reducing end.

Within the context of the *P. bryantii* β -mannan utilisation locus the localisation of PbCE7A and PbCExA suggests that acetylated oligosaccharides, formed through the action of PbGH26A-GH5A, are imported into the periplasm. Once there, the observed specificities of PbCExA and PbCE7A may serve to facilitate the actions of other periplasmic *exo*-glycanases, including PbGH36A, PbGH26C, PbGH5B, and PbGH3A. Furthermore, the combination of low activity and broad substrate range displayed by PbCE7A is complementary to the combination of high activity and narrow substrate range displayed by PbCExA. Thus, our evidence suggests that each of these esterases fills a distinct niche in the catabolism of complex acetylated glucomannans.

5.4.2 PbCE7A is Part of a Distinct Group within CE7

The sequence alignment shown in Figure C-5 shows that PbCE7A possesses all of the characteristic sequence motifs of CE7 esterases. PbCE7A also shares secondary structure with TmAcE and 20-30% sequence identity with each characterised CE7 esterase in CAZy. Following from our functional characterisation of PbCE7A, it was unsurprising that PbCE7A and TmAcE, displaying activity with cephalosporins and not acetylated xylan, were phylogenetically

segregated in our analysis (Vincent *et al.*, 2003). Examination of the genomic contexts of extracted PbCE7A homologue sequences within the Joint Genome Institute Integrated Microbial Genome browser suggests that the three apparent groups of PbCE7A homologues do not share common function. Only the sequences most similar to PbCE7A (hits 1-78) share the same genomic context as PbCE7A, being co-localised with other genes encoding putative β -mannan or xylan degrading enzymes. On this basis, we predict that these enzymes form a distinct group within CE7 which share the ability to deacetylate both xylans and β -mannans. In support of this, a maximum likelihood tree generated from an alignment of this set of sequences is monophyletic (Figure C-4). The second clade of PbCE7A homologues (hits 79-246) contains genes similar to PBR_1566 from *P. bryantii* B₁4. This gene is co-localised in the genome with a PL1 pectate lyase and a TonB-dependent transporter, suggesting that the second clade contains CE7 pectin acetylesterases displaying activity related to the CE12 rhamnogalacturonan acetylesterase (Leeuwen *et al.*, 1992, Mølgaard *et al.*, 2000).

5.4.3 PbCExA Activity is Minimally Impacted by the Polarisation of H353 by C126

Only two structures of CE6 esterases have been determined: the unpublished structure of CAC0529 (PDB ID: 1ZMB) from *Clostridium acetobutylicum* and the structure of At4g34215 (PDB ID: 2APJ) from *Arabidopsis thaliana* which was published without functional characterisation (Bitto *et al.*, 2005). Thus, there is currently no CE6 member for which both structure and function are known. Notably, NanS, a currently unclassified SGNH hydrolase from *Escherichia coli* displaying 9-O-sialic acid esterase activity, bears significant structural similarity to At4g34215 and CAC0529 (Rangarajan *et al.*, 2011). NanS shares the QGSN motif I of CE6 which is found in place of the canonical GDSΦ (following the nomenclature of Aasland *et al.* (2002)) motif I of SGNH hydrolases. It employs an unusual catalytic dyad with no clearly identifiable residue which orients and polarizes the catalytic histidine residue (Berg *et al.*, 2002).

Similar to NanS and some CE2 enzymes (Montanier *et al.*, 2009b, Rangarajan *et al.*, 2011), PbCExA is missing the canonical aspartate or glutamate residue which orients and polarizes the imidazole group of the catalytic histidine. Evidence presented by Montanier *et al.* (2009b) suggests that CjCE2B and CtCE2 replace this functionality with a carbonyl oxygen from a backbone amide group. We observe that PbCExA replaces the canonical carboxylate group with the sulfhydryl group of Cys126. The functional significance of this arrangement is supported by its conservation across 119 of the 130 identified homologues of PbCExA (Figure

C-3). Since the hydroxyl sidechain of Ser114 is a hydrogen bond donor to His353 in the precatalytic state, His353 must be a hydrogen bond donor to Cys126. Theoretical calculations provide support for the favourability of such an interaction (Mazmanian *et al.*, 2016) and predict a pKa-lowering effect of a hydrogen bond to the sulfhydryl group of a cysteine residue (Naor and Jensen, 2004). Thus, we tested whether Cys126 participates in a catalytic triad as a thiolate which orients and polarizes His353 as part of our mutational analysis of the PbCExA active site.

5.4.4 PbCExA is a Founding Member of a New CE Family

Our search for homologues of PbCExA on the basis of sequence similarity returned no characterised enzymes. On the basis of changes in E-value, we identified 130 homologues of PbCExA within the RefSeq database (Figure C-3). The two clades formed by these sequences roughly recapitulate the phylogenetic relationships between the genera in which they have been identified, suggesting that they form a small family of functionally related enzymes. Regrettably, these homologues have all been assigned as putative sialate O-acetylesterases on the basis of similarity to NanS. This reflects a relationship between NanS, PbCExA, and CE6 esterases which is not unlike clan membership among glycoside hydrolases (Henrissat and Bairoch, 1996). The X-ray crystal structure comparison in Figure 5-3C shows that NanS, PbCExA, CAC0529, and at4g34215 all share a common fold, catalytic Ser-His pair, and variation on the SGNH motif which dictates active site structure. Furthermore, the inclusion of CE2 and CE3 in the alignment in Figure 5-1A shows variations in the core SGNH motifs and active site structure which clearly indicate that CE2 and CE3 are part of a group distinct from CE6, NanS, and PbCExA. Thus, our analysis likewise suggests that SGNH hydrolases should be classified into two groups, group I having GDS Φ in motif I and group II having QGSN in motif I, as noted in the analysis of NanS (Rangarajan et al., 2011). A maximum likelihood tree constructed from our alignment of the structurally or functionally characterised members of group II failed to segregate PbCExA, NanS and the rest of CE6 (Figure 5-1B). We attribute this to the overall low sequence identity (<25%) between the structurally characterised enzymes included in the alignment. Thus, we believe that a new organisation of group II SGNH hydrolases is needed within CAZy to segregate NanS, PbCExA, and the existing CE6 enzymes into three distinct families.

5.5 Conclusions

We report the characterisation of two acetylmannan/acetylxylan esterases from a βmannan utilisation locus which display complementary activities. Unlike other CE7 esterases, PbCE7A is a broad specificity enzyme, capable of hydrolyzing all of the acetate groups found on konjac glucomannan or acetylated xylan. In contrast, PbCExA is an acetylesterase which targets specific sites of acetylation within certain short acetylated oligosaccharides. The predicted localisation of these enzymes suggests that, alongside PbGH3A, PbGH5B, PbGH26C, and PbGH36A participate in the saccharification of oligosaccharides found in the periplasm during glucomannan catabolism. The novel Ser-His-Cys catalytic triad found in the active site of PbCExA raises several questions about the mechanism of PbCExA-like esterases. For example, it remains to be experimentally determined what impact Cys126 has on the rates of acyl-enzyme formation and hydrolysis in the catalytic cycle. Going forward, we believe that the data presented here support the classification of roughly 130 esterases into a new family of group II SGNH acetylmannan/acetylxylan esterases sharing a clan membership-like relationship to CE6 and NanS-like sialate *O*-acetylesterases.

Chapter 6: General Conclusions

6.1 PbGH5A and the Discovery of the *P. bryantii* β-Mannan Utilisation System

Decades of enzyme discovery and characterisation have produced a large collection of well-characterised carbohydrate-degrading enzymes displaying diverse chemistries and specificities (Lombard *et al.*, 2014). However, much remains to be discovered. Revisiting PbGH5A with modern tools, techniques, and a broader library of substrates and probes has allowed a deeper dive into what dictates enzyme specificity. It has also allowed us to probe the limits of function prediction based on amino acid sequence.

PbGH5A was the only cellulase found in a screen for cellulases from *Prevotella bryantii*, a ruminant gut bacterium (Matsushita et al., 1990). It was investigated heavily as a cellulase because of its initially identified CMCase activity. However, building on the more recent characterisations of PpXG5 (Gloster et al., 2007) and BoGH5A (Larsbrink et al., 2014a), two highly specific endo-xyloglucanases closely related to PbGH5A, we hypothesised that the enzyme was a poor cellulase because it was part of a molecular family containing members which specifically degrade the larger and more complex xyloglucan molecules associated with cellulose (Aspeborg et al., 2012). Chapter 3 describes the detailed analyses of the structure, function, and inhibition of PbGH5A, revealing the full-range of this enzyme's glucan-degrading capacity. Measuring apparent Michaelis-Menten kinetic constants for PbGH5A acting on a variety of oligosaccharides and polysaccharides containing β-glucan motifs revealed that PbGH5A is an exceptionally efficient mixed-linkage glucanase, rivalling the archetypical *Bacillus* licheninase. It displays an impressive catalytic flexibility, cleaving both $\beta(1,3)$ and $\beta(1,4)$ glucosidic bonds, while also accommodating the complex branching pattern of tamarind xyloglucan and the mannan-containing backbone of konjac glucomannan. Its uncommon mode of mixed-linkage glucan recognition gives it further potential value in the production of mixedlinkage glucan oligosaccharides and cellooligosaccharides. However, contrary to my thesis, a clade of functionally related GH5 enzymes could not be delineated on the basis of differences between PbGH5A and the known GH5 xyloglucanases.

The growth of next-generation sequencing technology and the coincident rise of the polysaccharide utilisation locus paradigm, shortly after David Wilson and James Russell's work

on PbGH5A came to a close, brought new light and new structure to our understanding of carbohydrate-degradation in Bacteroidetes (Martens *et al.*, 2009). David Wilson and James Russel put significant effort into determining the sequence of the DNA fragment from which PbGH5A had originally been isolated, yet this provided a limited view (Gardner *et al.*, 1997). The recent sequencing of the genome of *P. bryantii* B₁4 by the North American Consortium for Rumen Bacteria revealed the complete genetic context of PbGH5A: its fusion to a GH26 β mannanase domain, its association with a TonB-dependent transporter, and its association with a much larger collection of genes possessing activities devoted to the saccharification of complex β -mannans (Purushe *et al.*, 2010). What was an investigation of the function of a single gene became an investigation of a complex system virtually unique to *P. bryantii*.

Synthesizing the work of Russell and Wilson with new experimental data and functional data from several recently characterised homologues of genes from this system has enabled the construction of a complete model for the saccharification of complex galactomannans and glucomannans in *P. bryantii*. Exploration of the phylogenetics of various elements within this model made it possible for us to identify similar systems across the Bacteroidetes phylum and will aid in the prediction of β -mannan utilisation capacities in bacterial genomes going forward. Unexpectedly, we found the traditional use of homologous SusC/SusD-like genes to identify PULs targeting similar substrates to be ineffective. We put forward that the presence of a three-gene cluster, originally identified by Senoura et al. (2011), is the strongest predictor of β -mannan utilisation in Bacteroidetes.

Two new carbohydrate esterases have been identified within the *P. bryantii* β-mannan utilisation locus. PbCE7A, homologous to a CE7 antibiotic resistance gene, has been shown to be an acetylmannan/acetylxylan esterase which completely deacetylates the glucomannan oligosaccharides generated by the action of PbGH5A and PbGH26A. Similarly, PbCExA is shown to also be an acetylmannan/acetylxylan esterase. This enzyme displays high specificity towards short acetylated oligosaccharides. Mass spectrometry showed that this enzyme targets specific acetylated oligosaccharides. A crystal structure of this new carbohydrate esterase has been presented alongside a structure-guided phylogenetic analysis which supports the formation of a new carbohydrate esterase family.

As shown by the above example, continued enzyme characterisation efforts, with a focus on PULs and dark spots in phylogenetic trees, will be invaluable to sorting and understanding the diverse carbohydrate-degrading capacities and strategies within growing repositories of microbial genomes. The structure-function study and phylogeny of PbGH5A presented in Chapter 3 demonstrates the power and the limits of function prediction. The increase in the number of characterised enzymes in the CAZy database made it possible for us to generate a hypothesis not considered 20 years ago: that PbGH5A is not a cellulase, but a xyloglucanase. This hypothesis was based on its inclusion in GH5 subfamily 4, the only subfamily containing characterised xyloglucanases. While this led us in a good direction, our prediction was not supported by our experimental evidence. PbGH5A is a poly-specific enzyme, displaying high enzyme performance with many substrates containing β -glucan motifs. We hoped that understanding the structure-function relationship in PbGH5A would enable the prediction of the specificities of homologous enzymes. However, we found no defining structural feature or global sequence similarity metric which enabled the delineation of specific xyloglucanases from PbGH5A. Thus, while structure-function analyses of homologues are an invaluable starting point for enzyme characterisation, it appears that the precision with which we may predict the functions of enzymes on the basis of sequence is fundamentally limited.

There is a growing body of evidence which suggests that the abundance of *Prevotella* species in the gut is positively correlated with the quantity of non-starch polysaccharides being consumed (Wu et al., 2011, Smits et al., 2017). Furthermore, the consumption of non-starch polysaccharides is commonly associated with positive indicators of gut health (Kovatcheva-Datchary et al., 2015), suggesting a potential link between the abundance of Prevotella species and gut health. However, causal factors which may explain these observations have not yet been found. Thus, further investigations of the biology of Prevotella species in comparison to the biology of Bacteroides species are needed. B. ovatus and B. thetaiotaomicron have been established as model carbohydrate fermenters in the gut, however, no similarly established model Prevotella species exists. Significant efforts to develop an understanding of metabolic systems within P. bryantii are still needed. The recent sequencing of the genome of P. bryantii revealed an abundance of CAZymes with as-yet-unknown functions. The ability to predict and test the functions of these genes is hindered by a dearth of transcriptomic data and tools for genetic manipulation. The work of Dodd et al. (2010b) represents the only transcriptomic data collected for P. bryantii to date and extensive efforts to generate stable recombinant strains of P. bryantii have yet to yield the tools necessary to perform targeted reverse genetics experiments (Fields et

al., 1997, Accetto and Avguštin, 2001, Accetto *et al.*, 2005, Accetto and Avguštin, 2007, Accetto and Avguštin, 2011). Thus, going forward, the work of Dodd *et al.*(2010b) should be expanded with transcriptomic data collected for *P. bryantii* with a variety of defined carbon substrates similar to what has been collected for *B. ovatus* and *B. thetaiotaomicron* (Martens *et al.*, 2011). Efforts to develop genetic tools to manipulate *P. bryantii* should be expanded so that targeted knockouts and gene fusions can be constructed. Such data and tools will facilitate the discovery and characterisation of novel proteins and, eventually, lead to the development of a comprehensive understanding carbohydrate metabolism in *P. bryantii*. Furthermore, comparing *P. bryantii* with model *Bacteroides* species will contribute to a causal understanding of the relationship between *Prevotella* species abundance, gut health, and non-starch polysaccharide consumption.

6.2 EG16s are of Emerging Interest in Fundamental Biology

The second chapter of this thesis explored the structure-function relationships in a newlydiscovered group of plant glycoside hydrolases, dubbed "EG16s," which compose a recently discovered enzyme family related to the *xyloglucan endo-transglycosylases/hydrolases* (XTHs) involved in plant cell wall remodelling. As noted in the Introduction, plant genomes encode a plethora of glycoside hydrolases. Many of these genes, such as those involved in remodelling xyloglucan in the cell wall, are part of large families of functionally-related genes encoded as multiple orthologues within the same genome. Thus, it is uncommon to find a phylogenetically distinct group of enzymes, found in almost all plant genomes, which is only encoded once per genome. For this reason and because of their similarity to XTHs, EG16s were only recently identified (Eklöf et al., 2013). Close inspection showed that they are missing key structural elements which underpin XTH gene product function. Furthermore, a lack of introns, apparent post-translational modification sites, or signal peptides suggests that these enzymes are not controlled through many of the complex regulatory systems identified in eukaryotes to date. These enigmatic "EG16" genes presented a rare opportunity to understand what appears to be a "core" enzyme within plants; one which has been passed down without significant modification since the appearance of the first land plants.

Since little is known of their biology, a critical step in characterizing the EG16 family was the production of one of its members in sufficient quantity. The successful recombinant production of PtEG16 and VvEG16 in *E. coli* was invaluable to the first functional studies of

EG16s. Biochemical studies of these enzyme revealed a remarkably broad substrate specificity: they cleave both the linear mixed-linkage $\beta(1,3)/\beta(1,4)$ -glucans found in grassy monocots (e.g. wheat, corn), and the complex, branched (fucogalacto)xyloglucans which are enriched in dicots (e.g. poplar, grape). Wanting to understand the active site structure which underpins this enzyme specificity, a library of oligosaccharide substrates derived from plant materials was prepared. A high-performance liquid chromatography-based assay was developed to quantify interactions between the enzyme active site and this library of substrates. This allowed the construction of a functional map of the active site.

The fortunate crystallisation of VvEG16 facilitated the first structure-function study of a member of this family. Our understanding of the EG16 family was extended further by the solution of the tertiary structure of VvEG16 in complex with MLG and XyG-derived oligosaccharides. Overall, this multi-pronged biochemical approach, incorporating active-site variants, co-crystallisation, and kinetics on a library of substrates, created a detailed molecular understanding of the first representative of the EG16 family. The structure of VvEG16 also revealed key evolutionary changes which delineated plant GH16 *endo*-glucanases from bacterial licheninases during the emergence of land plants. Interestingly, the specific mode of substrate recognition employed by VvEG16 gives it the unique ability to induce the gelation of mixed-linkage glucan. Potential applications of this remarkable ability remain to be explored.

Going forward, we should seek to understand the context in which the EG16s act. Experiments are needed to determine where and when it is expressed. Expanding on the work of Yokoyama *et al.* (2010), transgenic plant models are needed to explore the timing and tissuespecificity of EG16 expression in crop plants. The subcellular localisation of EG16s should also be explored. Due to their enigmatic simplicity, it remains unknown where EG16s are naturally found within plant cells. The localisation of EG16s within the cell will inform future studies about the potential impacts of changes in EG16 expression. Furthermore, as additional plant genetic information becomes available, there will be opportunities to explore changes in EG16 structure and function throughout the evolution of plants. The combination of the structurefunction study presented here with a variety of *in planta* experiments and a survey of full range of species which encode EG16s will furnish plant biologists with important additional understanding of plants ranging from algae to cereals.

169

References

- Aasland, R., Abrams, C., Ampe, C., Ball, L.J., Bedford, M.T., Cesareni, G., Gimona, M., Hurley, J.H., Jarchau, T., Lehto, V.-P., Lemmon, M.A., Linding, R., Mayer, B.J., Nagai, M., Sudol, M., Walter, U. and Winder, S.J. (2002) Normalisation of Nomenclature for Peptide Motifs as Ligands of Modular Protein Domains. *FEBS Letters*, 513, 141-144.
- Accetto, T. and Avguštin, G. (2001) Non-specific DNAases from the rumen bacterium *Prevotella bryantii. Folia Microbiol*, **46**, 31-35.
- Accetto, T. and Avguštin, G. (2007) Studies on *Prevotella* Nuclease Using a System for the Controlled Expression of Cloned Genes in *P. bryantii* TC1-1. *Microbiology*, 153, 2281-2288.
- Accetto, T. and Avguštin, G. (2011) Inability of *Prevotella bryantii* to Form a Functional Shine-Dalgarno Interaction Reflects Unique Evolution of Ribosome Binding Sites in Bacteroidetes. *PLOS ONE*, 6, e22914.
- Accetto, T. and Avguštin, G. (2015) Polysaccharide Utilisation Locus and CAZYme Genome Repertoires Reveal Diverse Ecological Adaptation of *Prevotella* Species. *Syst. Appl. Microbiol.*, 38, 453-461.
- Accetto, T., Peterka, M. and Avguštin, G. (2005) Type II Restriction Modification Systems of *Prevotella bryantii* TC1–1 and *Prevotella ruminicola* 23 Strains and their Effect on the Efficiency of DNA Introduction via Electroporation. *FEMS Microbiology Letters*, 247, 177-183.
- Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C. and Zwart, P.H. (2010) *PHENIX*: a Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallographica Section D Biological Crystallography*, 66, 213-221.
- Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W.,
 Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H. and Adams, P.D.
 (2012) Towards Automated Crystallographic Structure Refinement with *phenix.refine*.
 Acta Crystallogr D Biol Crystallogr, 68, 352-367.

- Agirre, J., Iglesias-Fernández, J., Rovira, C., Davies, G.J., Wilson, K.S. and Cowtan, K.D. (2015) Privateer: Software for the Conformational Validation of Carbohydrate Structures. *Nat Struct Mol Biol*, **22**, 833-834.
- Ahn, Y.O., Saino, H., Mizutani, M., Shimizu, B.-i. and Sakata, K. (2007) Vicianin Hydrolase is a Novel Cyanogenic β-Glycosidase Specific to β-Vicianoside (6-O-α-l-Arabinopyranosyl-β-d-Glucopyranoside) in Seeds of *Vicia angustifolia*. *Plant Cell Physiol*, 48, 938-947.
- Albersheim, P., Darvill, A., Roberts, K., Sederoff, R. and Staehelin, A. (2010) *Plant Cell Walls* New York: Garland Science.
- Anthon, G.E. and Barrett, D.M. (2002) Determination of Reducing Sugars with 3-Methyl-2-Benzothiazolinonehydrazone. *Anal. Biochem.*, 305, 287-289.
- Ardèvol, A. and Rovira, C. (2015) Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from *ab Initio* Quantum Mechanics/Molecular Mechanics Dynamic Simulations. *J Am Chem Soc*, **137**, 7528-7547.
- Ariza, A., Eklöf, J.M., Spadiut, O., Offen, W.A., Roberts, S.M., Besenmatter, W., Friis,
 E.P., Skjøt, M., Wilson, K.S., Brumer, H. and Davies, G. (2011) Structure and
 Activity of *Paenibacillus polymyxa* Xyloglucanase from Glycoside Hydrolase Family 44.
 Journal of Biological Chemistry, 286, 33890-33900.
- Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., Keduas, V. and Notredame, C. (2006) *Expresso*: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucl. Acids Res.*, 34, W604-608.
- Armstrong, H.E., Armstrong, E.F. and Horton, E. (1908) Studies on Enzyme Action. XII.-The Enzymes of Emulsin. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 80, 321-331.
- Arnal, G., Attia, M.A., Asohan, J. and Brumer, H. (2017) A Low-Volume, Parallel Copper-Bicinchoninic Acid (BCA) Assay for Glycoside Hydrolases. In *Protein-Carbohydrate Interactions*: Humana Press, New York, NY, pp. 3-14.
- Arnal, G., Bastien, G., Monties, N., Abot, A., Anton Leberre, V., Bozonnet, S., O'Donohue,
 M. and Dumon, C. (2015) Investigating the Function of an Arabinan Utilisation Locus
 Isolated from a Termite Gut Community. *Appl Environ Microbiol*, 81, 31-39.

- Arpat, A.B., Waugh, M., Sullivan, J.P., Gonzales, M., Frisch, D., Main, D., Wood, T., Leslie, A., Wing, R.A. and Wilkins, T.A. (2004) Functional Genomics of Cell Elongation in Developing Cotton Fibers. *Plant Mol Biol*, 54, 911-929.
- Aspeborg, H., Coutinho, P.M., Wang, Y., Brumer, H. and Henrissat, B. (2012) Evolution, Substrate Specificity and Subfamily Classification of Glycoside Hydrolase Family 5 (GH5). *BMC Evolutionary Biology*, **12**, 186.
- Aspinall, G.O. (1969) Gums and Mucilages. In Advances in Carbohydrate Chemistry and Biochemistry (Wolfrom, M.L., Tipson, R.S. and Horton, D. eds): Academic Press, pp. 333-379.
- Aspinall, G.O. and Cañas-Rodriguez, A. (1958) Sisal Pectic Acid. J. Chem. Soc., 0, 4020-4027.
- Aspinall, G.O., Hirst, E.L., Percival, E.G.V. and Williamson, I.R. (1953) The Mannans of Ivory Nut (*Phytelephas macrocarpa*). Part I. The Methylation of Mannan A and Mannan B. J. Chem. Soc., 0, 3184-3188.
- Attia, M., Stepper, J., Davies, G.J. and Brumer, H. (2016) Functional and Structural Characterisation of a Potent GH74 *Endo*-Xyloglucanase from the Soil Saprophyte *Cellvibrio japonicus* Unravels the First Step of Xyloglucan Degradation. *FEBS Journal*, 283, 1701-1719.
- Avguštin, G., Wallace, R.J. and Flint, H.J. (1997) Phenotypic Diversity among Ruminal Isolates of *Prevotella ruminicola*: Proposal of *Prevotella brevis* sp. nov., *Prevotella bryantii* sp. nov., and *Prevotella albensis* sp. nov. and Redefinition of *Prevotella ruminicola*. *Int J Syst Bacteriol*, 47, 284-288.
- Bacic, A. and Stone, B.A. (1981) Chemistry and Organisation of Aleurone Cell Wall Components From Wheat and Barley. *Functional Plant Biol.*, 8, 475-495.
- Bågenholm, V., Reddy, S.K., Bouraoui, H., Morrill, J., Kulcinskaja, E., Bahr, C.M.,
 Aurelius, O., Rogers, T., Xiao, Y., Logan, D.T., Martens, E.C., Koropatkin, N.M.
 and Stålbrand, H. (2017) Galactomannan Catabolism Conferred by a Polysaccharide
 Utilisation Locus of *Bacteroides ovatus*. J Biol Chem, 292, 229-243.
- Bakolitsa, C., Xu, Q., Rife, C.L., Abdubek, P., Astakhova, T., Axelrod, H.L., Carlton, D., Chen, C., Chiu, H.-J., Clayton, T., Das, D., Deller, M.C., Duan, L., Ellrott, K., Farr, C.L., Feuerhelm, J., Grant, J.C., Grzechnik, A., Han, G.W., Jaroszewski, L., Jin,

K.K., Klock, H.E., Knuth, M.W., Kozbial, P., Krishna, S.S., Kumar, A., Lam, W.W., Marciano, D., McMullan, D., Miller, M.D., Morse, A.T., Nigoghossian, E., Nopakun, A., Okach, L., Puckett, C., Reyes, R., Tien, H.J., Trame, C.B., van den Bedem, H., Weekes, D., Hodgson, K.O., Wooley, J., Elsliger, M.-A., Deacon, A.M., Godzik, A., Lesley, S.A. and Wilson, I.A. (2010) Structure of BT_3984, a Member of the SusD/RagB Family of Nutrient-Binding Molecules. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, **66**, 1274-1280.

- Barbeyron, T., Gerard, A., Potin, P., Henrissat, B. and Kloareg, B. (1998) The kappa-Carrageenase of the Marine Bacterium Cytophaga drobachiensis. Structural and Phylogenetic Relationships within Family-16 Glycoside Hydrolases. Mol Biol Evol, 15, 528-537.
- Bardsley, W.G., Leff, P., Kavanagh, J. and Waight, R.D. (1980) Deviations from Michaelis-Menten Kinetics. The Possibility of Complicated Curves for Simple Kinetic Schemes and the Computer Fitting of Experimental Data for Acetylcholinesterase, Acid Phosphatase, Adenosine Deaminase, Arylsulphatase, Benzylamine Oxidase, Chymotrypsin, Fumarase, Galactose Dehydrogenase, *beta*-Galactosidase, Lactate Dehydrogenase, Peroxidase and Xanthine Oxidase. *Biochemical Journal*, 187, 739-765.
- **Barnes, W.C. and Blakeney, A.B.** (1974) Determination of Cereal Alpha Amylase Using a Commercially Available Dye-labelled Substrate. *Starch Stärke*, **26**, 193-197.
- Barras, F., Bortoli-German, I., Bauzan, M., Rouvier, J., Gey, C., Heyraud, A. and Henrissat, B. (1992) Stereochemistry of the Hydrolysis Reaction Catalyzed by Endoglucanase Z from *Erwinia chrysanthemi*. *FEBS Letters*, **300**, 145-148.
- Baumann, M.J., Eklof, J.M., Michel, G., Kallas, A.M., Teeri, T.T., Czjzek, M. and Brumer,
 H. (2007) Structural Evidence for the Evolution of Xyloglucanase Activity from
 Xyloglucan *Endo*-Transglycosylases: Biological Implications for Cell Wall Metabolism.
 Plant Cell, 19, 1947-1963.
- Becnel, J., Natarajan, M., Kipp, A. and Braam, J. (2006) Developmental Expression Patterns of *Arabidopsis XTH* Genes Reported by Transgenes and Genevestigator. *Plant Mol Biol*, 61, 451-467.
- Berg, J.M., Tymoczko, J.L., Stryer, L., Berg, J.M., Tymoczko, J.L. and Stryer, L. (2002) Biochemistry 5th edn.: W H Freeman.

- Bergen, L.A.H.v., Alonso, M., Palló, A., Nilsson, L., Proft, F.D. and Messens, J. (2016) Revisiting Sulfur H-Bonds in Proteins: The Example of Peroxiredoxin AhpE. *Scientific Reports*, 6, 30369.
- Berger, E., Jones, W.A., Jones, D.T. and Woods, D.R. (1989) Cloning and sequencing of an endoglucanase (end1) gene from Butyrivibrio fibrisolvens H17c. Mol. Gen. Genet., 219, 193-198.
- Bhalla, A., Bischoff, K.M. and Sani, R.K. (2015) Highly Thermostable Xylanase Production from A Thermophilic *Geobacillus sp.* Strain WSUCF1 Utilizing Lignocellulosic Biomass. *Front Bioeng Biotechnol*, 3.
- Bibi, Z., Ansari, A., Zohra, R.R., Aman, A. and Ul Qader, S.A. (2014) Production of Xylan Degrading *Endo*-1,4-β-Xylanase from Thermophilic *Geobacillus stearothermophilus* KIBGE-IB29. *Journal of Radiation Research and Applied Sciences*, 7, 478-485.
- Biely, P. (2012) Microbial Carbohydrate Esterases Deacetylating Plant Polysaccharides. *Biotechnology Advances*, **30**, 1575-1588.
- **Bischof, R.H., Ramoni, J. and Seiboth, B.** (2016) Cellulases and beyond: the first 70 years of the enzyme producer *Trichoderma reesei*. *Microbial Cell Factories*, **15**, 106.
- Bitto, E., Bingman, C.A., McCoy, J.G., Allard, S.T.M., Wesenberg, G.E. and Phillips, G.N.,
 Jr. (2005) The Structure at 1.6 Å Resolution of the Protein Product of the At4g34215
 Gene from Arabidopsis thaliana. Acta Crystallogr D Biol Crystallogr, 61, 1655-1661.
- Black, T.S., Kiss, L., Tull, D. and Withers, S.G. (1993) N-Bromoacetyl-Glycopyranosylamines as Affinity Labels for a β-Glucosidase and a Cellulase. *Carbohydr. Res.*, **250**, 195-202.
- Blatch, G.L. and Lässle, M. (1999) The Tetratricopeptide Repeat: a Structural Motif Mediating Protein-Protein Interactions. *Bioessays*, **21**, 932-939.
- Blouzard, J.-C., Coutinho, P.M., Fierobe, H.-P., Henrissat, B., Lignon, S., Tardif, C., Pagès,
 S. and de Philip, P. (2010) Modulation of Cellulosome Composition in *Clostridium cellulolyticum*: Adaptation to the Polysaccharide Environment Revealed by Proteomic and Carbohydrate-Active Enzyme Analyses. *Proteomics*, 10, 541-554.
- Boraston, Alisdair B., Bolam, David N., Gilbert, Harry J. and Davies, Gideon J. (2004) Carbohydrate-Binding Modules: Fine-Tuning Polysaccharide Recognition. *Biochemical Journal*, 382, 769-781.

- Bourquin, V., Nishikubo, N., Abe, H., Brumer, H., Denman, S., Eklund, M., Christiernin, M., Teeri, T.T., Sundberg, B. and Mellerowicz, E.J. (2002) Xyloglucan Endotransglycosylases Have a Function during the Formation of Secondary Cell Walls of Vascular Tissues. *Plant Cell*, 14, 3073-3088.
- Brayer, G.D., Sidhu, G., Maurus, R., Rydberg, E.H., Braun, C., Wang, Y., Nguyen, N.T., Overall, C.M. and Withers, S.G. (2000) Subsite Mapping of the Human Pancreatic α-Amylase Active Site through Structural, Kinetic, and Mutagenesis Techniques. *Biochemistry*, **39**, 4778-4791.
- Breiten, B., Lockett, M.R., Sherman, W., Fujita, S., Al-Sayah, M., Lange, H., Bowers, C.M., Heroux, A., Krilov, G. and Whitesides, G.M. (2013) Water Networks Contribute to Enthalpy/Entropy Compensation in Protein–Ligand Binding. *J Am Chem Soc*, 135, 15579-15584.
- Bronnenmeier, K., Kern, A., Liebl, W. and Staudenbauer, W.L. (1995) Purification of *Thermotoga maritima* Enzymes for the Degradation of Cellulosic Materials. *Appl Environ Microbiol*, 61, 1399-1407.
- Brook, I. (2002) Microbiology of Polymicrobial Abscesses and Implications for Therapy. J Antimicrob Chemother, 50, 805-810.
- Brunecky, R., Alahuhta, M., Bomble, Y.J., Xu, Q., Baker, J.O., Ding, S.Y., Himmel, M.E. and Lunin, V.V. (2012) Structure and Function of the *Clostridium thermocellum* Cellobiohydrolase A X1-Module Repeat: Enhancement through Stabilisation of the CbhA Complex. *Acta Crystallogr D Biol Crystallogr*, 68, 292-299.
- Buchanan, M., Burton, R.A., Dhugga, K.S., Rafalski, A.J., Tingey, S.V., Shirley, N.J. and Fincher, G.B. (2012) *Endo*-(1,4)-β-Glucanase Gene Families in the Grasses: Temporal and Spatial Co-Transcription of Orthologous Genes. *BMC Plant Biology*, **12**, 235.
- Buckeridge, M.S. (2010) Seed Cell Wall Storage Polysaccharides: Models to Understand Cell Wall Biosynthesis and Degradation. *Plant Physiology*, **154**, 1017-1023.
- Burmeister, W.P., Cottaz, S., Rollin, P., Vasella, A. and Henrissat, B. (2000) High Resolution X-ray Crystallography Shows That Ascorbate Is a Cofactor for Myrosinase and Substitutes for the Function of the Catalytic Base. *Journal of Biological Chemistry*, 275, 39385-39393.

- **Burton, R.A. and Fincher, G.B.** (2009) (1,3;1,4)-β-D-Glucans in Cell Walls of the Poaceae, Lower Plants, and Fungi: A Tale of Two Linkages. *Mol. Plant*, **2**, 873-882.
- Burton, R.A., Gidley, M.J. and Fincher, G.B. (2010) Heterogeneity in the Chemistry, Structure and Function of Plant Cell Walls. *Nat Chem Biol*, **6**, 724-732.
- Caffall, K.H. and Mohnen, D. (2009) The Structure, Function, and Biosynthesis of Plant Cell Wall Pectic Polysaccharides. *Carbohydr. Res.*, **344**, 1879-1900.
- Cameron, E.A., Kwiatkowski, K.J., Lee, B.-H., Hamaker, B.R., Koropatkin, N.M. and Martens, E.C. (2014) Multifunctional Nutrient-Binding Proteins Adapt Human Symbiotic Bacteria for Glycan Competition in the Gut by Separately Promoting Enhanced Sensing and Catalysis. *mBio*, 5, e01441-01414.
- Cameron, E.A., Maynard, M.A., Smith, C.J., Smith, T.J., Koropatkin, N.M. and Martens,
 E.C. (2012) Multidomain Carbohydrate-binding Proteins Involved in *Bacteroides* thetaiotaomicron Starch Metabolism. J Biol Chem, 287, 34614-34625.
- Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V. and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucl. Acids Res.*, **37**, D233-D238.
- Cantarel, B.L., Lombard, V. and Henrissat, B. (2012) Complex Carbohydrate Utilisation by the Healthy Human Microbiome. *PLOS ONE*, **7**, e28742.
- Carpita, N.C. and Gibeaut, D.M. (1993) Structural Models of Primary Cell Walls in Flowering Plants: Consistency of Molecular Structure with the Physical Properties of the Walls During Growth. *The Plant Journal*, **3**, 1-30.
- Carpita, N.C. and McCann, M.C. (2000) The Cell Wall. In *Biochemistry and Molecular Biology of Plants*. Somerset, NJ: John Wiley & Sons, Inc., pp. 55-108.
- Cass, A.E.G., Davis, G., Francis, G.D., Hill, H.A.O., Aston, W.J., Higgins, I.J., Plotkin,
 E.V., Scott, L.D.L. and Turner, A.P.F. (1984) Ferrocene-Mediated Enzyme Electrode for Amperometric Determination of Glucose. *Anal. Chem.*, 56, 667-671.
- Castro, E.d., Sigrist, C.J.A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E., Bairoch, A. and Hulo, N. (2006) ScanProsite: Detection of *PROSITE* Signature Matches and ProRule-Associated Functional and Structural Residues in Proteins. *Nucl. Acids Res.*, 34, W362-W365.

- Charalampakis, G., Dahlén, G., Carlén, A. and Leonhardt, Å. (2013) Bacterial markers vs. clinical markers to predict progression of chronic periodontitis: a 2-yr prospective observational study. *Eur J Oral Sci*, **121**, 394-402.
- Charnock, S.J., Lakey, J.H., Virden, R., Hughes, N., Sinnott, M.L., Hazlewood, G.P., Pickersgill, R. and Gilbert, H.J. (1997) Key Residues in Subsite F Play a Critical Role in the Activity of *Pseudomonas fluorescens* Subspecies *cellulosa* Xylanase A Against Xylooligosaccharides but Not Against Highly Polymeric Substrates such as Xylan. *Journal of Biological Chemistry*, 272, 2942-2951.
- Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S. and Richardson, D.C. (2010) *MolProbity*: All-Atom Structure Validation for Macromolecular Crystallography. *Acta Crystallogr D Biol Crystallogr*, 66, 12-21.
- Coleman, D.J., Studler, M.J. and Naleway, J.J. (2007) A Long-Wavelength Fluorescent Substrate for Continuous Fluorometric Determination of Cellulase Activity: Resorufin-β-D-Cellobioside. *Anal. Biochem.*, **371**, 146-153.
- Corradini, C., Cavazza, A., Bignardi, C., Corradini, C., Cavazza, A. and Bignardi, C. (2012) High-Performance Anion-Exchange Chromatography Coupled with Pulsed Electrochemical Detection as a Powerful Tool to Evaluate Carbohydrates of Food Interest: Principles and Applications. *International Journal of Carbohydrate Chemistry*, *International Journal of Carbohydrate Chemistry*, 2012, 2012, e487564.
- Cosgrove, D.J. (2005) Growth of the Plant Cell Wall. Nat Rev Mol Cell Biol, 6, 850-861.
- Couger, M.B., Youssef, N.H., Struchtemeyer, C.G., Liggenstoffer, A.S. and Elshahed, M.S. (2015) Transcriptomic Analysis of Lignocellulosic Biomass Degradation by the Anaerobic Fungal Isolate *Orpinomyces sp.* Strain C1A. *Biotechnology for Biofuels*, 8, 208.
- Couturier, M., Roussel, A., Rosengren, A., Leone, P., Stålbrand, H. and Berrin, J.-G. (2013) Structural and Biochemical Analyses of Glycoside Hydrolase Families 5 and 26 β-(1,4)-Mannanases from *Podospora anserina* Reveal Differences upon Manno-oligosaccharide Catalysis. *J Biol Chem*, **288**, 14624-14635.

- Crost, E.H., Tailford, L.E., Gall, G.L., Fons, M., Henrissat, B. and Juge, N. (2013) Utilisation of Mucin Glycans by the Human Gut Symbiont *Ruminococcus gnavus* Is Strain-Dependent. *PLOS ONE*, 8, e76341.
- Cummings, J.H. and Macfarlane, G.T. (1997) Clinical Nutrition Scientific Publications Role of intestinal bacteria in nutrient metabolism. *JPEN J Parenter Enteral Nutr*, **21**, 357-365.
- Cuskin, F., Lowe, E.C., Temple, M.J., Zhu, Y., Cameron, E.A., Pudlo, N.A., Porter, N.T., Urs, K., Thompson, A.J., Cartmell, A., Rogowski, A., Hamilton, B.S., Chen, R., Tolbert, T.J., Piens, K., Bracke, D., Vervecken, W., Hakki, Z., Speciale, G., Munōz-Munōz, J.L., Day, A., Peña, M.J., McLean, R., Suits, M.D., Boraston, A.B., Atherly, T., Ziemer, C.J., Williams, S.J., Davies, G.J., Abbott, D.W., Martens, E.C. and Gilbert, H.J. (2015) Human Gut Bacteroidetes can Utilize Yeast Mannan through a Selfish Mechanism. *Nature*, 517, 165-169.
- Czjzek, M., Cicek, M., Zamboni, V., Bevan, D.R., Henrissat, B. and Esen, A. (2000) The Mechanism of Substrate (Aglycone) Specificity in β-Glucosidases is Revealed by Crystal Structures of Mutant Maize β-Glucosidase-DIMBOA, -DIMBOAGlc, and -Dhurrin Complexes. *Proc Natl Acad Sci U S A*, 97, 13555-13560.
- D'Elia, J.N. and Salyers, A.A. (1996) Effect of Regulatory Protein Levels on Utilisation of Starch by *Bacteroides thetaiotaomicron*. J Bacteriol, 178, 7180-7186.
- Dalrymple, B.P., Cybinski, D.H., Layton, I., McSweeney, C.S., Xue, G.-P., Swadling, Y.J. and Lowry, J.B. (1997) Three *Neocallimastix patriciarum* Esterases Associated with the Degradation of Complex Polysaccharides are Members of a New Family of Hydrolases. *Microbiology*, 143, 2605-2614.
- Davé, V. and McCarthy, S.P. (1997) Review of Konjac Glucomannan. Journal of Environmental Polymer Degradation; Dordrecht, 5, 237-241.
- **Davies, G. and Henrissat, B.** (1995) Structures and Mechanisms of Glycosyl Hydrolases. *Structure*, **3**, 853-859.
- Davies, G.J., Gloster, T.M. and Henrissat, B. (2005) Recent Structural Insights into the Expanding World of Carbohydrate-Active Enzymes. *Curr. Opin. Struct. Biol.*, 15, 637-645.
- **Davies, G.J. and Sinnott, M.L.** (2008) Sorting the Diverse: the Sequence-based Classifications of Carbohydrate-Active Enzymes. *The Biochemist*, **30**, 26-32.

- Davies, G.J., Wilson, K.S. and Henrissat, B. (1997) Nomenclature for Sugar-Binding Subsites in Glycosyl Hydrolases. *Biochemical Journal*, 321, 557-559.
- DeBoy, R.T., Mongodin, E.F., Fouts, D.E., Tailford, L.E., Khouri, H., Emerson, J.B.,
 Mohamoud, Y., Watkins, K., Henrissat, B., Gilbert, H.J. and Nelson, K.E. (2008)
 Insights into Plant Cell Wall Degradation from the Genome Sequence of the Soil
 Bacterium *Cellvibrio japonicus*. J Bacteriol, 190, 5455-5463.
- Desmet, T., Cantaert, T., Gualfetti, P., Nerinckx, W., Gross, L., Mitchinson, C. and Piens,
 K. (2007) An Investigation of the Substrate Specificity of the Xyloglucanase Cel74A
 from *Hypocrea jecorina*. *FEBS Journal*, 274, 356-363.
- Desvaux, M. (2005) Clostridium cellulolyticum: Model Organism of Mesophilic Cellulolytic Clostridia. FEMS Microbiol Rev, 29, 741-764.
- Dias, F.M.V., Vincent, F., Pell, G., Prates, J.A.M., Centeno, M.S.J., Tailford, L.E., Ferreira, L.M.A., Fontes, C.M.G.A., Davies, G.J. and Gilbert, H.J. (2004) Insights into the Molecular Determinants of Substrate Specificity in Glycoside Hydrolase Family 5 Revealed by the Crystal Structure and Kinetics of *Cellvibrio mixtus* Mannosidase 5A. *J Biol Chem*, 279, 25517-25526.
- Dodd, D., Kiyonari, S., Mackie, R.I. and Cann, I.K.O. (2010a) Functional Diversity of Four Glycoside Hydrolase Family 3 Enzymes from the Rumen Bacterium *Prevotella bryantii* B₁4. *J Bacteriol*, **192**, 2335-2345.
- Dodd, D., Moon, Y.-H., Swaminathan, K., Mackie, R.I. and Cann, I.K.O. (2010b) Transcriptomic Analyses of Xylan Degradation by *Prevotella bryantii* and Insights into Energy Acquisition by Xylanolytic Bacteroidetes. *Journal of Biological Chemistry*, 285, 30261-30273.
- **Doner, L.W. and Irwin, P.L.** (1992) Assay of Reducing End-groups in Oligosaccharide Homologues with 2,2′ -Bicinchoninate. *Anal. Biochem.*, **202**, 50-53.
- dos Santos, C.R., Cordeiro, R.L., Wong, D.W.S. and Murakami, M.T. (2015) Structural Basis for Xyloglucan Specificity and α -D-Xylp(1 \rightarrow 6)-d-Glcp Recognition at the -1 Subsite within the GH5 Family. *Biochemistry*, **54**, 1930-1942.
- Ducros, V., Czjzek, M., Belaich, A., Gaudin, C., Fierobe, H.P., Belaich, J.P., Davies, G.J. and Haser, R. (1995) Crystal Structure of the Catalytic Domain of a Bacterial Cellulase Belonging to Family 5. *Structure*, **3**, 939-949.

- Ducros, V.M.A., Zechel, D.L., Murshudov, G.N., Gilbert, H.J., Szabó, L., Stoll, D., Withers, S.G. and Davies, G.J. (2002) Substrate Distortion by a β-Mannanase: Snapshots of the Michaelis and Covalent-Intermediate Complexes Suggest a B2,5 Conformation for the Transition State. *Angew. Chem. Int. Ed.*, **41**, 2824-2827.
- Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y. and Liang, J. (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucl. Acids Res.*, 34, W116-W118.
- **Ebringerová, A.** (2005) Structural Diversity and Application Potential of Hemicelluloses. *Macromol. Symp.*, **232**, 1-12.
- Edstrom, R.D. and Phaff, H.J. (1964) Eliminative Cleavage of Pectin and of Oligogalacturonide Methyl Esters by Pectin *trans*-Eliminase. *Journal of Biological Chemistry*, 239, 2409-2415.
- Eisenthal, R., Danson, M.J. and Hough, D.W. (2007) Catalytic efficiency and *k*_{cat}/*K*_M: a useful comparator? *Trends Biotechnol.*, **25**, 247-249.
- Eklöf, J.M. and Brumer, H. (2010) The XTH Gene Family: An Update on Enzyme Structure, Function, and Phylogeny in Xyloglucan Remodeling. *Plant Physiology*, 153, 456-466.
- Eklöf, J.M., Ruda, M.C. and Brumer, H. (2012) Distinguishing Xyloglucanase Activity in *endo-* β (1 \rightarrow 4)Glucanases. *Meth. Enzymol.*, **510**, 97-120.
- Eklöf, J.M., Shojania, S., Okon, M., McIntosh, L.P. and Brumer, H. (2013) Structure-Function Analysis of a Broad Specificity *Populus trichocarpa Endo-β-glucanase Reveals* an Evolutionary Link between Bacterial Licheninases and Plant *XTH* Gene Products. *Journal of Biological Chemistry*, **288**, 15786-15799.
- Ekstrom, A., Taujale, R., McGinn, N. and Yin, Y. (2014) PlantCAZyme: a Database for Plant Carbohydrate-Active Enzymes. *Database (Oxford)*, **2014**.
- Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126-2132.
- Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and Development of Coot. Acta Crystallogr D Biol Crystallogr, 66, 486-501.
- Eschenfeldt, W., Lucy, S., Millard, C., Joachimiak, A. and Mark, I.D. (2009) A Family of LIC Vectors for High-Throughput Cloning and Purification of Proteins. In *High*

Throughput Protein Expression and Purification (Doyle, S. ed: Humana Press, pp. 105-115.

- Eschenfeldt, W.H., Makowska-Grzyska, M., Stols, L., Donnelly, M., Jedrzejczak, R. and Joachimiak, A. (2013) New LIC Vectors For Production of Proteins from Genes Containing Rare Codons. J Struct Funct Genomics, 14, 135-144.
- Faure, E., Belaich, A., Bagnara, C., Gaudin, C. and Belaich, J.-P. (1989) Sequence Analysis of the *Clostridium cellulolyticum* Endoglucanase-A-Encoding Gene, *celCCA. Gene*, 84, 39-46.
- Feng, S., Bagia, C. and Mpourmpakis, G. (2013) Determination of Proton Affinities and Acidity Constants of Sugars. J. Phys. Chem. A, 117, 5211-5219.
- Fenger, T.H. and Brumer, H. (2015) Synthesis and Analysis of Specific Covalent Inhibitors of Endo-Xyloglucanases. ChemBioChem, 16, 575-583.
- Fersht, A.R. (1974) Catalysis, Binding and Enzyme-Substrate Complementarity. Proceedings of the Royal Society of London. Series B, Biological Sciences, 187, 397-407.
- **Fibriansah, G., Masuda, S., Koizumi, N., Nakamura, S. and Kumasaka, T.** (2007) The 1.3 Å Crystal Structure of a Novel *Endo*-β-1,3-Glucanase of Glycoside Hydrolase Family 16 from Alkaliphilic *Nocardiopsis sp.* strain F96. *Proteins*, **69**, 683-690.
- Fields, M.W., Russell, J.B. and Wilson, D.B. (1997) A Mutant of *Prevotella ruminicola* B₁4 Deficient in β-1,4-Endoglucanase and Mannanase Activities. *FEMS Microbiology Letters*, **154**, 9-15.
- Filippo, C.D., Cavalieri, D., Paola, M.D., Ramazzotti, M., Poullet, J.B., Massart, S., Collini, S., Pieraccini, G. and Lionetti, P. (2010) Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *PNAS*, 107, 14691-14696.
- Fincher, G.B. (1975) Morphology and Chemical Composition of Barley Endosperm Cell Walls. *Journal of the Institute of Brewing*, 81, 116-122.
- **Fincher, G.B.** (2009) Exploring the Evolution of (1,3;1,4)-β-D-Glucans in Plant Cell Walls: Comparative Genomics can Help! *Current Opinion in Plant Biology*, **12**, 140-147.
- Flint, H.J. and Bayer, E.A. (2008) Plant Cell Wall Breakdown by Anaerobic Microorganisms from the Mammalian Digestive Tract. *Annals of the New York Academy of Sciences*, 1125, 280-288.

- Flint, H.J., Scott, K.P., Duncan, S.H., Louis, P. and Forano, E. (2012) Microbial Degradation of Complex Carbohydrates in the Gut. *Gut Microbes*, **3**, 289-306.
- Foong, F., Hamamoto, T., Shoseyov, O. and Doi, R.H. (1991) Nucleotide Sequence and Characteristics of Endoglucanase Gene *engB* from *Clostridium cellulovorans*. J. Gen. *Microbiol.*, 137, 1729-1736.
- Forment, J., Gadea, J., Huerta, L., Abizanda, L., Agusti, J., Alamar, S., Alos, E., Andres, F., Arribas, R., Beltran, J.P., Berbel, A., Blazquez, M.A., Brumos, J., Canas, L.A., Cercos, M., Colmenero-Flores, J.M., Conesa, A., Estables, B., Gandia, M., Garcia-Martinez, J.L., Gimeno, J., Gisbert, A., Gomez, G., Gonzalez-Candelas, L., Granell, A., Guerri, J., Lafuente, M.T., Madueno, F., Marcos, J.F., Marques, M.C., Martinez, F., Martinez-Godoy, M.A., Miralles, S., Moreno, P., Navarro, L., Pallas, V., Perez-Amador, M.A., Perez-Valle, J., Pons, C., Rodrigo, I., Rodriguez, P.L., Royo, C., Serrano, R., Soler, G., Tadeo, F., Talon, M., Terol, J., Trenor, M., Vaello, L., Vicente, O., Vidal, C., Zacarias, L. and Conejero, V. (2005) Development of a Citrus Genome-Wide EST Collection and cDNA Microarray as Resources for Genomic Studies. *Plant Mol Biol*, 57, 375-391.
- Franco, P.J. and Wilson, T.H. (1999) Arg-52 in the Melibiose Carrier of *Escherichia coli* Is Important for Cation-Coupled Sugar Transport and Participates in an Intrahelical Salt Bridge. *J Bacteriol*, 181, 6377-6386.
- Franková, L. and Fry, S.C. (2013) Biochemistry and Physiological Roles of Enzymes that 'Cut and Paste' Plant Cell-Wall Polysaccharides. J Exp Bot, 64, 3519-3550.
- Fry, S.C. (1989) The Structure and Functions of Xyloglucan. J Exp Bot, 40, 1-11.
- Fry, S.C., Mohler, K.E., Nesselrode, B.H.W.A. and Franková, L. (2008a) Mixed-Liinkage β-Glucan:Xyloglucan Endotransglucosylase, a Novel Wall-Remodelling Enzyme from *Equisetum* (Horsetails) and Charophytic Algae. *The Plant Journal*, **55**, 240-252.
- Fry, S.C., Nesselrode, B.H.W.A., Miller, J.G. and Mewburn, B.R. (2008b) Mixed-Linkage (1 \rightarrow 3,1 \rightarrow 4)- β -D-Glucan is a Major Hemicellulose of *Equisetum* (Horsetail) Cell Walls. *New Phytologist*, **179**, 104-115.
- Fry, S.C., York, W.S., Albersheim, P., Darvill, A., Hayashi, T., Joseleau, J.-P., Kato, Y.,
 Lorences, E.P., Maclachlan, G.A., McNeil, M., Mort, A.J., Grant Reid, J.S., Seitz,
 H.U., Selvendran, R.R., Voragen, A.G.J. and White, A.R. (1993) An Unambiguous

Nomenclature for Xyloglucan-Derived Oligosaccharides. *Physiologia Plantarum*, **89**, 1-3.

- Fujii, Y. (1999) Crystal Structure of an IRF-DNA Complex Reveals Novel DNA Recognition and Cooperative Binding to a Tandem Repeat of Core Sequences. *The EMBO Journal*, 18, 5028-5041.
- Fuwa, H. (1954) A New Method for Microdetermination of Amylase Activity by the Use of Amylose as the Substrate. *J Biochem*, 41, 583-603.
- Gaiser, O.J., Piotukh, K., Ponnuswamy, M.N., Planas, A., Borriss, R. and Heinemann, U. (2006) Structural Basis for the Substrate Specificity of a *Bacillus* 1,3-1,4-β-Glucanase. *Journal of Molecular Biology*, **357**, 1211-1225.
- Gallegos, M.T., Schleif, R., Bairoch, A., Hofmann, K. and Ramos, J.L. (1997) AraC/XylS Family of Transcriptional Regulators. *Microbiol. Mol. Biol. Rev.*, **61**, 393-410.
- Gardner, R.G., Wells, J.E., Fields, M.W., Wilson, D.B. and Russell, J.B. (1997) A Prevotella ruminicola B₁4 Operon Encoding Extracellular Polysaccharide Hydrolases. Curr Microbiol, 35, 274-277.
- Gardner, R.G., Wells, J.E., Russell, J.B. and Wilson, D.B. (1995) The cellular location of Prevotella ruminicola beta-1,4-D-endoglucanase and its occurrence in other strains of ruminal bacteria. *Appl Environ Microbiol*, **61**, 3288-3292.
- Garron, M.L. and Cygler, M. (2010) Structural and Mechanistic Classification of Uronic Acid-Containing Polysaccharide Lyases. *Glycobiology*, **20**, 1547-1573.
- Geisler-Lee, J., Geisler, M., Coutinho, P.M., Segerman, B., Nishikubo, N., Takahashi, J.,
 Aspeborg, H., Djerbi, S., Master, E., Andersson-Gunnerås, S., Sundberg, B.,
 Karpinski, S., Teeri, T.T., Kleczkowski, L.A., Henrissat, B. and Mellerowicz, E.J.
 (2006) Poplar Carbohydrate-Active Enzymes. Gene Identification and Expression
 Analyses. *Plant Physiology*, 140, 946-962.
- Gerttula, S., Zinkgraf, M., Muday, G.K., Lewis, D.R., Ibatullin, F.M., Brumer, H., Hart, F., Mansfield, S.D., Filkov, V. and Groover, A. (2015) Transcriptional and Hormonal Regulation of Gravitropism of Woody Stems in *Populus*. *Plant Cell*, 27, 2800-2813.
- Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A. and Smith, H.O. (2009) Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat Meth*, 6, 343-345.

- Gibson, Q.H., Swoboda, B.E.P. and Massey, V. (1964) Kinetics and Mechanism of Action of Glucose Oxidase. *Journal of Biological Chemistry*, 239, 3927-3934.
- Gilbert, H.J., Knox, J.P. and Boraston, A.B. (2013) Advances in Understanding the Molecular Basis of Plant Cell Wall Polysaccharide Recognition by Carbohydrate-Binding Modules. *Curr. Opin. Struct. Biol.*, 23, 669-677.
- Gilbert, H.J., Stålbrand, H. and Brumer, H. (2008) How the Walls Come Crumbling Down: Recent Structural Biochemistry of Plant Polysaccharide Degradation. *Current Opinion in Plant Biology*, 11, 338-348.
- Gilkes, N.R., Warren, R.A., Miller, R.C. and Kilburn, D.G. (1988) Precise Excision of the Cellulose Binding Domains from Two *Cellulomonas fimi* Cellulases by a Homologous Protease and the Effect on Catalysis. *Journal of Biological Chemistry*, 263, 10401-10407.
- Gille, S. and Pauly, M. (2012) *O*-Acetylation of Plant Cell Wall Polysaccharides. *Front Plant Sci*, **3**.
- Glenwright, A.J., Pothula, K.R., Bhamidimarri, S.P., Chorev, D.S., Baslé, A., Firbank, S.J., Zheng, H., Robinson, C.V., Winterhalter, M., Kleinekathöfer, U., Bolam, D.N. and van den Berg, B. (2017) Structural Basis for Nutrient Acquisition by Dominant Members of the Human Gut Microbiota. *Nature*, 541, 407-411.
- Gloster, T.M., Ibatullin, F.M., Macauley, K., Eklöf, J.M., Roberts, S., Turkenburg, J.P.,
 Bjørnvad, M.E., Jørgensen, P.L., Danielsen, S., Johansen, K.S., Borchert, T.V.,
 Wilson, K.S., Brumer, H. and Davies, G.J. (2007) Characterisation and Threedimensional Structures of Two Distinct Bacterial Xyloglucanases from Families GH5 and
 GH12. Journal of Biological Chemistry, 282, 19177-19189.
- Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks,
 W., Hellsten, U., Putnam, N. and Rokhsar, D.S. (2012) *Phytozome*: a Comparative
 Platform for Green Plant Genomics. *Nucl. Acids Res.*, 40, D1178-D1186.
- Gorvitovskaia, A., Holmes, S.P. and Huse, S.M. (2016) Interpreting *Prevotella* and *Bacteroides* as Biomarkers of Diet and Lifestyle. *Microbiome*, **4**, 15.
- Greenblatt, J. and Schleif, R. (1971) Arabinose C Protein: Regulation of the Arabinose Operon *in vitro*. *Nature New Biology*, 233, 166-170.

- Greer, D.H. and Weston, C. (2010) Heat Stress Affects Flowering, Berry Growth, Sugar Accumulation and Photosynthesis of *Vitis vinifera* cv. Semillon Grapevines Grown in a Controlled Environment. *Functional Plant Biol.*, 37, 206-214.
- Grohmann, K., Mitchell, D.J., Himmel, M.E., Dale, B.E. and Schroeder, H.A. (1989) The Role of Ester Groups in Resistance of Plant Cell Wall Polysaccharides to Enzymatic Hydrolysis. *Appl Biochem Biotechnol*, **20-21**, 45.
- Gusakov, A.V., Kondratyeva, E.G. and Sinitsyn, A.P. (2011) Comparison of Two Methods for Assaying Reducing Sugars in the Determination of Carbohydrase Activities. *International Journal of Analytical Chemistry*, 2011, e283658.
- Hahn, M., Pons, J., Planas, A., Querol, E. and Heinemann, U. (1995) Crystal Structure of *Bacillus licheniformis* 1,3-1,4-β-D-Glucan 4-Glucanohydrolase at 1.8 Å Resolution. *FEBS Letters*, 374, 221-224.
- Haldane, J.B.S. (1930) Enzymes: The MIT Press.
- Hamamoto, T., Foong, F., Shoseyov, O. and Doi, R.H. (1992) Analysis of Functional Domains of Endoglucanases from *Clostridium cellulovorans* by Gene Cloning, Nucleotide Sequencing and Chimeric Protein Construction. *Molec. Gen. Genet.*, 231, 472-479.
- Han, Y., Zhu, Q., Zhang, Z., Meng, K., Hou, Y., Ban, Q., Suo, J. and Rao, J. (2015) Analysis of *Xyloglucan Endotransglycosylase/Hydrolase (XTH)* Genes and Diverse Roles of Isoenzymes during Persimmon Fruit Development and Postharvest Softening. *PLOS ONE*, 10, e0123668.
- Hara, Y., Yokoyama, R., Osakabe, K., Toki, S. and Nishitani, K. (2014) Function of Xyloglucan Endotransglucosylase/Hydrolases in Rice. Ann Bot, 114, 1309-1318.
- Hatfield, R.D. and Nevins, D.J. (1987) Hydrolytic Activity and Substrate Specificity of an Endoglucanase from *Zea mays* Seedling Cell Walls. *Plant Physiology*, 83, 203-207.
- Hayashi, T. (1989) Xyloglucans in the Primary Cell Wall. *Annual Review of Plant Physiology and Plant Molecular Biology*, **40**, 139-168.
- Hayashi, T., Ogawa, K. and Mitsuishi, Y. (1994) Characterisation of the adsorption of Xyloglucan to Cellulose. *Plant Cell Physiol*, 35, 1199-1205.
- Hehemann, J.-H., Correc, G., Barbeyron, T., Helbert, W., Czjzek, M. and Michel, G. (2010) Transfer of Carbohydrate-Active Enzymes from Marine Bacteria to Japanese Gut Microbiota. *Nature*, 464, 908-912.

- Hehemann, J.-H., Correc, G., Thomas, F., Bernard, T., Barbeyron, T., Jam, M., Helbert,
 W., Michel, G. and Czjzek, M. (2012) Biochemical and Structural Characterisation of
 the Complex Agarolytic Enzyme System from the Marine Bacterium *Zobellia* galactanivorans. J Biol Chem, 287, 30571-30584.
- Hemsworth, G.R., Johnston, E.M., Davies, G.J. and Walton, P.H. (2015) Lytic Polysaccharide Monooxygenases in Biomass Conversion. *Trends Biotechnol.*, 33, 747-761.
- Henriksson, K., Teleman, A., Suortti, T., Reinikainen, T., Jaskari, J., Teleman, O. and Poutanen, K. (1995) Hydrolysis of Barley $(1\rightarrow 3)$, $(1\rightarrow 4)$ - β -D-Glucan by a Cellobiohydrolase II Preparation from *Trichoderma reesei*. *Carbohydrate Polymers*, **26**, 109-119.
- Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical Journal*, 280, 309-316.
- Henrissat, B. and Bairoch, A. (1996) Updating the Sequence-Based Classification of Glycosyl Hydrolases. *Biochemical Journal*, **316**, 695-696.
- Hervé, C., Rogowski, A., Blake, A.W., Marcus, S.E., Gilbert, H.J. and Knox, J.P. (2010) Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. *PNAS*, **107**, 15293-15298.
- Hess, M., Sczyrba, A., Egan, R., Kim, T.-W., Chokhawala, H., Schroth, G., Luo, S., Clark, D.S., Chen, F., Zhang, T., Mackie, R.I., Pennacchio, L.A., Tringe, S.G., Visel, A., Woyke, T., Wang, Z. and Rubin, E.M. (2011) Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen. *Science*, 331, 463-467.
- Hirano, K., Kurosaki, M., Nihei, S., Hasegawa, H., Shinoda, S., Haruki, M. and Hirano, N. (2016) Enzymatic Diversity of the *Clostridium thermocellum* Cellulosome is Crucial for the Degradation of Crystalline Cellulose and Plant Biomass. *Scientific Reports*, 6, 35709.
- Hoffmann, E.d. and Stroobant, V. (2007) Mass Spectrometry: Principles and Applications 3 edition edn. Chichester: Wiley-Interscience.
- Holm, L. and Rosenström, P. (2010) Dali Server: Conservation Mapping in 3D. Nucl. Acids Res., 38, W545-W549.

- Hotelier, T., Renault, L., Cousin, X., Negre, V., Marchot, P. and Chatonnet, A. (2004) ESTHER, the Database of the α/β - Hydrolase Fold Superfamily of Proteins. *Nucl. Acids Res.*, **32**, D145-D147.
- **Howard, S. and Withers, S.G.** (1998) Bromoketone C-Glycosides, a New Class of β-Glucanase Inactivators. *J Am Chem Soc*, **120**, 10326-10331.
- Hrmova, M., Farkas, V., Lahnstein, J. and Fincher, G.B. (2007) A Barley Xyloglucan Xyloglucosyl Transferase Covalently Links Xyloglucan, Cellulosic Substrates, and (1,3;1,4)-β-D-Glucans. *Journal of Biological Chemistry*, **282**, 12951-12962.
- Huggett, A.S.G. and Nixon, D.A. (1957) Use of Glucose Oxidase, Peroxidase, and O-Dianisidine in Determination of Blood and Urinary Glucose. *The Lancet*, 270, 368-370.
- Huggins, C. and Lapides, J. (1947) Chromogenic Substrates IV. Acyl Esters of P-Nitrophenol as Substrates for the Colorimetric Determination of Esterase. *Journal of Biological Chemistry*, 170, 467-482.
- Hughes, S. and Johnson, D.C. (1983) Triple-Pulse Amperometric Detection of Carbohydrates After Chromatographic Separation. *Analytica Chimica Acta*, 149, 1-10.
- Huson, D.H. and Scornavacca, C. (2012) *Dendroscope 3*: An Interactive Tool for Rooted Phylogenetic Trees and Networks. *Syst Biol*, 61, 1061-1067.
- Ibatullin, F.M., Banasiak, A., Baumann, M.J., Greffe, L., Takahashi, J., Mellerowicz, E.J. and Brumer, H. (2009) A Real-Time Fluorogenic Assay for the Visualisation of Glycoside Hydrolase Activity in Planta. *Plant Physiology*, **151**, 1741-1750.
- Ibatullin, F.M., Baumann, M.J., Greffe, L. and Brumer, H. (2008) Kinetic Analyses of Retaining *Endo*-(Xylo)glucanases from Plant and Microbial Sources Using New Chromogenic Xylogluco-Oligosaccharide Aryl Glycosides[†]. *Biochemistry*, 47, 7762-7769.
- Jacob, F. & Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* 3, 318–356.
- Jeffrey, G.A. and Saenger, W. (1994a) Hydrogen Bonding in Carbohydrates. In *Hydrogen Bonding in Biological Structures*: Springer, Berlin, Heidelberg, pp. 169-219.
- Jeffrey, G.A. and Saenger, W. (1994b) Theoretical Calculations of Hydrogen-Bond Geometries. In Hydrogen Bonding in Biological Structures: Springer, Berlin, Heidelberg, pp. 71-93.

- Jermyn, M.A. (1952) Fungal Cellulases I. General Properties of Unpurified Enzyme Preparations From *Aspergillus Oryzae*. *Aust. Jnl. Of Bio. Sci.*, **5**, 409-432.
- Johansson, P., Brumer, H., Baumann, M.J., Kallas, A.M., Henriksson, H., Denman, S.E., Teeri, T.T. and Jones, T.A. (2004) Crystal Structures of a Poplar Xyloglucan Endotransglycosylase Reveal Details of Transglycosylation Acceptor Binding. *Plant Cell*, 16, 874-886.
- Johnson, K.A. and Goody, R.S. (2011) The Original Michaelis Constant: Translation of the 1913 Michaelis–Menten Paper. *Biochemistry*, **50**, 8264-8269.
- Johnson, K.G., Fontana, J.D. and MacKenzie, C.R. (1988) Measurement of Acetylxylan Esterase in *Streptomyces*. In *Meth. Enzymol.*: Academic Press, pp. 551-560.
- Juncker, A.S., Willenbrock, H., von Heijne, G., Brunak, S., Nielsen, H. and Krogh, A. (2003) Prediction of Lipoprotein Signal Peptides in Gram-Negative Bacteria. *Protein Sci*, 12, 1652-1662.
- Kabisch, A., Otto, A., König, S., Becher, D., Albrecht, D., Schüler, M., Teeling, H., Amann,
 R.I. and Schweder, T. (2014) Functional Characterisation of Polysaccharide Utilisation
 Loci in the Marine Bacteroidetes '*Gramella forsetii*' KT0803. *ISME J*, 8, 1492-1502.
- Kabsch, W. (2010) XDS. Acta Crystallographica Section D Biological Crystallography, 66, 125-132.
- Kaewthai, N., Gendre, D., Eklöf, J.M., Ibatullin, F.M., Ezcurra, I., Bhalerao, R.P. and Brumer, H. (2013) Group III-A XTH Genes of Arabidopsis Encode Predominant Xyloglucan Endohydrolases That Are Dispensable for Normal Growth. *Plant Physiology*, 161, 440-454.
- Kailemia, M.J., Ruhaak, L.R., Lebrilla, C.B. and Amster, I.J. (2014) Oligosaccharide Analysis By Mass Spectrometry: A Review Of Recent Developments. *Anal. Chem.*, 86, 196-212.
- Kaiser, P.M. (1980) Substrate Inhibition as a Problem of Non-Linear Steady State Kinetics with Monomeric Enzymes. *Journal of Molecular Catalysis*, 8, 431-442.
- Kaoutari, A.E., Armougom, F., Gordon, J.I., Raoult, D. and Henrissat, B. (2013) The Abundance and Variety of Carbohydrate-Active Enzymes in the Human Gut Microbiota. *Nat Rev Micro*, **11**, 497-504.

- Kapaev, R.R. and Toukach, P.V. (2016) Simulation of 2D NMR Spectra of Carbohydrates Using *GODESS* Software. J. Chem. Inf. Model., 56, 1100-1104.
- Karplus, M. (1963) Vicinal Proton Coupling in Nuclear Magnetic Resonance. J Am Chem Soc, 85, 2870-2871.
- Kastle, J.H. (1906) The Influence of Chemical Constitution on the Lipolytic Hydrolysis of Ethereal Salts: US Government Printing Office.
- Kato, K. and Matsuda, K. (1969) Studies on the Chemical Structure of Konjac Mannan. *Agricultural and Biological Chemistry*, **33**, 1446-1453.
- Katoh, K. and Standley, D.M. (2014) MAFFT: Iterative Refinement and Additional Methods. Methods Mol. Biol., 1079, 131-146.
- Katsuraya, K., Okuyama, K., Hatanaka, K., Oshima, R., Sato, T. and Matsuzaki, K. (2003) Constitution of konjac glucomannan: chemical analysis and 13C NMR spectroscopy. *Carbohydrate Polymers*, 53, 183-189.
- Keeler, J. (2010) Understanding NMR Spectroscopy 2 edition edn. Chichester, U.K: Wiley.
- Kelley, L.A. and Sternberg, M.J.E. (2009) Protein Structure Prediction on the Web: a Case Study Using the *Phyre* Server. *Nat Protoc*, 4, 363-371.
- **Kempton, J.B. and Withers, S.G.** (1992) Mechanism of *Agrobacterium* β-Glucosidase: Kinetic Studies. *Biochemistry*, **31**, 9961-9969.
- Kengen, S.W., Luesink, E.J., Stams, A.J. and Zehnder, A.J. (1993) Purification and Characterisation of an Extremely Thermostable β-Glucosidase from the Hyperthermophilic Archaeon *Pyrococcus furiosus*. *European Journal of Biochemistry*, 213, 305-312.
- Kiemle, S.N., Zhang, X., Esker, A.R., Toriz, G., Gatenholm, P. and Cosgrove, D.J. (2014) Role of (1,3)(1,4)-β-Glucan in Cell Walls: Interaction with Cellulose. *Biomacromolecules*, **15**, 1727-1736.
- Kim, T., Joo, J.C. and Yoo, Y.J. (2012) Hydrophobic interaction network analysis for thermostabilisation of a mesophilic xylanase. J. Biotechnol., 161, 49-59.
- Kim, Y.S., Lee, J.H., Yoon, G.M., Cho, H.S., Park, S.-W., Suh, M.C., Choi, D., Ha, H.J., Liu, J.R. and Pai, H.-S. (2000) CHRK1, a Chitinase-Related Receptor-Like Kinase in Tobacco. *Plant Physiology*, **123**, 905-916.

Kitz, R. and Wilson, I.B. (1962) Esters of Methanesulfonic Acid as Irreversible Inhibitors of Acetylcholinesterase. *Journal of Biological Chemistry*, 237, 3245-3249.

Kleywegt, G.J. (2000) Validation of Protein Crystal Structures. Acta Cryst. D, 56, 249-265.

- Kong, Y., Peña, M.J., Renna, L., Avci, U., Pattathil, S., Tuomivaara, S.T., Li, X., Reiter,
 W.-D., Brandizzi, F., Hahn, M.G., Darvill, A.G., York, W.S. and O'Neill, M.A.
 (2015) Galactose-Depleted Xyloglucan is Dysfunctional and Leads to Dwarfism in
 Arabidopsis. Plant Physiology, 167, 1296-1306.
- Koropatkin, N.M., Cameron, E.A. and Martens, E.C. (2012) How Glycan Metabolism Shapes the Human Gut Microbiota. *Nat Rev Microbiol*, **10**, 323-335.
- Koropatkin, N.M., Martens, E.C., Gordon, J.I. and Smith, T.J. (2008) Starch Catabolism by a Prominent Human Gut Symbiont is Directed by the Recognition of Amylose Helices. *Structure*, 16, 1105-1115.
- Koshland, D.E. (1953) Stereochemistry and the Mechanism of Enzymatic Reactions. *Biological Reviews*, **28**, 416-436.
- **Koshland, D.E.** (2002) The Application and Usefulness of the Ratio k_{cat}/K_M . *Bioorganic Chemistry*, **30**, 211-213.
- Kovatcheva-Datchary, P., Nilsson, A., Akrami, R., Lee, Ying S., De Vadder, F., Arora, T., Hallen, A., Martens, E., Björck, I. and Bäckhed, F. (2015) Dietary Fiber-Induced Improvement in Glucose Metabolism Is Associated with Increased Abundance of *Prevotella. Cell Metabolism*, 22, 971-982.
- Krissinel, E. and Henrick, K. (2007) Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology*, **372**, 774-797.
- Labourel, A., Jam, M., Jeudy, A., Hehemann, J.-H., Czjzek, M. and Michel, G. (2014) The β-Glucanase ZgLamA from Zobellia galactanivorans Evolved a Bent Active Site Adapted for Efficient Degradation of Algal Laminarin. *Journal of Biological Chemistry*, 289, 2027-2042.
- Lairson, L.L., Henrissat, B., Davies, G.J. and Withers, S.G. (2008) Glycosyltransferases: Structures, Functions, and Mechanisms. *Annual Review of Biochemistry*, 77, 521-555.
- Langer, G., Cohen, S.X., Lamzin, V.S. and Perrakis, A. (2008) Automated Macromolecular Model Building for X-ray Crystallography using *ARP/wARP* Version 7. *Nat Protoc*, 3, 1171-1179.

- Larsbrink, J., Izumi, A., Ibatullin, Farid M., Nakhai, A., Gilbert, Harry J., Davies,
 Gideon J. and Brumer, H. (2011) Structural and Enzymatic Characterisation of a
 Glycoside Hydrolase Family 31 α-Xylosidase from *Cellvibrio japonicus* Involved in
 Xyloglucan Saccharification. *Biochemical Journal*, 436, 567-580.
- Larsbrink, J., Rogers, T.E., Hemsworth, G.R., McKee, L.S., Tauzin, A.S., Spadiut, O.,
 Klinter, S., Pudlo, N.A., Urs, K., Koropatkin, N.M., Creagh, A.L., Haynes, C.A.,
 Kelly, A.G., Cederholm, S.N., Davies, G.J., Martens, E.C. and Brumer, H. (2014a) A
 Discrete Genetic Locus Confers Xyloglucan Metabolism in Select Human Gut
 Bacteroidetes. *Nature*, 506, 498-502.
- Larsbrink, J., Thompson, A.J., Lundqvist, M., Gardner, J.G., Davies, G.J. and Brumer, H. (2014b) A Complex Gene Locus Enables Xyloglucan Utilisation in the Model Saprophyte Cellvibrio japonicus. Molecular Microbiology, 94, 418-33.
- Larsbrink, J., Zhu, Y., Kharade, S.S., Kwiatkowski, K.J., Eijsink, V.G.H., Koropatkin, N.M., McBride, M.J. and Pope, P.B. (2016) A Polysaccharide Utilisation Locus from *Flavobacterium johnsoniae* Enables Conversion of Recalcitrant Chitin. *Biotechnology for Biofuels*, 9, 260.
- Lee, J., Burns, T.H., Light, G., Sun, Y., Fokar, M., Kasukabe, Y., Fujisawa, K., Maekawa, Y. and Allen, R.D. (2010) *Xyloglucan Endotransglycosylase/hydrolase* Genes in Cotton and their Role in Fiber Elongation. *Planta*, 232, 1191-1205.
- Lee, N., Francklyn, C. and Hamilton, E.P. (1987) Arabinose-Induced Binding of AraC Protein to *araI2* Activates the *araBAD* Operon Promoter. *Proc Natl Acad Sci U S A*, 84, 8814-8818.
- Leeuwen, M.J.F.S.-v., Broek, L.A.M.v.d., Schols, H.A., Beldman, G. and Voragen, A.G.J. (1992) Rhamnogalacturonan Acetylesterase: a Novel Enzyme from *Aspergillus aculeatus*, Specific for the Deacetylation of Hairy (Ramified) Regions of Pectins. *Appl Microbiol Biotechnol*, **38**, 347-349.
- Lei, M., Mechref, Y. and Novotny, M.V. (2009) Structural Analysis of Sulfated Glycans by Sequential Double-Permethylation using Methyl Iodide and Deuteromethyl Iodide. J Am Soc Mass Spectrom, 20, 1660-1671.
- Levasseur, A., Drula, E., Lombard, V., Coutinho, P.M. and Henrissat, B. (2013) Expansion of the Enzymatic Repertoire of the CAZy Database to Integrate Auxiliary Redox Enzymes. *Biotechnology for Biofuels*, **6**, 41.
- Lever, M. (1973) Colorimetric and Fluorometric Carbohydrate Determination with *p*-Hydroxybenzoic Acid Hydrazide. *Biochemical Medicine*, **7**, 274-281.
- Levine, M.N., Lavis, L.D. and Raines, R.T. (2008) Trimethyl Lock: A Stable Chromogenic Substrate for Esterases. *Molecules*, 13, 204-211.
- Levisson, M., Han, G.W., Deller, M.C., Xu, Q., Biely, P., Hendriks, S., Ten Eyck, L.F.,
 Flensburg, C., Roversi, P., Miller, M.D., McMullan, D., von Delft, F., Kreusch, A.,
 Deacon, A.M., van der Oost, J., Lesley, S.A., Elsliger, M.-A., Kengen, S.W.M. and
 Wilson, I.A. (2012) Functional and Structural Characterisation of a Thermostable Acetyl
 Esterase from *Thermotoga maritima*. *Proteins*, 80, 1545-1559.
- Levitt, M.H. (2008) *Spin Dynamics: Basics of Nuclear Magnetic Resonance* 2 edition edn. Chichester, England ; Hoboken, NJ: Wiley.
- Li, L.-L., Taghavi, S., McCorkle, S.M., Zhang, Y.-B., Blewitt, M.G., Brunecky, R., Adney,
 W.S., Himmel, M.E., Brumm, P., Drinkwater, C., Mead, D.A., Tringe, S.G. and
 Lelie, D.v.d. (2011) Bioprospecting Metagenomics of Decaying Wood: Mining for New
 Glycoside Hydrolases. *Biotechnology for Biofuels*, 4, 23.
- **Libertini, E., Li, Y. and McQueen-Mason, S.J.** (2004) Phylogenetic Analysis of the Plant *Endo*-β-1,4-Glucanase Gene Family. *J Mol Evol*, **58**, 506-515.
- Lombard, V., Bernard, T., Rancurel, C., Brumer, H., Coutinho, Pedro M. and Henrissat,
 B. (2010) A Hierarchical Classification of Polysaccharide Lyases for Glycogenomics.
 Biochemical Journal, 432, 437-444.
- Lombard, V., Ramulu, H.G., Drula, E., Coutinho, P.M. and Henrissat, B. (2014) The Carbohydrate-Active Enzymes Database (CAZy) in 2013. *Nucl. Acids Res.*, **42**, D490-D495.
- Lopez-Casado, G., Urbanowicz, B.R., Damasceno, C.M.B. and Rose, J.K.C. (2008) Plant Glycosyl Hydrolases and Biofuels: a Natural Marriage. *Current Opinion in Plant Biology*, **11**, 329-337.

- Lopez, M., Bizot, H., Chambat, G., Marais, M.-F., Zykwinska, A., Ralet, M.-C., Driguez, H. and Buléon, A. (2010) Enthalpic Studies of Xyloglucan–Cellulose Interactions. *Biomacromolecules*, **11**, 1417-1428.
- Luis, A.S., Briggs, J., Zhang, X., Farnell, B., Ndeh, D., Labourel, A., Baslé, A., Cartmell, A., Terrapon, N., Stott, K., Lowe, E.C., McLean, R., Shearer, K., Schückel, J., Venditto, I., Ralet, M.-C., Henrissat, B., Martens, E.C., Mosimann, S.C., Abbott, D.W. and Gilbert, H.J. (2017) Dietary Pectic Glycans are Degraded by Coordinated Enzyme Pathways in Human Colonic *Bacteroides*. *Nature Microbiology*, 1.
- Maekaji, K. (1978) Determination of Acidic Component of Konjac Mannan. *Agricultural and Biological Chemistry*, **42**, 177-178.
- Malet, C., Jimenez-Barbero, J., Bernabe, M., Brosa, C. and Planas, A. (1993) Stereochemical Course and Structure of the Products of the Enzymic Action of *Endo*-1,3-1,4-Beta-D-Glucan 4-Glucanohydrolase from *Bacillus licheniformis*. *Biochemical Journal*, 296, 753-758.
- Manchenko, G. (2002) Handbook of Detection of Enzymes on Electrophoretic Gels. In *CRC Press.*
- Maris, A., Kaewthai, N., Eklöf, J.M., Miller, J.G., Brumer, H., Fry, S.C., Verbelen, J.-P. and Vissenberg, K. (2011) Differences in Enzymic Properties of Five Recombinant *xyloglucan endotransglucosylase/hydrolase (XTH)* Proteins of *Arabidopsis thaliana*. J *Exp Bot*, 62, 261-271.
- Mark, P., Baumann, M.J., Eklöf, J.M., Gullfot, F., Michel, G., Kallas, Å.M., Teeri, T.T.,
 Brumer, H. and Czjzek, M. (2009) Analysis of *nasturtium* TmNXG1 Complexes by
 Crystallography and Molecular Dynamics Provides Detailed Insight into Substrate
 Recognition by Family GH16 Xyloglucan *Endo*-Transglycosylases and *Endo*-Hydrolases.
 Proteins, 75, 820-836.
- Mark, P., Zhang, Q., Czjzek, M., Brumer, H. and Ågren, H. (2011) Molecular Dynamics Simulations of a Branched Tetradecasaccharide Substrate in the Active Site of a Xyloglucan *Endo*-Transglycosylase. *Molecular Simulation*, **37**, 1001-1013.
- Markovič, O. and Janeček, Š. (2001) Pectin Degrading Glycoside Hydrolases of Family 28: Sequence-Structural Features, Specificities and Evolution. *Protein Eng Des Sel*, 14, 615-631.

- Martens, E.C., Kelly, A.G., Tauzin, A.S. and Brumer, H. (2014) The Devil Lies in the Details: How Variations in Polysaccharide Fine-Structure Impact the Physiology and Evolution of Gut Microbes. *Journal of Molecular Biology*, 426, 3851-3865.
- Martens, E.C., Koropatkin, N.M., Smith, T.J. and Gordon, J.I. (2009) Complex Glycan Catabolism by the Human Gut Microbiota: The Bacteroidetes Sus-like Paradigm. *Journal* of Biological Chemistry, 284, 24673-24677.
- Martens, E.C., Lowe, E.C., Chiang, H., Pudlo, N.A., Wu, M., McNulty, N.P., Abbott, D.W., Henrissat, B., Gilbert, H.J., Bolam, D.N. and Gordon, J.I. (2011) Recognition and Degradation of Plant Cell Wall Polysaccharides by Two Human Gut Symbionts. *PLOS Biology*, 9, e1001221.
- Martinez, D., Berka, R.M., Henrissat, B., Saloheimo, M., Arvas, M., Baker, S.E., Chapman, J., Chertkov, O., Coutinho, P.M., Cullen, D., Danchin, E.G.J., Grigoriev, I.V., Harris, P., Jackson, M., Kubicek, C.P., Han, C.S., Ho, I., Larrondo, L.F., Leon, A.L.d., Magnuson, J.K., Merino, S., Misra, M., Nelson, B., Putnam, N., Robbertse, B., Salamov, A.A., Schmoll, M., Terry, A., Thayer, N., Westerholm-Parvinen, A., Schoch, C.L., Yao, J., Barabote, R., Nelson, M.A., Detter, C., Bruce, D., Kuske, C.R., Xie, G., Richardson, P., Rokhsar, D.S., Lucas, S.M., Rubin, E.M., Dunn-Coleman, N., Ward, M. and Brettin, T.S. (2008) Genome Sequencing and Analysis of the Biomass-Degrading Fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nature Biotechnology*, 26, 553.
- Matsushita, O., Russell, J.B. and Wilson, D.B. (1990) Cloning and Sequencing of a Bacteroides ruminicola B₁4 Endoglucanase Gene. J Bacteriol, 172, 3620-3630.
- Mazmanian, K., Sargsyan, K., Grauffel, C., Dudev, T. and Lim, C. (2016) Preferred Hydrogen-Bonding Partners of Cysteine: Implications for Regulating Cys Functions. J. Phys. Chem. B, 120, 10288-10296.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) *Phaser* Crystallographic Software. *Journal of Applied Crystallography*, 40, 658-674.
- McDermid, K.P., Forsberg, C.W. and MacKenzie, C.R. (1990) Purification and Properties of an Acetylxylan Esterase from *Fibrobacter succinogenes* S85. *Appl Environ Microbiol*, 56, 3805-3810.

- McGregor, N., Arnal, G. and Brumer, H. (2017a) Quantitative Kinetic Characterisation of Glycoside Hydrolases Using High-Performance Anion-Exchange Chromatography (HPAEC). In *Protein-Carbohydrate Interactions*: Humana Press, New York, NY, pp. 15-25.
- McGregor, N., Morar, M., Fenger, T.H., Stogios, P., Lenfant, N., Yin, V., Xu, X.,
 Evdokimova, E., Cui, H., Henrissat, B., Savchenko, A. and Brumer, H. (2016)
 Structure-Function Analysis of a Mixed-linkage β-Glucanase/Xyloglucanase from the
 Key Ruminal Bacteroidetes *Prevotella bryantii* B₁4. *Journal of Biological Chemistry*,
 291, 1175-1197.
- McGregor, N., Yin, V., Tung, C.-C., Van Petegem, F. and Brumer, H. (2017b) Crystallographic Insight into the Evolutionary Origins of Xyloglucan Endotransglycosylases and Endohydrolases. *The Plant Journal*, **89**, 651-670.
- McPhillips, T.M., McPhillips, S.E., Chiu, H.-J., Cohen, A.E., Deacon, A.M., Ellis, P.J.,
 Garman, E., Gonzalez, A., Sauter, N.K., Phizackerley, R.P., Soltis, S.M. and Kuhn,
 P. (2002) Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines. *Journal of Synchrotron Radiation*, 9, 401-406.
- Merceron, R., Foucault, M., Haser, R., Mattes, R., Watzlawick, H. and Gouet, P. (2012) The Molecular Mechanism of Thermostable α-Galactosidases AgaA and AgaB Explained by X-ray Crystallography and Mutational Studies. *Journal of Biological Chemistry*, 287, 39642-39652.
- Mewalal, R., Mizrachi, E., Mansfield, S.D. and Myburg, A.A. (2014) Cell Wall-Related Proteins of Unknown Function: Missing Links in Plant Cell Wall Development. *Plant Cell Physiol*, 55, 1031-1043.
- Mewis, K., Taupp, M. and Hallam, S.J. (2011) A High Throughput Screen for Biomining Cellulase Activity from Metagenomic Libraries. *J Vis Exp*.
- Miao, S., McCarter, J.D., Grace, M.E., Grabowski, G.A., Aebersold, R. and Withers, S.G. (1994) Identification of Glu340 as the Active-Site Nucleophile in Human Glucocerebrosidase by use of Electrospray Tandem Mass Spectrometry. *Journal of Biological Chemistry*, 269, 10975-10978.

Michaelis, L. and Menten, M.L. (1913) Die Kinetik der Invertinwirkung. Biochem. z, 49, 352.

- Michel, G., Chantalat, L., Duee, E., Barbeyron, T., Henrissat, B., Kloareg, B. and Dideberg,
 O. (2001) The κ-Carrageenase of *P. carrageenovora* Features a Tunnel-Shaped Active Site: A Novel Insight in the Evolution of Clan-B Glycoside Hydrolases. *Structure*, 9, 513-525.
- Miller, G.L. (1959) Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Anal. Chem.*, **31**, 426-428.
- Minic, Z. and Jouanin, L. (2006) Plant Glycoside Hydrolases Involved in Cell Wall Polysaccharide Degradation. *Plant Physiology and Biochemistry*, 44, 435-449.
- Minor, W., Cymborowski, M., Otwinowski, Z. and Chruszcz, M. (2006) HKL-3000: the Integration of Data Reduction and Structure Solution – from Diffraction Images to an Initial Model in Minutes. *Acta Crystallographica Section D Biological Crystallography*, 62, 859-866.
- Mølgaard, A., Kauppinen, S. and Larsen, S. (2000) Rhamnogalacturonan Acetylesterase Elucidates the Structure and Function of a New Family of Hydrolases. *Structure*, 8, 373-383.
- Montanier, C., Bueren, A.L.v., Dumon, C., Flint, J.E., Correia, M.A., Prates, J.A., Firbank,
 S.J., Lewis, R.J., Grondin, G.G., Ghinet, M.G., Gloster, T.M., Herve, C., Knox, J.P.,
 Talbot, B.G., Turkenburg, J.P., Kerovuo, J., Brzezinski, R., Fontes, C.M.G.A.,
 Davies, G.J., Boraston, A.B. and Gilbert, H.J. (2009a) Evidence that Family 35
 Carbohydrate-Binding Modules Display Conserved Specificity but Divergent Function. *PNAS*, 106, 3065-3070.
- Montanier, C., Money, V.A., Pires, V.M.R., Flint, J.E., Pinheiro, B.A., Goyal, A., Prates, J.A.M., Izumi, A., Stålbrand, H., Morland, C., Cartmell, A., Kolenova, K., Topakas, E., Dodson, E.J., Bolam, D.N., Davies, G.J., Fontes, C.M.G.A. and Gilbert, H.J. (2009b) The Active Site of a Carbohydrate Esterase Displays Divergent Catalytic and Noncatalytic Binding Functions. *PLOS Biology*, 7, e1000071.
- Montoro-García, S., Gil-Ortiz, F., García-Carmona, F., Polo, L.M., Rubio, V. and Sánchez-Ferrer, Á. (2011) The Crystal Structure of the Cephalosporin Deacetylating Enzyme Acetyl Xylan Esterase bound to Paraoxon Explains the Low Sensitivity of this Serine Hydrolase to Organophosphate Inactivation. *Biochem. J.*, 436, 321-330.

- Moraïs, S., Lamed, R. and Bayer, E. (2012) Affinity Electrophoresis as a Method for Determining Substrate-Binding Specificity of Carbohydrate-Active Enzymes for Soluble Polysaccharides. In *Biomass Conversion* (Himmel, M.E. ed: Humana Press, pp. 119-127.
- Moreira, L.R.S. and Filho, E.X.F. (2008) An Overview of Mannan Structure and Mannan-Degrading Enzyme Systems. *Appl Microbiol Biotechnol*, **79**, 165-178.
- Mosi, R., Sham, H., Uitdehaag, J.C.M., Ruiterkamp, R., Dijkstra, B.W. and Withers, S.G. (1998) Reassessment of Acarbose as a Transition State Analogue Inhibitor of Cyclodextrin Glycosyltransferase. *Biochemistry*, **37**, 17192-17198.
- Murphy Iv, F.V. (1999) The Structure of a Chromosomal High Mobility Group Protein-DNA Complex Reveals Sequence-Neutral Mechanisms Important for Non-Sequence-Specific DNA Recognition. *The EMBO Journal*, **18**, 6610-6618.
- Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method. Acta Crystallographica Section D Biological Crystallography, 53, 240-255.
- Mutenda, K.E. and Matthiesen, R. (2007) Analysis of Carbohydrates by Mass Spectrometry. In *Mass Spectrometry Data Analysis in Proteomics* (Matthiesen, R. ed: Humana Press, pp. 289-301.
- Nakamura, A.M., Nascimento, A.S. and Polikarpov, I. (2017) Structural Diversity of Carbohydrate Esterases. *Biotechnology Research and Innovation*, 1, 35-51.
- Namchuk, M.N. and Withers, S.G. (1995) Mechanism of *Agrobacterium* β-Glucosidase: Kinetic Analysis of the Role of Noncovalent Enzyme/Substrate Interactions. *Biochemistry*, **34**, 16194-16202.
- Naor, M.M. and Jensen, J.H. (2004) Determinants of Cysteine pKa Values in Creatine Kinase and Alpha1-Antitrypsin. *Proteins*, 57, 799-803.
- Ndeh, D., Rogowski, A., Cartmell, A., Luis, A.S., Baslé, A., Gray, J., Venditto, I., Briggs, J.,
 Zhang, X., Labourel, A., Terrapon, N., Buffetto, F., Nepogodiev, S., Xiao, Y., Field,
 R.A., Zhu, Y., O'Neill, M.A., Urbanowicz, B.R., York, W.S., Davies, G.J., Abbott,
 D.W., Ralet, M.-C., Martens, E.C., Henrissat, B. and Gilbert, H.J. (2017) Complex
 Pectin Metabolism by Gut Bacteria Reveals Novel Catalytic Functions. *Nature*, 544, 65.

- Neugebauer, H., Herrmann, C., Kammer, W., Schwarz, G., Nordheim, A. and Braun, V. (2005) ExbBD-Dependent Transport of Maltodextrins through the Novel MalA Protein across the Outer Membrane of *Caulobacter crescentus*. *J Bacteriol*, **187**, 8300-8311.
- Nishitani, K. (1998) Construction and restructuring of the cellulose-xyloglucan framework in the apoplast as mediated by the xyloglucan-related protein family—A hypothetical scheme. *J. Plant Res.*, **111**, 159.
- Noinaj, N., Guillier, M., Barnard, T.J. and Buchanan, S.K. (2010) TonB-Dependent Transporters: Regulation, Structure, and Function. *Annual review of microbiology*, 64, 43-60.
- Notenboom, V., Birsan, C., Nitz, M., Rose, D.R., Warren, R.A.J. and Withers, S.G. (1998) Insights into Transition State Stabilisation of the β-1,4-Glycosidase Cex by Covalent Intermediate Accumulation in Active Site Mutants. *Nature Structural and Molecular Biology*, **5**, 812.
- Paetzel, M., Karla, A., Strynadka, N.C.J. and Dalbey, R.E. (2002) Signal Peptidases. Chem. Rev., 102, 4549-4580.
- Palackal, N., Lyon, C.S., Zaidi, S., Luginbühl, P., Dupree, P., Goubet, F., Macomber, J.L.,
 Short, J.M., Hazlewood, G.P., Robertson, D.E. and Steer, B.A. (2007) A
 Multifunctional Hybrid Glycosyl Hydrolase Discovered in an Uncultured Microbial
 Consortium from Ruminant Gut. *Appl Microbiol Biotechnol*, 74, 113-124.
- Pannu, N.S., Murshudov, G.N., Dodson, E.J. and Read, R.J. (1998) Incorporation of Prior Phase Information Strengthens Maximum-Likelihood Structure Refinement. Acta Crystallographica Section D Biological Crystallography, 54, 1285-1294.

Pauling, L. (1948) Nature of Forces between Large Molecules of Biological Interest. In Nature.

- Pauly, M., Gille, S., Liu, L., Mansoori, N., Souza, A.d., Schultink, A. and Xiong, G. (2013) Hemicellulose Biosynthesis. *Planta*, 238, 627-642.
- Pauly, M. and Keegstra, K. (2016) Biosynthesis of the Plant Cell Wall Matrix Polysaccharide Xyloglucan. Annual Review of Plant Biology, 67, 235-259.
- Peng, F.Y., Reid, K.E., Liao, N., Schlosser, J., Lijavetzky, D., Holt, R., Martínez Zapater, J.M., Jones, S., Marra, M., Bohlmann, J. and Lund, S.T. (2007) Generation of ESTs in *Vitis vinifera* Wine Grape (Cabernet Sauvignon) and Table Grape (Muscat Hamburg)

and Discovery of New Candidate Genes with Potential Roles in Berry Development. *Gene*, **402**, 40-50.

- Petersen, T.N., Brunak, S., von Heijne, G. and Nielsen, H. (2011) SignalP 4.0: Discriminating Signal Peptides from Transmembrane Regions. Nat Meth, 8, 785-786.
- Pettolino, F.A., Walsh, C., Fincher, G.B. and Bacic, A. (2012) Determining the Polysaccharide Composition of Plant Cell Walls. *Nat Protoc*, 7, 1590-1607.
- Planas, A. (2000) Bacterial 1,3-1,4-β-Glucanases: Structure, Function and Protein Engineering.
 Biochimica et Biophysica Acta (BBA) Protein Structure and Molecular Enzymology,
 1543, 361-382.
- Planas, A., Juncosa, M., Cayetano, A. and Querol, E. (1992) Studies on *Bacillus licheniformis Endo*-β-1,3-1,4-D-Glucanase: Characterisation and Kinetic Analysis. *Appl Microbiol Biotechnol*, 37, 583-589.
- Pourcher, T., Bassilana, M., Sarkar, H.K., Kaback, H.R. and Leblanc, G. (1990) The Melibiose/Na+ Symporter of *Escherichia coli*: Kinetic and Molecular Properties. *Phil. Trans. R. Soc. Lond. B*, **326**, 411-423.
- Price, M.N., Dehal, P.S. and Arkin, A.P. (2010) FastTree 2 Approximately Maximum-Likelihood Trees for Large Alignments. PLOS ONE, 5.
- Purushe, J., Fouts, D.E., Morrison, M., White, B.A., Mackie, R.I., Coutinho, P.M., Henrissat, B., Nelson, K.E. and Bacteria, t.N.A.C.f.R. (2010) Comparative Genome Analysis of *Prevotella ruminicola* and *Prevotella bryantii*: Insights into Their Environmental Niche. *Microb Ecol*, **60**, 721-729.
- Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D.R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.-M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H.B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., Doré, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Bork, P., Ehrlich, S.D. and Wang, J. (2010) A Human Gut Microbial Gene Catalog Established by Metagenomic Sequencing. *Nature*, 464, 59-65.

- Quaile, A.T., Stogios, P.J., Egorova, O., Evdokimova, E., Valleau, D., Nocek, B., Kompella, P.S., Peisajovich, S., Yakunin, A.F., Ensminger, A.W. and Savchenko, A. (2018) The *Legionella pneumophila* Effector Ceg4 is a Phosphotyrosine Phosphatase that Attenuates Activation of Eukaryotic MAPK Pathways. *Journal of Biological Chemistry*, jbc.M117.812727.
- Rangarajan, E.S., Ruane, K.M., Proteau, A., Schrag, J.D., Valladares, R., Gonzalez, C.F.,
 Gilbert, M., Yakunin, A.F. and Cygler, M. (2011) Structural and Enzymatic
 Characterisation of NanS (YjhS), a 9-O-Acetyl N-Acetylneuraminic Acid Esterase from *Escherichia coli* O157:H7. Protein Sci, 20, 1208-1219.
- Ravachol, J., Borne, R., Tardif, C., de Philip, P. and Fierobe, H.-P. (2014) Characterisation of All Family-9 Glycoside Hydrolases Synthesised by the Cellulosome-Producing Bacterium *Clostridium cellulolyticum*. J Biol Chem, 289, 7335-7348.
- Redmond, M. and Fielder, D. (2006) Extraction and Purification Method for Cereal *beta*-Glucan: Ceapro Inc.
- Reese, E.T. (1956) Enzymatic Hydrolysis of Cellulose. Appl Microbiol, 4, 39-45.
- Reese, E.T., Siu, R.G.H. and Levinson, H.S. (1950) The Biological Degradation of Soluble Cellulose Derivatives and its Relationship to the Mechanism of Cellulose Hydrolysis. J Bacteriol, 59, 485-497.
- Reeves, A.R., Wang, G.R. and Salyers, A.A. (1997) Characterisation of Four Outer Membrane Proteins that Play a Role in Utilisation of Starch by *Bacteroides thetaiotaomicron*. J *Bacteriol*, **179**, 643-649.
- **Rempel, B.P. and Withers, S.G.** (2008) Covalent Inhibitors of Glycosidases and their Applications in Biochemistry and Biology. *Glycobiology*, **18**, 570-586.
- Robert, X. and Gouet, P. (2014) Deciphering Key Features in Protein Structures with the New ENDscript Server. Nucl. Acids Res., 42, W320-W324.
- Roberts, T.L. and Rosenkrantz, H. (1966) Lipase and Acetyl Esterase Activities in Intact Bacillus Coagulans Spores. *Can. J. Biochem.*, **44**, 677-685.
- Rocklin, R.D. and Pohl, C.A. (1983) Determination of Carbohydrates by Anion Exchange Chromatography with Pulsed Amperometric Detection. *Journal of Liquid Chromatography*, 6, 1577-1590.

- Rodgers, M.E. and Schleif, R. (2009) Solution Structure of the DNA Binding Domain of AraC Protein. *Proteins*, **77**, 202-208.
- Rogowski, A., Briggs, J.A., Mortimer, J.C., Tryfona, T., Terrapon, N., Lowe, E.C., Baslé,
 A., Morland, C., Day, A.M., Zheng, H., Rogers, T.E., Thompson, P., Hawkins, A.R.,
 Yadav, M.P., Henrissat, B., Martens, E.C., Dupree, P., Gilbert, H.J. and Bolam,
 D.N. (2015) Glycan Complexity Dictates Microbial Resource Allocation in the Large
 Intestine. *Nature Communications*, 6, 7481.
- Rose, I.A., Grunberg-Manago, M., Korey, S.R. and Ochoa, S. (1954) Enzymatic Phosphorylation of Acetate. *Journal of Biological Chemistry*, **211**, 737-756.
- Rose, J.K.C., Braam, J., Fry, S.C. and Nishitani, K. (2002) The XTH Family of Enzymes Involved in Xyloglucan Endotransglucosylation and Endohydrolysis: Current Perspectives and a New Unifying Nomenclature. *Plant Cell Physiol*, 43, 1421-1435.
- Rose, J.K.C. and Lee, S.-J. (2010) Straying off the Highway: Trafficking of Secreted Plant Proteins and Complexity in the Plant Cell Wall Proteome. *Plant Physiology*, 153, 433-436.
- **Rosenthal, A.L. and Saifer, A.** (1973) Continuous UV Monitoring of Fluorogenic Substrates: I. Kinetic Analysis of N-Acetyl-β-D-Hexosaminidases. *Anal. Biochem.*, **55**, 85-92.
- Rowe, S.E., Campbell, C., Lowry, C., O'Donnell, S.T., Olson, M.E., Lindgren, J.K., Waters,
 E.M., Fey, P.D. and O'Gara, J.P. (2016) AraC-Type Regulator Rbf Controls the
 Staphylococcus epidermidis Biofilm Phenotype by Negatively Regulating the *icaADBC* Repressor SarR. J Bacteriol, 198, 2914-2924.
- Rowley, J., Decker, S.R., Michener, W. and Black, S. (2013) Efficient Extraction of Xylan from Delignified Corn Stover using Dimethyl Sulfoxide. *3 Biotech*, **3**, 433-438.
- **Rupp, B.** (2009) *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology* 1 edition edn. New York: Garland Science.
- Russell, R.R., Aduse-Opoku, J., Sutcliffe, I.C., Tao, L. and Ferretti, J.J. (1992) A Binding Protein-Dependent Transport System in *Streptococcus mutans* Responsible for Multiple Sugar Metabolism. *Journal of Biological Chemistry*, 267, 4631-4637.
- Sampedro, J., Gianzo, C., Iglesias, N., Guitián, E., Revilla, G. and Zarra, I. (2012) AtBGAL10 Is the Main Xyloglucan β-Galactosidase in *Arabidopsis*, and Its Absence

Results in Unusual Xyloglucan Subunits and Growth Defects. *Plant Physiology*, **158**, 1146-1157.

- Sato, S., Kato, T., Kakegawa, K., Ishii, T., Liu, Y.-G., Awano, T., Takabe, K., Nishiyama,
 Y., Kuga, S., Sato, S., Nakamura, Y., Tabata, S. and Shibata, D. (2001) Role of the
 Putative Membrane-Bound *Endo*-1,4-β-Glucanase KORRIGAN in Cell Elongation and
 Cellulose Synthesis in *Arabidopsis thaliana*. *Plant Cell Physiol*, 42, 251-263.
- Schagerlöf, H., Nilsson, C., Gorton, L., Tjerneld, F., Stålbrand, H. and Cohen, A. (2009) Use of ¹⁸O Water and ESI-MS Detection in Subsite Characterisation and Investigation of the Hydrolytic Action of an Endoglucanase. *Anal Bioanal Chem*, **394**, 1977-1984.
- Scheer, W.D., Lehmann, H.P. and Beeler, M.F. (1978) An Improved Assay for Hexokinase Activity in Human Tissue Homogenates. *Anal. Biochem.*, 91, 451-463.
- Scheible, W.-R. and Pauly, M. (2004) Glycosyltransferases and Cell Wall Biosynthesis: Novel Players and Insights. *Current Opinion in Plant Biology*, 7, 285-295.
- Scheller, H.V. and Ulvskov, P. (2010) Hemicelluloses. *Annual Review of Plant Biology*, 61, 263-289.
- Schols, H.A., Bakx, E.J., Schipper, D. and Voragen, A.G.J. (1995) A Xylogalacturonan Subunit Present in the Modified Hairy Regions of Apple Pectin. *Carbohydr. Res.*, 279, 265-279.
- Schramm, V.L. (2011) Enzymatic Transition States, Transition-State Analogs, Dynamics, Thermodynamics, and Lifetimes. *Annual review of biochemistry*, **80**, 703-732.
- Scott, K.P., Duncan, S.H. and Flint, H.J. (2008) Dietary Fibre and the Gut Microbiota. *Nutrition Bulletin*, **33**, 201-211.
- Selig, M.J., Adney, W.S., Himmel, M.E. and Decker, S.R. (2009) The Impact of Cell Wall Acetylation on Corn Stover Hydrolysis by Cellulolytic and Xylanolytic Enzymes. *Cellulose*, 16, 711-722.
- Senoura, T., Ito, S., Taguchi, H., Higa, M., Hamada, S., Matsui, H., Ozawa, T., Jin, S., Watanabe, J., Wasaki, J. and Ito, S. (2011) New Microbial Mannan Catabolic Pathway that Involves a Novel Mannosylglucose Phosphorylase. *Biochemical and Biophysical Research Communications*, 408, 701-706.

- Shi, Y., Wang, Y.-F., Jayaraman, L., Yang, H., Massagué, J. and Pavletich, N.P. (1998) Crystal Structure of a Smad MH1 Domain Bound to DNA: Insights on DNA Binding in TGF-β Signaling. *Cell*, **94**, 585-594.
- Shipman, J.A., Berleman, J.E. and Salyers, A.A. (2000) Characterisation of Four Outer Membrane Proteins Involved in Binding Starch to the Cell Surface of *Bacteroides thetaiotaomicron. J Bacteriol*, **182**, 5365-5372.
- Silva, L.a.O., Terrasan, C.R.F. and Carmona, E.C. (2015) Purification and Characterisation of Xylanases from *Trichoderma inhamatum*. *Electronic Journal of Biotechnology*, 18, 307-313.
- Simmons, T.J., Mohler, K.E., Holland, C., Goubet, F., Franková, L., Houston, D.R., Hudson, A.D., Meulewaeter, F. and Fry, S.C. (2015) Hetero-*trans*-β-Glucanase, an Enzyme Unique to *Equisetum* plants, Functionalises Cellulose. *The Plant Journal*, n/an/a.
- Simmons, T.J., Uhrín, D., Gregson, T., Murray, L., Sadler, I.H. and Fry, S.C. (2013) An Unexpectedly Lichenase-Stable Hexasaccharide from Cereal, Horsetail and Lichen Mixed-Linkage β-Glucans (MLGs): Implications for MLG Subunit Distribution. *Phytochemistry*, **95**, 322-332.
- Simone, G.D., Alterio, V. and Supuran, C.T. (2013) Exploiting the Hydrophobic and Hydrophilic Binding Sites for Designing Carbonic Anhydrase Inhibitors. *Expert Opinion* on Drug Discovery, 8, 793-810.
- Singh, M.K. and Manoj, N. (2016) An Extended Loop in CE7 Carbohydrate Esterase Family is Dispensable for Oligomerisation but Required for Activity and Thermostability. *Journal* of Structural Biology, **194**, 434-445.
- Sjostrom, S.L., Joensson, H.N. and Svahn, H.A. (2013) Multiplex Analysis of Enzyme Kinetics and Inhibition by Droplet Microfluidics using Picoinjectors. *Lab Chip*, **13**, 1754-1761.
- Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J. and Klenk, D.C. (1985) Measurement of Protein using Bicinchoninic Acid. Anal. Biochem., 150, 76-85.
- Smits, S.A., Leach, J., Sonnenburg, E.D., Gonzalez, C.G., Lichtman, J.S., Reid, G., Knight, R., Manjurano, A., Changalucha, J., Elias, J.E., Dominguez-Bello, M.G. and

Sonnenburg, J.L. (2017) Seasonal Cycling in the Gut Microbiome of the Hadza Hunter-Gatherers of Tanzania. *Science*, **357**, 802-806.

Smogyi, M. (1952) Notes on Sugar Determination. J Biol Chem, 195, 19-23.

- Sonnenburg, E.D., Sonnenburg, J.L., Manchester, J.K., Hansen, E.E., Chiang, H.C. and Gordon, J.I. (2006) A Hybrid Two-Component System Protein of a Prominent Human Gut Symbiont Couples Glycan Sensing *in vivo* to Carbohydrate Metabolism. *PNAS*, 103, 8834-8839.
- Sonnenburg, E.D., Zheng, H., Joglekar, P., Higginbottom, S.K., Firbank, S.J., Bolam, D.N. and Sonnenburg, J.L. (2010) Specificity of Polysaccharide use in Intestinal Bacteroides Species Determines Diet-Induced Microbiota Alterations. *Cell*, 141, 1241-1252.
- Spadiut, O., Ibatullin, F.M., Peart, J., Gullfot, F., Martinez-Fleites, C., Ruda, M., Xu, C., Sundqvist, G., Davies, G.J. and Brumer, H. (2011) Building Custom Polysaccharides in Vitro with an Efficient, Broad-Specificity Xyloglucan Glycosynthase and a Fucosyltransferase. J Am Chem Soc, 133, 10892-10900.
- Speciale, G., Jin, Y., Davies, G.J., Williams, S.J. and Goddard-Borger, E.D. (2016) YihQ is a Sulfoquinovosidase that Cleaves Sulfoquinovosyl Diacylglyceride Sulfolipids. *Nat Chem Biol*, 12, 215-217.
- Stahl, E. and Kaltenbach, U. (1961) Dünnschicht-Chromatographie: VI. Mitteilung. Spurenanalyse von Zuckergemischen auf Kieselgur Geschichten. *Journal of Chromatography A*, 5, 351-355.
- Stam, M.R., Danchin, E.G.J., Rancurel, C., Coutinho, P.M. and Henrissat, B. (2006) Dividing the Large Glycoside Hydrolase Family 13 into Subfamilies: Towards Improved Functional Annotations of α-Amylase-Related Proteins. *Protein Engineering, Design and Selection*, **19**, 555-562.
- Stein, N. (2008) CHAINSAW: a Program for Mutating PDB Files Used as Templates in Molecular Replacement. Journal of Applied Crystallography, 41, 641-643.
- Street, I.P., Rupitz, K. and Withers, S.G. (1989) Fluorinated and Deoxygenated Substrates as Probes of Transition State Structure in Glycogen Phosphorylase. *Biochemistry*, 28, 1581-1587.
- Studier, F.W. (2005) Protein Production by Auto-Induction in High Density Shaking Cultures. Protein Expression and Purification, 41, 207-234.

- Studier, F.W. and Moffatt, B.A. (1986) Use of Bacteriophage T7 RNA Polymerase to Direct Selective High-Level Expression of Cloned Genes. *Journal of Molecular Biology*, 189, 113-130.
- Sulzenbacher, G., Driguez, H., Henrissat, B., Schülein, M. and Davies, G.J. (1996) Structure of the *Fusarium oxysporum* Endoglucanase I with a Nonhydrolyzable Substrate Analogue: Substrate Distortion Gives Rise to the Preferred Axial Orientation for the Leaving Group. *Biochemistry*, 35, 15280-15287.
- Sundqvist, G., Stenvall, M., Berglund, H., Ottosson, J. and Brumer, H. (2007) A General, Robust Method for the Quality Control of Intact Proteins using LC–ESI-MS. *Journal of Chromatography B*, 852, 188-194.
- Takeda, T., Furuta, Y., Awano, T., Mizuno, K., Mitsuishi, Y. and Hayashi, T. (2002) Suppression and Acceleration of Cell Elongation by Integration of Xyloglucans in Pea Stem Segments. *Proc Natl Acad Sci U S A*, **99**, 9055-9060.
- Tamura, K., Hemsworth, G.R., Déjean, G., Rogers, T.E., Pudlo, N.A., Urs, K., Jain, N., Davies, G.J., Martens, E.C. and Brumer, H. (2017) Molecular Mechanism by which Prominent Human Gut Bacteroidetes Utilize Mixed-Linkage Beta-Glucans, Major Health-Promoting Cereal Polysaccharides. *Cell Reports*, 21, 417-430.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol, 30, 2725-2729.
- Tauzin, A.S., Kwiatkowski, K.J., Orlovsky, N.I., Smith, C.J., Creagh, A.L., Haynes, C.A., Wawrzak, Z., Brumer, H. and Koropatkin, N.M. (2016) Molecular Dissection of Xyloglucan Recognition in a Prominent Human Gut Symbiont. *mBio*, 7, e02134-02115.
- Teather, R.M. and Wood, P.J. (1982) Use of Congo Red-Polysaccharide Interactions in Enumeration and Characterisation of Cellulolytic Bacteria from the Bovine Rumen. *Appl Environ Microbiol*, 43, 777-780.
- Tenkanen, M., Makkonen, M., Perttula, M., Viikari, L. and Teleman, A. (1997) Action of *Trichoderma reesei* Mannanase on Galactoglucomannan in Pine Kraft Pulp. J. *Biotechnol.*, 57, 191-204.
- Tenkanen, M., Puls, J., Rättö, M. and Viikari, L. (1993) Enzymatic Deacetylation of Galactoglucomannans. *Appl Microbiol Biotechnol*, **39**, 159-165.

- Terrapon, N., Lombard, V., Drula, É., Lapébie, P., Al-Masaudi, S., Gilbert, H.J. and Henrissat, B. (2017) PULDB: the Expanded Database of Polysaccharide Utilisation Loci. Nucl. Acids Res.
- Tropea, J.E., Cherry, S. and Waugh, D.S. (2009) Expression and Purification of Soluble His6 Tagged *TEV* Protease. In *High Throughput Protein Expression and Purification* (Doyle,
 S.A. ed: Humana Press, pp. 297-307.
- Tseng, C.-W., Ko, T.-P., Guo, R.-T., Huang, J.-W., Wang, H.-C., Huang, C.-H., Cheng, Y.-S., Wang, A.H.J. and Liu, J.-R. (2011) Substrate Binding of a GH5 from the Ruminal Fungus *Piromyces rhizinflata*. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, 67, 1189-1194.
- Tsukagoshi, H., Nakamura, A., Ishida, T., Touhara, K.K., Otagiri, M., Moriya, S., Samejima, M., Igarashi, K., Fushinobu, S., Kitamoto, K. and Arioka, M. (2014) Structural and Biochemical Analyses of Glycoside Hydrolase Family 26 β-Mannanase from a Symbiotic Protist of the Termite *Reticulitermes speratus*. *Journal of Biological Chemistry*, **289**, 10843-10852.
- Tull, D., Burgoyne, D.L., Chow, D.T., Withers, S.G. and Aebersold, R. (1996) A Mass Spectrometry-Based Approach for Probing Enzyme Active Sites: Identification of Glu 127 in *Cellulomonas fimi* Exoglycanase as the Residue Modified by N-Bromoacetyl Cellobiosylamine. *Anal. Biochem.*, 234, 119-125.
- Tuomivaara, S.T., Yaoi, K., O'Neill, M.A. and York, W.S. (2015) Generation and Structural Validation of a Library of Diverse Xyloglucan-Derived Oligosaccharides, Including an Update on Xyloglucan Nomenclature. *Carbohydr. Res.*, 402, 56-66.
- Tyler, L., Bragg, J.N., Wu, J., Yang, X., Tuskan, G.A. and Vogel, J.P. (2010) Annotation and Comparative Analysis of the Glycoside Hydrolase Genes in *Brachypodium distachyon*. *BMC Genomics*, **11**, 600.
- Vagin, A.A., Steiner, R.A., Lebedev, A.A., Potterton, L., McNicholas, S., Long, F. and Murshudov, G.N. (2004) *REFMAC5* Dictionary: Organisation of Prior Chemical Knowledge and Guidelines for its Use. *Acta Crystallographica Section D Biological Crystallography*, **60**, 2184-2195.
- van den Brink, J. and de Vries, R.P. (2011) Fungal Enzyme Sets for Plant Polysaccharide Degradation. *Appl Microbiol Biotechnol*, **91**, 1477-1492.

- Varghese, J.N., Garrett, T.P., Colman, P.M., Chen, L., Høj, P.B. and Fincher, G.B. (1994) Three-Dimensional Structures of Two Plant β-Glucan Endohydrolases with Distinct Substrate Specificities. *Proc Natl Acad Sci U S A*, **91**, 2785-2789.
- Varki, A., Cummings, R.D., Aebi, M., Packer, N.H., Seeberger, P.H., Esko, J.D., Stanley, P., Hart, G., Darvill, A., Kinoshita, T., Prestegard, J.J., Schnaar, R.L., Freeze, H.H., Marth, J.D., Bertozzi, C.R., Etzler, M.E., Frank, M., Vliegenthart, J.F., Lütteke, T., Perez, S., Bolton, E., Rudd, P., Paulson, J., Kanehisa, M., Toukach, P., Aoki-Kinoshita, K.F., Dell, A., Narimatsu, H., York, W., Taniguchi, N. and Kornfeld, S. (2015) Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology*, 25, 1323-1324.
- Varrot, A., Frandsen, T.P., Ossowski, I.v., Boyer, V., Cottaz, S., Driguez, H., Schülein, M. and Davies, G.J. (2003) Structural Basis for Ligand Binding and Processivity in Cellobiohydrolase Cel6A from Humicola insolens. *Structure*, **11**, 855-864.
- Vasur, J., Kawai, R., Andersson, E., Igarashi, K., Sandgren, M., Samejima, M. and Ståhlberg, J. (2009) X-ray Crystal Structures of *Phanerochaete chrysosporium* Laminarinase 16A in Complex with Products from Lichenin and Laminarin Hydrolysis. *FEBS Journal*, 276, 3858-3869.
- Venditto, I., Najmudin, S., Luís, A.S., Ferreira, L.M.A., Sakka, K., Knox, J.P., Gilbert, H.J. and Fontes, C.M.G.A. (2015) Family 46 Carbohydrate-Binding Modules Contribute to the Enzymatic Hydrolysis of Xyloglucan and β-1,3–1,4-Glucans through Distinct Mechanisms. *Journal of Biological Chemistry*, **290**, 10572-10586.
- Vikramathithan, J., Kumar, G.N., Muthuraman, P. and Srikumar, K. (2010) Purification and Characterisation of Thermophilic Xylanase Isolated from the Xerophytic *Cereus pterogonus sp. Protein J.*, 29, 481-486.
- Vincent, F., Charnock, S.J., Verschueren, K.H.G., Turkenburg, J.P., Scott, D.J., Offen,
 W.A., Roberts, S., Pell, G., Gilbert, H.J., Davies, G.J. and Brannigan, J.A. (2003)
 Multifunctional Xylooligosaccharide/Cephalosporin C Deacetylase Revealed by the
 Hexameric Structure of the *Bacillus subtilis* Enzyme at 1.9A Resolution. *Journal of Molecular Biology*, 330, 593-606.
- Vincken, J.P., York, W.S., Beldman, G. and Voragen, A.G. (1997) Two General Branching Patterns of Xyloglucan, XXXG and XXGG. *Plant Physiology*, **114**, 9-13.

- Vinueza, N.R., Gallardo, V.A., Klimek, J.F., Carpita, N.C. and Kenttämaa, H.I. (2013) Analysis of Xxyloglucans by Ambient Chloride Attachment Ionisation Tandem Mass Spectrometry. *Carbohydrate Polymers*, **98**, 1203-1213.
- Wakarchuk, W.W., Kilburn, D.G., Jr, R.C.M. and Warren, R.A.J. (1986) The Molecular Cloning and Expression of a Cellobiase Gene from an *Agrobacterium* in *Escherichia coli*. *Molec. Gen. Genet.*, **205**, 146-152.
- Walseth, C.S. (1952) The Influence of the Fine Structure of Cellulose on the Action of Cellulases. *TAPPI*, 35, 233-238.
- Wang, H.-L., Yeh, K.-W., Chen, P.-R., Chang, C.-h., Chen, J.-M. and Khoo, K.-H. (2006)
 Isolation and Characterisation of a Pure Mannan from *Oncidium* (cv. *Gower Ramsey*)
 Current Pseudobulb during Initial Inflorescence Development. *Bioscience, Biotechnology, and Biochemistry*, **70**, 551-553.
- Wang, T.-Y., Chen, H.-L., Lu, M.-Y.J., Chen, Y.-C., Sung, H.-M., Mao, C.-T., Cho, H.-Y., Ke, H.-M., Hwa, T.-Y., Ruan, S.-K., Hung, K.-Y., Chen, C.-K., Li, J.-Y., Wu, Y.-C., Chen, Y.-H., Chou, S.-P., Tsai, Y.-W., Chu, T.-C., Shih, C.-C.A., Li, W.-H. and Shih, M.-C. (2011) Functional Characterisation of Cellulases Identified from the Cow Rumen Fungus *Neocallimastix patriciarum* W5 by Transcriptomic and Secretomic Analyses. *Biotechnology for Biofuels*, 4, 24.
- Waugh, D.S. (2011) An Overview of Enzymatic Reagents for the Removal of Affinity Tags. Protein expression and purification, 80, 283-293.
- Wheatley, R.W., Lo, S., Jancewicz, L.J., Dugdale, M.L. and Huber, R.E. (2013) Structural Explanation for Allolactose (*lac* Operon Inducer) Synthesis by *lacZ* β-Galactosidase and the Evolutionary Relationship between Allolactose Synthesis and the lac Repressor. *J Biol Chem*, 288, 12993-13005.
- Wilkinson, A.J., Fersht, A.R., Blow, D.M. and Winter, G. (1983) Site-Directed Mutagenesis as a Probe of Enzyme Structure and Catalysis: Tyrosyl-tRNA Synthetase Cysteine-35 to Glycine-35 Mutation. *Biochemistry*, 22, 3581-3586.
- Willför, S., Sundberg, K., Tenkanen, M. and Holmbom, B. (2008) Spruce-Derived Mannans
 A Potential Raw Material for Hydrocolloids and Novel Advanced Natural Materials.
 Carbohydrate Polymers, 72, 197-210.

- Wilson, S.M., Ho, Y.Y., Lampugnani, E.R., Meene, A.M.L.V.d., Bain, M.P., Bacic, A. and Doblin, M.S. (2015) Determining the Subcellular Location of Synthesis and Assembly of the Cell Wall Polysaccharide (1,3; 1,4)-β-D-Glucan in Grasses. *Plant Cell*, 27, 754-771.
- Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan,
 R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., McNicholas, S.J., Murshudov,
 G.N., Pannu, N.S., Potterton, E.A., Powell, H.R., Read, R.J., Vagin, A. and Wilson,
 K.S. (2011) Overview of the CCP4 suite and Current Developments. *Acta Crystallographica Section D Biological Crystallography*, 67, 235-242.
- Withers, S.G., Rupitz, K., Trimbur, D. and Warren, R.A.J. (1992) Mechanistic Consequences of Mutation of the Active Site Nucleophile Glu 358 in *Agrobacterium* β-Glucosidase. *Biochemistry*, **31**, 9979-9985.
- Wolfenden, R., Lu, X. and Young, G. (1998) Spontaneous Hydrolysis of Glycosides. J Am Chem Soc, 120, 6814-6815.
- Wong, K.K.Y. and Saddler, J.N. (1992) *Trichoderma* Xylanases, Their Properties and Application. *Critical Reviews in Biotechnology*, **12**, 413-435.
- Wood, P.J., Weisz, J. and Blackwell, B.A. (1994) Structural Studies of (1-->3),(1-->4)-β-D-Glucans by ¹³C-Nuclear Magnetic Resonance Spectroscopy and by Rapid Analysis of Cellulose-Like Regions Using High-Performance Anion-Exchange Chromatography of Oligosaccharides Released by Lichenase. *Cereal chemistry*, **71**, 301-307.
- Woodward, J.R. and Fincher, G.B. (1982) Substrate Specificities and Kinetic Properties of Two (1 \rightarrow 3), (1 \rightarrow 4)- β -D-Glucan *Endo*-Hydrolases from Germinating Barley (*Hordeum vulgare*). *Carbohydr. Res.*, **106**, 111-122.
- Woodward, J.R., Fincher, G.B. and Stone, B.A. (1983) Water-Soluble $(1\rightarrow 3)$, $(1\rightarrow 4)$ - β -D-Glucans from Barley (*Hordeum vulgare*) Endosperm. II. Fine structure. *Carbohydrate Polymers*, **3**, 207-225.
- Wu, G.D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.-Y., Keilbaugh, S.A., Bewtra, M., Knights, D., Walters, W.A., Knight, R., Sinha, R., Gilroy, E., Gupta, K., Baldassano, R., Nessel, L., Li, H., Bushman, F.D. and Lewis, J.D. (2011) Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes. *Science*, 334, 105-108.
- **Xue, X. and Fry, S.C.** (2012) Evolution of Mixed-Linkage $(1 \rightarrow 3, 1 \rightarrow 4)$ - β -D-Glucan (MLG) and Xyloglucan in *Equisetum* (horsetails) and other Monilophytes. *Ann Bot*, mcs018.

- Yalpani, M. and Hall, L.D. (1982) Some Chemical and Analytical Aspects of Polysaccharide Modifications. II. A High-Yielding, Specific Method for the Chemical Derivatisation of Galactose-Containing Polysaccharides: Oxidation with Galactose Oxidase Followed by Reductive Amination. J. Polym. Sci. Polym. Chem. Ed., 20, 3399-3420.
- Yan, X.-X., An, X.-M., Gui, L.-L. and Liang, D.-C. (2008) From Structure to Function: Insights into the Catalytic Substrate Specificity and Thermostability Displayed by *Bacillus subtilis* Mannanase BCman. *Journal of Molecular Biology*, 379, 535-544.
- Yang, G., Yang, G., Aprile, L., Turturo, V., Pucciarelli, S., Pucciarelli, S. and Miceli, C.
 (2013) Characterisation and Comparative Analysis of Psychrophilic and Mesophilic α Amylases from *Euplotes* Species: A Contribution to the Understanding of Enzyme
 Thermal Adaptation. *Biochemical and Biophysical Research Communications*, 438, 715 720.
- Ye, X., Yuan, S., Guo, H., Chen, F., Tuskan, G.A. and Cheng, Z.-M. (2012) Evolution and Divergence in the Coding and Promoter Regions of the *Populus* Gene Family Encoding Xyloglucan Endotransglycosylase/Hydrolases. *Tree Genetics & Genomes*, 8, 177-194.
- Yip, V.L.Y. and Withers, S.G. (2006) Breakdown of oligosaccharides by the process of elimination. *Current Opinion in Chemical Biology*, **10**, 147-155.
- Yokoyama, R., Uwagaki, Y., Sasaki, H., Harada, T., Hiwatashi, Y., Hasebe, M. and Nishitani, K. (2010) Biological Implications of the Occurrence of 32 Members of the *XTH (xyloglucan endotransglucosylase/hydrolase)* Family of Proteins in the Bryophyte *Physcomitrella patens. The Plant Journal*, 64, 645-656.
- York, W.S., Harvey, L.K., Guillen, R., Albersheim, P. and Darvill, A.G. (1993) Structural Analysis of Tamarind Seed Xyloglucan Oligosaccharides using β-Galactosidase Digestion and Spectroscopic Methods. *Carbohydr. Res.*, 248, 285-301.
- York, W.S., van Halbeek, H., Darvill, A.G. and Albersheim, P. (1990) Structural Analysis of Xyloglucan Oligosaccharides by ¹H-N.M.R. Spectroscopy and Fast-Atom-Bombardment Mass Spectrometry. *Carbohydr. Res.*, 200, 9-31.
- Yoshida, S., Park, D.S., Bae, B., Mackie, R., Cann, I.K.O. and Nair, S.K. (2011) Structural and Functional Analyses of a Glycoside Hydrolase Family 5 Enzyme with an Unexpected β -Fucosidase Activity. *Biochemistry*, **50**, 3369-3375.

- Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S.C., Ester, M., Foster, L.J. and Brinkman, F.S.L. (2010) *PSORTb 3.0*: Improved Protein Subcellular Localisation Prediction with Refined Localisation Subcategories and Predictive Capabilities for all Prokaryotes. *Bioinformatics*, 26, 1608-1615.
- **Yuan, J., Martinez-Bilbao, M. and Huber, R.E.** (1994) Substitutions for Glu-537 of β-Galactosidase from *Escherichia coli* Cause Large Decreases in Catalytic Activity.
- Yuan, Y., Teng, Q., Zhong, R. and Ye, Z.-H. (2016) Roles of Arabidopsis TBL34 and TBL35 in Xylan Acetylation and Plant Growth. *Plant Science*, 243, 120-130.
- Zhang, J.X., Martin, J. and Flint, H.J. (1994) Identification of Non-Catalytic Conserved Regions in Xylanases Encoded by the *xynB* and *xynD* Genes of the Cellulolytic Rumen Anaerobe *Ruminococcus flavefaciens*. *Mol. Gen. Genet.*, **245**, 260-264.
- Zhang, Y.H.P., Cui, J., Lynd, L.R. and Kuang, L.R. (2006) A Transition from Cellulose Swelling to Cellulose Dissolution by o-Phosphoric Acid: Evidence from Enzymatic Hydrolysis and Supramolecular Structure. *Biomacromolecules*, 7, 644-648.
- Zheng, Y., Anderson, S., Zhang, Y. and Garavito, R.M. (2011) The Structure of Sucrose Synthase-1 from *Arabidopsis thaliana* and Its Functional Implications. *J Biol Chem*, 286, 36108-36118.
- Zhou, M., Diwu, Z., Panchuk-Voloshina, N. and Haugland, R.P. (1997) A Stable Nonfluorescent Derivative of Resorufin for the Fluorometric Determination of Trace Hydrogen Peroxide: Applications in Detecting the Activity of Phagocyte NADPH Oxidase and Other Oxidases. *Anal. Biochem.*, 253, 162-168.
- Zhou, Q., Rutland, M.W., Teeri, T.T. and Brumer, H. (2007) Xyloglucan in Cellulose Modification. *Cellulose*, 14, 625-641.
- Zverlov, V.V. and Velikodvorskaya, G.A. (1990) Cloning the *Clostridium thermocellum* Thermostable Laminarinase Gene in *Escherichia coli*: The Properties of the Enzyme thus Produced. *Biotechnol Lett*, **12**, 811-816.

Appendices

Appendix A Supporting Information for Chapter 2: Crystallographic Insight into the Evolutionary Origins of Xyloglucan *Endo*-Transglycosylases and *Endo*-Hydrolases



Figure A-1: Protein sequence alignment of EG16s with Licheninases and *XTH* gene products. EG16s are named according to the first letters of their genus and species names. For each EG16, the common name (where available) for the plant of origin is given along with the accession code for the gene product. Regions of high similarity are shown in red. Secondary structure features, defined based on the crystal structure of VvEG16(ΔV152,E89A) in complex with cellooligosaccharides, are displayed and named above the alignment. Conserved cysteines among monocot or dicot EG16s are highlighted in yellow. The licheninases loop, XEH loop and catalytic strand are labelled below the sequences, stars mark the catalytic nucleophile, general acid/base and "helper" aspartic acid.



Figure A-2: Intact MS and SDS-PAGE analyses of purified VvEG16 and its mutants. A) Reconstructed intact mass spectra of VvEG16, VvEG16(Δ V152) and VvEG16(Δ V152,E89A). The purified proteins had masses of 23432.7 (expected 23432.1), 23330.0 Da (expected 23331.0), and 23272.6 Da (expected 23272.9) respectively, verifying that the protein sequence was correct, demonstrating that no post-translational modifications were present after TEV protease cleavage of the His6 affinity tag and suggesting that all cysteine residues were reduced. B) SDS-PAGE analysis of purified VvEG16, VvEG16(Δ V152) and VvEG16 (Δ V152,E89A) after TEV protease cleavage.



Figure A-3: Comparison of wild-type VvEG16 and VvEG16 (Δ V152) activities. A) Overlay of wild-type VvEG16 and VvEG16 (Δ V152) acting bMLG and tXyG. B) Kinetics of VvEG16(Δ V152) acting on various polysaccharide substrates. C) Comparison of wild-type VvEG16 and VvEG16(Δ V152) acting XXXG-CNP. D) Comparison of wild-type VvEG16 and VvEG16(Δ V152) acting GGG-CNP. Each data point represents the average of two replicates. Activity of VvEG16(Δ V152) on a variety of polysaccharides at a variety of concentrations. Kinetics were modelled to extract apparent k_{cat} and K_M values based on the Michaelis-Menten curves shown. Error bars represent the standard deviation of three technical replicates.



Figure A-4: Dependence of VvEG16(ΔV152) activity on pH and temperature. A) Activity of VvEG16(ΔV152) on 1 mg/mL tXyG in buffers at various pH values measured using the BCA assay. Error bars represent that standard deviation of three technical replicates. B) Activity of VvEG16(ΔV152) on 1 mg/mL tXyG in pH 6.0 sodium citrate buffer at various temperatures measured using the BCA assay. Each data point represents the average of two technical replicates. Reactions were incubated for 15 minutes with no reducing agent present prior to quenching by the addition of BCA reagent. C) Thermal stability of VvEG16(ΔV152) in dithiothreitol-containing buffer. Activity on 1 mM XXXG-CNP was measured at room temperature at regular intervals during incubation of the enzyme at various temperatures in pH 5.5 sodium citrate.



Figure A-5: Identification of oligosaccharide series produced by action of VvEG16(Δ V152) on bMLG. Oligosaccharides were separated by size using a SEC column and analysed by HPAEC-PAD (panel A) and MALDI-TOF (panel B) to determine the mass corresponding to each HPAEC-PAD peak. Panel C depicts a schematic of bMLG breakdown by VvEG16. bMLG polysaccharide (top) is initially cleaved by VvEG16 at β 1,4 bonds following cellotriose motifs. The resulting oligosaccharides (second from top) contain repeating β 1,3-linked cellotriose motifs terminated at the non-reducing end with primarily β 1,3-linked glucose (~90%)

and less commonly β1,3-linked cellobiose (~5%) or cellotriose motifs (~3%). This gives rise to the three nested series of peaks observed in panel A. Continued digestion of these oligosaccharides gives the three oligosaccharides shown (second from bottom). Finally, slow hydrolysis of the pentasaccharide and cellotriose give the final products observed in the limit-digest (figure 2C,D): glucose, cellobiose and G3GGG which cannot be further broken down by VvEG16.



Figure A-6: Identification of the MLGO produced by the action of VvEG16(Δ V152) on bMLG (Glc4-MLGO) by HPAEC-PAD. A) 10 μ M Glc4-MLGO was mixed 1:1 with water or 10 μ M standards of G3GGG, GG3GG and GGG3G and run with gradient B. Notably, the Glc4-MLGO peak co-eluted with G3GGG and grew when G3GGG was added. B) Further confirmation of the identity of Glc4-MLGO was obtained by partial digestion with an *exo*-acting β -glucosidase. Only G, GG and GGG were observed as products, thus confirming that the

 $\beta(1,3)$ linkage was found between the two non-reducing-terminal glucose residues.



Figure A-7: Identification of the MLGO produced by the action of VvEG16(ΔV152) on bMLG (Glc4-MLGO) by LC-MS. A-C) Extracted ion chromatogram (m/z = 684.0-684.5), MS and CID MSMS spectra of The Glc4-MLGO peak eluting at 17.2 minutes, respectively. D-F) Extracted ion chromatogram (m/z = 684.0-684.5), MS and CID MS/MS spectra of commercial G3GGG, respectively.



Figure A-8: Time-dependent MLG hydrolysis by VvEG16(ΔV152). A) HPAEC-PAD chromatogram of the early products of MLG hydrolysis by VvEG16. B) HPAEC-PAD chromatogram of the licheninase digestion of the MLG shown in A). Licheninase-digested native MLG is shown in black and licheninase-digested
VvEG16-digested MLG is shown in red. C) HPAEC-PAD chromatogram of the gel-forming MLG produced by VvEG16 digestion. D) HPAEC-PAD chromatogram of the licheninase digestion of the MLG showed in C). Licheninase-digested native MLG is shown in black and licheninase-digested VvEG16-digested MLG is shown in black



Figure A-9: Kinetic data and Michaelis-Menten curves used to determine the k_{cat} and K_M values for the hydrolysis of various chromogenic substrates by VvEG16(Δ V152). Error bars are shown as the standard deviation of three technical replicates.



Figure A-10: Kinetic data and Michaelis-Menten curves used to determine the k_{cat} and K_M values for the hydrolysis of various oligosaccharide substrates by VvEG16(Δ V152). Error bars are shown as the standard deviation of three technical replicates.



Figure A-11: Mass spectra of products produced by the hydrolysis of cello-oligosaccharides by VvEG16(ΔV152) in H₂¹⁸O. A) cellotriose, B) cellotetraose, C) cellopentaose, D) cellohexaose. The degree of labelling calculated from peak integrations can be found in table A.2-2.



Figure A-12: Homo- and hetero-transglycosylation of oligosaccharides by VvEG16(Δ V152). A) The HPAEC-PAD chromatogram of the action of VvEG16(Δ V152) on 1 mM cellotetraose over time. The increase of the

peak at 9.6 minutes is indicative of the formation of cellohexaose. B) The HPAEC-PAD chromatogram of VvEG16(Δ V152) acting on 100 μ M XXXGXXXG in the presence of 5 mM cellobiose over time. A small peak is observed forming at a retention time of 9.3 minutes and, later, being hydrolysed. Though not isolated, this peak is putatively assigned as XXXGGG.



Figure A-13: Structure of VvEG16(ΔV152,E89A) in complex with a cellooligosaccharide. A) The asymmetric unit of this complex contains one protein molecule, shown as both a white surface representation and a cartoon representation coloured according to B-factors. The protein molecule bound to a molecule modelled as GGGG in the negative subsites and glucose in the positive subsites (both shown in green). B) and C) show the interactions between VvEG16(ΔV152,E89A) and cellooligosaccharides as in figure 2-5.



Figure A-14: Structure of VvEG16(ΔV152,C22S,C188S) in complex with a xyloglucan oligosaccharide. A) The asymmetric unit of this complex contains one protein molecule, shown as both a white surface representation and a cartoon representation coloured according to B-factors. The protein molecule bound to a molecule modelled as XGXG in the negative subsites (shown in green). The location of the C22S mutation is noted with an arrow; the C188S mutation is located on the back of the enzyme from this perspective. B) and

C) show the interactions between VvEG16(ΔV152, C22S,C188S) and XGXG as in figure 2-5.



Figure A-15: Superimposition of the carbohydrate ligands bound within the active site of each experimentally determined structure of VvEG16. Carbohydrates from all six chains (PDB IDs 5DZE (Magenta), 5DZF-A (Green), 5DZF-B (Cyan), 5DZG-A (Yellow), 5DZF-B (Salmon), and 5SV8 (White)) are shown.

Activity	GenBank Accession Codes
GH7 Cellulase	AAX28897.1, AAA65586.1, BAA09786.1, AEO58196.1, ABY56790.1
XTH	ABL75361.1, AAC09388.1, ACS13756.1, CAA48324.1, AAD31572.1
EG16	XM_002273939.3, XM_002301283.2, NM_001156874.1, AK356496.1,
	XM_002970606.1
Licheninase	CAA39426.1, AAB05759.1, CAA40547.1, AFX61538.1, AAA24896.1
Laminarinase	AAC38290.1, AAC44371.1, BAE54302.1, AAD35118.1, CAZ95067.1
Chitin	AAB65002.1, EAL03096.1, EAL02960.1, EAL01616.1, CAA67525.1
Transglycosylase	
Agarase	BAC99022.1, BAD29947.1, EDY95404.1, AAF21821.1, CAZ98378.1
Porphyranase	CBM41184.1, CBM41183.1, CBM41185.1, CBM41182.1, EDY95423.1
Kappa-	CAA50624.1, AGS43006.1, CAZ94309.1, BAJ61957.1, ADD92366.1
Carrageenase	

Table A-1: Sequences used for generation of the GH16 phylogenetic tree.

Table A-2: Primer sequences used for cloning and mutagenesis in Chapter 2.

Target	Primer Name	Primer Sequence		
VvEG16(ΔV152)	Vv1-F	5'TCGAAAACCTGTACTTCCAAGGCATGGCG		
	Vv1-R	GATCCGAGCTTG3'		
		5'CAGTGGTGGTGGTGGTGGTGGTGGTGGTCAC		
		TCCACTGCGGTACCAAC3'		
VvEG16(ΔV152,	Vv2-F	5'AGAAATTCGAAATCGATTGCGTCCTGGCTT		
E89A) and	Vv2-R	TTGTCCC3'		
VvEG16(E89A)		5'GGGACAAAAGCCAGGACGCAATCGATTTC		
		GAATTTCT3'		
VvEG16((ΔV152,	Vv3-F	5'GCGGTCGATTACTCTCCCGAATCTTGC 3'		
C22S)	Vv3-R	5'GCAAGATTCGGGGAGAGTAATCGACCGC 3'		
VvEG16((ΔV152,	Vv4-F	5'GGTCCTTACGTTGGTTCCGATGCGCCGTAT		
C22S,C188S)	Vv4-R	ATC 3'		
		5'GATATACGGCGCATCGGAACCAACGTAAG		
		GACC 3'		
pET24a(+)	pET24-F	5'GCCTTGGAAGTACAGGTTTTC3'		
_	pET24-R	5'TGACTCGAGCACCACCAC3'		
P15TvLIC	p15Tv-F	5'CAAAGCCCGAAAGGAAGC3'		
	p15Tv-R	5'GCCCTGGAAATACAAGTTTTC3'		
PpXG5-pET21a	PpXG-F	5'GTTATCCGGTCGTGATTGGCGAATTTGGAT		
(G323E)	PpXG-R	CGATTGATAAAAC3'		
		5'GTTTTATCAATCGATCCAAATTCGCCAATC		
		ACGACCGGATAAC3'		

	VvEG16(ΔV152,E89A):	VvEG16(ΔV152,E89A):	VvEG16(ΔV152,E89A):	VvEG16(ΔV152,C22S,			
	GGGG complex	bMLGO complex	tXyGO complex	C188S): tXyGO complex			
PDB Code	5DZE	5DZF	5DZG	5SV8			
DATA COLLECTION							
Wavelength (Å)	0.9795	0.8266	1.0332	1.0332			
Space group	$P2_{1}2_{1}2_{1}$	P4 ₃ 2 ₁ 2	$P2_12_12_1$	$P2_12_12_1$			
Cell dimensions							
a, b, c, (Å)	45.15, 52.02, 83.06	74.33, 74.33, 149.28	41.3, 79.0, 132.6	46.2, 51.9, 82.6			
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90			
Resolution, Å	44.09-0.97 (1.00-0.97) ^b	42.97-1.55 (1.69-1.65)	39.41-1.79 (1.83-1.79)	27.54-1.588 (1.62-1.59)			
$\mathbf{R}_{merge}{}^{a}$	0.065 (0.901)	0.20 (1.25)	0.097 (0.317)	0.19 (0.69)			
$I / \sigma(I)$	16.34 (1.25)	10.85 (1.43)	12.2 (3.6)	12.2 (2.8)			
Completeness, %	95.6 (69.0)	100 (97.0)	98.2 (78.3)	99.5 (91.2)			
Redundancy	6	12.1	5.9	6.2			
		REFINEMENT					
No. of reflections:							
working, test	106175, 5588	48580, 2558	38822, 2056	27417, 1382			
<i>R</i> -factor/free <i>R</i> -factor ^c	0.136/0.150 (0.350/0.366)	0.155/0.200	0.171/0.217 (0.225/0.285)	0.152/0.179			
		(0.304/0.376)		(0.220/0.259)			
No. of refined atoms							
Protein	3361	3208	3092	3113			
Ligand	114	146	215	63			
Solvent	284	463	401	228			
<i>B</i> -factors							
Protein	9.8	20.3	20.9	10.6			
Ligand	13.0	21.0	28	18.4			
Solvent	23.0	35.2	32.4	21.9			
RMSD							
Bond lengths, Å	0.009	0.008	0.011	0.010			
Bond angles, °	1.53	1.25	1.50	1.39			
Ramachandran Plot:							
Favoured, %	96.6	97.3	97.5	98.0			
Allowed, %	2.3	2.7	2.5	2.0			
Disallowed, %	1.1	0	0	0			

Table A-3: X-ray diffraction data statistics for VvEG16 crystal complexes.

 ${}^{a}R_{\text{merge}} = \sum_{h}\sum_{i}|I_{i}(h) - \langle I(h)\rangle/\sum_{h}\sum_{i}I_{i}(h)$, where $I_{i}(h)$ and $\langle I(h)\rangle$ are the *i*th and mean measurement of the intensity of reflection *h*. ^bFigures in parentheses throughout the table indicate the values for the outer shells of the data. ${}^{c}R = \Sigma |F_{p}^{obs} - F_{p}^{calc}| / \Sigma F_{p}^{obs}$, where F_{p}^{obs} and F_{p}^{calc} are the observed and calculated structure factor amplitudes, respectively.
Construct	Production result
SmEG16	Strong production into inclusion bodies
PpEG16	Strong production into inclusion bodies
PpEG16-sfGFP	Weak production into inclusion bodies
PpEG16-SUMO	Strong production into inclusion bodies
BdEG16	Strong production into inclusion bodies
	Soluble-aggregate production, precipitation after removal of sfGFP
BdEG16-sfGFP	tag
	Soluble-aggregate production, precipitation after removal of SUMO
BdEG16-SUMO	tag
VvEG16	Strong soluble production
VvEG16-sfGFP	Slow soluble production
VvEG16-SUMO	Strong soluble production
VvEG16(ΔV152)	Strong soluble production
VvEG16(Δ V152)-sfGFP	Weak soluble production
VvEG16(ΔV152)-SUMO	Strong soluble production
VvEG16(ΔV152,E89A)	Strong soluble production
VvEG16(E89A)	Strong soluble production

Table A-4: EG16 expression constructs screened for protein production

Substrate	% ¹⁸ O	
Product	Labelled	Error
Cellotriose		
Glc	0.4	1.4
GG	35.2	2.8
Cellotetraose		
Glc	3.2	2.8
GG	51.2	0.9
GGG	89.4	1.6
Cellopentaose		
GG	17.6	2.6
GGG	79.0	0.8
Cellohexaose		
GG	15.6	2.6
GGG	50.7	0.8
GGGG	80.5	1.1

Residue Name	Conformation	Average B-factor	RSCC ¹
VvF	EG16(ΔV152,E89A): GGGG complex (5D)	ZE)
GLC-1	${}^{4}C_{1}$	8.4	0.94
BGC-1	${}^{4}C_{1}$	6.5	0.96
BGC-2	${}^{4}C_{1}$	8.4	0.95
BGC-3	${}^{4}C_{1}$	43.0	0.71
BGC-4	${}^{4}C_{1}$	21.7	0.66
VvE	G16(AV152,E89A)	: bMLGO complex (5D	(ZF)
GLC-1	${}^{4}C_{1}$	16.4	0.95
GLC-2	${}^{4}C_{1}$	16.8	0.93
BGC-1	${}^{4}C_{1}$	13.8	0.96
BGC-2	${}^{4}C_{1}$	16.9	0.94
BGC-3	${}^{4}C_{1}$	24.5	0.92
BGC-4	${}^{4}C_{1}$	24.8	0.92
BGC-5	${}^{4}C_{1}$	25.8	0.90
BGC-6	${}^{4}C_{1}$	21.7	0.95
BGC-7	${}^{4}C_{1}$	16.6	0.95
BGC-8	${}^{4}C_{1}$	15.0	0.94
BGC-9	${}^{4}C_{1}$	19.4	0.94
BGC-10	${}^{4}C_{1}$	32.7	0.87
BGC-11	${}^{4}C_{1}$	28.6	0.77
VvE	CG16(ΔV152,E89A): tXyGO complex (5D2	ZG)
GLC-1	${}^{4}C_{1}$	15.4	0.90
GLC-2	${}^{4}C_{1}$	19.9	0.88
BGC-1	${}^{4}C_{1}$	17.3	0.94
BGC-2	${}^{4}C_{1}$	19.7	0.94
BGC-3	${}^{4}C_{1}$	27.3	0.90
BGC-4	${}^{4}C_{1}$	32.8	0.79
BGC-5	${}^{4}C_{1}$	28.5	0.87
BGC-6	${}^{4}C_{1}$	23.3	0.90
BGC-7	${}^{4}C_{1}$	18.7	0.87
BGC-8	${}^{4}C_{1}$	15.7	0.92
BGC-9	${}^{4}C_{1}$	27.5	0.88
BGC-10	${}^{4}C_{1}$	19.9	0.93
XYS-1	${}^{4}C_{1}$	33.3	0.83
XYS-2	${}^{4}C_{1}$	26.4	0.83
XYS-3	^{2}E	44.3	0.71
XYS-4	${}^{4}C_{1}$	42.1	0.80
XYS-5	$^{0}S_{2}$	38.1	0.79
XYS-6	^{2}E	37.1	0.80
XYS-7	${}^{4}C_{1}$	26.5	0.86
XYS-8	$^{2}S_{0}$	34.7	0.78
XYS-9	${}^{4}C_{1}$	35.5	0.83

Table A-6: Privateer validation results for VvEG16 crystal complexes.

Residue Name	Conformation	Average B-factor	RSCC ¹
VvEG1	l6(ΔV152,C22S,C1	88S): tXyGO complex	(5SV8)
BGC-1	${}^{4}C_{1}$	7.9	0.93
BGC-2	${}^{4}C_{1}$	9.1	0.89
BGC-3	${}^{4}C_{1}$	13.9	0.87
BGC-4	${}^{4}C_{1}$	19.7	0.84
XYS-1	${}^{4}C_{1}$	33.6	0.66
XYS-2	${}^{4}C_{1}$	31.7	0.75

³RSCC, short for Real Space Correlation Coefficient, measures the agreement between model and positive omit density. A RSCC below 0.8 is typically considered poor.

Appendix B Supporting Information for Chapter 4: Structural and Biochemical Analyses of β-Mannan Recognition Machinery from *Prevotella bryantii* B₁4



Figure B-1: SDS-PAGE following inM pull-down. Following incubations of SGBP-A, SGBP-B, and SGBP-C at 0.5 mg/mL with 50 mg/mL inM, the supernatant ("Free") and inM-extracted ("Bnd") fractions were separated using a 4-20% denaturing "stain-free" SDS-PAGE gel (Bio-Rad) and visualized by fluorescence imaging following UV-exposure to activate the stain. Molecular weight indicators were imaged using white light and are annotated according to the BluElf protein ladder manual (FroggaBio). Though run on the same gel, the central panel represents a shorter camera exposure relative to the left panel. The arrow marks an unknown contaminant band pair observed in the SGBP-A pulldown samples which is not the protein of interest.



Figure B-2: ITC of PbSGBP-B and PbSGBP-C with oligosaccharides. Ligands used were A,B) glucomannan oligosaccharides produced by the action of PbGH26A-GH5A on kGM. C,D) galactomannan oligosaccharides produced by the action of PbGH26A-GH5A on cGM, and E,F) mannohexaose.



Figure B-3: Alignment of BoMan26A with PbGH26B. The positions of the catalytic glutamic acid residues from BoMan26A are indicated with arrows. This figure was prepared using ESPript 3.0 (Robert and Gouet, 2014).



Figure B-4: Fluorescent gel images of GFP-PbGH26B CA-PAGE. GFP-PbGH26B was run on four separate native PAGE gels containing the indicated polysaccharides. Two lanes of each gel were loaded with identical samples. The protein was imaged using the native GFP fluorescence of the fusion protein.



Figure B-5: Alignment of SusD homologues with PbSGBP-A. The conserved tryptophan residues, W82 and W306, are marked with arrows. This figure was prepared using ESPript 3.0 (Robert and Gouet, 2014).



Figure B-6: CA-PAGE of PbSGBP-A fused to *E. coli* maltose binding protein. A) BSA B) MBP-PbSGBP-A.



Figure B-7: PbSGBP-B mutants run on CA-PAGE gels containing A) no polysaccharide B) 0.05% tXyG C)
0.05% kGM or D) 0.05% cGM. Each lane was loaded with 5 μg of the variant indicated above the lane. Gels were stained with Coomassie brilliant blue and imaged with white light.



Figure B-8: PbSGBP-C mutants run on CA-PAGE gels containing A) no polysaccharide B) 0.05% tXyG C)
0.05% kGM or D) 0.05% cGM. Each lane was loaded with 5 μg of the variant indicated above the lane. Gels were stained with Coomassie brilliant blue and imaged with white light.



Figure B-9: A) pH-activity and B) temperature-activity plots for PbGH26A. Data in panel A are fit with an asymmetric double sigmoid function.



Figure B-10: Michaelis-Menten plots showing the relationship between substrate concentration and activity for PbGH26A (B,D) or PbGH26A-GH5A (A,C) hydrolysing galactomannan (A,B) or glucomannan (C,D).



Figure B-11: LC-MS of oligosaccharides produced by the action of PbGH26A on inM.



Figure B-12: HPAEC-PAD traces demonstrating the activities of PbEpiA and PbGH130A. A) PbEpiA acting on mannobiose B) PbEpiA acting on cellobiose C) PbEpiA acting on lactose D) PbGH130A and PbEpiA acting on mannobiose.



Figure B-13: Michaelis-Menten plots showing the relationship between substrate concentration and activity for PbGH26A hydrolysing β-mannooligosaccharides of variable length at different cleavage sites.

Purpose	Name	Sequence (5'-3')
Cloning	P15Tv-F-1	CAAAGCCCGAAAGGAAGC
-	p15Tv-F-2	TGATGACGAAGCTTGCGG
	p15Tv-R-1	GCCCTGGAAATACAAGTTTTC
	PbGH26A-F	CTTGTATTTCCAGGGCGATACTTATGTAGTCAATAAGGCAG
	PbGH26A-R	TTCCTTTCGGGCTTTGTTCATCACGGGTAATCAC
	PbGH26A-5A-R	TTCCTTTCGGGCTTTGTTAATTATTATAAGCTTCCTGCATAG
	PbGH26B-F-1	GAGAAAACTTGTATTTCCAGGGCATGGTGTCTTGTCGTCATC
	PbGH26B-F-2	GAGAAAACTTGTATTTCCAGGGCCGTCATCATGAACGAAAG
	PbGH26B-R	GGCCGCAAGCTTCGTCATCATCAGTATTTAAGATACAATCCG
	PbSGBP-A-F	CTTGTATTTCCAGGGCAAGCCGGTAGAGGATAGC
	PbSGBP-A-R	TTCCTTTCGGGCTTTGTTAATAGCTATATGTATTACGCACATC
	PbSGBP-B-F	CTTGTATTTCCAGGGCCAGCCTGATAAATTTGAACTG
	PbSGBP-B-R	TTCCTTTCGGGCTTTGTTATTTATTAGGCACAACACG
	PbSGBP-C-F	CTTGTATTTCCAGGGCGATGACGATTTGAATACTGATC
	PbSGBP-C-R	TTCCTTTCGGGCTTTGTTACTTCAATGTAATCTTAGTGATTG
	pMCSG53-F	ATTGGATTGGAAGTACAGG
	pMCSG53-R	ATTGGAAGTGGATAACGG
	PbGH3A-F	ACCTGTACTTCCAATCCAATCAGTTACGAGCTGACAATATTG
	PbGH3A-R	ATCCGTTATCCACTTCCAATTTATTTATTTCTTCTCAATAAGTTAAGC
	PbGH5B-F	ACCTGTACTTCCAATCCAATCATGATGGTAAATTTTGGCG
		ATCCGTTATCCACTTCCAATTTATCGAATAGAAGATAAGATTTTACAA
	PbGH5B-R	TAC
	PbGH26C-F	ACCTGTACTTCCAATCCAATGCCTATAGATTATTGAATCGTC
	PbGH26C-R	ATCCGTTATCCACTTCCAATCTTAGCAAAATGCTTAAAATCAG
	PbGH36A-F	ACCTGTACTTCCAATCCAATAATATCACGATTCAGTCTAAAAAC
	PbGH36A-R	ATCCGTTATCCACTTCCAATTTATTGAGCTGTTAAGTATAAAACTC
	PbTF-F	ACCTGTACTTCCAATCCAATATGAGTGAAGTATTACATGAAATAAC
	PbTF-R	ATCCGTTATCCACTTCCAATTCACGTCATTTGTCTGATG
	PbEpiA-F	CTTGTATTTCCAGGGCATGAAATCAGAGGTGAAGG
	PbEpiA-R	GCAAGCTTCGTCATCATCATAAGCGTTCCAATATTTC
	PbGH130A-F	CTTGTATTTCCAGGGCATGGAACAATTTTCAGAAAAAATTG
	PbGH130A-R	GCAAGCTTCGTCATCATTACTGATTAATAATCTTCTTATTCTTGAC
Mutagenesis	B PbSGBPB-W242A-F	TAATTCTGGAGCGCATGATCGTATTATC
	PbSGBPB-W242A-R	CCTAACACTTCATTCGTAAC
	PbSGBPB-W274A-F	TAATGGTAAAGCGAATGATGAATTCTCTTTCGAATAC
	PbSGBPB-W274A-R	GCAGCCATAGTAACACCA
	PbSGBPB-W287A-F	GGCTGGTAATGCGGCTAATCCTG
	PbSGBPB-W287A-R	CAGTATTCGAAAGAGAATTC
	PbSGBPC-W524A-F	GAAGACTATTGCGACAGGTAGCTTTG
	PbSGBPC-W524A-R	TCTGCAGAATTTACAATGATC
	PbSGBPC-W533A-F	TTGTGCAAATGCGGCAGGCAATC
	PbSGBPC-W533A-R	TCAAAGCTACCTGTCCAAATAG
	PbSGBPC-W541A-F	AGATCTAGCTGCGGGTGGATATGATTG
	PbSGBPC-W541A-R	TGATTGCCTGCCCAATTTG
	PbSGBPC-W546A-F	TGGATATGATGCGTCTACTGTTAAAGCTG
	PbSGBPC-W546A-R	CCCCAAGCTAGATCTTGATTG

Table B-1: Primers used for the cloning and mutagenesis of the *P. bryantii* MUL genes.

Purpose	Name	Sequence (5'-3')
	PbSGBPC-W568A-F	AAGTGCTAGTGCGGTATGTATTAGTG
	PbSGBPC-W568A-R	GTATCGTTTGCTGTAAAATAG
	PbSGBPC-W579A-F	GGGAGATAAAGCGGACAATTTACCTAATG
	PbSGBPC-W579A-R	TGACGAACACTAATACATAC
Sequencing	PbGH26A-5A-F1	GATGTGGCATTTCAATGTGC
	PbGH26A-5A-F2	TGAACAAGCTTGGAATGCTG
	PbSGBP-A-F1	ACATTGGAAAAGGCAATGGA
	PbSGBP-C-F1	GAAGCAAAGACGGGACAGAT
	PbGH3A-F1	ATCCAAAGTCAGGGGGTAGG
	PbGH3A-F2	GTCGCCCTCTTTGGTGTAAA
	PbGH36A-F1	TTTTCCATCAAGCGATAGCA
	PbGH36A-F2	CGTTTCGGAGCTTTAACCAC

Alkaliflexus imshenetskii (DSM 15055)BacteroidiaAlkaline lakeYes (AlkimDRAFT_1886)15055)BacteroidiaGutYes (BACOVA_02097)Bacteroides ovatus bacteroidesBacteroidiaGutYes (BACOVA_02097)Bacteroides thetatotaomicron (VPI-5482)BacteroidiaGutNo(VPI-5482) Bacteroides uniformis (ATCC 8492))BacteroidiaGutYes (BACUNI_00387)Chrimophaga pinensis (DSM 2588) Chryseolinea serpens (DSM 24574)SphingobacteriaPine litterYes (Cpin_5396)Dysgonomonas gadei (DSM 22836) Flavobacterium anbustiense (CGMCC 16859)BacteroidiaGall bladderYes (HMPREF9455_00425)Jysgonomonas polaribacter (DSM 17595)BacteroidiaSoilYes (HMPREF9455_00425)Flavobacteria obsciteriaSoilYes (IP94DRAFT_01793)Flavobacteria obsciteriaSeawaterNoNoYes (DSM 17595)FlavobacteriaPolaribacter olokanomisti (DSM 17204)FlavobacteriaSeawaterPrevotella bryantii (DSM 11371)BacteroidiaGutYes (PBR_0351)Prevotella copri paludiviens (DSM 17968)BacteroidiaGutNoPrevotella paludiviens (DSM 17968)BacteroidiaGutNoPrevotella prevotella	Species	Class	Source	SEMP-containing β- Mannan Utilisation Locus? (JGI Genetag)
imshenetskii (DSM 15055)acteroides Bacteroides ovatus BacteroidesBacteroidiaGut(AlkimDRAFT_1886) (BACOVA_02097)Bacteroides thetaiotamicron (VPI-5482) 	Alkaliflexus	Bacteroidia	Alkaline lake	Yes
15055) Bacteroides ovatus (ATCC 8483) Bacteroides thatiotaomicron (VPI-5482) Bacteroides BacteroidesBacteroidiaGutYes (BACOVA_02097) NoWiformis (ATCC 8492))BacteroidiaGutYes (BACUNI_00387) uniformis (ATCC 8492))BacteroidiaGutYes (BACUNI_00387)Chitinophaga prinensis (DSM 2588) Chryseolinea serpens (DSM 24574)Sphingobacteria BacteroidiaPine litterYes (Cpin_5396)Dysgonomonas gadei (DSM 22836) Dysgonomonas mossii (DSM 22836)BacteroidiaGall bladderYes (HMPREF9455_00425)Prevonnas onssii (DSM 22836)BacteroidiaAbdominal fluidYes (BO49DRAFT_01793)Flavobacterium of camella forsetii (DSM 17595)Flavobacteria FlavobacteriaSeawaterNoPrevotella bryantii DSM 11371)BacteroidiaGutYes (PBR_0351)Prevotella copri DSM 18205) Prevotella Prevotella Prevotella PlatibacterBacteroidiaGutNoPrevotella Prevotella Prevotella Prevotella Prevotella Prevotella Prevotella PacteroidiaGutNoPrevotella Prevotella Prevotella Prevotella Prevotella Prevotella PrevotellaSoilYes Prevotella PacteroidiaPrevotella Prevotella Prevotella Prevotella Prevotella PacteroidiaGutNoPrevotella Prevotella Prevotella Prevotella PacteroidiaGutNoPrevotella Prevotella Prevotella Prevotella PacteroidiaGutNoPrevotella Prevotella Prevotella Prevotella 	imshenetskii (DSM			(AlkimDRAFT 1886)
Bacteroides ovatus (ATCC 8483) Bacteroides thetaiotaomicrom (VPI-5482)BacteroidiaGutYes (BACOVA_02097) NoBacteroides thetaiotaomicrom (VPI-5482)BacteroidiaGutYes (BACUNI_00387)Bacteroides uniformis (ATCC 8492))BacteroidiaGutYes (BACUNI_00387)Chitinophaga pinensis (DSM 2588) (Chryseolinea serpens (DSM 24574)SphingobacteriaPine litterYes (Cpin_5396)Dysgonomonas gadei (DSM 22836) Flavobacterium anhuiense (CGMCC 1.6859)SacteroidiaGall bladderYes (HMPREF9455_00425)Polaribacter obscher (DSM 17595)FlavobacteriaSoilYes (IP94DRAFT_2989)Polaribacter olostinia (DSM 11371)FlavobacteriaSeawaterNoPrevotella bryantii (DSM 11371)BacteroidiaGutYes (PBR_0351)Prevotella bryantii (DSM 11371)BacteroidiaGutNoPrevotella bryantii (DSM 11371)BacteroidiaGutNoPrevotella bryantii (DSM 11371)BacteroidiaGutNoPrevotella bryantii (DSM 11371)BacteroidiaGutNoPrevotella bryantii (DSM 11371)BacteroidiaGutNoPrevotella bryantii (DSM 18205)BacteroidiaGutNoPrevotella bryantii (DSM 18205)BacteroidiaGutNoPrevotella (DSM 18205)BacteroidiaGutNoPrevotella copri (DSM 18205)BacteroidiaGutNoPrevotella copri (DSM 18205)BacteroidiaGut	15055)			· · · · · · · · · · · · · · · · · · ·
(ATCC 8483) Bacteroides thetaiotaomicron (VPI-5482) Bacteroides uniformis (ATCC 8492))BacteroidiaGut(BACOVA_02097) NoBacteroides uniformis (ATCC 8492))BacteroidiaGutYes (BACUNI_00387)Chitinophaga pinensis (DSM 2588) Chryseolinea serpens (DSM 24574)SphingobacteriaPine litterYes (Cpin_5396)Chyseolinea serpens (DSM 24574)SphingobacteriaGall bladderYesDysgonomonas gadei (DSM 22836) Dysgonomas mossii (DSM 22836)BacteroidiaGall bladderYesFlavobacterium anhuiense (CGMCC 16859)FlavobacteriaSoilYesFlavobacteria (DSM 17595)FlavobacteriaSeawaterNoPolaribacter dokdonensis (DSM 17204)FlavobacteriaSeawaterNoPrevotella bryantii (DSM 11371)BacteroidiaGutYes (PBR_0351)Prevotella bryantii (DSM 11371)BacteroidiaGutNoPrevotella bryantii (DSM 11371)BacteroidiaGutNoPrevotella ruminicola (ATCC 19189)BacteroidiaGutNoPrevotella ruminicola (ATCC 19189)BacteroidiaGutNoPrevotella ruminicola (ATCC 19189)BacteroidiaGutNoPrevotella ruminicola or itsBacteroidiaOral/GutYesPrevotella oris DSGBacteroidiaOral/GutYesPrevotella oris DSGBacteroidiaOral/GutYesPrevotella oris DSGBacteroidiaOral/GutYes <td>Bacteroides ovatus</td> <td>Bacteroidia</td> <td>Gut</td> <td>Yes</td>	Bacteroides ovatus	Bacteroidia	Gut	Yes
Bacteroides thetaiotaomicron (VPI-5482)BacteroidiaGutNoVelse Bacteroides uniformis (ATCC 8492))BacteroidiaGutYes (BACUNI_00387)Chitinophaga pinensis (DSM 2588) Chryseolinea serpens (DSM 24574)SphingobacteriaPine litterYes (Cpin_5396)Dysgonomonas gadei (DSM 22836)BacteroidiaGall bladderYes (HMPREF9455_00425)Dysgonomonas gadei (DSM 22836)BacteroidiaAbdominal fluidYes (BO4PDRAFT_01793)Flavobacterium anhuiense (CGMCC 16859)FlavobacteriaSoilYes (IP94DRAFT_2989)I.6859) Polaribacter PolaribacterFlavobacteriaSeawaterNoPrevotella bryantii (DSM 11371)BacteroidiaGutYes (PBR_0351)Prevotella bryantii (DSM 11371)BacteroidiaGutNoPrevotella paladivivens (DSM 17968)BacteroidiaGutNoPrevotella (DSM 15006)BacteroidiaGutNoPrevotella (DSM 15606)BacteroidiaGutNoPrevotella oris (DSM 15606)BacteroidiaGutNoPrevotella oris (DSM 15606)BacteroidiaGutNoPrevotella oris (DSM 15606)BacteroidiaGutNoPrevotella oris (DSM 15606)BacteroidiaOral/GutYes (PAPCHEP 420_0400)Prevotella oris (DSM 15606)BacteroidiaOral/GutYes (HMPREF9420_0400)	(ATCC 8483)			(BACOVA 02097)
InteriorationDistributionDistributionInteriorationDistributionDistributionInteriorationBacteroidiaGutYes (BACUNI_00387)InteriorationSphingobacteriaPine litterYes (Cpin_5396)InteriorationSphingobacteriaPine litterYes (Cpin_5396)InteriorationCytophagiaSoilNo24574)DysgonomonasBacteroidiaGall bladderYesgadei (DSM 22836)BacteroidiaGall bladderYesDysgonomonasBacteroidiaAbdominal fluidYesmossii (DSM 22836)FlavobacteriaSoilYesInterse (CGMCCFlavobacteriaSoilYesInterse (CGMCCFlavobacteriaSeawaterNoInterse (CGMCCFlavobacteriaSeawaterNoPolaribacterFlavobacteriaSeawaterNoT204)BacteroidiaGutNoPrevotella bryantiiBacteroidiaGutNoPrevotella copriBacteroidiaGutNoPrevotellaBacteroidiaGutNoPrevotellaBacteroidiaGutNoPrevotellaBacteroidiaGutNoPrevotellaBacteroidiaGutNoPrevotella solivaeBacteroidiaGutNoPrevotella risBacteroidiaOralYesPrevotella arisBacteroidiaOralYesPrevotella arisBacteroidiaOral/GutYesPrevotella aris<	Racteroides	Bacteroidia	Gut	No
Interaction Interaction BacteroidesBacteroidiaGutYes (BACUNI_00387)BacteroidesSphingobacteriaPine litterYes (Cpin_5396)MagendaSphingobacteriaPine litterYes (Cpin_5396)ChrispeolineaCytophagiaSoilNoserpens (DSMBacteroidiaGall bladderYesgadei (DSM 22836)BacteroidiaGall bladderYesDysgonomonasBacteroidiaAbdominal fluidYesgadei (DSM 22836)FlavobacteriaSoil(HMPREF9455_00425)DysgonomonasBacteroidiaAbdominal fluidYesanhuiense (CGMCCFlavobacteriaSoil(IP94DRAFT_01793)FlavobacteriaSeawaterNo(DSM 17595)PolaribacterFlavobacteriaSeawaterNodokdonensis (DSMFlavobacteriaSeawaterNo17204)BacteroidiaGutYes (PBR_0351)Prevotella bryantiiBacteroidiaGutNo(DSM 18205)BacteroidiaGutNoPrevotella copriBacteroidiaGutNo17968)BacteroidiaGutNoPrevotellaBacteroidiaGutNoPrevotellaBacteroidiaGutNoPrevotellaBacteroidiaGutNoPrevotellaBacteroidiaGutNoPrevotella copriBacteroidiaGutNoPrevotella solivaeBacteroidiaOralYesPrevotella solivaeBacteroidiaOral	thetaiotaomicron	2		
Bacteroides uniformis (ATCC 8492))BacteroidiaGutYes (BACUNI_00387)Chitinophaga pinensis (DSM 2588)SphingobacteriaPine litterYes (Cpin_5396)Chryseolinea serpens (DSM 24574)SyhingobacteriaSoilNoDysgonomonas gadei (DSM 22836)BacteroidiaGall bladderYes (HMPREF9455_00425)Dysgonomonas monas gadei (DSM 22836)BacteroidiaAbdominal fluidYes (BO49DRAFT_01793)Flavobacterium anhuiense (CGMCC (DSM 17595)FlavobacteriaSoilYes (BO49DRAFT_2989)1.6859) Gramella forsetii (DSM 17595)FlavobacteriaSeawaterNoPolaribacter dokdonensis (DSM 17204)FlavobacteriaSeawaterNoPrevotella copri (DSM 18205)BacteroidiaGutYes (PBR_0351)Prevotella anuinicola (ATCC 19189)BacteroidiaGutNoPrevotella (DSM 15006)BacteroidiaGutNoPrevotella alivae (DSM 15606)BacteroidiaGutNoPrevotella oris (DSM 15606)BacteroidiaOralYes (PADAFT_004)Prevotella orisBacteroidiaOral/GutYes (PADAFT_004)	(VPI-5482)			
DationalDataDataTes (Differingenees)uniformis (ATCC 8492))SphingobacteriaPine litterYes (Cpin_5396)Chriseolnea serpens (DSM 24574)SphingobacteriaSoilNoDysgonomonas gadei (DSM 22836)BacteroidiaGall bladderYes (BO49DRAFT_01793)Dysgonomonas mossii (DSM 22836)BacteroidiaAbdominal fluidYes (BO49DRAFT_01793)Flavobacterium anhuiense (CGMCC 1.6859)FlavobacteriaSoilYes (BO49DRAFT_01793)Flavobacteria (DSM 17595)FlavobacteriaSeawaterNoPrevotella forsetii (DSM 11371)BacteroidiaGutYes (PBR_0351)Prevotella copri (DSM 11371)BacteroidiaGutNoPrevotella forse (DSM 11371)BacteroidiaGutNoPrevotella forse (DSM 11371)BacteroidiaGutNoPrevotella pryantii (DSM 11370)BacteroidiaGutNoPrevotella pryantii (DSM 11370)BacteroidiaGutNoPrevotella paludivivens (DSM (DSM 11966)BacteroidiaGutNoPrevotella prevotella salivae (DSM 15606)BacteroidiaGutNoPrevotella oris (DSM 15606)BacteroidiaOralYes (HMPREF9420_0400)Prevotella orisBacteroidiaOral/GutYes (FMPREF9420_0400)	Racteroides	Bacteroidia	Gut	Yes (BACUNI 00387)
Ringonius (ITICC) R4921)SphingobacteriaPine litterYes (Cpin_5396)Christinophaga pinensis (DSM 2588)CytophagiaSoilNoChryseolinea serpens (DSM 24574)CytophagiaSoilNo24574)BacteroidiaGall bladderYes (HMPREF9455_00425)Dysgonomonas gadei (DSM 22836)BacteroidiaAbdominal fluidYes (BO49DRAFT_01793)Flavobacterium anhuiense (CGMCCFlavobacteriaSoilYes (IP94DRAFT_2989)1.6859)FlavobacteriaSeawaterNoOramella forsetii (DSM 1759)FlavobacteriaSeawaterNoPolaribacter dokdonensis (DSM 17204)FlavobacteriaSeawaterNoPrevotella bryantii (DSM 11371)BacteroidiaGutYes (PBR_0351)Prevotella copri paludivivens (DSM 17968)BacteroidiaGutNoPrevotella paludivivens (DSM 17968)BacteroidiaGutNoPrevotella ruminicola (ATCC 19189)BacteroidiaGutNoPrevotella salivae (DSM 15606)BacteroidiaGutNoPrevotella arbivae (DSM 15606)BacteroidiaGutNoPrevotella oris (DSM 15606)BacteroidiaGutNoPrevotella oris (DSM 15606)BacteroidiaOralYes (FMPREF9420_0400)Prevotella oris (DSM 15606)BacteroidiaOral/GutYes (FMPREF9420_0400)	uniformis (ATCC	Bueteroraia	Cut	
Chilinophaga pinensis (DSM 2588)SphingobacteriaPine litterYes (Cpin_5396)Chryseolinea serpens (DSM 24574)CytophagiaSoilNoDysgonomonas gadei (DSM 22836)BacteroidiaGall bladderYes (HMPREF9455_00425)Dysgonomonas mossii (DSM 22836)BacteroidiaAbdominal fluidYes (BO49DRAFT_01793)Flavobacterium anhuiense (CGMCC 1.6859)FlavobacteriaSoilYes (IP94DRAFT_2989)Gramella forsetii (DSM 17595)FlavobacteriaSeawaterNoPolaribacter dokdonensis (DSM 17204)FlavobacteriaSeawaterNoPrevotella copri platudivivens (DSM 17968)BacteroidiaGutYes (BO45DRAFT_00691)Prevotella uminicola (ATCC 19189)BacteroidiaGutNoPrevotella salivae (DSM 15606)BacteroidiaOralYes (HMPREF9420_0400)Prevotella oris DSM 15606)BacteroidiaOral/GutYes (HMPREF9420_0400)	8492))			
Simmer PrinceSpring of activityFille FilleFille Fillepinensis (DSM 2588)CytophagiaSoilNo24574)BacteroidiaGall bladderYesgadei (DSM 22836)BacteroidiaAbdominal fluidYespysgonomonasBacteroidiaAbdominal fluidYesmossii (DSM 22836)FlavobacteriaSoilYesflavobacteriumFlavobacteriaSoilYesanhuiense (CGMCCFlavobacteriaSeawaterNo(DSM 17595)FlavobacteriaSeawaterNoPolaribacterFlavobacteriaSeawaterNodokdonensis (DSMBacteroidiaGutYes (PBR_0351)(DSM 11371)BacteroidiaGutNoPrevotella bryantiiBacteroidiaGutNo(DSM 11205)BacteroidiaGutNoPrevotellaBacteroidiaGutNoPrevotellaBacteroidiaGutNoPrevotellaBacteroidiaGutNoPrevotellaBacteroidiaGutNoPrevotella salivaeBacteroidiaGutNoPrevotella salivaeBacteroidiaOralYes(DSM 15606)BacteroidiaOralYesPrevotella orisBacteroidiaOralYesPrevotella orisBacteroidiaOralYesPrevotella orisBacteroidiaOralYesPrevotella orisBacteroidiaOralYesPrevotella orisBacteroidiaOral </td <td>Chitinophaga</td> <td>Sphingobacteria</td> <td>Pine litter</td> <td>Yes (Cpin 5396)</td>	Chitinophaga	Sphingobacteria	Pine litter	Yes (Cpin 5396)
PrincipleCytophagiaSoilNoserpens (DSM 24574)CytophagiaSoilNoDysgonomonas gadei (DSM 22836)BacteroidiaGall bladderYes (HMPREF9455_00425)Dysgonomonas mossii (DSM 22836)BacteroidiaAbdominal fluidYes (BO49DRAFT_01793)Flavobacterium anhuiense (CGMCC 1.6859)FlavobacteriaSoilYes (IP94DRAFT_2989)1.6859)FlavobacteriaSeawaterNoOrsmella forsetii (DSM 17595)FlavobacteriaSeawaterNoPolaribacter dokdonensis (DSM 17204)FlavobacteriaSeawaterNoPrevotella bryantii (DSM 11371)BacteroidiaGutYes (BO45DRAFT_00691)Prevotella paludivivens (DSM 17968)BacteroidiaGutNoPrevotella (DSM 15606)BacteroidiaGutNoPrevotella salivae (DSM 15606)BacteroidiaOralYes (HMPREF9420_0400)Prevotella orisBacteroidiaOral/GutYes (HMPREF9420_0400)	ninensis (DSM 2588)	Spiningoodeteria		105 (0pm_0000)
Serpens (DSM 24574)SoftNoDysgonomonas gadei (DSM 22836)BacteroidiaGall bladderYes (HMPREF9455_00425)Dysgonomonas mossii (DSM 22836)BacteroidiaAbdominal fluidYes (BO49DRAFT_01793)Flavobacterium anhuiense (CGMCC 1.6859)FlavobacteriaSoilYes (IP94DRAFT_2989)Gramella forsetii (DSM 17595)FlavobacteriaSeawaterNoPolaribacter dokdonensis (DSM 17204)FlavobacteriaSeawaterNoPrevotella bryantii (DSM 11371)BacteroidiaGutYes (PBR_0351)Prevotella copri (DSM 18205)BacteroidiaGutNoPrevotella paludivivens (DSM 1798)BacteroidiaGutNoPrevotella paludivivens (DSM (DSM 15606)BacteroidiaGutNoPrevotella oris (DSM 15606)BacteroidiaOralYes (Parcotella orisPrevotella orisBacteroidiaOralYes (PSM 15606)Prevotella orisBacteroidiaOralYes (HMPREF9420_0400)	Chryseolinea	Cytophagia	Soil	No
Scipits (BMA)BacteroidiaGall bladderYesDysgonomonas gadei (DSM 22836)BacteroidiaAbdominal fluidYesDysgonomonas mossii (DSM 22836)BacteroidiaAbdominal fluidYesFlavobacterium anhuiense (CGMCCFlavobacteriaSoilYesI.6859)FlavobacteriaSeawaterNoGramella forsetii (DSM 17595)FlavobacteriaSeawaterNoPolaribacter dokdonensis (DSM 17204)FlavobacteriaSeawaterNoPrevotella bryantii (DSM 11371)BacteroidiaGutYesPrevotella copri paludivivens (DSM 17968)BacteroidiaGutNoPrevotella paludivivens (DSM 19199)BacteroidiaGutNoPrevotella salivae (DSM 15606)BacteroidiaOralYesPrevotella oris (DSM 15606)BacteroidiaOralYesPrevotella orisBacteroidiaOralYesPrevotella orisBacteroidiaOralYes	sernens (DSM	Ojtophugiu	Soli	
Dysgonomonas gadei (DSM 22836)BacteroidiaGall bladderYes (HMPREF9455_00425)Dysgonomonas mossii (DSM 22836)BacteroidiaAbdominal fluidYes (BO49DRAFT_01793)Flavobacterium anhuiense (CGMCC 1.6859)FlavobacteriaSoilYes (IP94DRAFT_2989)I.6859)FlavobacteriaSeawaterNoGramella forsetii (DSM 17595)FlavobacteriaSeawaterNoPolaribacter dokdonensis (DSM 17204)FlavobacteriaSeawaterNoPrevotella bryantii (DSM 11371)BacteroidiaGutYes (BO45DRAFT_00691)Prevotella paludivivens (DSM 17968)BacteroidiaGutNoPrevotella (DSM 15606)BacteroidiaGutNoPrevotella salivae (DSM 15606)BacteroidiaOralYes (HMPREF9420_0400)Prevotella orisBacteroidiaOral/GutYes (HMPREF9420_0400)	24574)			
Dysolominal gadei (DSM 22836)DistributionDistributionHumplef 9455_00425)Dysgonomonas mossii (DSM 22836)BacteroidiaAbdominal fluidYes (BO49DRAFT_01793)Flavobacterium anhuiense (CGMCC 1.6859)FlavobacteriaSoilYes (IP94DRAFT_2989)I.6859)FlavobacteriaSeawaterNoOralibacter dokdonensis (DSM 17204)FlavobacteriaSeawaterNoPrevotella bryantii (DSM 11371)BacteroidiaGutYes (BO49DRAFT_2989)Prevotella copri paludivivens (DSM 17968)BacteroidiaGutNoPrevotella (DSM 15606)BacteroidiaGutNoPrevotella salivae (DSM 15606)BacteroidiaGutNoPrevotella oris (DSM 15606)BacteroidiaOralYes (HMPREF9420_0400)Prevotella oris (DSM 15606)BacteroidiaOral/GutYes (Paludivivens (DSM 15606)	Dysgonomonas	Bacteroidia	Gall bladder	Yes
Jose (Control Decorrection)BacteroidiaAbdominal fluidYesDysgonomas mossii (DSM 22836)BacteroidiaAbdominal fluidYesFlavobacterium anhuiense (CGMCCFlavobacteriaSoilYes1.6859)FlavobacteriaSeawaterNoGramella forsetii (DSM 17595)FlavobacteriaSeawaterNoPolaribacter dokdonensis (DSM 17204)FlavobacteriaSeawaterNoPrevotella bryantii (DSM 11371)BacteroidiaGutYes (PBR_0351)Prevotella copri paludivivens (DSM 17968)BacteroidiaGutNoPrevotella paludivivens (DSM 1798)BacteroidiaGutNoPrevotella paludivivens (DSM 17968)BacteroidiaGutNoPrevotella ruminicola (ATCC 19189)BacteroidiaOralYesPrevotella salivae (DSM 15606)BacteroidiaOralYesPrevotella virisBacteroidiaOral/GutYes	gadei (DSM 22836)	Ductororan		(HMPREF9455_00425)
DescriptionDescriptionProcessing <td>Dysgonomonas</td> <td>Bacteroidia</td> <td>Abdominal fluid</td> <td>Yes</td>	Dysgonomonas	Bacteroidia	Abdominal fluid	Yes
Initial (John Lloso)FlavobacteriaSoil(John Lloso)Flavobacterium anhuiense (CGMCC 1.6859)FlavobacteriaSoilYes (IP94DRAFT_2989)Gramella forsetii (DSM 17595)FlavobacteriaSeawaterNoPolaribacter dokdonensis (DSM 17204)FlavobacteriaSeawaterNoPrevotella bryantii (DSM 11371)BacteroidiaGutYes (PBR_0351)Prevotella copri (DSM 18205)BacteroidiaGutNoPrevotella paludivivens (DSM 17968)BacteroidiaGutNoPrevotella (DSM 15606)BacteroidiaGutNoPrevotella salivae (DSM 15606)BacteroidiaOralYes (HMPREF9420_0400)Prevotella orisBacteroidiaOral/GutYes (HMPREF9420_0400)	mossii (DSM 22836)	Bueteroraia		(BO49DRAFT 01793)
AntoolationalFairoolationalSolaFileanhuiense (CGMCC	Flavobacterium	Flavobacteria	Soil	Yes
InitiationControl of the control of the c	anhuiense (CGMCC	1 hav oblacteria	Soli	(IP94DRAFT 2989)
InstantFlavobacteriaSeawaterNoGramella forsetii (DSM 17595)FlavobacteriaSeawaterNoPolaribacter dokdonensis (DSM 17204)FlavobacteriaSeawaterNoPrevotella bryantii (DSM 11371)BacteroidiaGutYes (PBR_0351)Prevotella copri (DSM 18205)BacteroidiaGutNoPrevotella paludivivens (DSM 17968)BacteroidiaSoilYes (BO45DRAFT_00691)Prevotella ruminicola (ATCC 19189)BacteroidiaGutNoPrevotella oris (DSM 15606)BacteroidiaOralYes (HMPREF9420_0400)Prevotella oris (DSM 10711)BacteroidiaOral/GutYes	1 6859)			
Oranical posteriaFlavobacteriaSeawaterNo(DSM 17595)FlavobacteriaSeawaterNoPolaribacterFlavobacteriaSeawaterNodokdonensis (DSM 17204)BacteroidiaGutYes (PBR_0351)Prevotella bryantii (DSM 11371)BacteroidiaGutNoPrevotella copri (DSM 18205)BacteroidiaGutNoPrevotella paludivivens (DSM 17968)BacteroidiaSoilYesPrevotella ruminicola (ATCC 19189)BacteroidiaGutNoPrevotella aslivae (DSM 15606)BacteroidiaOralYesPrevotella oris (DSM 10711)BacteroidiaOral/GutYes	Gramella forsetii	Flavobacteria	Seawater	No
Polaribacter dokdonensis (DSM 17204)FlavobacteriaSeawaterNoPrevotella bryantii (DSM 11371)BacteroidiaGutYes (PBR_0351)Prevotella copri (DSM 18205)BacteroidiaGutNoPrevotella paludivivens (DSM 17968)BacteroidiaSoilYes (BO45DRAFT_00691)Prevotella ruminicola (ATCC 19189)BacteroidiaGutNoPrevotella oris (DSM 15606)BacteroidiaOralYes (HMPREF9420_0400)Prevotella oris (DSM 10011)BacteroidiaOral/GutYes	(DSM 17595)			
AdvisitionFunctionalDeamaterFunctionaldokdonensis (DSM17204)BacteroidiaGutYes (PBR_0351)Prevotella bryantii (DSM 11371)BacteroidiaGutNoPrevotella copri (DSM 18205)BacteroidiaGutNoPrevotella paludivivens (DSMBacteroidiaSoilYes (BO45DRAFT_00691)17968) Prevotella ruminicola (ATCC 19189)BacteroidiaGutNoPrevotella salivae (DSM 15606)BacteroidiaOralYes (HMPREF9420_0400)Prevotella oris (DSM 10011)BacteroidiaOral/GutYes	Polaribacter	Flavobacteria	Seawater	No
InterventionConstruction17204)Prevotella bryantiiBacteroidiaGutYes (PBR_0351)Prevotella copri (DSM 11371)BacteroidiaGutNoPrevotella copri (DSM 18205)BacteroidiaGutYesPrevotella paludivivens (DSM 17968)BacteroidiaSoilYes (BO45DRAFT_00691)Prevotella ruminicola (ATCC 19189)BacteroidiaGutNoPrevotella salivae (DSM 15606)BacteroidiaOralYes (HMPREF9420_0400)Prevotella oris (DSM 10711)BacteroidiaOral/GutYes	dokdonensis (DSM			
Prevotella bryantii (DSM 11371)BacteroidiaGutYes (PBR_0351)Prevotella copri (DSM 18205)BacteroidiaGutNoPrevotella paludivivens (DSM 17968)BacteroidiaSoilYes (BO45DRAFT_00691)Prevotella ruminicola (ATCC 19189)BacteroidiaGutNoPrevotella salivae (DSM 15606)BacteroidiaOralYes (HMPREF9420_0400)Prevotella orisBacteroidiaOral/GutYes	17204)			
(DSM 11371)Dataset of dataOutDescriptionPrevotella copri (DSM 18205)BacteroidiaGutNoPrevotella paludivivens (DSM 17968)BacteroidiaSoilYes (BO45DRAFT_00691)Prevotella ruminicola (ATCC 19189)BacteroidiaGutNoPrevotella salivae (DSM 15606)BacteroidiaOralYes (HMPREF9420_0400)Prevotella oris (DSM 10711)BacteroidiaOral/GutYes (CDD 4 UT 01001)	Prevotella brvantii	Bacteroidia	Gut	Yes (PBR 0351)
Prevotella copri (DSM 18205)BacteroidiaGutNoPrevotella paludivivens (DSM 17968)BacteroidiaSoilYes (BO45DRAFT_00691)Prevotella ruminicola (ATCC 19189)BacteroidiaGutNoPrevotella salivae (DSM 15606)BacteroidiaOralYes (HMPREF9420_0400)Prevotella orisBacteroidiaOral/GutYes	(DSM 11371)	2		
InstructionData Foundation(DSM 18205)PrevotellaPrevotellaBacteroidiapaludivivens (DSM17968)PrevotellaPrevotellaBacteroidiaGutNoruminicola (ATCC19189)BacteroidiaPrevotella salivaeBacteroidia(DSM 15606)BacteroidiaPrevotella orisBacteroidiaOral/GutYes(DSM 107111)	Prevotella copri	Bacteroidia	Gut	No
Prevotella paludivivens (DSM 17968)BacteroidiaSoilYes (BO45DRAFT_00691)Prevotella ruminicola (ATCC 19189)BacteroidiaGutNoPrevotella salivae (DSM 15606)BacteroidiaOralYes (HMPREF9420_0400)Prevotella orisBacteroidiaOral/GutYes (ENDEADED ADED ADED ADED ADED ADED ADED A	(DSM 18205)			
paludivivens (DSM)InternationalInternationalInternational17968)PrevotellaBacteroidiaGut(BO45DRAFT_00691)PrevotellaBacteroidiaGutNo19189)Prevotella salivaeBacteroidiaOralYes(DSM 15606)BacteroidiaOral/GutYesPrevotella orisBacteroidiaOral/GutYes	Prevotella	Bacteroidia	Soil	Yes
17968) Prevotella ruminicola (ATCC 19189)BacteroidiaGutNoPrevotella salivae (DSM 15606)BacteroidiaOralYes (HMPREF9420_0400)Prevotella oris (DSM 107111)BacteroidiaOral/GutYes (HMPREF9420_0400)	paludivivens (DSM			(BO45DRAFT 00691)
Prevotella ruminicola (ATCC 19189)BacteroidiaGutNoPrevotella salivae (DSM 15606)BacteroidiaOralYes (HMPREF9420_0400)Prevotella oris CDSM 10711)BacteroidiaOral/GutYes	17968)			
ruminicola (ATCC 19189)Prevotella salivae (DSM 15606)BacteroidiaOralYes (HMPREF9420_0400)Prevotella orisBacteroidiaOral/GutYes	Prevotella	Bacteroidia	Gut	No
19189) Prevotella salivae (DSM 15606)BacteroidiaOralYes (HMPREF9420_0400)Prevotella oris CDSM 10711)BacteroidiaOral/GutYes	ruminicola (ATCC			
Prevotella salivae (DSM 15606)BacteroidiaOralYes (HMPREF9420_0400)Prevotella orisBacteroidiaOral/GutYes	19189)			
(DSM 15606) Prevotella orisBacteroidiaOral/Gut(HMPREF9420_0400) Yes00000000000000000000000000000000000	Prevotella salivae	Bacteroidia	Oral	Yes
Prevotella oris Bacteroidia Oral/Gut Yes	(DSM 15606)			(HMPREF9420 0400)
	Prevotella oris	Bacteroidia	Oral/Gut	Yes
(DSM 18/11) (D468DRAF1_01281)	(DSM 18711)			(D468DRAFT_01281)

Table B-2: Genomes searched for putative SEMP-containing β -mannan utilisation loci.

Species	Class	Source	SEMP-containing β- Mannan Utilisation Locus? (JGI Genetag)
Prevotella denticola	Bacteroidia	Oral	No
(DSM 20614)			
Prevotella	Bacteroidia	Oral	No
multiformis (DSM			
16608)			
Saccharicrinis	Cytophagia	Shark gill	Yes
fermentans (DSM			(cytfeDRAFT_1117)
9555)			-

Appendix C Supporting Information for Chapter 5: Characterisation of Two Mannan/Xylan Acetylesterases from *Prevotella bryantii* B₁4



Figure C-1: HPAEC-PAD analysis of the XOs mixture.



Figure C-2: Plots of change in E-value vs. hit number for NCBI RefSeq BLAST hits using PBR_0354 (top) or PBR_0355 (bottom) as search sequences



Clade I



Figure C-3: Phylogenetic tree of PbCExA homologues. Bootstrap values are shown next to each node with

support greater than 50%.



Figure C-4: Phylogenetic tree of PbCE7A homologues. Bootstrap values are shown next to each node with

support greater than 50%.



Figure C-5: Molecular phylogeny of PbCE7A. Structure-guided alignment of PbCE7A and six characterised CE6 esterases. The three residues forming the catalytic triad of CE7 are marked with arrows. Secondary structure (mapped from the TmAcE) is shown above each row of sequence data. Panel A was prepared using ESPript 3.0 (Robert and Gouet, 2014).



Figure C-6: Acetonitrile tolerance, pH optimum and thermal stability of PbCExA and PbCE7A. Conditions tolerated by PbCE7A. C) pH-activity curve of PbCExA showing an optimal pH range of 7 to 8. B)
Temperature-residual activity plot showing the loss of PbCE7A activity above 40°C. C) Activity-acetonitrile concentration plot showing tolerance of up to 12% v/v acetonitrile by PbCE7A.



Figure C-7: Michaelis-Menten plots showing the relationship between substrate concentration and initial hydrolytic rates for PbCExA and PbCE7A hydrolysing A) PNP-Acetate or B) PNP-Butyrate.

Name	Sequence
pMCSG53-F	ATTGGATTGGAAGTACAGG
pMCSG53-R	ATTGGAAGTGGATAACGG
PbCE7A-F	ACCTGTACTTCCAATCCAATTCATTCGTTCAACAAATGTTTAC
PbCE7A-R	ATCCGTTATCCACTTCCAATCAAATTCGTGGAACAAATATAATG
PbCExA-F	ACCTGTACTTCCAATCCAATGTCAAACTTCCCCACATG
PbCExA-R	ATCCGTTATCCACTTCCAATTTACCAGTTGTCAGTTCG

Table C-1: Primers used for the cloning of PbCExA and PbCE7A.

Data collection	
Space group	$P2_{1}2_{1}2_{1}$
Cell dimensions	
a, b, c (Å)	70.94, 145.67, 163.10
Resolution, Å	66.50 - 2.28
$\mathbf{R}_{merge}{}^{a}$	0.258 (2.733) ^b
R _{pim}	0.069 (0.723)
$I / \sigma(I)$	11.8 (1.3)
Completeness, %	100 (100)
Redundancy	14.9 (15.1)
CC _{1/2}	0.997 (0.512)
Refinement	
Resolution, A	66.50 - 2.28
No. of unique reflections:	77807, 3962
working, test	
<i>R</i> -factor/free <i>R</i> -factor ^c	17.9/22.5 (32.2/41.9)
No. of refined atoms,	
molecules	
Protein	11199
Solvent	102
Water	1069
<i>B</i> -factors	
Protein	50.0
Solvent	79.5
Water	52.2
r.m.s.d.	
Bond lengths, Å	0.003
Bond angles, °	0.568

Table C-2: X-ray diffraction data collection and refinement statistics for PbCExA.

 ${}^{a}R_{\text{merge}} = \sum_{h} \sum_{i} |I_{i}(h) - \langle I(h) \rangle / \sum_{h} \sum_{i} I_{i}(h)$, where $I_{i}(h)$ and $\langle I(h) \rangle$ are the *i*th and mean

measurement of the intensity of reflection h.

^bFigures in parentheses indicate the values for the outer shells of the data.

 ${}^{c}R = \Sigma |F_{p}{}^{obs} - F_{p}{}^{calc}|/\Sigma F_{p}{}^{obs}$, where $F_{p}{}^{obs}$ and $F_{p}{}^{calc}$ are the observed and calculated structure factor amplitudes, respectively.

Appendix D Additional Publications Not Included in this Thesis

McGregor, N., Arnal, G., and Brumer, H. (2017). Quantitative Kinetic Characterisation of Glycoside Hydrolases Using High-Performance Anion-Exchange Chromatography (HPAEC). In *Protein-Carbohydrate Interactions*, (Humana Press, New York, NY), pp. 15–25.

Fanuel, M., Garajova, S., Ropartz, D., McGregor, N., Brumer, H., Rogniaux, H., Berrin, J-G. (2017) The *Podospora anserina* Lytic Polysaccharide Monooxygenase PaLPMO9H Catalyses Oxidative Cleavage of Diverse Plant Cell Wall Matrix Glycans. *Biotechnology for Biofuels* 10, 63.

Robb, M., Hobbs, J.K., Woodiga, S. A., Shapiro-Ward, S., Suits, M. D. L., McGregor, N., Brumer, H., Yesilkaya, H., King, S. J., Boraston, A. B. (2017) Molecular Characterisation of N-glycan Degradation and Transport in *Streptococcus pneumoniae* and Its Contribution to Virulence. *PLOS Pathogens* **13**, e1006090.