Elucidation of structure and substrate-specificity of a glycoside hydrolase from family 81 and a carbohydrate binding module from family 56

by

Alexander Fillo
BSc, University of Victoria, 2012

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Biochemistry and Microbiology

Supervisory Committee

Elucidation of structure and substrate-specificity of a glycoside hydrolase from family 81 and a carbohydrate binding module from family 56

by

Alexander Fillo
BSc, University of Victoria, 2012

Supervisory Committee

Dr. Alisdair B. Boraston (Department of Biochemistry and Microbiology)
Supervisor

Dr. Martin Boulanger (Department of Biochemistry and Microbiology)
Departmental Member

Dr. Juergen Ehlting (Department of Biology)
Outside Member

Abstract

Supervisory Committee
Dr. Alisdair B. Boraston, Department of Biochemistry and Microbiology
Supervisor
Dr. Martin Boulanger, Department of Biochemistry and Microbiology
Departmental Member
Dr. Juergen Ehlting, Department of Biology
Outside Member

The degradation of carbohydrates is essential to many biological processes such as cell wall remodelling, host-pathogen defense, and energy synthesis in the form of ATP. Several of these processes utilize carbohydrate-active enzymes to accomplish these goals. Studying the degradation of polysaccharides by carbohydrate-active enzymes synthesized by microbes has allowed us to further understand biomass conversion. A portion of these polysaccharides consists of β-1,3-linked glucose (i.e. β-1,3-glucan), which is found in plants, fungi, and brown macroalgae. The hydrolysis of β-1,3-glycosidic linkages is catalyzed by β-1,3-glucanases, which are present in six different glycoside hydrolase (GH) families: 16, 17, 55, 64, 81, and 128. These enzymes play important biological roles including carbon utilization, cell wall modeling, and pathogen defense. This study focuses on a gene from *Bacillus halodurans* encoding for a multi-modular protein (*Bh*Lam81) consisting of a glycoside hydrolase from family 81 (*Bh*GH81), a carbohydrate-binding module (CBM) from family 6 (*Bh*CBM6), and a CBM from family 56 (*Bh*CBM56). Previously, thorough structural and substrate-specific characterization has been carried out on *Bh*CBM6. This CBM binds the non-reducing end of β-1,3-glucan. A member of CBM family 56 has been shown to recognize and bind the insoluble β-1,3-glucan, pachyman, however it is structurally uncharacterized. A glycoside hydrolase belonging to family 81 from *Saccharomyces cerevisiae* has been previously shown to degrade the β-1,3-glucans, laminarin and pachyman, however the structure of this enzyme was not determined. Recently, a member of

GH family 81 has been structurally characterized; however, substrate-specificity was not determined in that study. Therefore, this study concentrated on two goals: Determining the substrate-specificity of *Bh*GH81 and *Bh*CBM56, and solving the structure of *Bh*GH81 and *Bh*CBM56 in order to gain insight into the molecular details of how they recognize and act on their substrate(s). The deoxyribonucleic acid (DNA) encoding for these modules were dissected by restriction digest from *B. halodurans* genomic DNA and recombinantly expressed in *Escherichia coli* (*E. coli*) as separate constructs. Both *Bh*GH81 and *Bh*CBM56 were purified and their crystal structures obtained. *Bh*GH81 and *Bh*CBM56 were solved to 2.5 Å resolution by single-wavelength anomalous dispersion (SAD) and to 1.7 Å resolution by multi-wavelength anomalous dispersion (MAD), respectively. In order to determine the substrate-specificity of *Bh*GH81 and *Bh*CBM56 and speculate on the molecular details of how they recognize and act on their substrate(s), substrate-specificity tests were combined with structural analysis for both of these modules. By using qualitative depletion assays, quantitative depletion assays, and affinity electrophoresis, it was revealed that *Bh*CBM56 binds both insoluble and soluble β-1,3-glucan. The crystal structure of *Bh*CBM56 revealed that it is a β-sandwich composed of two antiparallel β-sheets consisting of five β-strands each. By comparing *Bh*CBM56 to a β-1,3-glucan binding protein from *Plodia interpunctella* (βGRP) a putative substrate-binding cleft on the concave side of the β-sandwich created by a platform of hydrophobic residues surrounded by several polar and charged residues was revealed. This comparison also allowed for speculation of the amino acids (W1015, H965, and D963) that are potentially essential for recognition of β-1,3-glucan substrates by *Bh*CBM56. Activity of *Bh*GH81 on β-1,3-glucans was confirmed by both thin-layer chromatography and product analysis using high performance anion exchange chromatography. The high performance anion exchange chromatography of *Bh*GH81 hydrolysis

suggested it has both exo and endo modes of action. The crystal structure of *Bh*GH81 revealed that it consists of domains A, B, and C: A β-sandwich domain (A), a linker domain (B), and an $(\alpha/\alpha)_6$-barrel domain (C). This structure revealed a putative substrate-binding cleft on one side of the $(\alpha/\alpha)_6$-barrel with a blind canyon active site topology. It also revealed two putative catalytic residues, E542 and E546. All GHs from family 81 characterized so far, hydrolyze β-1,3-glucan in an endo acting manner. By comparing the structure of *Bh*GH81 acquired in this study to a cellulase from *Thermobifida fusca*, which has an endo-processive mode of action, we can speculate that *Bh*GH81 also has an endo-processive mode of action. The structural and biochemical analysis of *Bh*GH81 and *Bh*CBM56 in this study has aided in further understanding the molecular details both GH family 81 and CBM family 56 proteins, as well as the degradation of β-1,3-glucan by multimodular enzymes. Understanding these molecular details could be important for industrial applications such as, engineering a microbial platform for more efficient biofuel production.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

CBM: carbohydrate-binding module

DNA: deoxyribonucleic acid

GH: glycoside hydrolase

IMAC: immobilized metal affinity column

KDa: kiloDaltons

K_d: dissociation constant

LB: Luria Bertani

MAD: multi-wavelength anomalous diffraction

N_o: binding capacity

PDB: Protein data bank

RMSD: root mean square deviation

SAD: single-wavelength anomalous dispersion

SDS-PAGE: sodium dodecylsulfate polyacrylamide gel electrophoresis

SEC: size-exclusion chromatography

TLC: thin-layer chromatography

Acknowledgments

Primarily, I would like to thank my supervisor, Dr. Alisdair Boraston, who allowed me the amazing opportunity to pursue a graduate degree in his lab. He has provided me with invaluable knowledge and supervision throughout my Masters degree. He revealed to me that working in a lab can be rewarding, exciting, and fun. I envy his ability to stay focussed and calm throughout stressful times and I hope to take these qualities with me into my future.

I would also like to thank the other members my committee: Juergen Ehlting and Martin Boulanger for their guiadance thoughout my Masters degree. I would like to specifically thank Juergen Ehlting, who gave me the first incredible opportunity of performing research in a lab. Without his contribution to my laboratory experience, my pursuit of Masters degree may have not been possible.

I would like to thank the members of the lab who have helped guide me through my Masters degree. Firstly, I would like to thank Craig Robb who I think of as a mentor and a friend. Also Ben Pluvinage, Ilit Noach, Melissa Cid, Andrew Hettle, and Kaleigh Giles for their help and support. A special thanks to Michelle Lee, who I met in Dr. Boraston's Proteins class, and who has supported me and become my partner.

Dedication

I would like to thank my family for believing in me and supporting me throughout my undergraduate and graduate experience. I would like to specifically thank my father who passed away in 2011. For a plumber, he had an amazing knowledge of science. He loved to have conversations with me about science and the cosmos, which is one of the main reasons I developed an interest in science.

I would also like to thank my mother for supporting me no matter what. I hope that my endeavours are successful so I can ensure she has a comfortable and joyful life.

Chapter 1

Introduction

1.1 Carbohydrates in Nature

Carbohydrates are the most abundant macromolecular structures found in nature. They are found in a surplus of places, such as bacterial cell surfaces and biofilms, plant cell walls, insect exoskeletons, and mammalian cell surfaces. Additionally, carbohydrates are the most complex macromolecular structure on the planet (Wang et al., 2007) and are organized into four chemical groups that differ in length: monosaccharides, disaccharides, oligosaccharides, and polysaccharides. These saccharides are linked together by glycosidic bonds synthesized by connecting the anomeric hydroxyl (bonded to the anomeric carbon) of the glycosyl donor to the alcoholic hydroxyl of the glycosyl acceptor. In a carbohydrate molecule, the anomeric carbon is the carbonyl carbon.

These saccharides can have various properties that result in a great diversity of polysaccharide structures. Monosaccharides can either be five membered rings (furanose) or six membered rings (pyranose). There are a number of monosaccharides that exist is nature such as glucose, fructose, rhamnose, xylose, arabinose, galactose, ribose, deoxyribose, mannose, and uronic acid. In addition to these monosaccharides having several hydroxyl groups that can form bonds with anomeric hydroxyls of other saccharides, the anomeric stereochemistry can be either α or β. In D sugars, the stereochemistry of the anomeric centre can be in the axial or equatorial position, resulting in an α and β bond, respectively. In L sugars, the stereochemistry for α and β bonds is reversed. Carbohydrates become increasingly complex as all of these different bonds and monosaccharides are combined to create a polysaccharide chain. Some polysaccharide chains consist of many different bonds and monosaccharide types. These chains can also have

branch points, adding further complexity to these chains. All of these properties contribute to the formation of complex and diverse structures and roles of carbohydrates.

1.1.2 Carbohydrates in plants and fungi

The prevalent contributors of terrestrial biomass are the carbohydrates found in the plant cell wall (Duchesne, & Larson, 1989). The plant cell wall consists mainly of cellulose, a linear a polymer of β-1,4 linked glucose. Cellulose exists in two forms in nature: amorphous and crystalline. Crystalline cellulose polymers associate with themselves through inter- and intramolecular hydrogen bonds and van der Waals forces to create cellulose fibrils and microfibrils. It is these hydrogen bonds that make crystalline cellulose recalcitrant to degradation by enzymes secreted from various organisms (Tomme et al., 1995). Additionally, the robust crystalline form of cellulose is particularly insoluble and responsible for the majority of the tensile strength of the cell wall. Amorphous forms of cellulose lack these hydrogen bonds and forces and are therefore, more readily degraded by secreted enzymes. The plant cell wall contains a number of other carbohydrates called hemicellulose, which includes xylan (β-1,4-linked xylose), mannan (β-1,4-linked mannose), mixed β-1,3-1,4-linked glucan, and β-1,3-linked glucan (Tomme et al., 1995). The plant cell wall also contains pectins, which consist of heteropolysaccharides composed of an α-1,4-D-galacturonic acid backbone with arabinose, galactose, and rhamnose substituents (Tomme et al., 1995). The plant cell wall is a major contributor to biomass and consists of a complex web formed by cellulose with regions attached to hemicellulose and pectin (Tomme et al., 1995). This complexity makes the cell wall recalcitrant to environmental stress and biological attack.

Another glucose polysaccharide that is a major contributor to terrestrial and aquatic biomass is β-1,3-glucan. This polymer has several sources, mainly from the plant and fungi kingdoms. β-1,3-Glucans have a variety of biological roles which are dependent on their structure, solubility in water, degree of polymerization, degree of branching, and conformation. For example, evidence has suggested that pachyman from *Poria cocos* is an insoluble β-1,3-glucan with a degree of polymerization of approximately 255 with a few β-1,6-linkages and three to six β-1,2-linked branch points (Hoffman, 1971). β-1,3-Glucan in fungi plays a structural role in the fungal cell wall along with β-1,6-glucans, mannoproteins, and chitin (Cheng, Hong, Liu, & Meng, 2009). In species of higher plants, a β-1,3-glucan called callose precedes the deposition of cellulose (Samuels, Giddings, & Staehelin, 1995) and plays an important role in the development and response to many biotic and abiotic stresses (Chen & Kim, 2009). The structure of callose consists of a linear glucose polymer constructed mostly with β-1,3-linkages and some β-1,6-linkages (Aspinall & Kessler, 1957). β-1,3-Glucan also exists in kingdoms other than plants and fungi (Staehelin & Hepler, 1996). For example laminarin plays a role as a storage polysaccharide in brown macroalgae, which belongs to the kingdom chromista (Beattie et al., 1961). Electrospray-ionisation-mass spectrometry analysis of laminarin has revealed that it is soluble, has a mean degree of polymerization of 25, can contain up to four β-1,6-linked branches per molecule, and has an average of 1.3 branches per molecule (Read, Currie, & Bacic, 1996). All β-1,3-glucans analyzed so far have been shown to be triple-helical structures via several methods, such as solid state ^{13}C nuclear magnetic resonance spectroscopy and x-ray fiber diffraction (Kanagawa et al., 2011).

1.2 Carbohydrate-active enzymes

The complexity and abundance of carbohydrates allows for a variety of biological roles. Thus, organisms from all biological kingdoms synthesize enzymes that function to degrade carbohydrates. These enzymes contribute to a number of biological functions, such as metabolism of glycoconjugates, oligosaccharides and polysaccharides, degradation of the host cell wall and storage polysaccharides in plant pathogens (Zerillo et al., 2013), and remodeling of the cell wall (Adams, 2004). Due to the variety of biological roles they are involved in, these enzymes are of particular interest to scientific researchers and are called carbohydrate-active enzymes. There are four types of carbohydrate-active enzymes: glycoside hydrolases (GHs), polysaccharide lyases, carbohydrate esterases, and glycosyl transferases. Since hydrolysis is the most common reaction involved in the degradation of carbohydrates, GHs are the most abundant and best characterized (Davies, Gloster, & Henrissat, 2005) type of catabolic carbohydrate-active enzyme (Henrissat & Davies, 1997).

Glycoside hydrolases are extremely important in biomass turnover through degradation of the plant, fungal, and algal cell wall carbohydrates; therefore, they are essential in numerous biological and industrial processes. These GHs are often secreted by microbes. However, these cell walls can be recalcitrant to enzymatic attack because they can contain complex polysaccharide structures such as that of cellulose, pectin, and hemicellulose in the plant cell wall. Microbes overcome this recalcitrance by secreting enzymes with different modes of action. Two main modes of action exist among carbohydrate active enzymes: Endo (endoglycanases) modes of action and exo modes of action (exoglycanases). An endoglycanase cleaves internal glycosidic bonds of a polysaccharide chain, whereas an exoglycanase cleaves terminal glycosidic bonds of a polysaccharide chain. Enzymes can sometimes possess combinations of these two

modes of action such as, endo-processive enzymes. These enzymes have an initial endo mode of action followed by a processive exo mode of action. Processive hydrolysis is the ability of an enzyme to release the product while remaining strongly bound to the remaining polysaccharide chain.

Microbes have also evolved enzymatic schemes in which enzymes with different modes of action cooperate in order to degrade the complex carbohydrates in the cell wall. An example of this is seen with cellobiohydrolase II and endoglucanase I and II from *Trichoderma reesei*, where these enzymes were shown to work together exhibiting a synergistic effect on the degradation of cellulose (Henrissat, Driguez, & Schulein, 1985). A cellobiohydrolase is an exoglycanase that hydrolyzes cellulose and releases two glucose residues linked by β-1,4-linkages (cellobiose) as a product. Crystalline cellulose has very few reducing and non-reducing ends available to exoglycanases. Endoglycanases can cleave the internal β-1,4-linked glycosidic bonds, resulting in more free non-reducing and reducing ends. The shorter polysaccharide chains created by the endoglycanases can then be degraded by exoglycanases from the newly exposed ends. This would then allow the endoglycanases more access to longer polysaccharide chains. In this way, endoglycanases and exoglycanases can cooperate to allow for a synergistic degradation of complex polysaccharides contained in cell walls. This result indicates one of the several examples of how enzymes with different modes of action can cooperate to degrade polysaccharides that are resistant to enzymatic attack.

1.2.1 Glycoside hydrolases

There are currently 133 glycoside hydrolase families that are classified based on amino acid sequence similarities; these are publicly accessible from the carbohydrate-active enzymes

(CaZY) database (www.cazy.org) (Cantarel et al., 2009a). These families are grouped into 14 clans, GH-A to GH-N based on the specific three-dimensional (3D) fold of the enzyme and the mechanism with which the enzyme degrades it substrate (Figure 1) (Henrissat & Davies, 1997). The 3D folds currently known are: $(\alpha/\alpha)_6$, 6-fold-β- propeller, $(\beta/\alpha)_8$, β-helix, β-jelly roll, α + β, 5-fold-β-propeller. The most common clan among all 133 families is GH-A, in which the core of the enzyme exhibits an $(\beta/\alpha)_8$ fold. This consists of eight β-strands connected by α-helices forming an α/β-barrel (Yu, et al, 2013). Some families possess the same 3D fold, yet they are grouped into different clans because GH families that have similar fold can have a significantly different catalytic mechanism.

1.2.2 Mechanism of glycoside hydrolases

The mechanism of glycosidic bond hydrolysis by enzymes has been studied thoroughly. Most glycoside hydrolases share a common mechanism where two amino acids are responsible for catalyzing the hydrolysis of the glycosidic bond. The amino acids involved in catalysis are commonly glutamate or aspartate (MacCarter & Withers, 1994). This catalysis can occur via two mechanisms: the inverting mechanism or the retaining mechanism (Rye & Withers, 2000). The inverting mechanism is a one-step process, whereas the retaining mechanism is a two step process. Additionally, the distance between the two catalytic residues is longer in the inverting mechanism compared to the retaining mechanism. In the inverting mechanism, there is sufficient distance between the two catalytic residues (~10 Å) to accommodate the carbohydrate substrate and a water molecule. In this mechanism, one residue acts as an acid to protonate the glycosidic bond being cleaved, thereby facilitating the departure of the aglycone leaving group, while the other acts as a base to deprotonate a water molecule and facilitate its attack on the anomeric centre (Figure 2A). This mechanism results in direct displacement of the aglycone

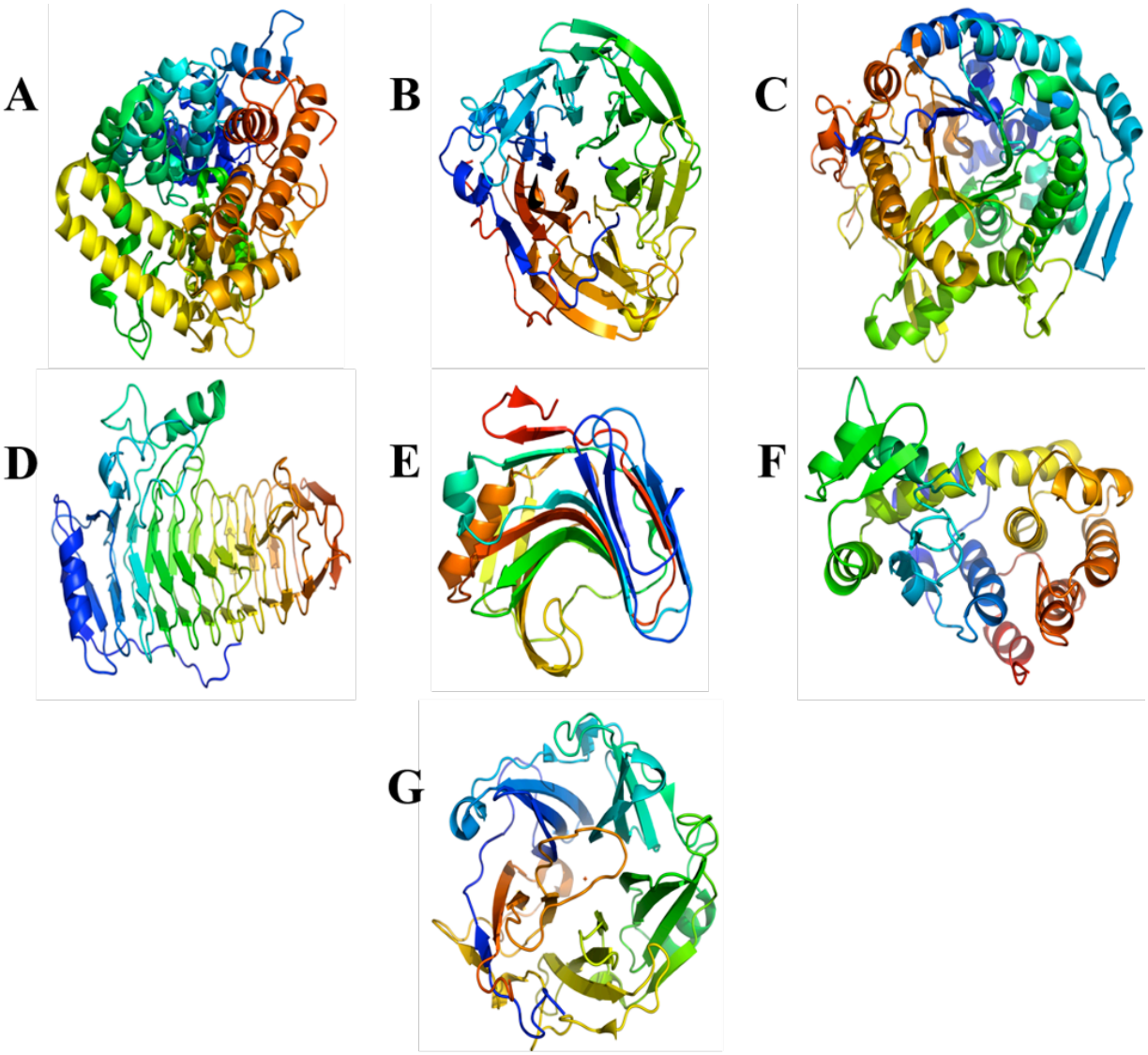

Figure 1: Folds observed among glycoside hydrolase families. (A) $(\alpha/\alpha)_6$ motif from *E. coli* trehelase (GH37) (PDB code 2JJB) (Cardona et al., 2009). (B) 6-fold-β-propeller from avian influenza N8 neuraminidase (GH34) (PDB code 2HT5) (Russell et al., 2006). (C) $(\beta/\alpha)_8$ motif from *Acidilobus saccharovorans* β-glycosidase (GH1) (PDB code 4HA3) (Trofimov et al., 2013). (D) β-helix motif from *Erwinia carotovora ssp. carotovora* polygalacturonase (GH28) (PDB code, 1BHE) (Pickersgill, Smith, Worboys, & Jenkins, 1998). (E) β-jellyroll motif from *Bacillus agaradhaerens* xylanase (GH11) (PDB code 1H4G) (Sabini et al., 1999). (F) α + β motif from *Bacillus circulans* chitosanase (GH46) (PDB code 1QGI) (Saito, Kita, Higuchi, & Chem, 1999). (G) 5-fold- β-propeller from *Podospora anserine* α-L-arabinofuranosidase (GH62) (PDB code 4N2Z) (Siguier et al., 2014).

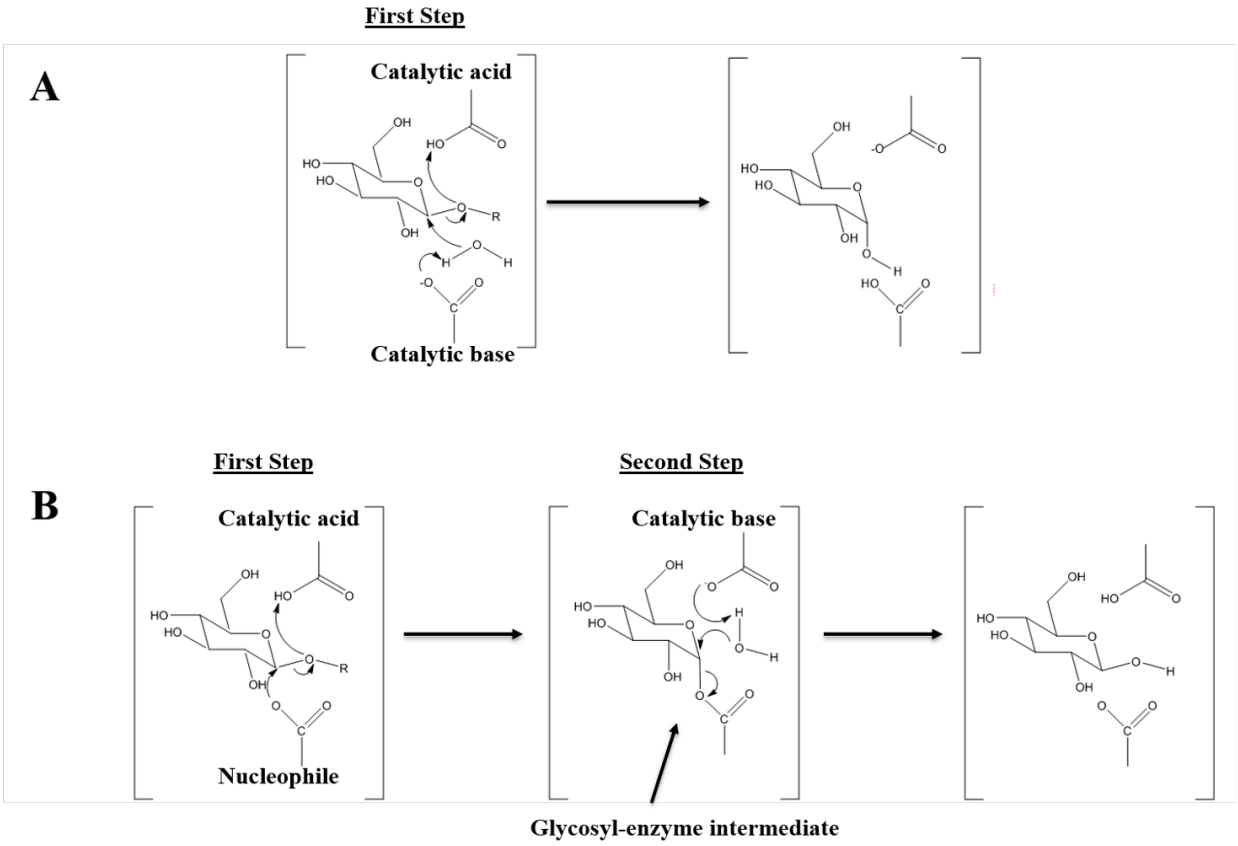

Figure 2: The two types of mechanisms of enzymatic β glycosidic bond hydrolysis: (A) The inverting mechanism, in which the acid catalyst protonates the glycosidic oxygen and the catalytic base deprotonates a water molecule and facilitates its attack on the anomeric centre in one step. This single step results in a product with reversed stereochemistry to the substrate. (B) The retaining mechanism, in which the acid catalyst protonates the glycosidic oxygen and the catalytic base acts as a nucleophile to attack the anomeric centre in the first step. This forms an enzyme sequestered intermediate. This intermediate is then hydrolyzed by a water molecule facilitated by the catalytic base and the amino acid that acts as a nucleophile from the first step, acts as the leaving group in the second step .

leaving group by water and inversion of the anomeric stereochemistry in one step (Figure 2A). In contrast, in the retaining mechanism, there is only enough space between the two catalytic residues (~5. Å) to accommodate the substrate. In the first step, one residue acts as an acid to protonate the glycosidic bond being cleaved, which again, facilitates the departure of the aglycone leaving group, while the other catalytic residue acts as a nucleophile to attack the anomeric centre (Figure 2B). This forms a covalent glycosyl enzyme intermediate (Figure 2B). The catalytic acid from the first step acts as a base in the second step to deprotonate a water molecule and facilitate its attack on the anomeric centre (Figure 2B). The catalytic residue forming the enzyme sequestered intermediate acts as the leaving group, releasing a hemiacetal product with retained anomeric stereochemistry (Figure 2B).

1.2.3 Active site topologies of glycoside hydrolases

The three dimensional space limitations of GHs allows them to accommodate appropriately shaped sugars and make carbohydrate degradation more efficient by restricting cleavage to either terminal (exoglycanase) or internal glycosidic bonds (endoglycanase).The active site topology dictates whether a GH is an exoglycanase or an endoglycanase There are three types of active site topologies: pocket or crater, cleft or groove, and tunnel (Davies & Henrissat, 1995). In the pocket or crater topology, one end of the polysaccharide chain (usually the non-reducing end) interacts with residues in the active site. For example, a glucoamylase from *Aspergillus awamori* has an overall doughnut shape where the hole of the doughnut contains hydrophobic residues that create a screen approximately halfway through, which forms two pockets on opposing sides of the screen. One of these pockets is an active site with residues

that interact with one end of a polysaccharide chain. More specifically, the residues Arg54 and Asp55, along with the two catalytic residues Glu179 and Glu180, interact with the non-reducing end of its substrate, starch (Figure 3A) (Aleshin et al., 1994). These active sites are large enough to accommodate a maximum of two sugar residues at one end of the polysaccharide chain. Because of this, these types of active site topologies are optimal for exo-glycanases that hydrolyze carbohydrates with many non-reducing ends such as a starch granule. However, they would not be efficient at hydrolyzing carbohydrates with very few or no non-reducing ends such as, cellulose.

In contrast, cleft or groove active sites are commonly trench-like allowing for a long polysaccharide chain to fit into them and interact with a number of residues along the trench. The catalytic residues then catalyze the hydrolysis of an internal glycosidic bond of the polysaccharide chain. For example, a GH from family 44 has a trench-like active site topology containing eight key residues that interact with cellohexaose, including the catalytic residues (E186 and E359) positioned between the third and fourth pyranose rings (Figure 3B) (Kitago et al., 2007). These types of topologies are optimal for endoglycanases that hydrolyze long polysaccharides to create shorter polysaccharides or oligosaccharides.

It is believed that the cleft or groove active site topology has evolved to create the tunnel active site topology by an amino acid loop folding over the trench-like active site. Tunnel active site topologies are optimal for processive hydrolysis of polysaccharide chains. The polysaccharide chain will fit somewhere inside this tunnel where a number of residues will interact with it to facilitate binding. The catalytic residues located somewhere inside the tunnel will catalyze the hydrolysis of a glycosidic bond, releasing the product. The polysaccharide chain

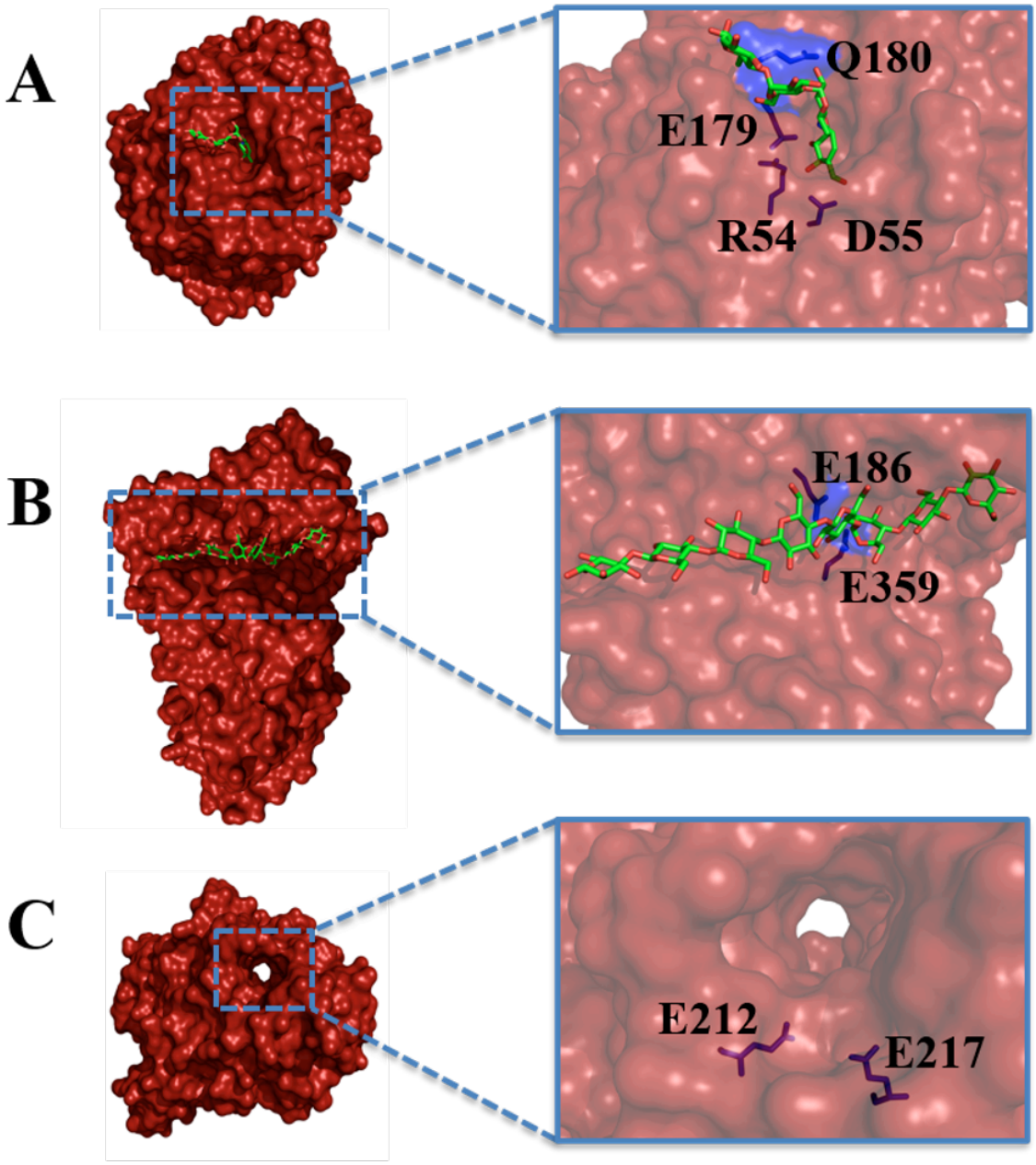

Figure 3: The three types of active site topologies in glycoside hydrolases: (A) Glucoamylase from *Asperigillus awamori* complexed with acarbose showing a typical pocket or crater active site topology of exo-glycanases and the residues interacting with the non-reducing end of the polysaccharide chain (E180 is mutated to Q180 in this structure) (GH15) (PDB code 1AGM) (Aleshin, Firsovs, & Honzatkoll, 1994). (B) Glycoside hydrolase from *Clostridium thermocellum* showing a typical trench-like active site topology of endo-glycanases and the catalytic residues which hydrolyze the internal glycosidic bonds of, cellooctaose (GH44) (PDB code 2EQD) (Kitago et al., 2007). (C) Cellobiohydrolase from *Trichoderma reesei* showing a typical tunnel shaped active site topology of processive glycoside hydrolases (GH7) (PDB code 1CEL) (Divne et al., 2014). The substrates are shown in green and the active site residues are shown in blue.

can then be moved in proximity to the catalytic residues and hydrolyzed again releasing a product identical to the first product. This process is repeated until the polysaccharide chain is broken down into simpler carbohydrates of uniform length. This type of topology is commonly found in cellobiohydrolases. For example, a cellobiohydrolase from *Trichoderma reesei* exhibits a tunnel shaped active site, which makes it more likely for a cellulose chain to remain attached after it is hydrolyzed and allow the enzyme to further hydrolyze the polysaccharide chain into cellobiose units (Figure 3C) (Divne et al., 2014).

1.2.4 β-1,3-Glucanases

A complex carbohydrate that plays an important role in storage, development, and maintenance of structure in several organisms is β-1,3-glucan. There are six known families of GHs that degrade β-1,3-glucans: glycoside hydrolase families 16, 55, 17, 64, 128 and 81. The fully characterized families of β-1,3-glucanases (16, 55, 17, and 64) exhibit a variety of modes of hydrolysis, mechanism, and structural folds (Figure 4A, B, D, and E). Even though there are members of the GH family 16 that hydrolyze β-1,3-glucan, there is a wide variety of substrate specificity within members of this family (Ilari et al., 2009). For example, a glycoside hydrolase belonging to family 16 from *Pseudoalteromonas carrageenovora* exhibits activity on κ-carrageenan both in solution and in solid state with an endoprocessive mechanism (Michel et al., 2001). All known members of this family that are active on β-1,3-glucan have an endo mode of action, retain the anomeric carbon configuration, exhibit a β-jelly roll fold (Fibriansah, Masuda, Koizumi, Nakamura, & Kumasaka, 2007), and belong to GH clan B. All members of family 55 GHs hydrolyze β-1,3-glucan, exhibit a β-helix type fold, and invert the anomeric configuration (Ishida et al., 2009). This family contains members that hydrolyze β-1,3-glucan with exo and endo-acting modes and has not yet been organized into a clan (Gajera, 2013; Ishida et al.,

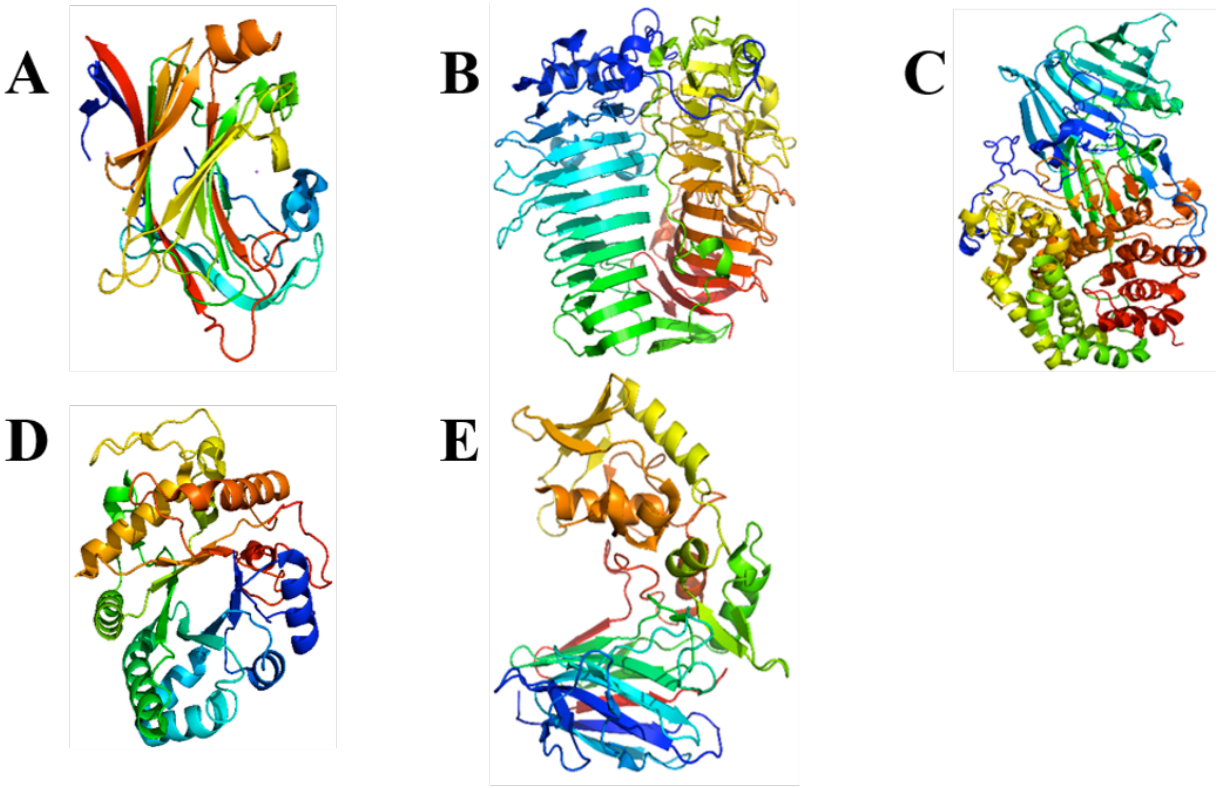

Figure 4: Crystal structures of β-1,3-glucanases from five different glycoside hydrolase families: (A) Endo-β-1,3-glucanase from *Pyrococcus furiosus* exhibiting a β-jelly roll fold (GH16) (PDB code 2VY0) (Fibriansah et al., 2007). (B) Exo-β -1,3-glucanase from *Phanerochaete chyrosporium* exhibiting a β-helix fold (GH55) (PDB code 3EQN) (Ishida et al., 2009). (C) β-1,3-glucanase from from *Rhizomucor miehei* exhibiting a core $(\alpha/\alpha)_6$ fold (GH81) (PDB code 4K35) (Zhou et al., 2013). (D) Endo-β-1,3-glucanase from *Hordeum vulgare* exhibiting an $(\beta/\alpha)_8$ fold (PDB code: 1GHS) (Varghese, et al., 1994). (E) Exo-β-1,3-glucanase from *Phanerochaete chyrosporium* exhibiting a β-barrel and mixed α/β domain (GH64) (PDB code 3DG0) (Wu et al., 2009b).

2009). GH family 55 members are common among species of filamentous fungi and it has been suggested that β-1,3-glucanases play a role in mobilization of cell wall β-glucans, particularly in yeast β-1,3-glucanases (Ishida et al., 2009). Thus, GHs from family 55 may be involved in morphogenetic changes such as, aggregation and mycelial strand formation in falimentous fungi (Stone, et al., 1992)**.** Many members of family 17 GHs hydrolyze β-1,3-glucan and have an endo-β-1,3-glucanase mode of action. All members of this family exhibit a $(\beta/\alpha)_8$ core enzyme fold (Varghese et al., 1994), belong to GH clan A, and use a retaining catalytic mechanism (Receveur-Bréchot et al., 2006; Rodríguez-Romero et al., 2014). GHs from family 17 are common among plant species where they have been implied in the protection of plants against potential pathogenic microbes by hydrolyzing the β-1,3-glucan commonly found in the fungal cell walls (Hrmova & Fincher, 1993). This family of endo-β-1,3-glucanases contain members that have been identified as allergens in pollen grain (Huecas, Villalba, & Rodríguez, 2001), natural rubber latex (Sunderasan et al., 1995), and fresh fruits (Wagner et al., 2004)**.** Many members of family 64 GHs hydrolyze β-1,3-glucan, have an endo-β-1,3-glucanase mode of action, and invert the anomeric carbon configuration. This family has a unique fold consisting of two domains: a β-barrel and a mixed α/β domain (Wu et al., 2009). This family does not currently belong to a particular clan.

The least characterized GH families are families 128 and 81. GH family 128 consists only of β-1,3-glucanases and is almost completely uncharacterized. A member of glycoside hydrolase from family 81, which are all β-1,3-glucanases, has been structurally characterized and revealed to have a core $(\alpha/\alpha)_6$ fold (Figure 4C). Researchers have revealed that this family exhibits an endo mode of action; however, there are no structural studies to support this.

Structural characterization of these two families could elucidate certain characteristics needed for a complete understanding of their molecular details such as the mode of action, mechanism, and structural fold. Glycoside hydrolase family 81 has been shown to invert the anomeric configuration (McGrath & Wilson, 2006). A study of a member of GH family 81 from *Saccharomyces cerevasiae* showed activity on laminarin and pachyman, which are soluble and insoluble β-1,3-glucan, respectively (Martín-Cuadrado et al., 2008). Interestingly, a family 81 glycoside hydrolase from *Glycine max* is a β-glucan elicitor receptor with no activity; however, it does bind the β-1,3-glucan, laminarin (Fliegmann, Mithofer, Wanner, & Ebel, 2004). These findings are crucial in characterizing GH81s because they provide information on the mode of action, mechanism, and structural fold. However, a more complete understanding would be elucidated if mode of action and structural characterization were done in a single study, on a single member. This would not only consolidate the previous findings for GH81s mode of action and structure, but also could explain the molecular details of GH81s mode of action through structural insight. Thus, further structural and biochemical analysis of a member from this family will shed light on β-1,3-glucan degradation. A superb candidate for further characterization of this family of GHs is a multimodular polypeptide consisting of 1020 amino acids encoded by an open reading frame in *B. halodurans*, which has a GH from family 81 as a catalytic module.

1.3 Multimodularity of glycoside hydrolases

Even though GHs are capable of degrading polysaccharides on their own, many GHs have non-catalytic modules attached via linker peptides. This is known as a multimodular GH, which consists of a catalytic module and one or more non-catalytic modules. A module is an amino acid sequence contained within a larger sequence that folds separately and has an independent

function. These modules can often retain stability, fold, and function despite being expressed independently. The limited papain digestion of cellobiohydrolase I and II from *Trichoderma reesei* was the first piece of evidence indicating GHs harbour modules with independent functions (Tilbeurgh, Tomme, & Claeyssens, 1986; Tomme et al., 1988). This experiment showed a cleavage of a polypeptide by papain digestion resulting in two polypeptides corresponding to N and C terminal domains. The specific activity of the N-terminal domain on cellulose was 50 % less compared to the non-truncated polypeptide (Tilbeurgh et al., 1986). The C-terminal domain displayed binding to crystalline cellulose; however no activity was observed indicating that the C-terminal domain had a function separate from the catalytic module. These results showed that these cellobiohydrolases have an N-terminal catalytic domain and a C-terminal carbohydrate binding domain that improves the overall efficiency of the enzyme. Bioinformatics has identified other modules that have sequence and secondary structure similarities to other protein motifs such as, dockerins, cohesins, and FN3 motifs. The most prominent are those that bind carbohydrates called carbohydrate-binding modules (CBMs). GHs can have multiple non-catalytic modules appended to them with each having their own function. For example, SpuA, a GH from family 13, consists of a catalytic module, two CBMs from family 41, a linker region, and two Ig-like domain modules (Figure 5) (Lammerts van Bueren, et al., 2011). Studying these multimodular enzymes can be difficult because they are not easy to recombinantly express. Therefore, scientists commonly dissect the modules and study them separately in order to piece together the overall function of the protein.

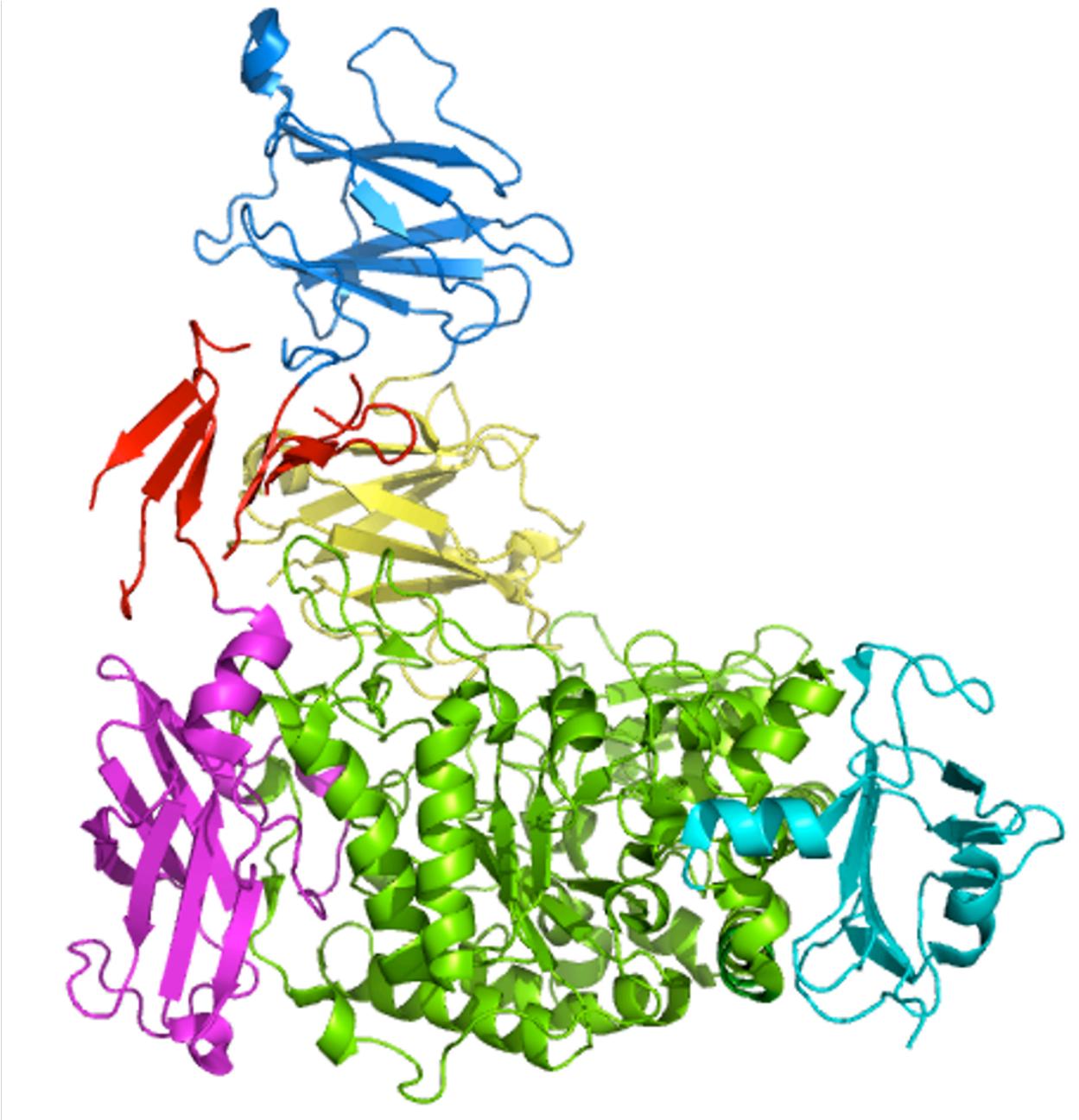

Figure 5: Multimodularity of a glycoside hydrolase shown by the virulence factor SpuA from *Streptococcus pneumoniae* (PDB code 2YA1) (Lammerts van Bueren, 2011). Two CBMs from family 41 shown in marine blue and yellow, a linker region shown in red, an Ig-like domain shown in purple, the catalytic module shown in green, and another Ig-like domain shown in cyan.

1.3.1 Carbohydrate binding modules

The most common type of non-catalytic modules found in GHs are CBMs (Boraston, Bolam, Gilbert, & Davies, 2004). CBMs increase the specific activity of the enzyme by binding the carbohydrate substrate. Biochemical analysis of CBMs resulted in a number of biotechnological applications and has added an immense amount of knowledge to protein-carbohydrate interactions. For example, CBMs from multimodular GHs belonging to families 2a, 6, and 29 have been used for analysis and detection of polysaccharides in plant cell walls, and this has ultimately helped with understanding the developmental and functional features of plant cell walls (McCartney, Gilbert, Bolam, Boraston, & Knox, 2004). Another biotechnological concern is the effect of glycosylation on the behavior of bacterial polypeptides that are not glycosylated in their native organism. This is of particular concern when utilizing cellulose binding CBMs as affinity tags for purification of a fused polypeptide companion expressed in a host that N-glycosylates. This is because N-glycosylation can interfere with binding of the CBM alone or when fused to a partner as demonstrated with Xyn10ACBM2a of *Cellulomonas fimi* (Boraston, McLean, Guarna, Amandaron-Akow, & Kilburn, 2001). Fortunately, one study has demonstrated that CBMs which lack the N-glycosylation sites such as, Cel5ACBM2a from *Cellulomas fimi*, can be used as effective affinity tags for purification of fused polypeptide companions after being expressed in a host that N-glycosylates (Boraston et al., 2001). Also, a variety of properties render biodegradable starch- and cellulose-based polymers appropriate for utilization in a diverse array of biomedical applications. Fusing of a cellulose-binding protein to a starch-binding domain has demonstrated that biotechnology can be used to create novel proteins with cross-bridging ability in model systems composed of insoluble or soluble starch

and cellulose (Levy, Paldi, & Shoseyov, 2004). These novel proteins could be used to increase the variety of physical and biological activities of starch- and cellulose-based polymers by designing chimeric polysaccharide-based biomaterials. Thus, characterization of CBMs is important in order to exploit these binding properties.

There are three roles of CBMs known in the degradation of carbohydrates: concentrating the catalytic module of the enzyme in close vicinity to the substrate surfaces, targeting the enzyme to its substrate, and anchoring the enzyme onto the surface of bacterial cells. By concentrating enzymes in proximity to their target substrates, CBMs allow for more efficient degradation of complex carbohydrates that exist in nature. The targeting role of CBMs allows for recognition of specific polysaccharides substructures such as, the insoluble (crystalline) and soluble (amorphous) regions in cellulose (Liu et al., 2011). The anchoring role of CBMs is used to keep a secreted glycoside hydrolase attached to bacterial cell wall. For example, evidence has suggested that a CBM35 appended to the exo-β-D-glucosaminidase CsxA in *Amycolatopsis orientalis*, binds either GlcA or or uronic acid on the bacterial cell wall, thereby tethering the enzyme to the surface of *A. orientalis* (Montanier et al., 2009).

Multimodular GHs can contain multiple CBMs from the same or different families appended to a single catalytic module. These may be present in tandem or separated by other modules. CBMs attached to each other have been shown to have an increased affinity for the substrate compared to single CBMs (Boraston et al., 2002, 2006) The first example of this was shown with two recombinantly fused CBMs, which had an increased affinity for cellulose compared to the single domain (Linder et al., 1996). In contrast to GHs having multiple CBMs that target one substrate, GHs may also have one CBM that targets multiple substrates. These

CBMs are capable of specifically binding a substrate with high affinity and bind to other sugars with a lower affinity (Charnock et al., 2002).

The first CBMs discovered were classified as cellulose binding domains because they bound cellulose, but some researchers began to discover that the ligands for these modules extended beyond cellulose and these domains were later called CBMs. These modules are now classified into 69 different families based on amino acid sequence similarity and are found on CaZY (www.cazy.org) (Cantarel et al., 2009). A new family of CBM is created when the carbohydrate binding activity of a putative CBM with no amino acid sequence similarity to other CBM families is demonstrated, and subsequent members are added to this family based on amino acid sequence similarity. This classification system of CBMs allows for organization of CBMs according to primary structure; however, it does not necessarily organize them based on carbohydrate-binding specificity. The binding specificity is determined by the differences within loops and side chains and thus, particular members of a specific family may or may not have the same binding specificity.

1.3.2 Structural classification of Carbohydrate-binding modules

CBMs have been grouped into fold families based on their 3D fold. There are seven fold families: β-sandwich, cysteine Knot, β-trefoil, OB fold, hevein fold, unique, and unique, contains hevein-like fold (Boraston et al., 2004). The β-sandwich is the most common and consists of two sandwiched β-sheets, each consisting of 3-6 antiparallel β-strands. The binding site in this fold is most often found on the concave surface of one of the β-sheets; however, some CBMs with this fold have binding sites on the convex side or at the tip of the β-sandwich on the loops that connect the two β-sheets. Most CBMs with β-sandwich type folds contain a metal ion that helps

maintain the structure of the protein and there are also CBMs from this fold family that contain metal ions that promote binding to their ligands. For example, CBM35 from a rhamnogalauronan acetyl esterase in *Clostridium thermocellum* requires a calcium ion to bind its ligand, 4,5-GalAα1,4Gal (Figure 6) (Montanier et al., 2009). This fold is very adaptable to carbohydrate-binding, which makes it a superb platform for protein-carbohydrate interactions with a variety of polysaccharides. CBMs are also classified based on their binding site topology, which reflects the macromolecular structure of their substrates. There are three binding types of CBMs: Type A, B, and C (Figure 7). Type A CBMs exhibit a planar hydrophobic ligand-binding surface that interacts with insoluble, highly crystalline polysaccharides such as, cellulose or chitin (Gilbert, Knox, & Boraston, 2013). The planar architectures of these binding sites are thought to be complementary to the flat surfaces presented by cellulose or chitin crystals (Figure 7A). Type A CBMs generally have residues with aromatic side chains that are oriented to form a platform that facilitates hydrophobic interactions with cellulose by stacking onto the pyranose rings of glucose and have little or no affinity for soluble carbohydrates (Tormo et al., 1996). Hydrogen bonding is thought to be unimportant for Type A CBM ligand binding since, mutation of residues involved in hydrogen bonding does not affect the affinity of the CBM for its substrate (McLean et al., 2000). Type B CBMs are most common due to their presence in plant enzymes that degrade the cell wall. Type B CBMs are classified as those that bind internally on glycan chains (endo-type) (Figure 7B) and Type C CBMs are classified as those that bind the ends of glycan chains (exo-type) (Figure 7C) (Gilbert, Knox, & Boraston, 2013). This classification of CBM binding types has allowed for unambiguous organization of all three types of CBMs based on their modes of binding.

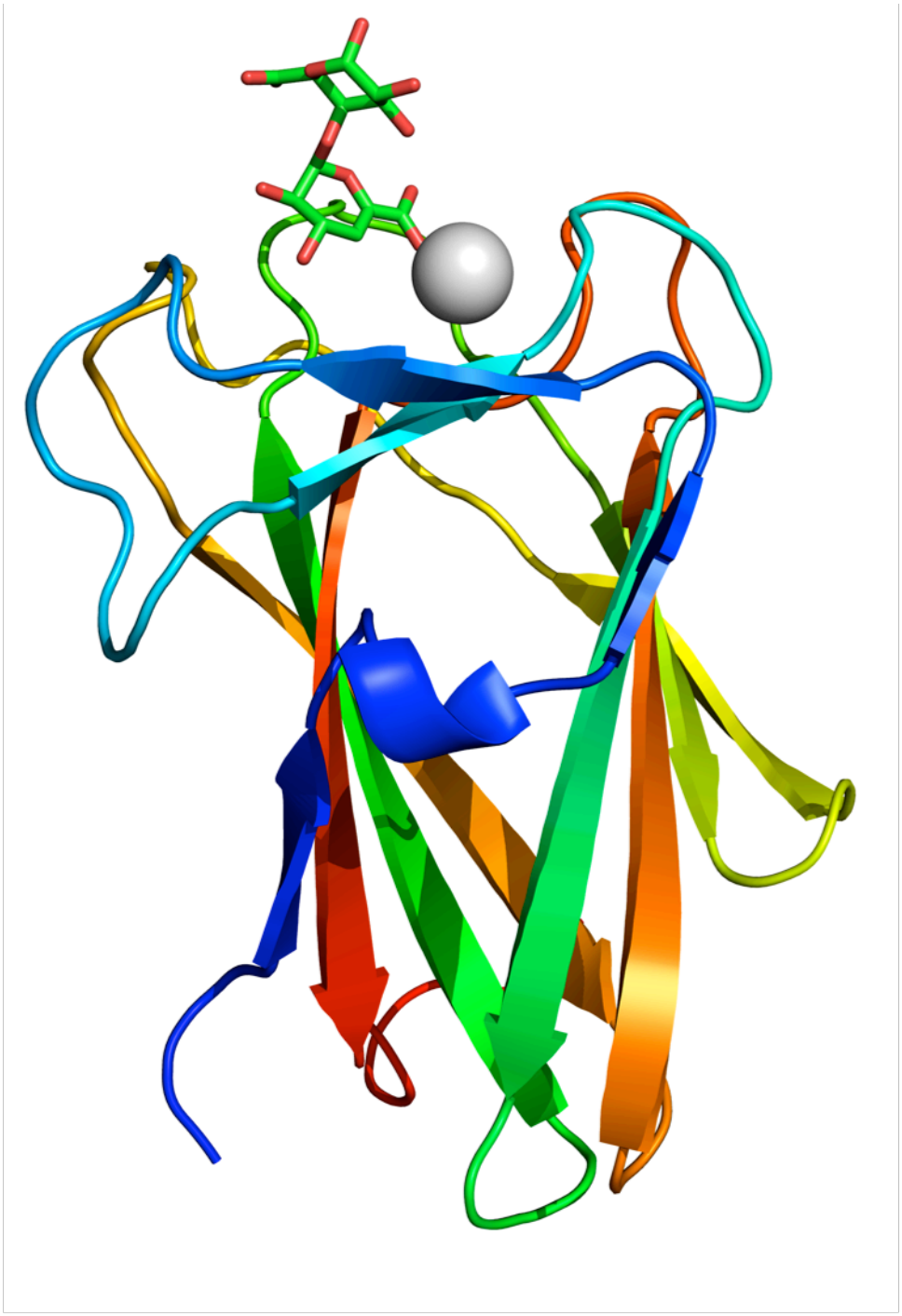

Figure 6: CBM35 from Clostridium thermocellum showing the use of a calcium ion in binding 4,5-Gal-Aα-1,4Gal (PDB code: 2WZ8) (Correia et al., 2010). The ligand, 4,5-GalAα1,4Gal, is shown in green and the calcium ion is represented as a grey sphere.

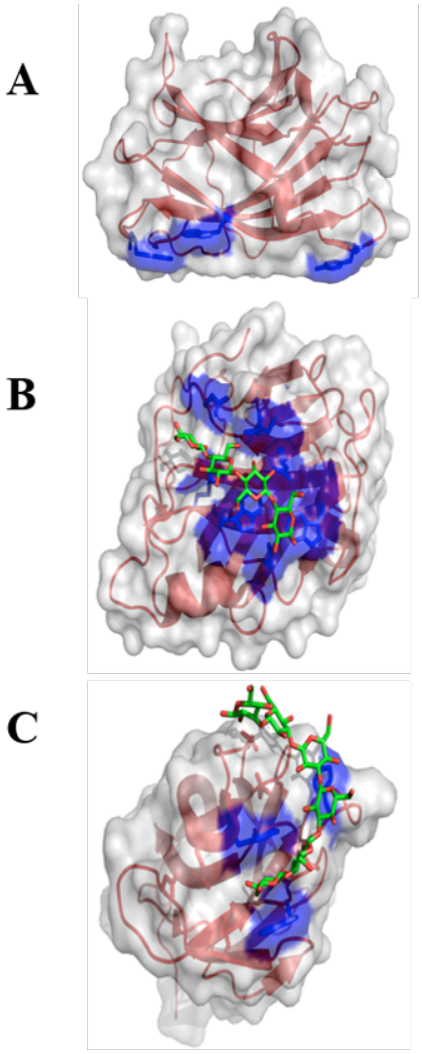

Figure 7: Three types of CBM binding site topologies. (A) Type A CBMs use residues in a planar architechture for binding to crystalline ligands: CBM3 from Clostridium thermocellum (PDB code 1NBC) (Tormo et al., 1996). (B) Type B CBMs have long binding site grooves for binding internally on long polysaccharide chains (endo-type): CBM17 from Clostridium cellulovorans (PDB code 1J84) (Notenboom et al., 2001). (C) Type C CBMs have shallow binding pockets for interaction with the ends of polysaccharide chains (exo-type): CBM6 from *B. halodurans* (PDB code 1EN2) (van Bueren, Morland, Gilbert, & Boraston, 2005).

1.4 Hypothesis, objectives, and significance

The details of how multimodular enzymes function is more complex compared to enzymes consisting of one module and the molecular details of how each module functions has become of interest to researchers. Only one multimodular β-1,3-glucanase has been characterized so far. This polypeptide is from *Schizosaccharomyces pombe* and consists of a putative endo-β-1,3-glucanase appended to two CBMs from family 52, which bind β-1,3-glucan and assist in cell separation (Martin-Cuadardo, 2003). Therefore, how CBMs contribute to the degradation of β-1,3-glucan by multimodular enzymes is not well understood in bacteria, which degrade β-1,3-glucan for carbon utilization rather than cell separation. In order to understand degradation of β-1,3-glucan by multimodular enzymes in bacteria, we have chosen to study an open reading frame (BH0236) in the genome of the alkalophilic bacterium, *B. halodurans*. BH0236 encodes for a 1020 amino acid enzyme, called *Bh*Lam81, which contains structurally uncharacterized modules. According to www.cazy.org (Cantarel et al., 2009), the encoded protein consists of a catalytic module and two CBMs (Figure 8). At the N-terminus lies the catalytic module (a GH from family 81) consisting of ~750 amino acids. Studies have shown that GHs from family 81 in *S. cerevisiae* (*Sc*Eng2) hydrolyze β-1,3-glucan in an endo acting manner (Martín-Cuadrado et al., 2008). The C-terminus of this polypeptide contains a module consisting of ~100 amino acids that has sequence similarity to CBM family 56 (*Bh*CBM56). Prior to this work, a member of CBM family 56 has been shown to bind the insoluble β-1,3-glucan, pachyman, in *B. circulans (Bc*CBM56) (Yamamoto, Ezure, Watanabe, Tanaka, & Aono, 1998). Between the C-terminus and N-terminus modules, lies a module consisting of ~140 amino acids that belongs to CBM family 6 (*Bh*CBM6). *Bh*CBM6 has been shown to bind the non-reducing end of soluble β-1,3-glucooligosaccharides via ultraviolet difference spectra, isothermal titration calorimetry, and x-

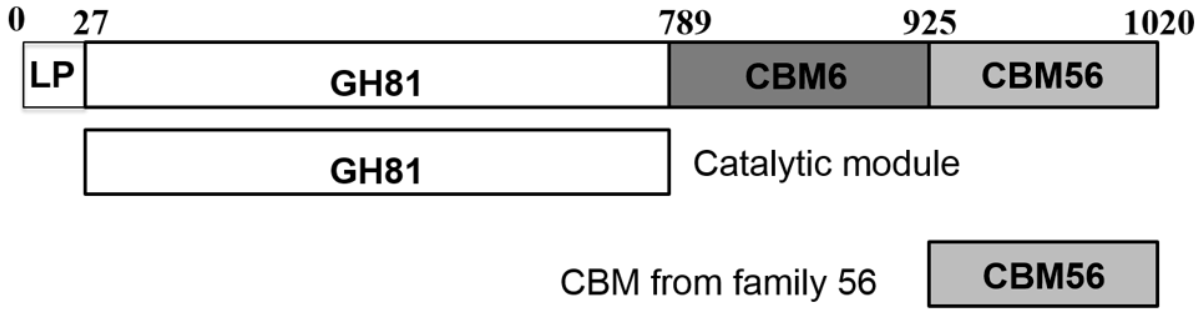

Figure 8: Multimodular schematic of the 1020 amino acid open reading frame from *B. halodurans* (BH0236). Above the schematic are amino acid numbers depicting the module boundaries and the leader peptide (LP). The constructs of the separate modules used in this study are also shown.

ray crystallography (Lammerts van Bueren, Morland, Gilbert, & Boraston, 2005). The binding of *Bh*CBM6 was tested against the insoluble β-1,3-glucan, pachyman, using standard depletion binding analysis; however, no binding was found (Lammerts van Bueren et al., 2005).

Based on the fact that a member of CBM family 56 and members of GH family 81 have been found to bind insoluble β-1,3-glucan and hydrolyze β-1,3-glucan, respectively; we hypothesize that BhCBM56 is an insoluble β-1,3-glucan binding module and BhGH81 is an enzyme that hydrolyzes β-1,3-glucan. The overall objective for studying this multimodular polypeptide was to characterize the molecular details of *Bh*CBM56's, *Bh*CBM6`s, and *Bh*GH81's function and present a model of how these modular components allow *Bh*Lam81 to more efficiently recognize and degrade its substrate in the environment. For this particular study, there were two objectives:

1. To determine the substrate specificity of *Bh*GH81 and use x-ray crystallography to provide insight into the molecular details of hydrolysis of *Bh*GH81.

2. To determine the substrate-specificity of *Bh*CBM56 and use x-ray crystallography to provide insight into the molecular details of binding by *Bh*CBM56.

The wide variety of molecular mechanisms and modes of action exhibited by all the GH families contributes to a significant portion of the complex carbohydrate biomass conversion. The majority of previous studies of multimodular proteins concentrate on characterizing catalytic modules; however, dissecting multimodular proteins and studying the modules separately will

allow for a more refined understanding of how multimodular GHs recognize and hydrolyze their substrates. Thus, by studying multimodular GHs, we gain insight into the degradation of β-1,3-glucan in the environment. Furthermore, a greater understanding gained by studying the degradation of β-1,3-glucan by these multimodular enzymes has potential applications such as, improving industrial processes for biofuel production or developing antifungal agents.

Chapter 2

2.0 Materials and Methods

2.1 Carbohydrates and polysaccharides

Laminarin, β-1,3-glucooligosaccharides, and curdlan were obtained from Megazyme Ireland Ltd. (Bray Co., Wicklow, Ireland). Pachyman was obtained from Calbiochem Co.

2.2 Cloning, expression, and purification of *Bh*GH81 and *Bh*CBM6

The DNA fragment (nucleotides encoding for *Bh*GH81 and *Bh*CBM56) were amplified by polymerase chain reaction from the genomic DNA of *B. halodurans* strain C-125. The 5' oligonucleotide and 3' oligonucleotide primers used for *Bh*GH81 were 5'GGCGGATCTAGACATGCGGTGAGCGTCGGG3' (XbaI cut site is underlined) and 5' TCCGCCCTCGAG**TTA**CGGCTCTGAAGGATCTGG 3' (XhoI cut site is underlined and the stop codon is in bold), respectively. These allowed for the amplification of the nucleotide sequence encoding amino acids 28–789 of *Bh*Lam81. The 5' oligonucleotide and 3' oligonucleotide primers used for *Bh*CBM56 were 5' GGCGGAGCTAGCCAAGGGAATGGCGAT 3' (NheI cut site is underlined) and 5' TCCGCCCTCGAG**TTA**ACGCGAATACGTAAA 3' (XhoI cut site is underlined and the stop codon shown in bold). These allowed for the amplification of the nucleotide sequence encoding amino acids 926 – 1020 of *Bh*Lam81. The *Bh*GH81 and *Bh*CBM56 amplification products were cut with the restriction enzymes XbaI/XhoI and NheI/XhoI, respectively. Digested *Bh*GH81 and *Bh*CBM56 were ligated into pET28a(+) that was previously digested with the restriction enzymes NheI and XhoI to give pET28-*Bh*GH81 and pET28-*Bh*CBM56 constructs. Bidirectional DNA sequencing confirmed the fidelity of all cloned inserts. Inserts were transformed into *E.coli*

BL21 (DE3) for protein expression. All expressed polypeptides contained a His_6 tag attached to the N-terminus by a thrombin cleavage site.

For each construct, four litres of Luria Bertani (LB) medium supplemented with 50 μg/ml of kanamycin were inoculated with *E. coli* BL21 bacterial cells containing the appropriate expression plasmid and were grown at 37 ˚C with shaking until they reached an optical density of 0.8 – 0.9 at 600 nm. Protein expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside to a final concentration of 0.5 mM and incubation was carried out overnight at 16 ˚C with shaking. Cells were collected by centrifugation and subjected to chemical lysis. Both proteins were purified from cell-free extracts by using Ni^{2+} immobilized metal ion affinity chromatography (IMAC). For both polypeptides, the His_6 tag was removed by thrombin cleavage, which consisted of an incubation overnight at 18 ˚C. A secondary purification was done for both polypeptides using size exclusion chromatography (SEC). A 2 mL sample of *Bh*GH81 or *Bh*CBM56 was injected into a HiPrep 16/60 Sephacryl S-200 HR column (GE Healthcare) and a HiPrep 16/60 Sephacryl S-100 HR column, respectively. Both polypeptides were eluted with 20 mM Tris-HCl, pH 8.0. Purified *Bh*GH81 and *Bh*CBM56 polypeptides were concentrated and exchanged into 20 mM Tris-HCl, pH 8.0, in a stirred ultrafiltration unit (Amicon, Beverly, MA) on 10 and 5 kilodaltons (KDa) molecular weight cut-off membranes, respectively. Purity assessed by sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) was > 95%.

Selenomethionine-labeled *Bh*GH81 was produced using *E. coli* BL21 (DE3). Four litres of SelenoMet Medium Base (Molecular Dimensions Ltd.) supplemented with SelenoMet Nutrient Mix (Molecular Dimensions Ltd.) and L-selenomethionine (40 mg/L) were inoculated with *E. coli* BL21 colonies containing the pET28-*Bh*GH81 construct taken from an LB agar

plate. These cultures were grown, induced, harvested; and the protein was purified and concentrated using the same method described previously.

2.3 β-1,3-glucooligosaccharide product analysis

Activity of *Bh*GH81 on laminarin was first demonstrated utilizing thin-layer chromatography (TLC). Digests were set up with enzyme and substrate concentrations of 250 nM and 1.0 mM, respectively. All digests were done in Tris-HCl , pH 8.0, and incubated at 37 ˚C overnight. Aliquots of all digests were removed and analyzed for the production of β-1,3-glucooligosaccharide products by loading them onto a TLC silica gel plate and subjected to 1-proponol/ethanol/water (2:1:1, v/v/v) as a solvent for two hours. Plates were dried and developed using orcinol sulfuric acid reagent (sulfuric acid/ethanol, 3:70 [v/v], orcinol 1%). In order to visualize the sugars, the silica plates were incubated at 90 ˚C for fifteen minutes.

Additionally, activity of *Bh*GH81 on β-1,3-glucooligosaccharides was demonstrated using high performance anion exchange chromatography. High performance anion exchange chromatography runs were performed using a DIONEX ICS3000 instrument equipped with a CarboPac PA-200 column and pulsed amperometric detection. Samples with a volume of twenty microlitre were loaded onto the column and eluted with a linear gradient of 0 – 300 mM sodium acetate in 100 mM NaOH over 40 min at flow rate of 0.5 ml/min. Carbohydrates incubated in parallel with the samples were used as control standards. The digest samples were β-1,3-glucooligosaccharides incubated with 250 nM of *Bh*GH81 at 37 ˚C for two hours. The reactions were stopped by the addition of one reaction volume of 0.2 M NaOH and centrifuged at 13000 rpm for 10 minutes at 4 ˚C prior to high performance anion exchange chromatography analysis.

Comparison of the unique peak elution positions with those of the standards identified the products released upon incubation of β-1,3-glucooligosaccharides with *Bh*GH81.

2.4 SDS-PAGE depletion assay

One-hundred micrograms of *Bh*CBM56 or the *Ha*Kdgf negative control was mixed with 0.5 - 10 mg of curdlan or pachyman in 20 mM Tris-HCl ,pH 8.0, to a final volume of 0.4 mL and incubated at 4 ˚C overnight. Samples were rotated end-over-end for 16 hours to allow the adsorption reaction to equilibrate and were centrifuged at 4 ˚C at 13000 rpm for 10 min to separate the pachyman from the unbound *Bh*CBM56. After removal of the supernatant, the pachyman pellets were washed three times by resuspension in 1 ml of 20 mM Tris-HCl, pH 8.0, and centrifugation at 13000 rpm for 10 min at 4 ˚C. SDS-PAGE loading buffer was added to each sample and incubated for 5 minutes at 90 ˚C. The adsorption of *Bh*CBM56 was then analyzed by loading 10 uL of each sample into an 15% SDS-PAGE gel.

2.5 Affinity electrophoresis

Qualitative binding of *Bh*CBM56 to soluble β-1,3-glucan (laminarin) was examined by affinity electrophoresis in 9% (w/v) polyacrylamide native gels. One gel had a laminarin incorporated into the polyacrylamide gel matrix at a concentration of 0.25%, while another gel did not contain laminarin. Electrophoresis was carried out for 4 hours at 4 ˚C, pH 8.8, at a constant voltage of 150 V. After electrophoresis the gels were stained with Coomassie Brilliant Blue R-250 for protein detection. The adsorption of *Bh*CBM56 to laminarin compared to the negative control (*Ha*Kdgf) was determined by measuring the migration distances of *Bh*CBM56 run with laminarin compared to *Bh*CBM56 run without laminarin relative to the bromophenol blue front:

r/R_0. Where r is the relative migration distance of *Bh*CBM56 in the presence of the polysaccharide in the gel matrix and R_0 is the relative migration distance of *Bh*CBM56 in the absence of the polysaccharide.

2.6 Binding isotherms

All pachyman used for binding isotherms was washed with 1 L of distilled water in a vacuum filter, dried, and lyophilized overnight prior to incubation with *Bh*CBM56. The amount of *Bh*CBM56 bound to pachyman was determined as follows: Thirteen *Bh*CBM56 samples were incubated with or without 1.5 mg of pachyman in a 1.5mL centrifuge tube with a range of concentrations of *Bh*CBM56 (2.4 uM - 34 uM) in 20 mM Tris-HCl, pH 8.0, to a total volume of 0.6 mL. Thirteen negative control samples were conducted at the same concentrations of *Bh*CBM56 without pachyman in parallel. Samples were incubated with end-over-end mixing at 4 ˚C for 2 hours. The pachyman was pelleted by centrifugation at 13000 rpm for 10 min at 4 ˚C. The concentration of bound *Bh*CBM56 was calculated from the difference between total and free *Bh*CBM56 concentrations, which were determined spectrophotometrically ($A_{280\ nm}$) from the supernatants of tubes without and with pachyman, respectively. Binding parameters (K_d and Bmax) were determined from the depletion isotherms (Plot of bound [B] versus free [F]) after fitting the raw data to a one-site binding model using non-linear regression.

2.7 Crystallization of *Bh*GH81 and *Bh*CBM56

Crystals of selenomethionine-labeled *Bh*GH81 (72 mg/mL) were obtained in 1.4 M NaH_2PO_4/0.35 M K_2HPO_4 and 0.0750 M Na_2HPO_4: citric acid, pH 3.2, at 18 ˚C using the vapor-phase diffusion technique from hanging drops. Crystals of *Bh*CBM56 (20 mg/mL) were obtained

in 21 % polyethylene glycol 3350 and 0.1 M Bis-tris/HCl, pH 5.5, at 18 ˚C using the vapor-phase diffusion technique from hanging drops

2.8 Data collection, structure solution, and refinement

CCP4, an integrated suite of programs that allows researchers to determine macromolecular structures, (Bailey, 1994) was used to perform all computing unless stated otherwise. Crystals of selenomethionine *Bh*GH81 were snap cooled in liquid nitrogen in the crystallization condition supplemented with 25% ethylene glycol as a cryo-protectant. Single-wavelength anomalous dispersion (SAD) data of *Bh*GH81 was collected on the beam line 08ID-1 at the Canadian Light Source (CLS, Saskatoon, Canada). The SAD dataset was processed using CLS data analysis software and SCALA (Leslie, 2006).

Crystals of *Bh*CBM56 were soaked in the crystallization condition supplemented with 25 % ethylene glycol as a cryo-protectant and 1 M KBr prior to being snap cooled in liquid nitrogen. A multi-wavelength (peak, 0.91962 Å, inflection, 0.92002 Å, remote, 0.90496 Å) anomalous dispersion (MAD) dataset of *Bh*CBM56 was collected on beam line BL11-1 at the Stanford Synchrotron Radiation Lightsource (SSRL). *Bh*CBM56 was processed using MOSFILM (Isorna et al., 2007) and AIMLESS (Evans, 2011).

Heavy atom substructure determination, phasing, and density modification was performed with autoSHARP (Vonrhein, Blanc, Roversi, Bricogne, & 2007) for both *Bh*GH81 and *Bh*CBM56. Sixteen selenium positions found corresponding to sixteen methionine residues present in one *Bh*GH81 molecule were used for phasing with the full 2.50 Å resolution data set. A total of five bromine positions present in the two *Bh*CBM56 molecules in the asymmetric unit were used for phasing with the full 1.70 Å resolution dataset. The phases resulted in density improvement,

which were of sufficient quality for ARP-WARP to build a virtually complete model of both *Bh*CBM56 and *Bh*GH81. The remainder of the model building of both structures was done using COOT (Emsley & Cowtan, 2004), and refinement of atomic coordinates was done using REFMAC (Murshudov et al., 1997). Water molecules were added using COOT and checked manually after refinement. Data collection, processing statistics, and final model statistics are given in Table 1 and 2 for *Bh*GH81 and *Bh*CBM56, respectively. Stereochemical analysis of both the refined structures were performed with PROCHECK and SFCHECK in CCP4 (Vaguine, Richelle, & Wodak, 1999; Laskowski, MacArthur, Moss, & Thornton, 1993). The Ramachandran plot showed adequate stereochemistry with 100 % and 99.4 % of the residues in the favoured conformations and no residues modeled in disallowed orientations for *Bh*CBM56 and *Bh*GH81, respectively.

Chapter 3

3.0 Results

3.1 Cloning and purification of *Bh*CBM56 and *Bh*GH81

The *Bh*CBM56 and *Bh*GH81 modules from *Bh*Lam81 were amplified from *Bacillus halodurans* genomic DNA and cloned separately into a pET28a vector containing an N-terminal His_6-tag. Both these constructs, which were transformed into *E. coli* and recombinantly produced, required two purification steps to obtain adequate purity for crystallization. The first purification step, immobilized metal affinity chromatography (IMAC), revealed polypeptides of approximately 11 and 81 kDa for *Bh*CBM56 and *Bh*GH81, respectively (Figure 9A). The second purification step, size exclusion chromatography (SEC), showed that *Bh*CBM56 and *Bh*GH81 eluted after ~87 mL and 55 mL, Tris-HCl, pH 8.0, respectively (Figure 9B).

3.2 Hydrolytic specificity of *Bh*GH81

Since GHs from family 81 have been previously determined to be enzymes that hydrolyzes β-1,3-glucan, the hydrolysis of laminarin and β-1,3-glucooligosaccharides (laminaribiose and laminarihexaose) by *Bh*GH81 was visualized via thin layer chromatography. The insoluble β-1,3-glucan, pachyman, was also tested because a member of CBM family 56 had been shown to bind insoluble β-1,3-glucan (Yamamoto, et al., 1998).The results of this experiment revealed that *Bh*GH81 hydrolyzes laminarin, laminarihexaose, and pachyman into a single product of either glucose or laminaribiose (Figure 10A). *Bh*GH81 activity with laminaribiose was undiscernible since it migrated the same distance as glucose on the TLC (Figure 10A). In order to obtain a more detailed profile of the products released upon hydrolysis of β-1,3-glucooligosaccharides by *Bh*GH81, high performance anion exchange chromatography with pulsed amperometric

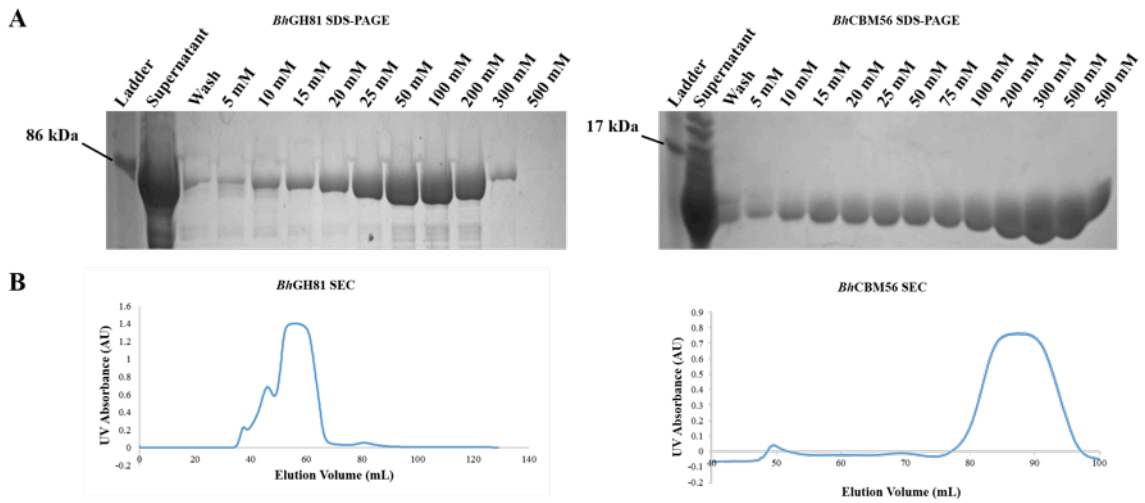

Figure 9: Purification of *Bh*CBM56 and *Bh*GH81. (A) SDS-PAGE gel images, which demonstrate sizes, imidazole fractions, and overall purity of both *Bh*GH81 (81 KDa) and *Bh*CBM56 (11 KDa). (B) Size Exclusion Chromatography profile of *Bh*CBM56 and *Bh*GH81. *Bh*CBM56 eluted at 87 mL whereas *Bh*GH81 eluted at 55 mL.

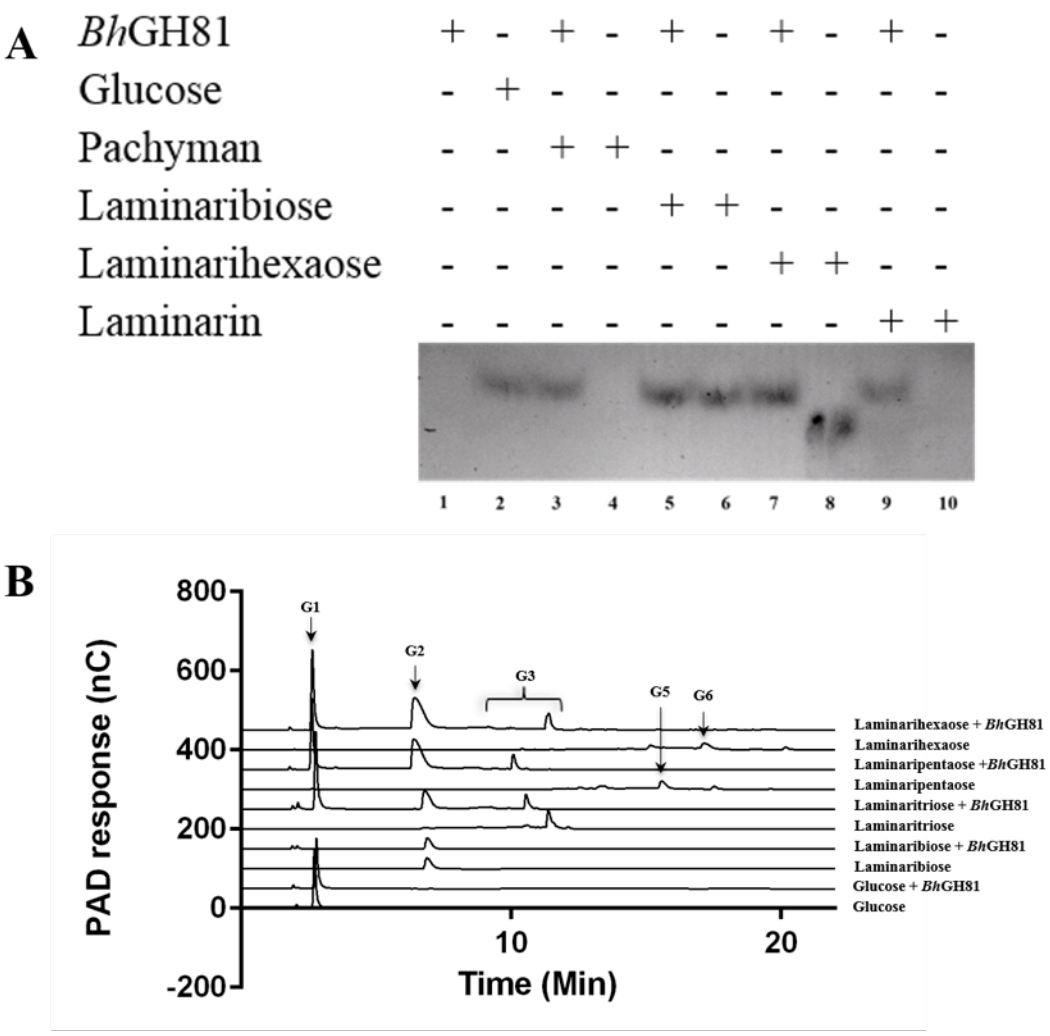

Figure 10: Product analysis of *Bh*GH81 hydrolysis of β-1,3-glucooligosaccharide. (A) TLC gel of *Bh*GH81 digestion. Lanes are loaded as follows: enzyme only control (*lane 1*), glucose standard (*lane 2*), *Bh*GH81 and pachyman (*lane 3*), pachyman (*lane 4*), *Bh*GH81 and laminaribiose (*lane 5*), laminaribiose (*lane 6*), *Bh*GH81 and laminarihexaose (*lane 7*), laminarihexaose (*lane 8*), *Bh*GH81 and laminarin (*lane 9*), laminarin (*lane 10*) (B) High performace anion exchange chromatography with pulsed amperometric detection (PAD) analysis of glucose (G1), laminaribiose (G2), laminaritriose (G3), laminaripentaose (G5), and laminarihexaose (G6). PAD response corresponds to the pulsed amperometric detector signal in nanocoulombs (nC). An analysis of individual sugar standards with unique elution positions: G1, G2, G3, G5, and G6. Product profiles of complete reactions between each β-1,3-glucooligosaccharide (G1, G2, G3, G5, and G6) and *Bh*GH81 are shown directly above each standard.

detection was used with defined β-1,3-glucooligosaccharides: glucose (G1), laminaribiose (G2), laminaritriose (G3), laminaripentaose (G5), and laminarihexaose (G6) (Figure 10B). High performance anion exchange chromatography of these five β-1,3-glucooligosaccharides (G1, G2, G3, G5, and G6) resulted five different retention times. *Bh*GH81 incubated with G1 resulted in one product with a retention time equivalent to the G1 standard. *Bh*GH81 incubated with G2 resulted in one product with a retention time equivalent to the G2 standard. *Bh*GH81 incubated with G3 resulted in three products with retention times equivalent to the G1, G2, and G3 standards. *Bh*GH81 incubated with G5 resulted in three products with retention times equivalent to the G1, G2, and G3 standards. *Bh*GH81 incubated with G6 resulted in three products with retention times equivalent to the G1, G2, and G3 standards. These results also indicated that *Bh*GH81 hydrolyzes G3, G5, and G6, and released G1 as a major product and G2 and G3 as minor products (Figure 10C).

3.3 Structure of *Bh*GH81

In order to understand the mechanism of hydrolysis of β-1,3-glucan by *Bh*GH81, the three-dimensional structure was solved by x-ray crystallography using SAD on a crystallized selenomethionine derivative (See Materials and Methods). *Bh*GH81 was solved in the space group $P2_12_12_1$ with a Matthews coefficient of 2.86 and a solvent content of approximately 57% (Figure 11A). The final model of *Bh*GH81 consisted of 718 amino acids and 332 water molecules (refinement statistics are given in Table 1). The structure of *Bh*GH81 is broken up into domains A, B, and C. Domain A contains a core of eight antiparallel β-strands forming a β-sandwich (Figure 11B). Domain C of *Bh*GH81 contains a core $(\alpha/\alpha)_6$ fold consisting of a double barrel of 14 α-helices

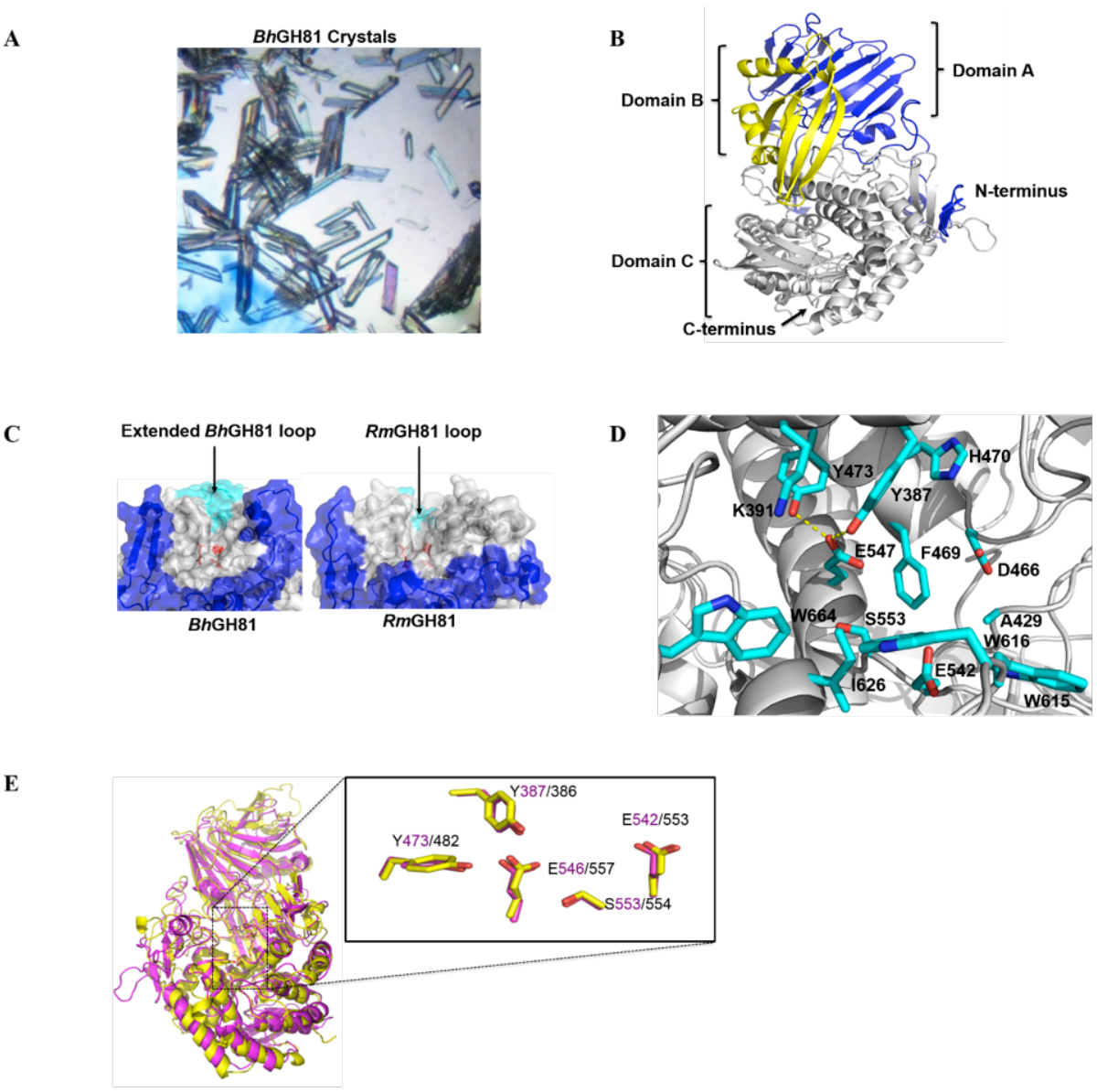

Figure 11: The three-dimensional crystal structure of *Bh*GH81 solved at 2.4 Å resolution using SAD and its comparison to the crystal structure of *Rm*GH81. (A) Selenomethionine derivative of *Bh*GH81 crystals required 1.4 M NaH_2PO_4/0.35 M K_2HPO_4 and 0.0750 M Na_2HPO_4: citric acid, pH 3.2, for nucleation and growth. (B) Domain A, B, and C of *Bh*GH81 are shown: β-sandwich domain (blue), a linker domain (yellow), and an $(\alpha/\alpha)_6$ domain (grey). (C) Surface representation of *Bh*GH81 and *Rm*GH81 putative catalytic cleft showing the blind end canyon topology of *Bh*GH81 created by an extended loop (cyan) and trench-like topology of *Rm*GH81 created by a shorter loop (cyan). (D) Surface exposed residues in the putative substrate-binding cleft of *Bh*GH81 that are potentially involved in interacting with the substrate or stabilizing the catalytic residues. (E) Structural alignment of *Bh*GH81 (magenta) and another GH81 crystal structure from *R. miehei* (yellow) with a close up of the putative key residues for catalysis in the active site on one side of the domain C.

Table 1: Data collection and model statistics for *Bh*GH81

	*Bh*GH81 Selenomethionine
Data collection	
Wavelength	0.97944 Å
Beamline ID	08ID-1
Space group	$P2_12_12_1$
Cell dimensions	
a, b, c (Å)	65.54, 94.29, 159.17
Resolution (Å)	81.1243 - 2.50 (2.60-2.50)*
R_{merge}	0.17 (0.48)
$I/\sigma I$	14.5 (7.3)
Completeness (%)	99.8 (99.9)
Redundancy	5.9 (6.0)
Refinement	
Resolution (Å)	2.50
No. reflections	33036
R_{work}/R_{free}	0.14/0.18
No. residues	
Protein	718
Waters	332
B-factors	
Protein	13.193
Water	16.686
R.m.s deviations	
Bond lengths (Å)	0.019
Bond angles (°)	2.079

* Values for highest resolution shell are shown in parentheses

(α8 - α21) with the C-terminus of the outer helix connected to the N-terminus of the inner helix (Figure 11B). Domain B acts as a linker region between Domain A and C and consists of two twisted β-sheets and three α-helices (Figure 11B). The crystal structure of *Bh*GH81 was compared to another GH of family 81 from *Rhizomucor miehei*, which also consists of domains A, B, and C, and cellulase E4 from *T. fusca*. A 3D alignment of *Rm*GH81 with *Bh*GH81 revealed an root mean square deviation (RMSD) value of 2.1 Å over 584 Cα indicating significant similarity between the two crystal structures. Similar to *Bh*GH81, *Rm*GH81 domains A, B, and C contain a core of two eight-stranded antiparallel β-sheets stacked on top of one another, two twisted anti-parallel β-strands, and a core $(\alpha/\alpha)_6$-barrel fold, respectively (Zhou et al., 2013).

*Bh*GH81 has a cleft located on one side of the $(\alpha/\alpha)_6$-barrel that may contain the catalytic centre. The putative active site of *Rm*GH81 is also located on one side of the $(\alpha/\alpha)_6$-barrel and exhibits a trench-like topology The putative active site of *Bh*GH81 contains two acidic residues (E542 and E546) that are conserved among several organisms with GHs from family 81. The surface representation of *Bh*GH81 reveals this putative substrate-binding cleft is open on one end and closed off at the other creating a blind canyon topology (Figure 11C). The putative catalytic residues (E542 and E546) are located near the blind end of the canyon. In contrast, the putative catalytic residues of *Rm*GH81 are located in the centre of a trench-like catalytic cleft that is open at both ends. A direct comparison of the putative active sites from *Bh*GH81 and *Rm*GH81 revealed that the amino acid loop, which connects helices α20 - α21 is 12 residues longer in *Bh*GH81 compared to the corresponding loop in *Rm*GH81. These 12 extra residues result in a blockage at one end of the active site, creating *Bh*GH81s blind canyon active site topology (Figure 11C). The amino acid sequence alignment of *Bh*GH81 with five other characterized GHs from family 81 reveals 45 conserved amino acids (Figure 12). Of these, 11 residues (Y387,

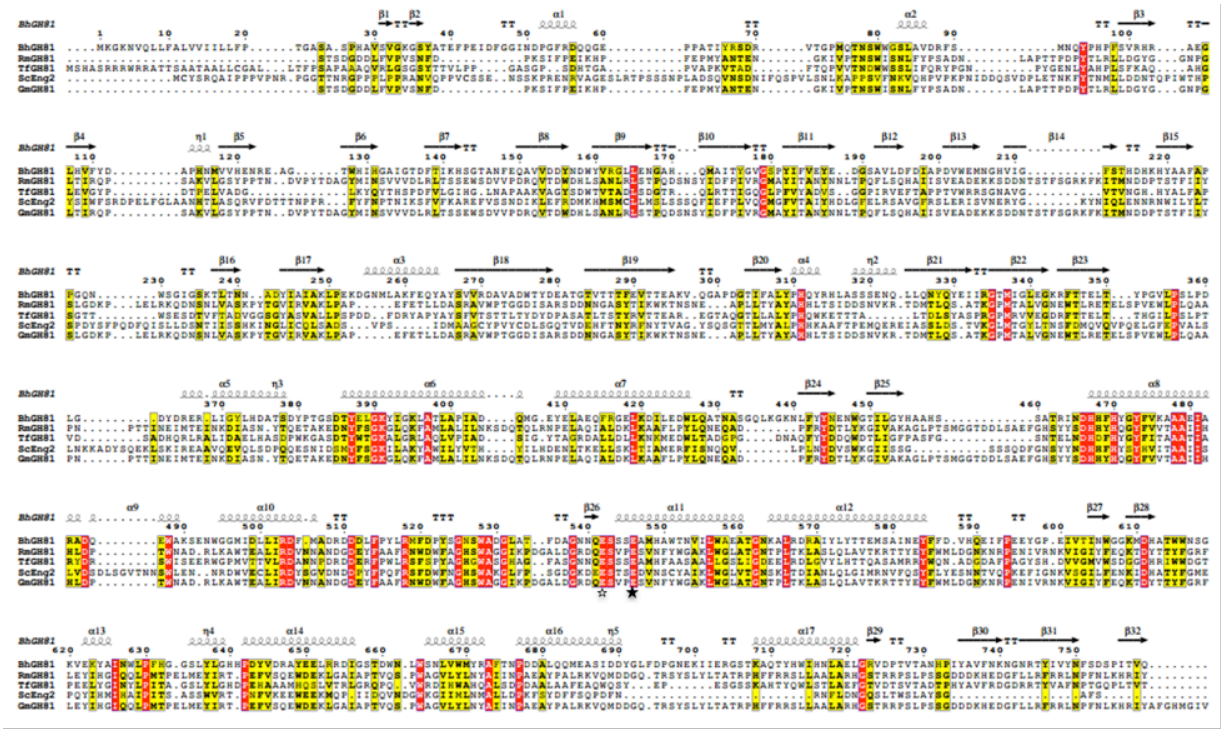

Figure 12: Structural sequence alignments of six GH family 81 proteins. Residues forming secondary structure are indicated above the sequence of *Bh*GH81. The identical residues are shown in white with a red background and the conservative changes are shown in black with a yellow background. The catalytic residues are marked by empty (proton donor) and filled (basic catalyst) stars. The sequences of *Bh*GH81, *A. fumigatus* EngA (*Af*EngA; AAF13033), *S. cerevisiae* Eng2 (*Sc*Eng2; AAB82378), *S. cerevisiae* Eng1 (*Sc*Eng1; CAA96349), *S. pombe* Eng1 (*Sp*Eng1; CAB57443), *S. pombe* Eng2 (*Sp*Eng2; CAA91245), *T. fusca* XY Lam81A (*Tf*Lam81A; AAZ56163), and *Glycine max* Gbp are aligned (*Gm*Gbp; BAA11407).

K391, Y473, D466, H470, A529, E543, S543, E546, I626, and W664) are located in the putative substrate-binding cleft and are surface exposed (Figure 11D). There are three conserved aromatic surface exposed residues (Y387, Y473, and W664) and three aromatic surface exposed residues that can be replaced by other aromatic residues (F469, W615, W616) in the putative substrate-binding cleft of *Bh*GH81 (Figure 11E). A 3D alignment of *Bh*GH81 and *Rm*GH81 reveals that Y387, Y473, E542, E546, and S543 align well with Y386, Y483, E553, E557, and S554, respectively (Figure 11E). In addition to similarity with *Rm*GH81, *Bh*GH81 also has structural similarities with an Cellulase E4 from *T. fusca*, which has been shown to have an initial endo mode of action followed by an exo processive mode of action (i.e. endo-processive). Not only are the overall structures similar, but E4 from *T. fusca* has a blind canyon topology, similar to *Bh*GH81 (Figure 13A, B, and C). The overall structure of Cellulase E4 consists of an $(\alpha/\alpha)_6$-barrel and two four-stranded antiparallel β-sheets stacked on top of one another.

3.4 *Bh*CBM56 binding specificity

A CBM from family 56 had exhibited binding to insoluble β-1,3-glucan (Yamamoto, et al., 1998). Therefore, in order to determine if *Bh*CBM56 binds insoluble and/or soluble β-1,3-glucans, multiple methods were used to test the binding specificity of *Bh*CBM56. The binding of *Bh*CBM56 to the insoluble β-1,3-glucans, pachyman or curdlan, was tested using SDS-PAGE as described in the materials and methods (Figure 14). The supernatant of the samples consisting of *Bh*CBM56 incubated with insoluble β-1,3-glucan was separated from the insoluble pellet. *Bh*CBM56 was present in both the supernatant and the insoluble pellet of all samples incubated with curdlan and pachyman at all amounts (10 mg – 0.5 mg). The insoluble fraction showed a

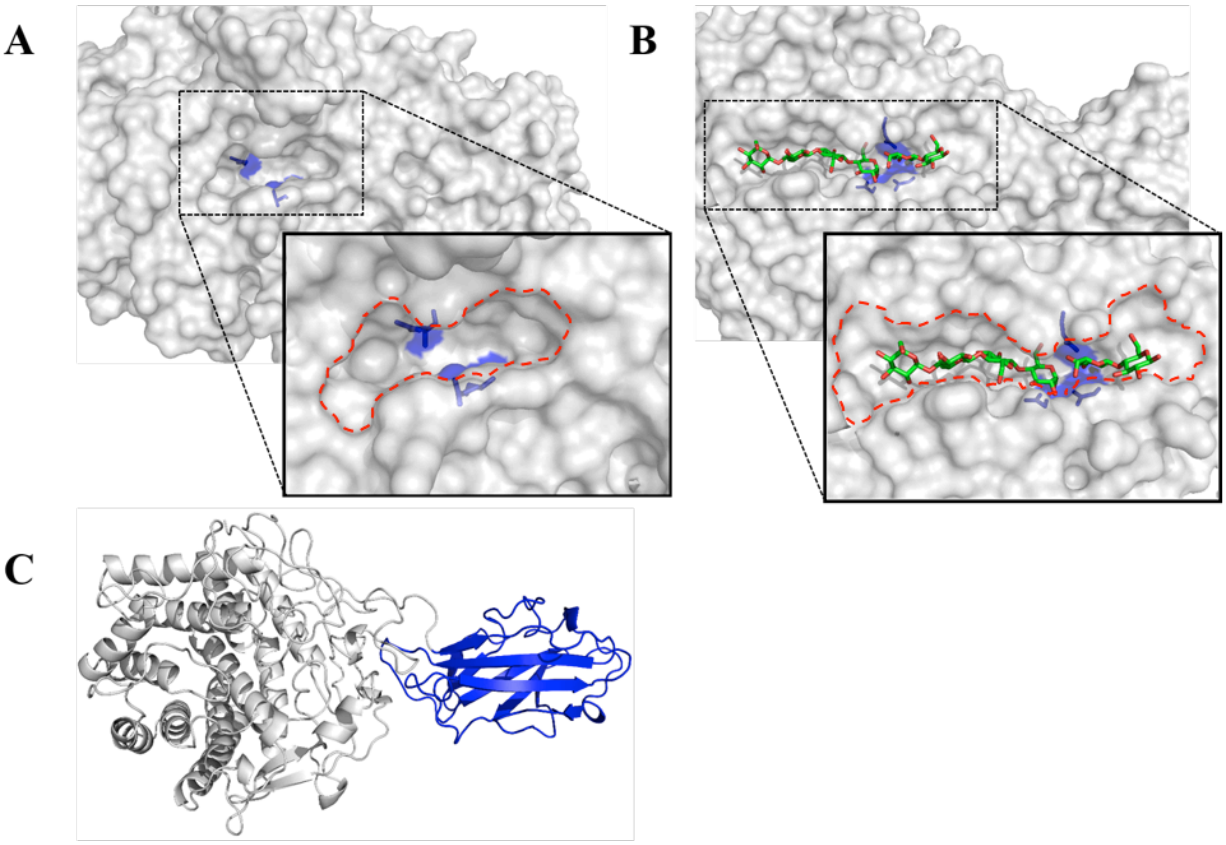

Figure 13: The comparison of E4 and *Bh*GH81 crystal structures. (A) The putative catalytic cleft of *Bh*GH81 (red dashes) showing a blind canyon topology with the catalytic residues shown in blue. (B) The catalytic cleft of E4 (red dashes) showing the blind canyon topology similar to *Bh*GH81's putative catalytic cleft with the catalytic residues shown in blue (Sakon et al., 1997). The active site of E4 is complexed with cellopentaose shown in green (PDB code:4TF4). (C) The overall structure of E4 showing the $(\alpha/\alpha)_6$-barrel in grey and the CBM from family 3 having an antiparallel β-sandwich fold in blue.

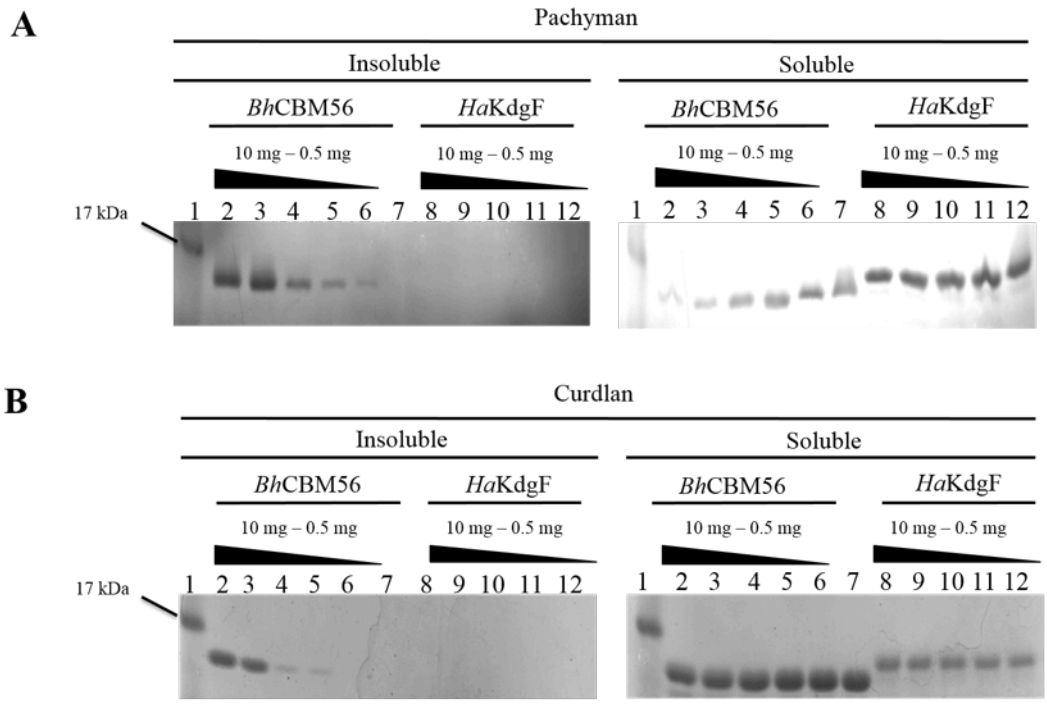

Figure 14: An SDS-PAGE gel showing binding to the insoluble β-1,3-glucans, pachyman and curdlan. Lanes are loaded as follows: Ladder (lane 1), *Bh*CBM56 with decreasing concentration of sugar (10 mg - 0.5 mg) (lanes 2-6), *Bh*CBM56 only negative control (lane 7), *Ha*Kdgf with decreasing concentration of sugar (10mg – 0.5mg) (lanes 8-12). Samples were run on 15 % polyacerylamide gels at 210 volts for 40 minutes. (A) Insoluble sugar pellet of pachyman treated samples (left) and supernatant of pachyman treated samples (right). (B) Insoluble sugar pellet of curdlan treated samples (left) and supernatant of curdlan treated samples (right).

clear decrease in protein concentration as the sugar concentration decreased from 10 mg – 0.5 mg with both curdlan and pachyman. Kdgf from *Halomonas* was used as a negative control since it was known that it would not bind β-1,3-glucan. The negative control experiments set up with Kdgf from *Halomonas* did not show bands in the insoluble pellet with curdlan nor pachyman, whereas the bands in the supernatant were uniform in concentration for Kdgf incubated with pachyman and curdlan. The presence of *Bh*CBM56 in the insoluble pellets when incubated with pachyman or curdlan indicated that this protein bound to both these insoluble β-1,3-glucans. The binding *Bh*CBM56 when incubated with curdlan and pachyman was also quantitatively tested (refer to methods) a using a standard depletion binding isotherm and plotted as bound CBM versus free CBM (Figure 15). *Bh*CBM56 bound with a dissociation constant of (K_d) of 4.06 μM^{-1} ± 0.46 μM^{-1} and a binding capacity (N_o) of 0.95 μmol of CBM/gram of pachyman ± 0.30 μmol/g.

The binding of *Bh*CBM56 to the soluble β-1,3-glucan, laminarin, was tested using native gel affinity electrophoresis (Figure 16). The r/R_0 values for the migration of *Bh*CBM56 and *Ha*Kdgf through the native gels were 0.52 and 0.98, respectively. These results determined that *Bh*CBM56 binds laminarin because the migration of *Bh*CBM56 through a native gel with laminarin incorporated was retarded. Together, not only do these results confirm that *Bh*CBM56 binds insoluble β-1,3-glucans, but they also suggest that *Bh*CBM56 may play a role in binding soluble β-1,3-glucans.

3.5 Structure of *Bh*CBM56

In order to gain insight into the mechanism of carbohydrate recognition by *Bh*CBM56, the three dimensional crystal structure of *Bh*CBM56 was solved to 1.7 Å resolution using MAD on a

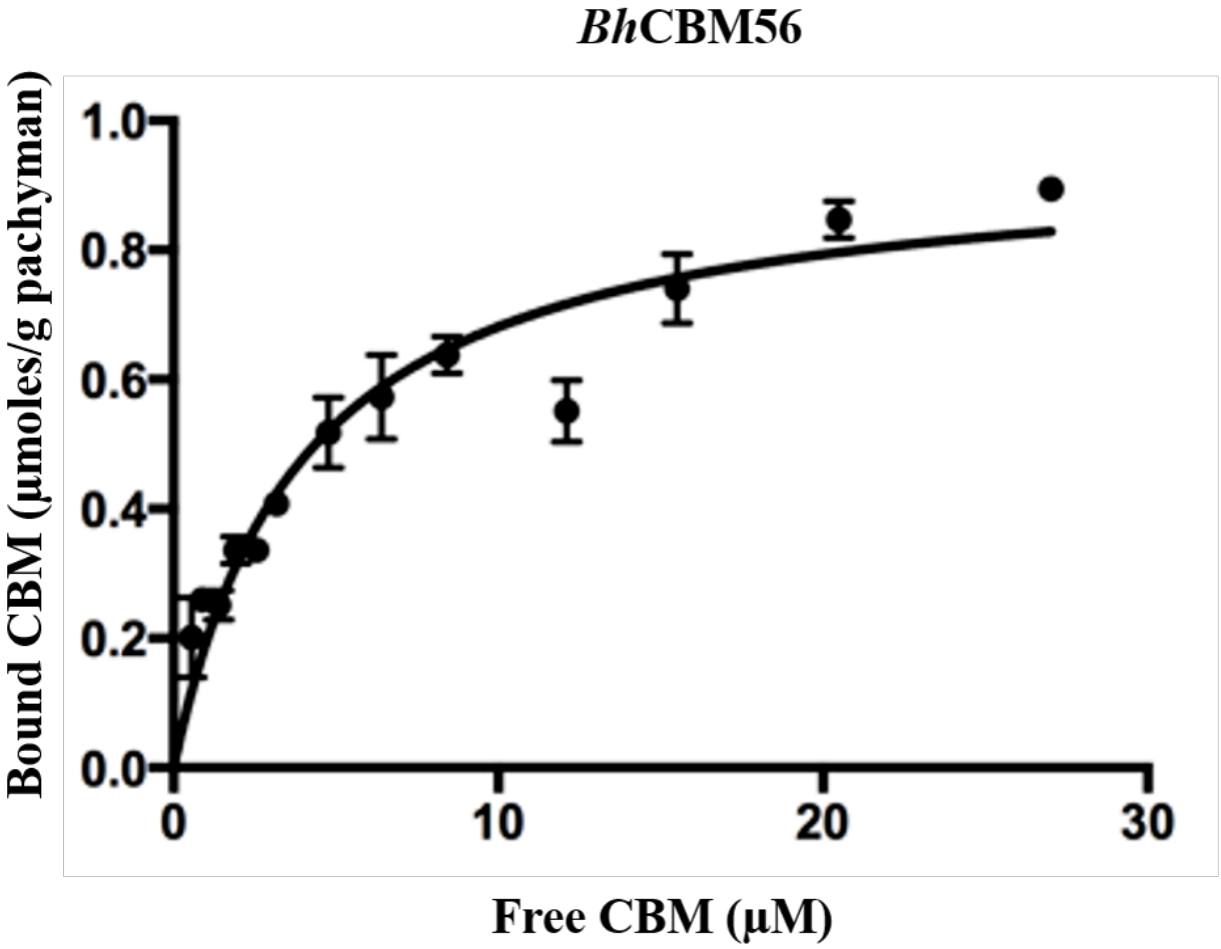

Figure 15: Depletion binding isotherm of *Bh*CBM56 to pachyman. Each data point is represented by a black circle. The solid line is the best fit line to a one-site binding model. Error bars are standard deviation values of three replicates.

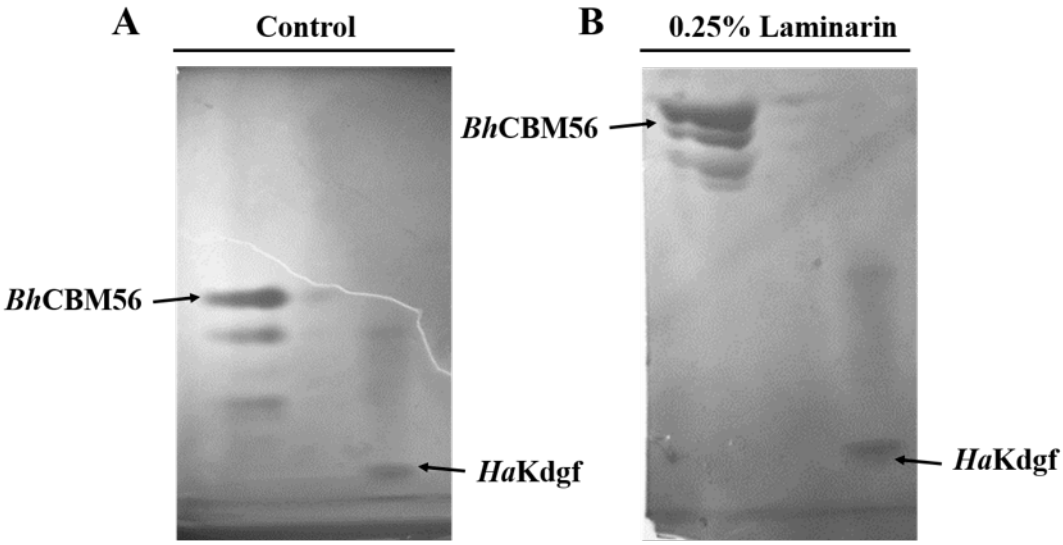

Figure 16: Affinity electrophoresis of *Bh*CBM56 and *Ha*Kdgf on a native polyacrylamide gel matrix with or without laminarin incorporated. Ten micrograms of samples were run at 150 volts for 3 hours and 50 minutes on 9% polyacrylamide gels. (A) Native gel without laminarin. (B) Electrophoresis in the presence of 0.25% (w/v) laminarin.

crystallized bromine derivative (see Materials and Methods). *Bh*CBM56 crystallized with two molecules in space group C2 with a Matthews coefficient of 2.06 and a solvent content of approximately 40.0% (Figure 17A). The final model consisted of two *Bh*CBM56 molecules (94 amino acids each) and 296 water molecules in the asymmetric unit (refinement statistics are given in Table 2). Like the majority of CBM families, *Bh*CBM56 adopts a β-sandwich type fold composed of two antiparallel β-sheets consisting of five β-strands each (Figure 17B). One β-sheet (β-strands 1,2,3,6, and 7) exhibits a concave surface exposed face, while the opposing β-sheet (β-strands 4, 5, 8, 9, and 10) exhibits a convex surface exposed face.

The *Bh*CBM56 structure was compared to a β-1,3-glucan recognition protein from *Plodia interpunctella* (βGRP) with similar properties (Kanagawa et al., 2011). Similar to *Bh*CBM56, βGRP is a β-sandwich composed of two antiparallel β-sheets, each consisting of four β-strands. Three-dimensional alignment of *Bh*CBM56 and βGRP has revealed structural similarity with an RMSD value of 2.1 Å over 82 residues out of 99. This βGRP protein binds to long structured β-1,3-glucan and functions to initiate an immune response (Kanagawa et al., 2011). Another similarity these two proteins share is that *Bh*CBM56 has an arrangement of residues on the convex side of the β-sandwich that is similar to those in the βGRP β-1,3-glucan binding site, which is created by a platform of hydrophobic residues surrounded by several polar and charged residues (Figure 17C and D). From this comparison, it can be speculated that three residues (W1015, H965, and D963) are potentially essential for carbohydrate recognition by *Bh*CBM56 (Figure 17E).

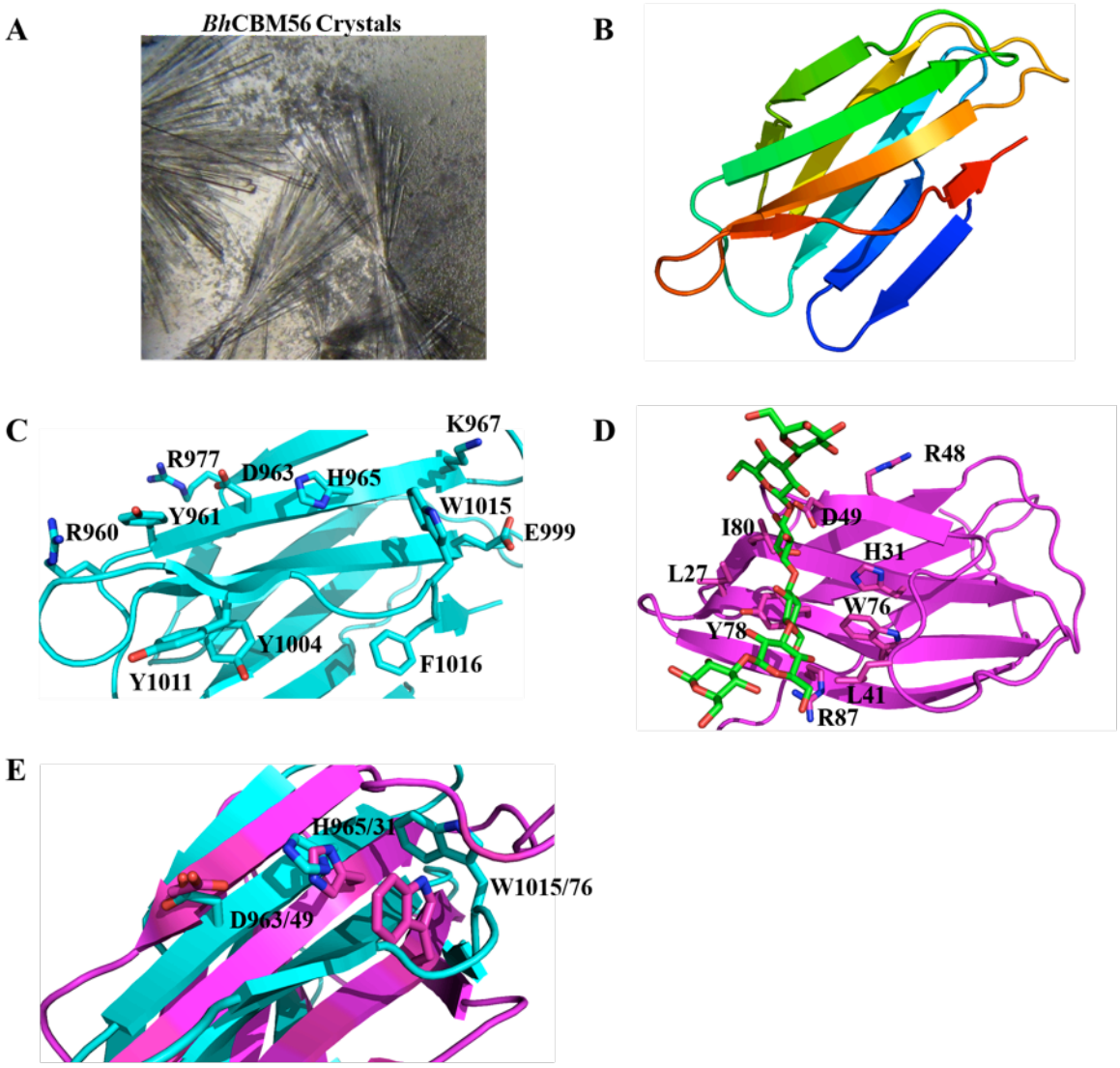

Figure 17: The Crystal structure of *Bh*CBM56 solved at 1.7 Å using MAD and its comparison to β-GRP complexed with laminarihexaose. (A) Native crystals used for bromide soaks required 21 % polyethylene glycol 3350 and 0.1 M Bis-tris/HCl, pH 5.5, for nucleation and growth (B) Cartoon representation of *Bh*CBM56 adopting the common β-sandwich type fold. The putative binding residues located on the convex side of the β-sandwich of (C) *Bh*CBM56 and (D) βGRP from *P. interpunctella*. (E) A 3D alignment of the βGRP protein from *P. interpunctella* with its binding residues (PDB code 3AQZ) and *Bh*CBM56 with its putative binding residues.

Table 2: Data collection and model statistics for *Bh*CBM56

	*Bh*CBM56 Bromide		
Data collection			
Wavelengths	0.91962 Å	0.92002 Å	0.90496 Å
Cell dimensions			
Beamline	BL11-1	BL11-1	BL11-1
Space Group	C2	C2	C2
a, b, c (Å)	76.33, 54.31, 58.20	76.30, 54.29, 58.18	76.34. 54.34, 57.96
α, β, γ (˚)	90.00, 130.90, 90.00	90.00, 131.25, 90.00	90.00, 131.25, 90.00
Resolution (Å)	39.39 -1.70	39.52 – 1.68	39.40 – 1.75
R_{merge}	0.05 (0.30)	0.04 (0.31)	0.041 (0.27)
$I/\sigma I$	61.1/10.9	37.0/5.8	37.6/7.3
Completeness (%)	100.0 (100.0)	99.2 (98.5)	99.1 (99.6)
Redundancy	30.2 (29.6)	8.4 (7.5)	8.4 (8.7)
Refinement			
Resolution (Å)	1.70		
No. reflections	18736		
R_{work}/R_{free}	0.15/0.20		
No. residues			
Protein	180		
Waters	293		
B-factors			
Protein	14.4 (A)		
	12.8 (B)		
Water	25.6		
R.m.s deviations			
Bond lengths (Å)	0.021		
Bond angles (˚)	1.945		

*Values for highest resolution shell are shown in parentheses

Chapter 4

4.0 Discussion

4.1 *Bh*GH81 has a unique mode of action

Sequence similarity revealed that the catalytic module of a multimodular protein from *B. halodurans* was a GH from family 81. This family of GHs has been previously determined to hydrolyze β-1,3-glucan in *S. cerevisiae* (Martín-Cuadardo et al., 2008) and *A. fumigatus* (Fontaine et al., 1997). TLC analysis of *Bh*GH81 hydrolytic specificity confirmed that it is active on the β-1,3-glucans, laminarin and pachyman (Figure 10A). This confirmed our hypothesis that *Bh*GH81 hydrolyzes β-1,3-glucan; however, the molecular details of how *Bh*GH81 hydrolyzed β-1,3-glucan remained unknown.

Previous work on two members of glycoside hydrolase family 81 from *S. cerevisiae* and *A. fumigatus* (*Sc*Eng2 and *Af*Eng1, respectively) have provided insight into the molecular details of hydrolysis by GHs from family 81. *Sc*Eng2 and *Af*Eng1 have been shown to be endo-acting enzymes (Fontaine et al., 1997; Martín-Cuadrado et al., 2008). *Sc*Eng2 incubated with reduced laminaripentaose (G5 with the glucose at the reducing end in the hemiacetal ring form [G5r]) released mainly G2r and G3 with smaller amounts of G3r and G2, indicating that this enzyme cleaves preferentially between the glucose residues at positions 2 and 3 from the reducing end. Reduced substrates longer than G5r (G6r and G7r) were cleaved at several positions, but with a minimum size of a trisaccharide from the non-reducing end. G4r was a poor substrate for *Sc*Eng2; however, it was slowly degraded to G2r and G2. G2 and G3 were not substrates for *Sc*Eng2. Similar results were obtained from *Af*Eng1 with the exception that small amounts of G2 was released from the non-reducing end upon incubation with substrates ranging from G5r and G8r. These results suggested that glycoside hydrolases from family 81 are able to bind to several

positions on β-1,3-glucooligosaccharides, but require a length of at least four glucose residues to do so and hydrolyze them in an endo-acting manner.

Given that these two members hydrolyze laminari-oligosaccharides in an endo-acting manner, it was expected that *Bh*GH81 is an endo-β-1,3-glucanase. In order to test this expectation, product analysis of *Bh*GH81 hydrolysis on β-1,3-glucooligosaccharides using high performance anion exchange chromatography was carried out (Figure 10B). These experiments revealed that this enzyme recognizes β-1,3-linked glucose polymers with a length greater than G2 and hydrolyzes them into G1, G2, and G3 (Figure 10B). These results were different from what had been determined previously for *Sc*Eng2 and *Af*Eng1 for two reasons (Figure 10B). Firstly, neither of these enzymes released glucose as a major product, whereas *Bh*GH81 did (Figure 10B). Secondly, G3 was not a viable substrate for *Sc*Eng2 or *Af*Eng1; however, *Bh*GH81 was able to hydrolyze G3 releasing G1, G2, and G3 as products (Figure 10B). The multiple products released by *Bh*GH81 is typical of an endo-acting enzyme, but glucose as a major product is indicative of an exo-acting enzyme. An explanation for these results could be that *Bh*GH81 is hydrolyzing β-1,3-glucan with an endo-processive mode of action as seen with Cellulase E4 from *T. fusca*. Together, with these results it can be speculated that *Bh*GH81's mode of action is unique as compared to other characterized GHs from family 81 because it can recognize oligosaccharides with a degree of polymerization greater than two and degrades them with an endo-processive mode of action.

4.2 Structural insights into a β-1,3-Glucanase from family 81

The product analysis of *Bh*GH81 suggested that it has a unique mode of action compared to other members of its family. This could also mean that the architecture of the catalytic site is

unique compared to other GHs from family 81. In order to determine if *Bh*GH81 has an active site that is unique, the enzyme was structurally characterized by x-ray crystallography. The crystal structure of *Bh*GH81 was solved to 2.5 Å resolution revealing that *Bh*GH81 consists of three domains: Domain A, B, and C.

An amino-acid sequence alignment with five other GHs from family 81 uncovered 45 conserved residues (Figure 12). Out of these, 11 (Y387, K391, Y473, D466, H470, A529, E542, S543, E546, I626, and W664) were located in the putative substrate binding cleft and surface exposed. Along the catalytic cleft, there are three conserved aromatic residues (Y387, Y473, and W664). There are also three residues in the substrate-binding cleft (F469, W615, W616) that can replaced by other aromatic residues among these five GHs from family 81. Since aromatic residues located in substrate-binding clefts have been known to make stacking interactions with the pyranose rings of glucose, these residues may be involved in interacting the β-1,3-glucan substrate.

Typically hydrolysis of glycosidic bonds occurs via a general acid catalysis mechanism and requires two acidic residues. These catalytic residues are typically conserved across most GH families. The only acidic conserved residues located in the catalytic cleft are E542, E546, and D466. Previous studies on *Sc*Eng2 have shown a drastic reduction in activity on laminarin when two acidic residues, E608 and E612 (E542 and E546 in *Bh*GH81), were mutated to glutamine (Martín-Cuadrado et al., 2008). D526 in *Sc*Eng2 (D466 in *Bh*GH81) is not crucial in catalytic activity (Martín-Cuadrado et al., 2008). This indicates that E542 and E546 are likely the catalytic residues in *Bh*GH81.

A member of GH family 81 has been shown to adopt an inverting mechanism (McGrath & Wilson, 2006) and the putative catalytic residues are 8.6 Å apart in *Bh*GH81. Thus, it is likely

that its catalytic mechanism involves a proton donor and a basic catalyst. During this general acid catalysis mechanism, carboxyl groups that function as basic catalysts are negatively charged, whereas carboxyl groups that function as proton donors are initially protonated. In order to achieve this, the structure of active sites have evolved to modulate the pKa of these amino acids with carboxylate groups. The only two conserved aromatic residues in the substrate-binding cleft of the only structurally characterized GH from family 81 (*Rm*GH81), Y386 and Y482 (Y387 and Y473 in *Bh*GH81), have been found to hydrogen bond to the putative basic catalyst, E557, (E546 in *Bh*GH81). Thus, Y386 and Y482 of *Rm*GH81 stabilize the basic catalyst by modulating its pKa (Zhou et al., 2013). Similarly to *Rm*GH81, Y473 and Y387 of *Bh*GH81 are positioned similarly to E546 (Y386 and Y483 in *Rm*GH81) and also form hydrogen bonds to this residue (E557 in *Rm*GH81), thereby modulating its pKa and suggesting that it is the basic catalyst (Figure 11E). Additionally, S543 of *Bh*GH81 is conserved among all GH81 family members (Figure 12). Structural characterization of *Rm*GH81 has suggested that this residue (S554 in *Rm*GH81) is involved in stabilizing the position of the putative basic catalyst by hydrogen bonding to the backbone nitrogen atom of E557 in *Rm*GH81 (Zhou et al., 2013). S543 of *Bh*GH81 is oriented in a similar way indicating that it may stabilize E546. These findings suggest that E546 is the basic catalyst of *Bh*GH81. E542 (E554 in *Rm*GH81) does not form any hydrogen bonds to other amino acids based on its position and is located in a hydrophobic environment, indicating that it is the proton donor (Figure 11E).

The catalytic centres of all characterized enzymes with an $(\alpha/\alpha)_6$-barrel fold is on one side of the barrel as seen in the GH family 8 xylanase from *Pseudoaltermonas haloplanktis* (Bott et al., 2008) and the GH family 15 glucoamylase from *Trichoderma reesei* (Van Petegem et al., 2003). The x-ray crystallography results of *Bh*GH81 revealed that the location of the putative

active site is on one side of the $(\alpha/\alpha)_6$ barrel and it has a blind canyon active site topology. The putative catalytic residues (E542 and E546) are located near the blind end of this active site. The distance between the blind end of the canyon and the catalytic residues is only large enough for a mono/disaccharide. However, the whole cleft is large enough to accommodate a long polysaccharide chain. The active site of *Bh*GH81 has a 23 amino acid loop located between helices α20 - α21 and is oriented in a way that it blocks one end of the active site, creating the blind canyon topology. Previous structural characterization of *Rm*GH81 also revealed that the putative active site is located on one side of the $(\alpha/\alpha)_6$ barrel. In contrast, the loop connecting helices α20 - α21 is 12 residues shorter in *Rm*GH81 resulting in a trench-like active site. Every GH81 enzyme characterized by product analysis has so far, hydrolyzed β-1,3-glucan in an endo-acting manner. We propose that *Bh*GH81 is unique and it is due to this extended loop that allows it to degrade β-1,3-glucan in an exo-acting manner.

An issue that remains to be resolved is that the product analysis of *Bh*GH81 incubated with β-1,3-glucooligosaccharides released multiple products, which is typical of an endo-acting enzyme. Unfortunately, further structural analysis and product analysis experiments are needed to resolve this issue. However, speculations can be made based on a comparison with an enzyme that has similar properties. A comparison of the crystal structures of *Bh*GH81 to cellulase E4 (i.e. E4; GH family 9) from *T. fusca* could explain why the product analysis results of *Bh*GH81 with β-1,3-glucooligosaccharides released products that are indicative of both an exo and an endo acting enzyme. E4 has an endo-processive mode of action. The crystal structure of E4 has a catalytic domain exhibiting an $(\alpha/\alpha)_6$ barrel fold and a CBM from family 3 having an antiparallel β-sandwich fold (Sakon et al., 1997). These overall features are similar to those observed in the crystal structure of *Bh*GH81 (Figure 13C). The substrate-binding cleft of the E4 enzyme runs

from helices α1 - α8. The catalytic residues of this enzyme (D55, D58, and E424) are located in the middle of this cleft (Figure 13A). The end of the substrate-binding cleft closest to helice α8 is blocked by a loop created by residues 245 - 255, however the end of the cleft closest to helice α1 connects smoothly with the surface of the CBM from family 3 (Figure 13B). This substrate-binding cleft is similar to that observed in *Bh*GH81, suggesting it may also have an endo-processive mode of action. For an initial endo mode of action to exist in *Bh*GH81, the catalytic residues must hydrolyze an internal glycosidic bond in a polysaccharide chain. However, the fact that the catalytic residues of *Bh*GH81 are located near the blind end of the catalytic cleft poses a problem for this theory. Therefore, in order for an initial endo mode of action to exist in *Bh*GH81, the polysaccharide substrate would have to enter the cleft at the end opposite to the catalytic residues and exit the cleft at the end with the catalytic residues. This would allow the catalytic residues to interact with the internal glycosidic bonds of a polysaccharide chain. Therefore, it is possible that *Bh*GH81 may hydrolyze β-1,3-glucan with an endo-processive mode of action as observed in E4. Further x-ray crystallography experiments involving an enzyme-product complex complemented with a time course product analysis of *Bh*GH81 hydrolysis on β-1,3-glucooligosaccharides using could elucidate this.

4.3 Ligand specificity of *Bh*CBM56

Sequence similarity revealed that *Bh*Lam81 from *B. halodurans* has a CBM from family 56 as one of its modules. The binding of a CBM from family 56 to insoluble β-1,3-glucan has already been shown (Yamamoto et al., 1998). In order to determine if *Bh*CBM56 binds insoluble β-1,3-glucan, an SDS-PAGE depletion assay was performed. The results from this experiment revealed that *Bh*CBM56 binds the insoluble β-1,3-glucans, curdlan and pachyman. In order to

obtain more quantitative binding data, binding isotherms of *Bh*CBM56 and pachyman were performed. This experiment revealed that *Bh*CBM56 had a relatively high affinity for insoluble β-1,3-glucan (K_d, 4.06 μM^{-1} ± 0.46 and N_o, 0.95 μmol of CBM/ gram of pachyman ± 0.30 μM). Additionally, binding of *Bh*CBM56 to laminarin was also shown via native gel affinity electrophoresis. These results not only confirmed our hypothesis that *Bh*CBM56 binds insoluble β-1,3-glucan, but revealed that it binds soluble β-1,3-glucan as well.

4.4 Structural insights into *Bh*CBM56

The binding isotherms and SDS-PAGE depletion assays for *Bh*CBM56 indicated that it binds both insoluble and soluble β-1,3-glucan. In order to gain insight into the binding mechanism of *Bh*CBM56, we solved its structure using x-ray crystallography. A structural homologue search using the DALI server showed that this CBM shares highest structural similarity to βGRP from *P. interpunctella* (PDB code, 3AQY-B; Z-score 9.8; Kanagawa et al., 2011). This βGRP protein functions to initiate an immune response in *Plodia interpunctella* upon binding fungal cell wall β-1,3-glucan. The crystal structure of βGRP revealed the presence of hydrophobic and charged residues mostly on the convex side of the β-sheet forming a large binding platform. Residues L41, L27, W76, Y78, and I80 contribute to binding laminarin by creating a solvent exposed hydrophobic patch. Four polar and charged residues of βGRP (H31, R48, D49, and R87) surround this hydrophobic patch and are involved in the interaction with O4 and O6 atoms of glucose residues in laminarin. The *Bh*CBM56 structure also has a cluster of hydrophobic residues that are mostly on its convex β-sheet (Y961, Y1004, Y1011, W1015, F1016, and Y1018). Of these, W1015 and F1016 are conserved with W93 and F98 from βGRP, respectively. However, neither W93 or F98 have been found to be involved in binding β -1,3-

glucan in βGRP. Six polar and charged residues exist around the hydrophobic patch in *Bh*CBM56 (R960, D963, H965, K967, R976, and E999). It is possible that some of these residues on the convex side of the β-sandwich make similar interactions with the β-1,3-glucans tested in this study.

β -1,3-Glucan has been proposed to be a triple-helical structure existing in many fungi, bacteria, plants, and algae and is known variously as pachyman, laminarin, callose, paramylon, laricinan, and curdlan. The x-ray crystallographic complex of βGRP and laminarin reveals that D49 and W76 interact with more than one strand of the triple helix, thereby stabilizing it (Kanagawa et al., 2011). D963 and W1015 of *Bh*CBM56 are not conserved with D49 and W76 of the βGRP protein, respectively. However, the alignment of these two proteins reveals that the side chains of D963 and W1015 of *Bh*CBM56 and D49 and W76 of βGRP are in close proximity, respectively (Figure 17E). Therefore, D963 and W1015 may interact with the triple helix of pachyman in a similar manner to D49 and W76 in the βGRP protein. Another essential residue that binds the triple-helical structure of laminarin in βGRP was found to be H31. This residue is not conserved with *Bh*CBM56, but the alignment between *Bh*CBM56 and βGRP shows that there is one histidine (H966) with its imidiazole side chain in a similar position to H31 of βGRP. Thus, D963, W1015, and H965 may be the residues essential for binding the triple-helical structure of pachyman. This comparison of *Bh*CBM56 with βGRP has revealed that the binding site for *Bh*CBM56 may be on the convex face of the β-sandwich created by a platform of hydrophobic residues surrounded by several polar and charged residues.

4.5 Conclusions

The objectives of this study were to determine the substrate-specificity of *Bh*GH81 and *Bh*CBM56 and speculate on the molecular details of how both of these modules recognize and act on their substrates. The results in this study confirm our hypothesis that *Bh*GH81 is a β-1,3-glucan hydrolyzing enzyme and *Bh*CBM56 is a β-1,3-glucan binding CBM. By identifying the substrate-specificity and attempting to determine the amino acids involved in binding by *Bh*CBM56, the results presented in this study for *Bh*CBM56 can be used to work towards determining the molecular details of how it recognizes and acts on its substrate. The high performance anion exchange chromatography results and the comparison of *Bh*GH81 to Cellulase E4 from *T. fusca* may indicate that *Bh*GH81 has an endo-processive mode of action, however all characterized glycoside hydrolases from family 81 are endo-β-1,3-glucanases. An exo acting enzyme typically releases one reaction product whereas an endo acting enzyme releases a variety of reaction products. The reaction products generated by the catalytic domain are not typical of an exo-mode or an endo-mode of action. *Bh*GH81 releases a variety of reaction products including G1, G2, and a small amount of G3, which is typical of an endo-mode of action. On the other hand, G1 is the most abundant reaction product, which is a typical of an exo-acting enzyme. The active site topologies of endo acting enzymes are typically long grooves that will fit long polysaccharide chains with the catalytic residues located somewhere in the middle of the groove. In contrast, the active sites of exo-acting enzymes have a pocket or crater active site topology that fit only mono or disaccharides. The structure of *Bh*GH81 reveals a blind canyon active site topology with the catalytic residues at the blind end of this canyon. This canyon could fit a long polysaccharide chain, yet there is a loop that creates a blockage near the catalytic residues. This type of active site topology has been observed prior to this study in E4, an

endo/exo cellulase from family 9. This enzyme initially hydrolyzed long chains of cellulose with an endo-processive mode of action. Therefore, it can be speculated that *Bh*GH81 is using a similar mode of action. This indicates the *Bh*GH81 is a unique member of its family in that it may be an endo-processive enzyme as opposed to its endo acting relatives from GH family 81. The initial endo mode of action would explain production of the larger G2 and G3 products, while G1 would be produced by the exo-mode of action. By identifying the products released when various β-1,3-glucooligosaccharides were subjected to hydrolysis by *Bh*GH81 and obtaining structural insight, the results presented in this study can be used to work towards characterizing the molecular details of how it recognizes and acts on its substrate(s). The elucidation of the molecular details of how both these modules recognize and act on their substrates can be used to work towards presenting a model of how the modular components of *Bh*Lam81 allow for more efficient recognition and degradation of its substrate(s) in the environment.

Research on the degradation of polysaccharides by multimodular carbohydrate-active enzymes has showed us that many of them have ancillary modules, which aid in increasing the efficiency of polysaccharide degradation. These ancillary modules are known as CBMs, which aid in degradation by binding carbohydrates. The majority of the forces allowing CBMs and their catalytic modules to interact with their substrates are hydrophobic stacking interactions between the sugar and aromatic amino acid side chains in the substrate-binding clefts. These amino acid side chains determines whether the terminal or the internal units of the polysaccharide chain interacts with the substrate-binding cleft of CBMs or their catalytic modules. The research on GHs from family 81 and CBMs from family 56 may be useful in predicting how other uncharacterized polypeptides from these families interact with their putative substrates based on

amino acid sequence. Studying these interactions and the degradation of their substrates may have applications in industry for engineering multimodular enzymes with enhanced ability to hydrolyze their substrates. As the need for alternative fuel sources will likely increase, this research may be important. There is current work being done to produce ethanol from alginate in brown macroalgae. This system uses a 36 kilobase pair DNA from *Vibrio splendidus* encoding enzymes for alginate transport and metabolism as well as an engineered system for extracellular alginate depolymerisation (Wargacki et al., 2012). This study was done in hope that a system could one day be engineered to generate a microbial platform that can degrade, uptake, metabolize, and synthesize ethanol from alginate. By studying β-1,3-glucanases and determining the molecular details of how they recognize and degrade their substrates, we could one day engineer a similar microbial platform for biofuel production from β-1,3-glucan.

.

References

1. Aleshin, A. E., Firsovs, L. M., & Honzatkoll, R. B. (1994). Refined Structure for the Complex of Acarbose with Glucoamylase. *Journal of Biological Chemistrty*, *269*(22), 15631–15639.

2. Aspinall G. O., & Kessler G. (1957). The structure of callose from the grape vine. *J. Soc. Chem & Ind*, 1296.

3. Bailey, S. (1994). The Ccp4 Suite - Programs for Protein Crystallography. *Acta Crystallographica Section D-Biological Crystallography 50*, 760-763.

4. Beattie, A., Hirst, E. L., & Percival, E. (1961). Studies on the metabolism of the chryosphyceae. Comparative structural investigations on leucosin (chrysolaminarin) separated from diatoms and laminarin from the brown algae. *Biochem. J 79(3)*, 531-537.

5. Boraston, A. B., Bolam, D. N., Gilbert, H. J., & Davies, G. J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J, 382*, 769- 781.

6. Boraston, A. B., Healey, M., Klassen, J., Ficko-Blean, E., Lammerts van Bueren, A., & Law, V. (2006). A structural and functional analysis of alpha-glucan recognition by family 25 and 26 carbohydrate-binding modules reveals a conserved mode of starch recognition. *The Journal of Biological Chemistry*, *281*(1), 587–98. doi:10.1074/jbc.M509958200

7. Boraston, A. B., McLean, B. W., Guarna, M. M., Amandaron-Akow, E., & Kilburn, D. G. (2001). A family 2a carbohydrate-binding module suitable as an affinity tag for proteins produced in Pichia pastoris. *Protein Expression and Purification*, *21*(3), 417–23. doi:10.1006/prep.2001.1393

8. Bott, R., Saldajeno, M., Cuevas, W., Ward, D., Scheffers, M., Aehle, W., & Hansson, H. (2008). Three-dimensional structure of an intact glycoside hydrolase family 15 glucoamylase from Hypocrea jecorina. *Biochemistry*, *47*(21), 5746–54. doi:10.1021/bi702413k

9. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research*, *37*(Database issue), D233–8. doi:10.1093/nar/gkn663

10. Cardona, F., Parmeggiani, C., Faggi, E., Bonaccini, C., Gratteri, P., Sim, L., & Goti, A. (2009). Total syntheses of casuarine and its 6-O-alpha-glucoside: complementary inhibition towards glycoside hydrolases of the GH31 and GH37 families. *Chemistry (Weinheim an Der Bergstrasse, Germany)*, *15*(7), 1627–36. doi:10.1002/chem.200801578

11. Charlwood, J., Birrell, H., & Camilleri, P. (1998). Efficient carbohydrate release, purification, and derivatization. *Analytical Biochemistry*, *262*(2), 197–200. doi:10.1006/abio.1998.2803

12. Charnock, S. J., Bolam, D. N., Nurizzo, D., Szabó, L., McKie, V. A., Gilbert, H. J., & Davies, G. J. (2002). Promiscuity in ligand-binding: The three-dimensional structure of a Piromyces carbohydrate-binding module, CBM29-2, in complex with cello- and mannohexaose. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(22), 14077–82. doi:10.1073/pnas.212516199.

13. Chen, X. Y., & Kim, J. Y. (2009). Callose synthesis in higher plants. *Plant Signaling & Behavior*, *4*(6), 489–92. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2688293&tool=pmcentrez&rendertype=abstract

14. Cheng, Y., Hong, T., Liu, C., & Meng, M. (2009) Cloning and functional characterization of a complex endo- β-1,3-glucanase from *Paenibacillus sp*. *Appl Microbiol Biotechnol*. *81*, 1051-1061.

15. Correia, M.A, Abbott, D. W., Gloster, T. M., Fernandes, V. O., Prates, J. A., Montenier, C., Dumon, C., Williamson, M. P., Tunnicliffe, R. B., Liu, Z., Flint, J. E., Davies, G. J., Henrissat, B., Coutinho, P. M., Fontes, C. M., & Gilbert, H. J. (2010) Signature active site architectures illuminate the molecular basis for ligand specificity in family 35 carbohydrate binding module. *Biochemistry*, *49,* 6193-6205.

16. Davies, G. J., Gloster, T. M., & Henrissat, B. (2005). Recent structural insights into the expanding world of carbohydrate-active enzymes. *Current Opinion in Structural Biology*, *15*(6), 637–45. doi:10.1016/j.sbi.2005.10.008.

17. Davies, G. J., & Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure*, *3*(9), 853–9.

18. Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Knowles, J. K. C., Teeri, T. T., & Jones, T. A. (2014). The Three-Dimensional of the Catalytic Core of Cellobiohydrolase from Trichoderma reesei. *Science*, *265*(5171), 524–528.

19. Duchesne, L. C., & Larson, D. W. (1989). Cellulose and the evolution of plant life. *BioScience*, *39(4),* 238-241.

20. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr, 60*, 2126-32.

21. Evans, P. R. (2011). An introduction to data reduction: space-group determination, scaline and intensity statistics. *Acta Cryst D, 67*, 282-292.

22. Fibriansah, G., Masuda, S., Koizumi, N., Nakamura, S., & Kumasaka, T. (2007). The 1.3 Å crystal structure of a novel endo-β-1,3-glucanase of glycoside hydrolase family 16 from alkophillic *Nocardiopsis* sp. strain F96. *Proteins, 69*(3), 683–690. doi:10.1002/prot.

23. Fliegmann, J., Mithofer, A., Wanner, G., & Ebel, J. (2004). An ancient enzyme domain hidden in the putative beta-glucan elicitor receptor of soybean may play an active part in the perception of pathogen-associated molecular patterns during broad host resistance. *The Journal of Biological Chemistry*, *279*(2), 1132–40. doi:10.1074/jbc.M308552200

24. Fontaine, T., Hartland, R. P., Beauvais, a, Diaquin, M., & Latge, J. P. (1997). Purification and characterization of an endo-1,3-beta-glucanase from Aspergillus fumigatus. *European Journal of Biochemistry / FEBS*, *243*(1-2), 315–21. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9030754

25. Gilbert, H. J., Knox, J. P., & Boraston, A. B. (2013). Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. *Current Opinion in Structural Biology*, *23*(5), 669–77. doi:10.1016/j.sbi.2013.05.005.

26. Henrissat, B. & Bairoch, A. (1996) Updating the sequence-based classification of glycosyl hydrolases. *Biochem J*, *316*, 695-696.

27. Henrissat, B., & Davies, G. J. (1997). Structural and sequenced-based classification of glycoside hydrolases. *Curr Opin Struct Biol, 7*, 637–644.

28. Henrissat, B., Driguez, H., Vlet, C., & Schulein, M. (1985). Synergism of cellulases from *Trichoderma reesei* in the degradation of cellulose. *Nature Biotechnology*, *3*, 722-726.

29. Huecas, S., Villalba, M., & Rodríguez, R. (2001). Ole e 9, a major olive pollen allergen is a 1,3-beta-glucanase. Isolation, characterization, amino acid sequence, and tissue specificity. *The Journal of Biological Chemistry*, *276*(30), 27959–66. doi:10.1074/jbc.M103041200

30. Hoffmann, G. C., & Simon, B. W. (1971). Structure and molecular size of pachyman. *Carbohydrate Research*, *20*, 185–188.

31. Hrmova, M., & Fincher, G. B. (1993). Leaves of barley (Hordeum vulgare), *461*, 453–461.

32. Ilari, A., Fiorillo, A., Angelaccio, S., Florio, R., Chiaraluce, R., van der Oost, J., & Consalvi, V. (2009). Crystal structure of a family 16 endoglucanase from the hyperthermophile *Pyrococcus furiosus*--structural basis of substrate recognition. *The FEBS Journal*, *276*(4), 1048–58. doi:10.1111/j.1742-4658.2008.06848.x.

33. Ishida, T., Fushinobu, S., Kawai, R., Kitaoka, M., Igarashi, K., & Samejima, M. (2009). Crystal structure of glycoside hydrolase family 55 {beta}-1,3-glucanase from the basidiomycete *Phanerochaete chrysosporium*. *The Journal of Biological Chemistry*, *284*(15), 10100–9. doi:10.1074/jbc.M808122200.

34. Isorna, P., Polaina, J., Latorre-García, L., Cañada, F. J., González, B., & Sanz-Aparicio, J. (2007). Crystal structures of Paenibacillus polymyxa beta-glucosidase B complexes reveal the molecular basis of substrate specificity and give new insights into the catalytic machinery of family I glycosidases. *Journal of Molecular Biology*, *371*(5), 1204–18. doi:10.1016/j.jmb.2007.05.082.

35. Kanagawa, M., Satoh, T., Ikeda, A., Adachi, Y., Ohno, N., & Yamaguchi, Y. (2011). Structural insights into recognition of triple-helical beta-glucans by an insect fungal receptor. *J. Biol. Chem*, *286*, 29158 – 29165

36. Kitago, Y., Karita, S., Watanabe, N., Kamiya, M., Aizawa, T., Sakka, K., & Tanaka, I. (2007). Crystal structure of Cel44A, a glycoside hydrolase family 44 endoglucanase from *Clostridium thermocellum*. *The Journal of Biological Chemistry*, *282*(49), 35703–11. doi:10.1074/jbc.M706835200

37. Lammerts van Bueren, A., Ficko-Blean, E., Pluvinage, B., Hehemann, J.-H., Higgins, M. a, Deng, L., & Boraston, A. B. (2011). The conformation and function of a multimodular glycogen-degrading pneumococcal virulence factor. *Structure*, *19*(5), 640–51. doi:10.1016/j.str.2011.03.001.

38. Lammerts van Bueren, A. L., Morland, C., Gilbert, H. J., & Boraston, A. B. (2005) Family 6 carbohydrate binding modules recognize the non-reducing end of β -1,3-linked glucans by presenting unique ligand binding surface. *The Journal of Biological Chemistry*, *280*, 530- 537.

39. Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, *26*(2), 283–291. doi:10.1107/S0021889892009944.

40. Leslie, A. G. W. (2006). The integration of macromolecular diffraction data. *Acta Crystallographica. Section D, Biological Crystallography*, *62*(1), 48–57. doi:10.1107/S0907444905039107.

41. Levy, I., Paldi, T., & Shoseyov, O. (2004). Engineering a bifunctional starch–cellulose cross-bridge protein. *Biomaterials*, *25*(10), 1841–1849. doi:10.1016/j.biomaterials.2003.08.041.

42. Linder, M., Salovuori, I., Ruohonen, L. & Teeri, T. T. (1996). Characterization of a double cellulose-binding domain. Synergistic high affinity binding to crystalline cellulose. *J Biol Chem*, *271*, 21268-72.

43. Liu, Y.-S., Baker, J. O., Zeng, Y., Himmel, M. E., Haas, T., & Ding, S.-Y. (2011). Cellobiohydrolase hydrolyzes crystalline cellulose on hydrophobic faces. *The Journal of Biological Chemistry*, *286*(13), 11195–201. doi:10.1074/jbc.M110.216556

44. Martin-Cuadardo, A., Fontaine, T., Esteban, P., del Dedo, J. E., de Medina-Redonda, M., del Rey, F., Latge, J. P., & de Aldana, C. R. (2007) Characterization of the endo- β -1,3-glucanase activity of *S. Cerevisiae* Eng2 and other members of the GH81 family. *Fungal genetics and Biology*, *45*, 542-553.

45. McCarter, J. D., & Withers, S. G. (1994) Mechanisms of enzymatic glycoside hydrolysis. *Current Opinion in Structural Biology*, *4*, 885-892.

46. McCartney, L., Gilbert, H. J., Bolam, D. N., Boraston, A. B., & Knox, J. P. (2004). Glycoside hydrolase carbohydrate-binding modules as molecular probes for the analysis of plant cell wall polymers. *Analytical Biochemistry*, *326*(1), 49–54. doi:10.1016/j.ab.2003.11.011

47. McGrath, C. E., & Wilson, D. B. (2006). Characterization of a *Thermobifida fusca* beta-1,3-glucanase (Lam81A) with a potential role in plant biomass degradation. *Biochemistry*, *45*(47), 14094–100. doi:10.1021/bi061757r.

48. McLean, B. W., Bray, M. R., Boraston, A. B., Gilkes, N. R., Haynes, C. A., & Kilburn, D. G. (2000). Analysis of binding of the family 2a carbohydrate-binding module from *Cellulomonas fimi* xylanase 10A cellulose: specificity and identification of the functionally important amino acid residues. *Protein Engineering*, *13*, 801-809.

49. Michel, G., Chantalat, L., Duee, E., Barbeyron, T., Henrissat, B., Kloareg, B., & Dideberg, O. (2001). The kappa-carrageenase of P. carrageenovora features a tunnel-shaped active site: a novel insight in the evolution of Clan-B glycoside hydrolases. Structure, 9(6), 513–25. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11435116

50. Montanier, C., Lammerts van Bueren, A. L., Dumon, C., Flint, J. E., Correia, M. A., Prates, J. A., & Gilbert, H. J. (2009). Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(9), 3065–70. doi:10.1073/pnas.0808972106

51. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*, *53*, 240-55.

52. Notenboom, V., Boraston, A. B., Chiu, P., Freelove, A. C., Kilburn, D. G., & Rose, D. R. (2001). Recognition of cello-oligosaccharides by a family 17 carbohydrate-binding module: an X-ray crystallographic, thermodynamic and mutagenic study. *Journal of Molecular Biology*, *314*(4), 797–806. doi:10.1006/jmbi.2001.5153.

53. Pickersgill, R., Smith, D., Worboys, K., Jenkins, J., & Jenkins, J. (1998). Crystal Structure of Polygalacturonase from *Erwinia carotovora ssp . J. Biol Chem, 273*(38), 24660-24664.

54. Read, S. M., Currie, G., & Bacic, A. (1996). Analysis of the structural heterogeneity of laminarin by electrospray-ionisation-mass spectrometry. *Carbohydrate Research*, *281*(2), 187–201. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8721145.

55. Receveur-Bréchot, V., Czjzek, M., Barre, A., Roussel, A., Peumans, W. J., Van Damme, & E. J. M., & Rougé, P. (2006). Crystal structure at 1.45-Å resolution of the major allergen endo-beta-1,3-glucanase of banana as a molecular basis for the latex-fruit syndrome. *Proteins*, *63*(1), 235–42. doi:10.1002/prot.20876

56. Rodríguez-Romero, A., Hernández-Santoyo, A., Fuentes-Silva, D., Palomares, L. A., Muñoz-Cruz, S., Yépez-Mulia, L., & Orozco-Martínez, S. (2014). Structural analysis of the endogenous glycoallergen Hev b 2 (endo-β-1,3-glucanase) from *Hevea brasiliensis* and its recognition by human basophils. *Acta Crystallographica. Section D, Biological Crystallography*, *70*(2), 329–41. doi:10.1107/S1399004713027673

57. Russell, R. J., Haire, L. F., Stevens, D. J., Collins, P. J., Lin, Y. P., Blackburn, G. M., & Skehel, J. J. (2006). The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. *Nature*, *443*(7107), 45–9. doi:10.1038/nature05114

58. Rye, C. S. & Withers, S. G. (2000) Glycosidase mechanisms. *Curr. Opin. Chem. Biol.*, *4*, 573-580.

59. Sabini, E., Sulzenbacher, G., Dauter, M., Dauter, Z., Jørgensen, P. L., Schülein, M., & Wilson, K. S. (1999). Catalysis and specificity in enzymatic glycoside hydrolysis: a 2,5B conformation for the glycosyl-enzyme intermediate revealed by the structure of the *Bacillus agaradhaerens* family 11 xylanase. *Chemistry & Biology*, *6*(7), 483–92. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10381409

60. Saito, J., Kita, A., Higuchi, Y., & Chem, J. B. (1999). Crystal structure of chitosanase from mechanism crystal structure of chitosanase from *Bacillus circulans* MH-K1 at 1 . 6-

Å resolution and its Substrate recognition mechanism. *J. Biol Chem*, *274*(43), 30818-30825.

61. Sakon, J., Irwin, D., Wilson, D. B., & Karplus, A. (1997). Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. *Nature Publishing Group, 4*(10), 810-818.

62. Samuels, A. L., Giddings, T. H., & Staehelin, L. A. (1995). Cytokinesis in tobacco BY-2 and root tip cells: a new model of cell plate formation in higher plants, *Journal of Cell Biology, 130*(6), 1345–1357.

63. Siguier, B., Haon, M., Nahoum, V., Marcellin, M., Burlet-Schiltz, O., Coutinho, P. M., & Dumon, C. (2014). First structural insights into α-L-arabinofuranosidases from the two GH62 glycoside hydrolase subfamilies. *The Journal of Biological Chemistry*, *289*(8), 5261–73. doi:10.1074/jbc.M113.528133

64. Staehelin, L. A., & Hepler, P. K. (1996). Cytokinesis in higher plants. *Minireview*, *84*, 821–824.

65. Stone, B. A., & Clarke, A. E. (1992). Chemistry and Biology of (1,3)- β –glucan, *La Trobe University Press, Bundoora, Australia*, 283-364.

66. Sunderasan E., Hamzah S., Hamid S., Ward W. A., Yeang H. Y., & Cardosa M. J. (1995) Allergenic proteins of *Hevea brasiliensis* latex fractions. *J Nat Rubb Res*, *10*, 82-99.

67. Tilbeurgh, H. V., Tomme, P., Claeyssens, M., Bhikhabhai, R., & Pattersson G. (1986). Limited proteolysis of the cellobiohydrolase I from *Trichoderma reesei*. *Federation of European Biochemical Societies*, *204*, 223-227.

68. Tomme, P., Van Tilbeurgh H., Pettersson, G., Van Damme, J., Vandekerckhove, J., Knowles, J., Teeri, T., & Claeyssens, M. (1988). Studies of the cellulolytic system of *Trichoderma reesei* QM9414. *Eur. J. Biochem*, *170*, 575-581.

69. Tomme, P., Warren, R. A. J., & Gikes, N. R. (1995). Cellulose hydrolysis by bacteria and fungi. *Advances in Microbial Physiology*, *37*, 1-66.

70. Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y., & Steitz, T. A. (1996). Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. *The EMBO Journal*, *15*(21), 5739–51. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=452321&tool=pmcentrez&rendertype=abstract

71. Trofimov, A. A., Polyakov, K. M., Tikhonov, A. V., Bezsudnova, E. Y., Dorovatovskii, P. V, Gumerov, V. M., & Popov, V. O. (2013). Structures of β-glycosidase from *acidilobus saccharovorans* in complexes with tris and glycerol. *Doklady. Biochemistry and Biophysics*, *449*(3), 99–101. doi:10.1134/S1607672913020129

72. Vaguine, A. A., Richelle, J., & Wodak, S. J. (1999). SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallographica. Section D, Biological Crystallography*, *55*(Pt 1), 191–205. doi:10.1107/S0907444998006684.

73. Van Petegem, F., Collins, T., Meuwis, M. A., Gerday, C., Feller, G., & Van Beeumen, J. (2003). The structure of a cold-adapted family 8 xylanase at 1.3 Å resolution. Structural adaptations to cold and investgation of the active site. *The Journal of Biological Chemistry*, *278*(9), 7531–9. doi:10.1074/jbc.M206862200.

74. Varghese, J. N., Garrett, T. P., Colman, P. M., Chen, L., Høj, P. B., & Fincher, G. B. (1994). Three-dimensional structures of two plant beta-glucan endohydrolases with distinct substrate specificities. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(7), 2785–9. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=43455&tool=pmcentrez&rendertype=abstract

75. Vonrhein, C., Blanc, E., Roversi, P., & Brincogne, G. (2007). Macromolecular Crystallography Protocols. *Volume 2: structure determination.* (pp. 149-158) Totowa, NJ.

76. Wagner, S., Radauer, C., Hafner, C., Fuchs, H., Jensen-Jarolim, E., Wüthrich, B., & Breiteneder, H. (2004). Characterization of cross-reactive bell pepper allergens involved in the latex-fruit syndrome. *Clinical and Experimental Allergy : Journal of the British*

Society for Allergy and Clinical Immunology, *34*(11), 1739–46. doi:10.1111/j.1365-2222.2004.02103.x.

77. Wang, C. C., Lee, J. C., Luo, S. Y., Kulkarni, S. S., Huang, Y. W., Lee, C. C., & Hung, S. C. (2007). Regioselective one-pot protection of carbohydrates. *Nature*, *446*(7138), 896–9. doi:10.1038/nature05730

78. Wargacki, A. J., Leonard, E., Win, M. N., Regitsky, D. D., Santos, C. N. S., Kim, P. B., & Yoshikuni, Y. (2012). An engineered microbial platform for direct biofuel production from brown macroalgae. *Science (New York, N.Y.)*, *335*(6066), 308–13. doi:10.1126/science.1214547.

79. Wu, H. M., Liu, S. W., Hsu, M. T., Hung, C. L., Lai, C. C., Cheng, W. C., & Wang, W.-C. (2009). Structure, mechanistic action, and essential residues of a GH-64 enzyme, laminaripentaose-producing beta-1,3-glucanase. *The Journal of Biological Chemistry*, *284*(39), 26708–15. doi:10.1074/jbc.M109.010983.

80. Yamamoto, M., Ezure, T., Watanabe, T., Tanaka, H., & Aono, R. (1998). C-terminal domain of β -1,3-glucanase H in *Bacillus circulans* IAM1165 has a role in binding insoluble β -1,3-glucan. *FEBS*, *433*, 41-43.

81. Yu, W., Jiang, Y., Pikis, A., Cheng, W., Bai, X., Ren, Y., Thompson, J., Zhou, C., & Chen, Y. (2013) Structural Insights into the Substrate Specificity of a 6-Phospho-β-glucosidase BglA-2 from *Streptococcus pneumoniae* TIGR4. *J. Biol. Chem*, *288*, 14949-14958.

82. Zhou, P., Chen, Z., Yan, Q., Yang, S., Hilgenfeld, R., & Jiang, Z. (2013). The structure of a glycoside hydrolase family 81 endo-β-1,3-glucanase. *Acta Crystallographica. Section D, Biological Crystallography*, *69*(10), 2027-38. doi:10.1107/S090744491301799X