STUDIES OF THERMOBIFIDA FUSCA LYTIC POLYSACCHARIDE MONOOXYGENASES AA10A AND AA10B FOR CLEAVAGE OF RECALCITRANT CELLULOSE

A Dissertation

Presented to the Faculty of the Graduate School of Cornell University

In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Nathan Kruer-Zerhusen

August 2016

STUDIES OF THERMOBIFIDA FUSCA LYTIC POLYSACCHARIDE MONOOXYGENASES AA10A AND AA10B FOR CLEAVAGE OF RECALCITRANT CELLULOSE

Nathan Kruer-Zerhusen, Ph. D.

Cornell University 2016

Cellulose recalcitrance remains a barrier to producing sustainable fuels and chemicals from renewable plant biomass. The cellulolytic bacterium *Thermobifida fusca* secretes lytic polysaccharide monooxygenases (LPMOs) to contribute to cellulose breakdown. LPMOs use an oxidative mechanism to synergize with hydrolytic cellulases. LPMOs may play an important role in overcoming cellulose recalcitrance, the physical properties that slow cellulose breakdown. Understanding LPMO structure and function will inform the optimization of industrial cellulase mixtures for use in biomass saccharification. The goal of my work is to explore the relationship between LPMO structure and oxidative cleavage using site-directed mutagenesis, and to determine if the mechanism of LPMO-cellulase synergism relies on changes to cellulose crystallinity.

Mutations to residues on the AA10A binding surface change both activity and binding properties, confirming the histidine brace for chelation of copper to be essential for activity. In addition to surface residues that mediate substrate interaction, the conserved cellulose binding domain was shown to be critical for activity of TfAA10B on crystalline cellulose. Kinetics of LPMO mutants were measured by performing secondary hydrolysis to simplify the product distribution to monomeric products, then quantifying using a rapid high performance liquid chromatography (HPLC) approach.

The role of LPMOs was initially thought to be amorphogensis, a disruption of crystalline cellulose structure, induced by oxidative cleavages of surface chains. This proposed amorphogenesis model of LPMO action and cellulase synergism was explored using Fourier transform infrared spectroscopy (FTIR). FTIR can penetrate residual substrates to detect changes in multiple measures of crystallinity caused by digestion. While past reports had observed significant changes to cellulose crystallinity imparted by cellulases, these changes were not reproduced after optimization of the measurement approach. Similarly, LPMOs did not induce changes to cellulose crystallinity during oxidative cleavage, suggesting an alternative mechanism of LPMO synergism with cellulases.

BIOGRAPHICAL SKETCH

Nathan Kruer-Zerhusen was born in 1989 in Erlanger, Kentucky to Tom Kruer and Karen Kruer-Zerhusen. He grew up with his older sister, Adriane, in the modernized family farmhouse. After graduating from Covington Latin high school, he briefly lived in Thailand as an exchange student. He moved to Sarasota, Florida to attend New College of Florida and graduated with his Bachelor of Arts degree in Environmental Science. At New College he chose to continue working on problems of sustainability and renewable energy. He received the opportunity to work towards a Ph.D. at Cornell University with Dr. David Wilson, in the rapidly growing field of oxidative cellulolytic enzymes.

ACKNOWLEDGMENTS

I have always felt incredibly fortunate for the opportunity to work with my advisor, Dr. David Wilson. It has been an honor to work with someone that has such an abundance of both kindness and scientific force, whose understanding and interactions shaped the growth of a field I believe in so strongly. He has always been eager to discuss new ideas and provide support when things didn't turn out as expected. Under his supervision I was afforded a lot of room to grow, and always had the freedom to explore my ideas. I would also like to thank my committee for their support.

Thank you to those I've worked with here at Cornell. Thanks to Maxim Kostylev for all the great poker games, and for teaching me new ways of getting things done. Thanks to Borja for believing in everything we worked on together. Everyone that has worked alongside me in the lab over the years; Tony, Uzma, and Kathleen for giving shape to the lab.

Special thanks to my parents, for providing so much support and all the many opportunities that prepared me for where I am. Thank you Emily for your constant love and support, and for putting up with two additional years of cold weather.

I am very grateful to the United States Department of Energy Bioenergy Research Center (BESC) for providing the funding for my work.

Table of Contents

Chapter 1. Introduction

1.1 Cellulose and cellulose-degrading enzymes

As the negative impacts of fossil fuels become clearer, renewable sources are needed to meet the growing demand for energy. Plant biomass represents a sustainable supply of feedstock material for conversion into liquid fuel and chemicals, including those used in plastics. Polysaccharides are produced by many organisms for both energy storage and structure. Structural cellulose is the most abundant of these polysaccharides, comprising up to 50% of plant dry weights (Lynd et al., 2002). Cellulose is advantageous for liquid fuel or chemical production because it can be hydrolyzed and converted to glucose monomers under moderate conditions using enzymatic cocktails. The chemical composition of cellulose is homogeneous, a lattice of linear glucose chains. Cellulose saccharification, breakdown into simple sugars, is constrained by cellulose structural stability and its location within complex cell walls. The cellulose microfibrils within lignocellulosic biomass are protected by enclosure within other polymers. This complex structure reduces the efficiency of enzymatic hydrolysis and increases the cost of plant biomass feedstock conversion to liquid fuels (Himmel et al., 2007a). While biomass pretreatment can expose the energy-dense cellulose, crystalline cellulose itself is also resistant to enzymatic breakdown. Cellulose recalcitrance remains a major barrier to using enzymatic biomass deconstruction to produce useful renewable energy at low cost.

Cellulolytic microbes secrete complex enzyme mixtures to release sugars from biomass for consumption. The secretomes of cellulolytic fungi currently used for industrial biomass saccharification contain cellulases, hemicellulases, and lignin oxidases (Adav et al., 2010). *Thermobifida fusca* is a ubiquitous cell wall degrading Gram-positive actinomycete found in aerobic soils that serves as a model organism for industrial fungal mixtures (Lykidis et al., 2007).

T. fusca can grow on crystalline cellulose as the sole source of carbon using a cellulase system that has been studied extensively (Wilson, 2004). *T. fusca* produces multiple cellulases that share similar structures to those produced by the industrially-relevant filamentous fungus *Hypocrea jecorina* (formerly *Trichoderma reesei*). In addition to cellulases, *T. fusca* also produces several redox-active auxiliary activity enzymes that contribute significantly to biomass breakdown. *T. fusca* serves as an effective model system to investigate the complex interactions between cellulases and auxiliary activity enzymes within fungal secretomes.

Lytic polysaccharide monooxygenases (LPMOs) are recently discovered auxiliary activity enzymes that have received significant attention in recent years (Hemsworth et al., 2015). LPMOs are abundant in both bacterial and fungal secretomes due to their ability to attack difficult to degrade cellulose within biomass. The novel oxidative mechanism of LPMOs provides a new approach to improving the efficiency of saccharification cocktails, and expands our understanding of biomass degradation strategies in nature. Many details of LPMO activity remain to be explored, including the role of conserved residues, the contribution of multiple LPMO domains, and the effect of LPMO activity on cellulose structure.

Cellulose is difficult to deconstruct due to its highly ordered structure. Cellulose is made up of repeating cellobiose units linked by β-1,4 glycosidic bonds (Lynd et al., 2002). Long cellulose polysaccharides can have a very high degree of polymerization (DP), with polymers made up of thousands of glucose monomers. These polysaccharide chains form parallel sheets, with the equatorial oxygen atoms of glucose rings forming inter- and intramolecular hydrogen bonds. The cellulose sheets created through noncovalent bonds are ordered into a lattice with adjacent sheets, interacting through van der Waals forces and planar stacking (Chundawat et al., 2011). Chains at the core of cellulose microfibrils form a crystalline lattice due to the

stabilization by this network of interactions. In comparison, surface chains are more solvent exposed, making them more disordered and amorphous. Cellulose microfibrils overlap to form a matrix that provides structural rigidity to plant cell walls. Cellulose is structurally stable due to the network of noncovalent interaction present at multiple scales.

These chemical interactions that provide structural stability also provide recalcitrance, or resistance to both chemical and enzymatic degradation (Chundawat et al., 2011). The tight packing of cellulose microfibers prevent access to the β-1,4-glycosidic bonds that link glucose components. Under ideal conditions, glycosidic bonds within intact cellulose have a theoretical half-life of 5 million years (Quinlan et al., 2011). Cellulose within plants that serves a structural role is characterized by a high degree of recalcitrance. Recalcitrance is defined as the chemical and physical properties preventing the microbial, enzymatic, or chemical degradation of polysaccharides within or derived from biomass. Recalcitrance is a major limitation to enzymatic saccharification of insoluble polysaccharides (Himmel et al., 2007a). Biomass recalcitrance is not fully understood, and is an aggregate phenomenon attributed to cellulose crystallinity and structural complexity, reduced accessibility due to hemicellulose and lignin, and unknown factors (Park et al., 2010).

Hemicelluloses are essential biomass polysaccharides that are more structurally complex but less recalcitrant than cellulose. Hemicelluloses are branching polymers made up of saccharide monomers including xylans, mannans and glucans that are readily digested by hemicellulases or acid (Adav et al., 2010). Within biomass, hemicellulose interacts directly with cellulose microfibers to form connections between adjacent fibers, providing a site for proteinaceous pectin cross-linking to other matrix components (Himmel et al., 2007a). Lignin is a more complex polymer of variable aromatic subunits that surrounds cellulose to impart

structural strength to biomass. Lignin is stabilized by extensive covalent bonding within its structure, and from interactions with hemicellulose (Zhao et al., 2012). During industrial saccharification, hemicellulose and lignin are removed from biomass before enzymatic digestion through thermal, chemical, or thermochemical pretreatment. Pretreatment of biomass using heat and chemical solvents improves enzymatic hydrolysis by exposing cellulose microfibers, but may also impede saccharification by negatively altering substrate structure and producing inhibitory chemicals.

The auxiliary activity enzymes produced by cellulolytic microbes are thought to contribute to overcoming the recalcitrance of lignocellulosic substrates. Several families of LPMOs cleave cellulose and other crystalline polysaccharides to enhance industrial cocktails. LPMOs were thought to introduce disorder in cellulose, increasing the amount of easily-degraded amorphous cellulose for cellulase attack. Understanding the structure and function of LPMOs, as well as the way they contribute to cellulase activity, are promising avenues for improvement of saccharification cocktails.

1.2 Cellulose and cellulase kinetics

Cellulases are a diverse group of glycoside hydrolases that contribute the majority of activity to degrade cellulose substrates (Vuong and Wilson, 2010). Lytic polysaccharide monooxygenases (LPMOs) improve the effectiveness of cellulolytic mixtures using a fundamentally different mechanism. Cellulases are the foundation of enzymatic biomass hydrolysis, but both interact with the same cellulose substrates and have kinetics affected by the same factors. Cellulases genes are found only in a small fraction of microorganisms and in some animals (Mba Medie et al., 2012), (Smant et al., 1998). Cellulases have multiple modes of

activity, categorized by the Carbohydrate Active enZYmes database (www.CAZy.org) into 14 distinct families based on structural folds (Levasseur et al., 2013). Some cellulases with different modes of attacking cellulose show synergism and also exhibit synergistic activity with LPMOs.

Cellulases have diverse modes of activity but share a common hydrolytic mechanism. Cellulases can perform random cleavages within cellulose chains (endo-), make multiple processive cleavages after end-on attack (exo-), or exhibit processive cleavage after endo- attack. The active site of cellulases can accommodate a single polysaccharide chain, either in a tunnel for exocellulases or in a cleft for endocellulases (Breyer and Matthews, 2001). Cellulases perform hydrolytic cleavage using the acid-base chemistry of conserved residues to catalyze the addition of water to polysaccharide β-1,4 glycosidic bonds (Walker and Wilson, 1991). Within the active site, conserved catalytic residues, typically glutamate or aspartic acid, interact with water and the scissile bond of the positioned polysaccharide chain to catalyze the cleavage reaction (Zechel and Withers, 2000). Hydrogen bonding networks in the cellulase active site adjust the pKa of the catalytic residues for optimal activity at the proper pH for organism growth (Zhou et al., 2004). Since the active site residues are buried within the cellulose structure, the separation and binding of a single cellulose chain is a limiting step for cellulose hydrolysis.

Cellulases often contain multiple modular domains, typically catalytic and binding domains separated by flexible a linker. Carbohydrate binding modules (CBMs) are small globular domains that anchor the catalytic domain to the insoluble substrate (Wilson, 2008). Conserved structural folds of CBMs are grouped into CAZy families. CBMs are found in diverse cell wall degrading enzymes including LPMOs, where they augment catalytic domain binding. CBMs that target cellulose have aromatic residues on a flat binding surface to form hydrophobic stacking interactions with pyranose rings (Hashimoto, 2006). CBM binding is tight but reversible

(Jung et al., 2003). Specialized CBMs such as the family 3c CBM of *T. fusca* Cel9A help extract individual cellulose chains into the active site, but CBMs are not thought to affect cellulose structure or crystallinity (Sakon et al., 1997).

Cellulases have nonlinear kinetics caused by changes to the insoluble polysaccharide substrate, rather than changes to the enzymes themselves. Cellulase rate constants decline because the substrate becomes less digestible over time, not because the cellulases are inactivated (Zhang et al., 1999). Cellulases have high activity when degrading soluble substrates such as short oligosaccharides and colorimetric analogs such as 2,4-dinitrophenyl-β-D-cellobioside. *T. fusca* cellulases exhibit classical Michaelis-Menten kinetics on these soluble and homogeneous substrates, producing typical K_m and V_{max} kinetic parameters (Vuong and Wilson, 2009). However, engagement with increasingly difficult substrates slows their kinetic rate and results in nonlinear kinetics (Zhang et al., 1999).

Classical Michaelis-Menten kinetics assumes substrate homogeneity, complicating its application to cellulase activity on insoluble substrates, so another kinetic model must be used to characterize this activity. Fractal-like kinetics describes systems with diffusion over heterogeneous three dimensional substrates such as cellulose (Zhang et al., 1999). Kinetic models that use fractal-like kinetics can accommodate changing rate constants that reflect changing cellulase-substrate interactions over the course of digestion (Väljamäe et al., 2003), (Kostylev and Wilson, 2013). LPMOs interact with the same heterogeneous and changing cellulose substrates as cellulases, and thus exhibit the same nonlinear kinetics which require fractal-like kinetics.

The relationship between cellulose recalcitrance and nonlinear kinetics is explored using model cellulose substrates with varying recalcitrance. Model cellulose lacks the hemicellulose

and lignin that influence cellulase activity on biomass, so only the structural properties of chemically similar cellulose affect cellulase kinetic rates. For both cellulases and LPMOs, the structure of cellulose dictates activity. Avicel® is a highly crystalline purified plant cellulose that is highly resistant to degradation by individual cellulases, slowing kinetic rates significantly (Watson et al., 2002a). Bacterial cellulose (BC) produced by *Acetobacter xylinum* allows for faster digestion despite high crystallinity, due to the large surface area from its ribbon-like structure (Lynd et al., 2002). Phosphoric acid swollen cellulose (PASC) is easily degraded due to the chain order disruption introduced by solubilization and recrystallization (Zhang et al., 2006). Model cellulose allows kinetic properties to be measured on a variety of substrates, enabling the use of the proper substrate to explore the effect of recalcitrance on activity.

Cellulases must extract and bind a single cellulose chain into their active site, and accessible substrate becomes limited as the substrate recalcitrance increases. Cellulase systems depend on mixtures of cellulases that have complementary interactions to overcome cellulose recalcitrance. Since multiple cellulases can interact with the same substrate in complementary ways, cellulase synergism results in the combination of cellulases producing activity that is higher than the sum of individual cellulases alone (Wilson and Kostylev, 2012). Synergism between cellulases is only observed on recalcitrant substrates, and does not require the cellulases to interact physically. The magnitude of synergism decreases as protein loading increases, with the largest effect at protein loading far below saturation on the substrate (Watson et al., 2002b). Enzymes including β-glucosidase and β-xylosidase assist cellulases by relieving product inhibition caused by disaccharide accumulation, but do not interact with the insoluble substrate (Kostylev and Wilson, 2012). LPMOs synergize effectively with hydrolytic cellulases, using an oxidative mechanism to interact with highly recalcitrant regions of cellulose.

Several models of cellulase synergism have been developed, each describing the mechanism through which cellulases create substrate for others to attack. The endo-exo model described by Wood and McCrae describes an endocellulase making cleavages on exposed surface chains. The new ends generated by these endocellulase cleavages can then be bound by exocellulases to cleave processively (Wood and McCrae, 1979). In the exo-exo model, exocellulases directed towards opposite ends of cellulose chains operate simultaneously, attacking different regions of the same substrate to clear impassible blockages (Tomme et al., 1990). An additional model of cellulase synergism on crystalline cellulose has been proposed which involves cellulose surface changes. In this model, a processive endocellulase cleaves an intact crystalline surface, and the resulting surface grooves are attacked by a processive exocellulase (Kostylev and Wilson, 2012). This mechanism for synergism on crystalline cellulose is enabled by the ability of processive endocellulase to attack crystalline surfaces, an ability shared with the oxidative mechanism of LPMOs.

Cellulose has a range of physical properties, and cellulase kinetics are determined by the ability of cellulases to overcome cellulose recalcitrance and heterogeneity. Synergistic cellulases overcome cellulose recalcitrance using complementary mechanisms, essential interactions that provide the foundation for the synergism of LPMOs with cellulases. The multiple models of hydrolase synergism are based on complementary activity producing enhanced substrate availability. The enhancement of saccharification by synergistic cellulose mixtures thus provides an important foundation for understanding the mechanistic basis of LPMO synergism.

1.3 Substrate recalcitrance and enzyme synergism

Early models of cellulase synergism focused on cellulose crystallinity as the source of recalcitrance that prevented cellulose breakdown (Eriksson, 1982). LPMOs differ from hydrolytic cellulases in their ability to attack crystalline cellulose, providing a novel avenue for degradation through synergism with cellulases. LPMOs fit into this model of synergism by inducing amorphogenesis, a structural change to cellulose that produces a more easily digested substrate (Arantes and Saddler, 2010; Kostylev and Wilson, 2013). Coughlan defined amorphogenesis as the changes to insoluble substrates that reduce crystallinity and increase substrate accessibility (Coughlan, 1985). Amorphogenesis is a physical dispersion, swelling, or delamination of ordered materials that weakens hydrogen bonding of cellulose chains, but leaves the chemical composition unchanged (Coughlan, 1985). Amorphogenesis reduces crystallinity, providing more cellulose active sites for hydrolytic cellulases. Fourier Transform infrared spectroscopy (FTIR) is a tool that can measure the crystallinity of cellulose in order to directly measure changes to cellulose crystallinity resulting from enzymatic degradation.

Substrate recalcitrance is an essential structural property of plant biomass that prevents breakdown into usable sugars (Park et al., 2010). Recalcitrance is beneficial for plants, as it enables plants to resist deconstruction and digestion by microbes or animals (Himmel et al., 2007b). Many components other than cellulose contribute to cell wall structural integrity, and recalcitrance of plant biomass is more complex than cellulose crystallinity (Zhao et al., 2012). However, there is a strong link between cellulose crystallinity and limited cellulase digestion, especially in model cellulose substrates. Recalcitrance is always observed to varying extents when cellulose microfibrils have highly crystalline structure (Beckham et al., 2011). The crystalline structure of cellulose has been proposed to be the primary challenge of biomass

deconstruction, and thus the key limiting factor to economically feasible enzymatic saccharification (Wyman et al., 2005).

Cellulase activity correlates with substrate crystallinity, as cellulose chains must be accessible to enzymes for hydrolysis to occur. When cellulose chains are buried in a crystalline lattice, or are inaccessible due to pores smaller than the diameter of cellulases, cellulases will have decreased kinetic rate on that substrate (Arantes and Saddler, 2010). This was demonstrated experimentally by showing the initial rate of endocellulases was inversely correlated with substrate crystallinity (Hall et al., 2010). Cellulases with synergistic mechanisms can interact to increase net activity despite substrate crystallinity. Cellulase synergism increases with increasing substrate recalcitrance, due to the greater benefit of creating new active sites. The degree of synergistic interaction (DSE) between cellulases is the highest when the cellulose substrate has a high degree of polymerization and medium crystallinity index (CI), when crystallinity is low or very high the DSE between cellulases declines (Jalak et al., 2012). Synergism improves if the amorphous regions of the substrate are removed, supporting this model of cellulase synergism (Jeoh et al., 2006).

Since cellulases remove cellulose during digestion, cellulose structure changes in meaningful ways. Cellulose structure changes to different extents depending on whether digestion is performed by a single cellulase or a synergistic mixture (Mansfield and Meder, 2003). The connection between cellulase activity and substrate recalcitrance (as crystallinity) can be measured with varying sensitivity using several techniques, including Fourier transform infrared spectroscopy (FTIR), X-ray diffraction (XRD), and nuclear magnetic resonance spectroscopy (NMR) (Park et al., 2010). Several reports have shown that differences in FTIR crystallinity indices correlate with cellulose digestibility (Chen et al., 2007). Cellulase digestion

was observed to increase cellulose crystallinity, using characteristic bonds in FTIR spectra to measure substrate order (Corgié et al., 2011). The link between FTIR crystallinity indexes and decreasing cellulase activity provided support that increasing substrate crystallinity correlates with increasing recalcitrance (Corgié et al., 2011). However, it remains unclear whether cellulose crystallinity causes recalcitrance or if it is only correlated. While crystallinity as a physical parameter correlates with reduced digestibility, FTIR can clarify whether changes to crystallinity caused by cellulase activity increase cellulose recalcitrance (Park et al., 2010).

The concept of amorphogenesis, disruption resulting from enzyme action on crystalline cellulose regions, was first proposed by Reese et. al in the C_1/C_x model of cellulase synergism. This model of cellulose hydrolysis involves an initial disrupting factor, C_1, that makes the crystalline cellulose more accessible to the degrading factors, C_x (Reese et al., 1950). This substrate change would increase cellulase access over a greater fraction of the substrate, providing a significant increase to both the kinetic rate and extent of digestion for microbial systems. This model predicts that microbial secretomes contain an enzymatic swelling factor that causes amorphogenesis of crystalline cellulose regions (Reese et al., 1950). However, the hydrolytic cellulases that have been investigated for initial substrate disruption do not appear to be the C_1 swelling factor.

Cellulase activity increases as cellulose crystallinity is disrupted, and other enzyme families have been investigated for their role in reducing substrate crystallinity to increase the amount of amorphous cellulose (Arantes and Saddler, 2010). Examples of enzymatic cellulose amorphogenesis exist in both plants and fungi. Expansins are a family of plant proteins that induce the loosening of plant cell walls to enable plant cell growth. Expansins have two polysaccharide binding domains, each with an open binding surface. Addition of plant expansins

to cellulase mixtures enhances crystalline cellulose degradation, with a greater effect at low cellulase loading (Arantes and Saddler, 2010). Expansin-like proteins are expressed by cellulolytic fungi, such as the swollenin family of *H. jecorina*. Swollenins enhance cellulase activity on recalcitrant cellulose through an unknown non-hydrolytic mechanism. The stimulation despite lack of hydrolysis suggests a change to the cellulose substrate to reduce recalcitrance and enable increased cellulase activity (Baker et al., 2000). This role would be valuable for synergistic secretomes, as chain separation from the insoluble substrate has been proposed as the rate limiting step of cellulose degradation (Wilson, 2009). Microbial cellulolytic systems would benefit from increased chain separation in recalcitrant substrates, created by enzymes such as expansins and swollenins that perform substrate disruption rather than sugar release. The presence of enzymes that change the structural properties of cellulose demonstrates that manipulating cellulose crystallinity can be an effective strategy for increasing cellulose reactivity.

Based on their synergistic interaction with hydrolytic cellulases, LPMOs are a strong candidate for a disrupting C_1 factor. LPMOs synergize with cellulases using a unique oxidative mechanism, but have limited product release on their own (Kostylev and Wilson, 2013). The mechanistic basis of LPMO stimulation of hydrolytic cellulases has been proposed to be amorphogenesis, swelling substrates to reduce crystallinity and increase digestibility (Arantes and Saddler, 2010). This hypothesis is important to understanding the role of LPMOs in cellulolytic mixtures, and more generally the substrate features that limit the activity of cellulase cocktails. If this mechanism of disruptive activity is correct, it provides an essential parameter to optimize enzyme mixtures that is mainly independent of biomass source and pretreatment. The ability of LPMOs to induce amorphogenesis in recalcitrant biomass may represent an essential

attribute of efficient and economically viable cellulase cocktails. The changes to cellulose structure caused by cellulase and LPMO digestion are measurable using FTIR, clarifying the current amorphogenesis model of LPMO activity.

1.4 LPMO historical context and discovery

The economic feasibility of industrial biomass saccharification is dependent on the efficiency of cellulase cocktails, minimizing the time to reach high extent of conversion. The cost of industrial cellulase mixtures has dropped significantly in recent years, coinciding with the introduction of LPMOs to enhance biomass breakdown (Cannella and Jørgensen, 2014). Despite thermochemical pretreatment of lignocellulosic biomass, crystalline cellulose microfibers remain resistant to enzymatic digestion. Synergistic cellulases can overcome some cellulose recalcitrance, but addition of novel mechanisms greatly improves conversion efficiency. A new avenue of research focuses on LPMOs based on their proven ability to attack recalcitrant substrates and synergize with hydrolytic cellulases.

During the long history of research into enzymatic biomass deconstruction, hydrolases including cellulases, hemicellulases, and esterases were thought to be the major actors (Wyman et al., 2004). Microbial redox-active enzymes were found to degrade lignin, and oxidative enzymes have more recently been found to play a significant role on insoluble polysaccharides (Palonen, 2004). These oxidative enzymes attack cellulose without a cleft-like active site or generating typical hydrolysis products, and were originally thought to be binding modules due to their similar size and compact globular shape. LPMOs have since been reclassified in the carbohydrate active enzymes (CAZy) database (Levasseur et al., 2013). Bacterial family AA10 auxiliary activity monooxygneases, formerly categorized as family 33, were first characterized as

active on chitin hydrolysis by aerobic bacteria (Vaaje-Kolstad et al., 2005a). Similar fungal LPMOs were initially classified as glycoside hydrolase 61 family (GH61), and have been relabeled as family AA9. The crystal structures of both families lack the typical cleft or tunnel active sites present in cellulases (Karkehabadi et al., 2008).

The discovery of oxidative cleavage by LPMOs was preceded by investigation of several related chitin-binding proteins (CHBs) in the late 1990s. These proteins are present in the extracellular enzyme systems of several gram-positive soil bacteria, in particular *Streptomyces spp.*, that have similar phenotypes, ecological roles, and crystalline polysaccharides substrate utilization as the gram-positive model organism *Thermobifida fusca*. These chitin-binding proteins were small, had varied substrate specificities, and were secreted in large amounts during growth on crystalline polysaccharides. Though they were first characterized as having a putative substrate binding function, they are now known to be LPMOs.

CHB1 of *Streptomyces olivaceoviridis*, an 18.7 kDa protein induced by growth on chitin, was the first of these proteins to be characterized. CHB1 binds only α-chitin, mediated in part by a conserved tryptophan based on spectroscopy and mutagenesis (Schnellmann et al., 1994). CHB2 of *Streptomyces reticuli* is very similar in size at 18.6 kDa, and also binds only α-chitin including chitin derived from fungi (Kolbe et al., 1998). CHB3 from *Streptomyces coelicolor* is a smaller CHB at 14.9 kDa, and binds to both α- and β- forms of chitin as well as soluble chitosan (a less crystalline form of chitin). CHB3 is related to the other CHBs, but shares less sequence identity with the others (Saito et al., 2001). CHBB of *Bacillus amyloliquefaciens,* 19.8 kDa, is similar to CHB3 in its capacity to bind both α and β-chitin (with a preference for the β form), and while it can bind with some affinity to cellulose it cannot bind soluble chitosan (Chu et al., 2001). Interest in these secretome components existed prior to the discovery of their LPMO

oxidative cleavage activity.

The first LPMO in the AA10 family shown to contribute to chitin digestion was smCBP21 from the chitinolytic gram-negative soil bacterium *Serratia marcescens*. smCBP21 is similar in size to the other chitin binding domains at 18.8 kDa and shows high sequence identity to CHB1 despite *S. marcescens* being phylogenetically distant from *Streptomyces spp.* The binding properties of smCBP21 to crystalline chitin are also similar, though smCBP21 binds preferentially to β-chitin (Suzuki et al., 1998). smCBP21 was the first member of this family whose crystal structure was solved, revealing a fibronectin-like fold with a bud-like extension to form a flat binding surface (Vaaje-Kolstad et al., 2005b). The structure revealed that many of the conserved aromatic residues thought to play a role in binding (similar to those found on the binding surface of cellulose-directed carbohydrate binding modules) were actually buried deep in the core of the protein. Several conserved polar surface residues were investigated through site directed mutagenesis and found to be critical for mediating smCBP21 binding (Vaaje-Kolstad et al., 2005a).

These chitin-binding proteins are similar in upregulation, secretion and binding to the much smaller antifungal protein Afp1 of *Streptomyces tendae*. Afp1 can bind to α-chitin and chitosan but not cellulose, akin to CHB3 of *Streptomyces coelicolor* (Bormann et al., 1999). However, the antifungal proteins are distinct from chitin binding proteins, with Afp1 having much smaller size at 9.8 kDa, different structural fold (having more sequence homology with hydrolase CBMs), and inherent antifungal ability which the other CHBs lack (Campos-Olivas et al., 2001). Afp1 is smaller than the CHBs, but is not as small as the antifungal chitin binding lectins secreted by gram positive bacteria (Bormann et al., 1999).

This coregulation between LPMOs (formerly CHBs) and antifungal proteins suggests that

LPMOs could play a role in antifungal host defense. CHBs such as CHB1 bind α-chitin and cover the fungal chitin to form a 'glue-like' covering, which may facilitate physical interactions between bacteria and fungal hyphae. Originally, the *S. olivaceoviridis* CHB1 was proposed to act antifungally through binding to the chitin at growing hyphal tips. The proposed role was non-hydrolytic disruption through binding to underlying microfibrils to release additional surface chains for exochitinase attack, similar to the nonhydrolytic disruption effect proposed for CBMs (Schnellmann et al., 1994). Bacterial exochitinases alone have been observed to effectively destroy chitinous fungal hyphae, so it is not known if additional disruption would be beneficial (Schrempf, 2001). Plants produce similar pathogenesis-related (PR) proteins, including the PR-4 family of small chitin-binding antifungal proteins (Theis and Stahl, 2004). The mechanism of fungal inhibition is unknown, but has been hypothesized to involve cell wall weakening at the fungal hyphae akin to the proposed bacterial CHB mechanism (Selitrennikoff, 2001). Plant PR-4 proteins are similar in size but are not structurally related to bacterial antifungal chitin binding proteins, and further investigation is required to explore whether they have LPMO activity (Bertini et al., 2012).

Our current understanding of LPMO structure and function is built upon investigation of diverse proteins with varied roles. The wide diversity of proteins that share structural similarity with LPMOs suggests important microbial roles beyond biomass deconstruction, which remain to be explored. Despite their sequence similarity, their varied binding properties show how surface residues impart binding diversity among LPMOs. Each protein is specific but limited in its substrate binding, usually targeted towards one form of crystalline chitin and most being unable to bind to cellulose or soluble chitosan. SoCHB1, SrCHB2, ScCHB3, BaCHBB, and SmCBP21 are all now categorized as AA10 proteins, while StAfp1 is not in that family

(http://www.cazy.org, 2013). It is unclear whether LPMOs contribute to antifungal host defense, or whether they are contributing to substrate breakdown. The extensive characterization of LPMO binding properties set the stage for the current exploration of LPMO oxidative cleavage of chitin and cellulose. In addition, they provide avenues of host/pathogen research that may offer insights into the mechanisms of LPMOs targeting biomass.

1.5 LPMO structure and mechanism

Fungal and bacterial LPMOs were investigated simultaneously after their oxidative cleavage activity was discovered. Both fungal and bacterial LPMOs were shown to have low endoglucanase activity before their essential cofactors were added (Karkehabadi et al., 2008). The LPMO-rich *T. terrestris* secretome significantly stimulated activity of *H. jecorina* cellulase mixtures that lacked strong LPMO activity (Harris et al., 2010). The first oxidative cleavage of a bacterial LPMO activity was also observed indirectly, smCBP21 stimulated chitinase activity when external reducing agents and molecular oxygen were present (Vaaje-Kolstad et al., 2010b). The first characterization of oxidative activity by a purified LPMO represented a major step forward in understanding (Vaaje-Kolstad et al., 2010b). Cellulose oxidative cleavage by a bacterial LPMO was first demonstrated by *Streptomyces coelicolor* A3 CelS2, which produced aldonic acid cello-oligosaccharides similar to those produced from chitin-active LPMOs (Forsberg et al., 2011). LPMOs were identified as metalloenzymes since EDTA completely inhibited oxidative activity. Despite initial confusion on the identity of the transition metal, copper was eventually found to be essential (Vaaje-Kolstad et al., 2012).

Initial exploration of the two *Thermobifida fusca* LPMOs showed their induction on multiple biomass substrates, and high binding ability on pure cellulose substrates(Moser et al.,

2008). These proteins were shown to be monooxygenases when they were incorporated into multi-enzyme cellulosomes. The patterns of oxidized products produced from Avicel cleavage indicated that the two LPMOs created differing products, putting them into separate subfamilies (as discussed further below) (Arfi et al., 2014). Additionally, both TfAA10A and TfAA10B produce oxidized products when incubated with bacterial cellulose (BC) in the presence of reducing agents (Forsberg et al., 2014). TfAA10B appears to have higher activity than TfAA10A, releasing significantly more soluble oxidized oligosaccharides when assayed alone on Avicel, mostly due to the presence of an additional cellulose binding module (Arfi et al., 2014). These oxidized products have a range of lengths when measured by high-performance anion-exchange chromatography (HPAEC), and this was confirmed using mass spectrometry (Forsberg et al., 2014). While *T. fusca* LPMOs are confirmed as monooxygenases, their kinetic parameters, domain interactions, and roles in the *T. fusca* secretome have not been fully explored.

The initial investigation into LPMOs in fungal crude mixtures showed that LPMOs were only active on pretreated biomass, and not on pure cellulose (Harris et al., 2010). LPMOs require external reducing power to provide electrons that maintain the catalytic cycle. In the case of biomass substrates the reducing agent was gallate, a small molecule cofactor released from pretreated corn stover (Quinlan et al., 2011). Either a small molecule reducing agent such as ascorbate, lignin components of pretreated biomass, or a protein partner such as the flavoprotein cellulose dehydrogenase (CDH) can serve as reducing agents (Cannella et al., 2012), (Langston et al., 2011) Cellulose dehyrogenase is produced by most cellulolytic fungi, and is able to activate multiple families of LPMOs (Hemsworth et al., 2015). An analogous enzymatic source of reducing power in bacteria, quinone-containing proteins with YceI-like domains, is being explored (Hemsworth et al., 2015). In experimental systems, small molecule reductants such as

glutathione and ascorbate provide reducing power to ensure maximum LPMO activity.

As LPMOs with diverse properties have been characterized, they have been grouped by their mechanistic similarities. The first oxidative cleavage products of smCBP21 were characterized using MALDI-TOF mass spectrometry, which showed both native and oxidized sugars with a range of lengths. These products were aldonic acids resulting from the opening of lactone rings after oxidation of carbohydrates at the C1 position (Vaaje-Kolstad et al., 2010b). Oxidation of the C4 carbon on the non-reducing end of the β-glycosidic bond creates 4-keto products, which are readily hydrolyzed to create the gem-diol form (Forsberg et al., 2014). As detection has become more sophisticated, LPMOs have been classified based on their product regioselectivity, providing fine-grained distinctions within LPMO families. Vu et. al categorized auxiliary activity monooxygenases into groups I-III based on regioselectivity, which is linked to their phylogeny (Vu et al., 2014). LPMOs that exhibit C1-oxidation preference are PMO type-I, C4 targeting LPMOs are type-II, and those that produce both are type-III (Vu et al., 2014). Many examples of LPMOs producing C1 and C4 oxidized products have been characterized, and C6 oxidation has been reported but not extensively explored (Hemsworth et al., 2015). In addition to cellulose and chitin, novel LPMO families may target other substrates. An AA13 LPMO has a surface binding groove to enable binding to the amylose chains of starch. Additionally, AA9 LPMOs have recently been shown to have activity on soluble cellooligosaccharides, a novel activity for a formerly insoluble substrate-directed family (Hemsworth et al., 2015).

LPMOs perform a novel type of redox chemistry, oxidation of an intact polysaccharide C-H bond. While computational models describe the LPMO oxidative mechanism, it has not been confirmed experimentally. LPMO oxidative cleavage proceeds by creating a highly reactive radical oxygen species that removes the exposed hydrogen adjacent to a β-1,4 glycosidic bond.

The catalytic cycle of LPMO oxidative cleavage begins with the binding of molecular oxygen to the active site copper and its activation to create a Cu(II)-superoxide intermediate (Kim et al., 2014). The superoxide gains two protons and two electrons from an external reducing agent, water is formed and dissociates, while the active Cu(II)-oxyl catalyst remains. Once positioned adjacent to the substrate hydrogen, the Cu(II)-oxyl abstracts hydrogen from the substrate carbon in the rate limiting step to form Cu(II)-OH. The carbon radical created by hydrogen abstraction is resolved through an 'oxygen rebound' where the copper bound OH group transfers to the substrate. Solvent water cleaves the newly oxidized bond via elimination, producing the C1 or C4 oxidized product (Kim et al., 2014). There is uncertainty as to whether the superoxide radical performs the abstraction of the substrate hydrogen directly, due to the unknown order of electron transfer from the reducing agent. Current models using density functional theory (DFT) show that the reaction pathway involving the Cu(II)-oxyl intermediate is energetically favorable and thus more likely (Kim et al., 2014). Both mechanism pathways result in the return of copper to the activated Cu(I) state, which may be hazardous if not controlled. In the absence of substrate, superoxide intermediates generated in solution dissociate as peroxide radicals that could damage enzyme structure (Kjaergaard et al., 2014).

1.6 LPMO inactivation and characterization

While LPMOs are similar to cellulases in domain arrangement and substrate preference, their enzymology is more complex due to their transition metal mediated redox chemistry. Enzyme mixtures containing high LPMO activity are inactivated due to these side products, which is prevented by the addition of a peroxidase (Scott et al., 2015). While cellulases and LPMOs share the same kinetic limitations due to a changing insoluble substrate, LPMOs also

require the presence of a cofactor. Finally, LPMOs create product distributions that are far more diverse than hydrolytic cellulases due to their pseudo-random cleavage and inability to cleave oligomers. These products must be simplified for accurate LPMO quantification. Additional experimental factors influence the measurement of LPMO activity.

The superoxide radical created from molecular oxygen at the start of a productive LPMO catalytic cycle is typically terminated by reduction and attack of the adjacent carbohydrate substrate (Kim et al., 2014). However, the active site copper can activate oxygen in the absence of substrate, potentially releasing the oxygen radical as a side product (Kittl et al., 2012). Computational models support this phenomenon, but the mechanism and specific peroxide species have not been confirmed (Kjaergaard et al., 2014). This LPMO side reaction was experimentally observed indirectly, suggesting hydrogen peroxide (H_2O_2) to be the likely candidate for the released oxygen radical (Kittl et al., 2012). Hydrogen peroxide inactivates cellulase activity at low concentrations, similar to the inhibition observed when high LPMO activity is present (Scott et al., 2015). The accumulation of oxygen radical side products poses a significant challenge to preserving enzyme activity in both industrial reactions and kinetic experiments.

The solution to enzyme inactivation by LPMO peroxide side products is detoxification using a suitable peroxidase. Catalase is an abundant and well characterized peroxidase produced by numerous organisms in response to oxidative stress, which dismutates H_2O_2 into water and molecular oxygen (Scott et al., 2015). Catalase is a monofunctional peroxidase with high substrate specificity for hydrogen peroxide, mediated by a porphyrin ring in each of the tetrameric monomers (Diaz et al., 2001), (Gruft et al., 1978). While it is effective at removing H_2O_2, catalase is sensitive to inhibition and inactivation by many things, including H_2O_2 for

some less stable catalases (Ghadermarzi et al., 1999). The *T. fusca* genome contains two catalase genes in addition to the genes for two secreted LPMOs (Lončar and Fraaije, 2015). The protective effect of catalase in experimental systems was demonstrated using a thermostable catalase from *Thermoascus aurantiacus* (Scott et al., 2015).

To show individual LPMO turnover rates, the initial kinetics of aldonic acid production were reported for GbpA, a chitin-acting LPMO (Loose et al., 2014). The product distribution of the *T. fusca* LPMOs was similar to other characterized LPMOs (Arfi et al., 2014). Time course data from *T. fusca* LPMOs have been reported, but no kinetic parameters were described (Forsberg et al., 2014). The two parameter nonlinear kinetic fit approach provides a useful tool to compare wild type and mutant LPMO activity under the same conditions (Kostylev and Wilson, 2013). This approach has been used to compare the activity of cellulase mutants to the wild type enzyme and for synergistic cellulase combinations (Kostylev et al., 2014; Shu et al., 2014). Nonlinear time course kinetics also described the synergism of TfAA10A and Cel48A, which showed a large initial increase of activity that declined rapidly (Kostylev and Wilson, 2013). In this work, the relative importance of *T. fusca* LPMO domains and surface residues are characterized using nonlinear kinetic fits.

Measuring the activity of LPMOs involves more components than cellulase reactions, including source of reducing power, copper occupancy, and protecting reducing agents from autooxidation. The kinetic rate of LPMOs is directly linked to reducing agent concentration and the efficiency of reducing electron transfer, presenting a challenge to measuring specific rates consistently (Vaaje-Kolstad et al., 2010a). The LPMO active site must be occupied by copper for activity, and binding of the reduced copper is very tight (Aachmann et al., 2012). However, the addition of exogenous copper to maintain the metalloprotein form accelerates the autooxidation

of reducing agents and the formation of reactive oxygen species (Loose et al., 2014). Optimal LPMO activity can be maintained with exogenous copper in conjunction with excess reducing agent by using catalase to slow reducing agent oxidation while ensuring copper saturation.

Cellulases produce soluble products in characteristic product distributions, ranging from one to four glucose units. In comparison, LPMOs create products with a wide range of lengths, complicating separation and quantification of LPMO products. Rapid quantification of activity is possible by first reducing the complexity of LPMO product distributions through secondary hydrolysis of the soluble fraction. The reduction of products to monomeric neutral and oxidized oligosaccharides enables consistent and rapid detection (Cannella et al., 2012). The pattern of LPMO product distribution is likely relevant to binding and the catalytic mechanism (as the pseudo-random distribution likely results from substrate binding preference), but has not yet been fully investigated. Secondary hydrolysis of the residual substrate using an exocellulase followed by detection of oxidized products provides a measure of the turnover.

This work aims to expand the understanding of the activity and role of lytic polysaccharide monooxygenases. To do this, mutant forms of the two *Thermobifida fusca* LPMOs were characterized to explore the relationship between conserved surface residues and LPMO activity. The role of conserved surface residues in binding and activity were related to their location in a new crystal structure of *Tf*AA10A. The multiple domains of TfAA10B were characterized to determine the importance of the CBM2 and X1 domains for binding and activity. In addition, Fourier transform infrared spectroscopy (FTIR) was used to investigate the amorphogenesis model of LPMO synergism. These experiments explore enzymatic mechanisms to overcome cellulose recalcitrance by measuring cellulose crystallinity after hydrolytic and oxidative cleavage.

REFERENCES

Aachmann FL, Sørlie M, Skjåk-Bræk G, Eijsink VGH, Vaaje-Kolstad G. 2012. NMR structure of a lytic polysaccharide monooxygenase provides insight into copper binding, protein dynamics, and substrate interactions. *Proc. Natl. Acad. Sci. U. S. A.* **109**.

Adav SS, Ng CS, Arulmani M, Sze SK. 2010. Quantitative iTRAQ secretome analysis of cellulolytic Thermobifida fusca. *J. Proteome Res.* **9**:3016–3024.

Arantes V, Saddler JN. 2010. Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis. *Biotechnol. Biofuels* **3**:2–11.

Arfi Y, Shamshoum M, Rogachev I, Peleg Y, Bayer E a. 2014. Integration of bacterial lytic polysaccharide monooxygenases into designer cellulosomes promotes enhanced cellulose degradation. *Proc. Natl. Acad. Sci. U. S. A.* **111**:9109–14.

Baker JO, King MR, Adney WS, Decker SR, Vinzant TB, Lantz SE, Nieves RE, Thomas SR, Li LC, Cosgrove DJ, Himmel ME. 2000. Investigation of the cell-wall loosening protein expansin as a possible additive in the enzymatic saccharification of lignocellulosic biomass. *Appl. Biochem. Biotechnol.* **84-86**:217–23.

Beckham GT, Matthews JF, Peters B, Bomble YJ, Himmel ME, Crowley MF. 2011. Molecular-level origins of biomass recalcitrance: decrystallization free energies for four common cellulose polymorphs. *J. Phys. Chem. B* **115**:4118–27.

Bertini L, Proietti S, Aleandri MP, Mondello F, Sandini S, Caporale C. 2012. Modular structure of HEL protein from Arabidopsis reveals new potential functions for PR-4 proteins. *Biol. Chem.* **393**:1–14.

Bormann C, Baier D, Hörr I, Raps C, Berger J, Jung G, Schwarz H. 1999. Characterization of a novel, antifungal, chitin-binding protein from Streptomyces tendae Tü901 that interferes with growth polarity. *J. Bacteriol.* **181**:7421–7429.

Breyer WA, Matthews BW. 2001. A structural basis for processivity. *Protein Sci.* **10**:1699–1711.

Campos-Olivas R, Hörr I, Bormann C, Jung G, Gronenborn AM. 2001. Solution structure, backbone dynamics and chitin binding of the anti-fungal protein from Streptomyces tendae TU901. *J. Mol. Biol.* **308**:765–82.

Cannella D, Hsieh CWC, Felby C, Jørgensen H. 2012. Production and effect of aldonic acids during enzymatic hydrolysis of lignocellulose at high dry matter content. *Biotechnol. Biofuels* **5**:26.

Cannella D, Jørgensen H. 2014. Do new cellulolytic enzyme preparations affect the industrial strategies for high solids lignocellulosic ethanol production? *Biotechnol. Bioeng.* **111**:59–68.

Chen Y, Stipanovic AJ, Winter WT, Wilson DB, Kim Y-J. 2007. Effect of digestion by pure cellulases on crystallinity and average chain length for bacterial and microcrystalline celluloses. *Cellulose* **14**:283–293.

Chu HH, Hoang V, Hofemeister J, Schrempf H. 2001. A Bacillus amyloliquefaciens ChbB protein binds beta- and alpha-chitin and has homologues in related strains. *Microbiology* **147**:1793–803.

Chundawat SPS, Bellesia G, Uppugundla N, da Costa Sousa L, Gao D, Cheh AM, Agarwal UP, Bianchetti CM, Phillips GN, Langan P, Balan V, Gnanakaran S, Dale BE. 2011. Restructuring the crystalline cellulose hydrogen bond network enhances its depolymerization rate. *J. Am. Chem. Soc.* **133**:11163–74.

Corgié SC, Smith HM, Walker LP. 2011. Enzymatic transformations of cellulose assessed by quantitative high-throughput fourier transform infrared spectroscopy (QHT-FTIR). *Biotechnol. Bioeng.* **108**:1509–20.

Coughlan MP. 1985. The Properties of Fungal and Bacterial Cellulases with Comment on their Production and Application. *Biotechnol. Genet. Eng. Rev.* **3**:39–110.

Diaz A, Rangel P, Montes de Oca Y, Lledias F, Hansberg W. 2001. Molecular and Kinetic Study of Catalase-I, A Durable Large Catalase of Neurospora crassa. *Free Radic. Biol. Med.* **31**:1323–1333.

Eriksson K-E. 1982. Degradation of cellulose. *Experientia* **38**:156–159.

Forsberg Z, Mackenzie AK, Sørlie M, Røhr ÅK, Helland R, Arvai AS, Vaaje-Kolstad G, Eijsink VGH. 2014. Structural and functional characterization of a conserved pair of bacterial cellulose-oxidizing lytic polysaccharide monooxygenases. *Proc. Natl. Acad. Sci. U. S. A.* **111**:8446–51.

Forsberg Z, Vaaje-Kolstad G, Westereng B, Bunæs AC, Stenstrøm Y, MacKenzie A, Sørlie M, Horn SJ, Eijsink VGH. 2011. Cleavage of cellulose by a CBM33 protein. *Protein Sci.* **20**:1479–83.

Ghadermarzi M, Moosavi-movahedi AA, Ghadermarzi M. 1999. Influence of different types of effectors on the kinetic parameters of suicide inactivation of catalase by hydrogen peroxide. *Biochim. Biophys. Acta* **1431**:30–36.

Gruft H, Ruck R, Traynor J. 1978. Properties of a unique catalase isolated from Aspergillus niger. *Can.J.Biochem.* **56**:916–919.

Hall M, Bansal P, Lee JH, Realff MJ, Bommarius AS. 2010. Cellulose crystallinity - A key predictor of the enzymatic hydrolysis rate. *FEBS J.* **277**:1571–1582.

Harris P V, Welner D, McFarland KC, Re E, Navarro Poulsen JC, Brown K, Salbo R, Ding H, Vlasenko E, Merino S, Xu F, Cherry J, Larsen S, Lo Leggio L. 2010. Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family. *Biochemistry* **49**:3305–16.

Hashimoto H. 2006. Recent structural studies of carbohydrate-binding modules. *Cell. Mol. Life Sci.* **63**:2954–67.

Hemsworth GR, Johnston EM, Davies GJ, Walton PH. 2015. Lytic Polysaccharide Monooxygenases in Biomass Conversion. *Trends Biotechnol.* **33**:747–761.

Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD. 2007a. Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* **315**:804–7.

Himmel ME, Ding S-Y, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD. 2007b. Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* **315**:804–807.

Jalak J, Kurašin M, Teugjas H, Väljamä P. 2012. Endo-exo synergism in cellulose hydrolysis revisited. *J. Biol. Chem.* **287**:28802–28815.

Jeoh T, Wilson DB, Walker LP. 2006. Effect of Cellulase Mole Fraction and Cellulose Recalcitrance on Synergism in Cellulose Hydrolysis and Binding. *Biotechnol. Prog.* **22**:270–7.

Jung H, Wilson DB, Walker LP. 2003. Binding and reversibility of Thermobifida fusca Cel5A, Cel6B, and Cel48A and their respective catalytic domains to bacterial microcrystalline cellulose. *Biotechnol. Bioeng.* **84**:151–9.

Karkehabadi S, Hansson H, Kim S, Piens K, Mitchinson C, Sandgren M. 2008. The first structure of a glycoside hydrolase family 61 member, Cel61B from Hypocrea jecorina, at 1.6 A resolution. *J. Mol. Biol.* **383**:144–54.

Kim S, Ståhlberg J, Sandgren M, Paton RS, Beckham GT. 2014. Quantum mechanical calculations suggest that lytic polysaccharide monooxygenases use a copper-oxyl, oxygen-rebound mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **111**:149–154.

Kittl R, Kracher D, Burgstaller D, Haltrich D, Ludwig R. 2012. Production of four Neurospora crassa lytic polysaccharide monooxygenases in Pichia pastoris monitored by a fluorimetric assay. *Biotechnol. Biofuels* **5**:79.

Kjaergaard CH, Qayyum MF, Wong SD, Xu F, Hemsworth GR, Walton DJ, Young NA, Davies GJ, Walton PH, Johansen KS, Hodgson KO, Hedman B, Solomon EI. 2014. Spectroscopic and computational insight into the activation of O2 by the mononuclear Cu center in polysaccharide monooxygenases. *Proc. Natl. Acad. Sci.* **111**:8797–8802.

Kolbe S, Fischer S, Becirevic A, Hinz P. 1998. The Streptomyces reticuli α-chitin-binding protein CHB2 and its gene. *Microbiology* **144**:1291–1297.

Kostylev M, Alahuhta M, Chen M, Brunecky R, Himmel ME, Lunin V V, Brady J, Wilson DB. 2014. Cel48A from Thermobifida fusca: structure and site directed mutagenesis of key residues. *Biotechnol. Bioeng.* **111**:664–73.

Kostylev M, Wilson D. 2012. Synergistic interactions in cellulose hydrolysis. *Biofuels* **3**:61–70.

Kostylev M, Wilson D. 2013. Two-parameter kinetic model based on a time-dependent activity coefficient accurately describes enzymatic cellulose digestion. *Biochemistry* **52**:5656–5664.

Langston JA, Shaghasi T, Abbate E, Xu F, Vlasenko E, Sweeney MD. 2011. Oxidoreductive cellulose depolymerization by the enzymes cellobiose dehydrogenase and glycoside hydrolase 61. *Appl. Environ. Microbiol.* **77**:7007–7015.

Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. 2013. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes.

Biotechnol. Biofuels **6**:41.

Lončar N, Fraaije MW. 2015. Not so monofunctional-a case of thermostable Thermobifida fusca catalase with peroxidase activity. *Appl. Microbiol. Biotechnol.* **99**:2225–2232.

Loose JSM, Forsberg Z, Fraaije MW, Eijsink VGH, Vaaje-Kolstad G. 2014. A rapid quantitative activity assay shows that the Vibrio cholerae colonization factor GbpA is an active lytic polysaccharide monooxygenase. *FEBS Lett.* **588**:3435–40.

Lykidis A, Mavromatis K, Ivanova N, Anderson I, Land M, DiBartolo G, Martinez M, Lapidus A, Lucas S, Coopeland A, Richardson P, Wilson DB, Kyrpides N. 2007. Genome sequence and analysis of the soil cellulolytic actinomycete Thermobifida fusca YX. *J. Bacteriol.* **189**:2477–86.

Lynd LR, Weimer PJ, Willem HVZ, Pretorius IS. 2002. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol. Mol. Biol. Rev.* **66**:506–577.

Mansfield S, Meder R. 2003. Cellulose hydrolysis–the role of monocomponent cellulases in crystalline cellulose degradation. *Cellulose* **10**:159–170.

Mba Medie F, Davies GJ, Drancourt M, Henrissat B. 2012. Genome analyses highlight the different biological roles of cellulases. *Nat. Rev. Microbiol.* **10**:227–34.

Moser F, Irwin D, Chen S, Wilson DB. 2008. Regulation and characterization of Thermobifida fusca carbohydrate-binding module proteins E7 and E8. *Biotechnol. Bioeng.* **100**:1066–77.

Palonen H. 2004. Role of lignin in the enzymatic hydrolysis of lignocellulose. *VTT Publ.*:1–80.

Park S, Baker JO, Himmel ME, Parilla P a, Johnson DK. 2010. Cellulose crystallinity index: measurement techniques and their impact on interpreting cellulase performance. *Biotechnol. Biofuels* **3**:10.

Quinlan RJ, Sweeney MD, Lo Leggio L, Otten H, Poulsen JCN, Johansen KS, Krogh KBRM, Jørgensen CI, Tovborg M, Anthonsen A, Tryfona T, Walter CP, Dupree P, Xu F, Davies GJ, Walton PH. 2011. Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc. Natl. Acad. Sci. U. S. A.* **108**:15079–84.

Reese ET, Siu RGH, Levinson HS. 1950. The biological degradation of soluble cellulose derivatives and its relationship to the mechanism of cellulose hydrolysis. *J. Bacteriol.* **59**:485–97.

Saito A, Miyashita K, Biuković G, Schrempf H. 2001. Characteristics of a Streptomyces coelicolor A3(2) extracellular protein targeting chitin and chitosan. *Appl. Environ. Microbiol.* **67**:1268–1273.

Sakon J, Irwin D, Wilson DB, Karplus PA. 1997. Structure and mechanism of endo/exocellulase E4 from Thermomonospora fusca. *Nat. Struct. Biol.* **4**:810–818.

Schnellmann J, Zeltins A, Blaak H, Schrempf H. 1994. The novel lectin-like protein CHB1 is encoded by a chitin-inducible Streptomyces olivaceoviridis gene and binds specifically to crystalline alpha-chitin of fungi and other organisms. *Mol. Microbiol.* **13**:807–19.

Schrempf H. 2001. Recognition and degradation of chitin by streptomycetes. *Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol.* **79**:285–289.

Scott BR, Huang HZ, Frickman J, Halvorsen R, Johansen KS. 2015. Catalase improves saccharification of lignocellulose by reducing lytic polysaccharide monooxygenase-associated enzyme inactivation. *Biotechnol. Lett.*

Selitrennikoff C. 2001. Antifungal proteins. *Appl. Environ. Microbiol.* **67**:2883–2894.

Shu Z, Wang Y, An L, Yao L. 2014. The Slowdown of the Endoglucanase Trichoderma reesei Cel5A-Catalyzed Cellulose Hydrolysis Is Related to Its Initial Activity. *Biochemistry* **53**:7650–7658.

Smant G, Stokkermans JP, Yan Y, de Boer JM, Baum TJ, Wang X, Hussey RS, Gommers FJ, Henrissat B, Davis EL, Helder J, Schots A, Bakker J. 1998. Endogenous cellulases in animals: isolation of beta-1, 4-endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proc. Natl. Acad. Sci. U. S. A.* **95**:4906–11.

Suzuki K, Suzuki M, Taiyoji M, Nikaidou N, Watanabe T. 1998. Chitin binding protein (CBP21) in the culture supernatant of Serratia marcescens 2170. *Biosci. Biotechnol. Biochem.* **62**:128–135.

Theis T, Stahl U. 2004. Antifungal proteins : targets, mechanisms and prospective applications. *Cell. Mol. Life Sci. C.* **61**:437–455.

Tomme P, Heriban V, Claeyssens M. 1990. Adsorption of two cellobiohydrolases from Trichoderma reesei to Avicel: evidence for “exo-exo” synergism and possible “loose complex” formation. *Biotechnol. Lett.* **12**:525–530.

Vaaje-Kolstad G, Bøhle LA, Gåseidnes S, Dalhus B, Bjørås M, Mathiesen G, Eijsink VGH. 2012. Characterization of the chitinolytic machinery of Enterococcus faecalis V583 and high-resolution structure of its oxidative CBM33 enzyme. *J. Mol. Biol.* **416**:239–54.

Vaaje-Kolstad G, Horn SJ, van Aalten DMF, Synstad B, Eijsink VGH. 2005a. The non-catalytic chitin-binding protein CBP21 from Serratia marcescens is essential for chitin degradation. *J. Biol. Chem.* **280**:28492–7.

Vaaje-Kolstad G, Houston DR, Riemen AHK, Eijsink VGH, van Aalten DMF. 2005b. Crystal structure and binding properties of the Serratia marcescens chitin-binding protein CBP21. *J. Biol. Chem.* **280**:11313–9.

Vaaje-Kolstad G, Westereng B, Horn SJ, Liu Z, Zhai H, Sørlie M, Eijsink VGH. 2010a. An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science* **330**:219–222.

Vaaje-Kolstad G, Westereng B, Horn SJ, Liu Z, Zhai H, Sørlie M, Eijsink VGH. 2010b. An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science (80-.).* **330**:219–222.

Väljamäe P, Kipper K, Pettersson G, Johansson G. 2003. Synergistic cellulose hydrolysis can be described in terms of fractal-like kinetics. *Biotechnol. Bioeng.* **84**:254–7.

Vu V V., Beeson WT, Phillips CM, Cate JHD, Marletta MA. 2014. Determinants of regioselective hydroxylation in the fungal polysaccharide monooxygenases. *J. Am. Chem. Soc.* **136**:562–565.

Vuong T V., Wilson DB. 2009. Processivity, Synergism, and Substrate Specificity of Thermobifida fusca Cel6B. *Appl. Environ. Microbiol.* **75**:6655–6661.

Vuong T V., Wilson DB. 2010. Glycoside hydrolases: catalytic base/nucleophile diversity. *Biotechnol. Bioeng.* **107**:195–205.

Walker LP, Wilson DB. 1991. Enzymatic hydrolysis of cellulose: an overview. *Bioresour. Technol.* **36**:3–14.

Watson DL, Wilson DB, Walker LP. 2002a. Synergism in Binary Mixtures of Thermobifida fusca Cellulases Cel6B , Cel9A , and Cel5A on BMCC and Avicel. *Appl. Biochem. Biotechnol.* **101**:97–111.

Watson DL, Wilson DB, Walker LP. 2002b. Synergism in binary mixtures of *Thermobifida fusca* cellulases Cel6B, Cel9A, and Cel5A on BMCC and Avicel. *Appl. Biochem. Biotechnol.* **101**:97–111.

Wilson DB. 2004. Studies of Thermobifida fusca plant cell wall degrading enzymes. *Chem. Rec.* **4**:72–82.

Wilson DB. 2008. Three microbial strategies for plant cell wall degradation. *Ann. N. Y. Acad. Sci.* **1125**:289–97.

Wilson DB. 2009. Cellulases and biofuels. *Curr. Opin. Biotechnol.* **20**:295–299.

Wilson D, Kostylev M. 2012. Cellulase processivity. In: Himmel, ME, editor. *Biomass Convers. Methods Protoc.* Totowa, NJ: Humana Press, Vol. 908, pp. 93–99.

Wood TM, McCrae SI. 1979. Synergism between enzymes involved in the solubilization of native cellulose. In: Brown, R, editor. *Hydrolys. Cellul. Mech. Enzym. Acid Catal.* Washington, DC: American Chemical Society.

Wyman CE, Decker SR, Himmel ME, Brady JW, Skopec CE. 2004. Hydrolysis of Cellulose and Hemicellulose. In: Dumitriu, S, editor. *Polysaccharides Struct. Divers. Funct. Versatility, Second Ed.* CRC Press.

Wyman CE, Decker SR, Himmel ME, Brady JW, Skopec CE. 2005. Hydrolysis of Cellulose and Hemicellulose. *Polysaccharides Struct. Divers. Funct. Versatility, Second Ed. Ed. by Sev. Dumitriu CRC Press 2004*.

Zechel DL, Withers SG. 2000. Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Acc. Chem. Res.* **33**:11–8.

Zhang S, Wolfgang DE, Wilson DB. 1999. Substrate heterogeneity causes the nonlinear kinetics of insoluble cellulose hydrolysis. *Biotechnol. Bioeng.* **66**:35–41.

Zhang Y, Cui J, Lynd L, Kuang L. 2006. A transition from cellulose swelling to cellulose

dissolution by o-phosphoric acid: evidence from enzymatic hydrolysis and supramolecular structure. *Biomacromolecules* **7**:644–648.

Zhao X, Zhang L, Liu D. 2012. Biomass recalcitrance. Part I: The chemical compositions and physical structures affecting the enzymatic hydrolysis of lignocellulose. *Biofuels, Bioprod. Biorefining* **6**:465–482.

Zhou W, Irwin DC, Escovar-Kousen J, Wilson DB. 2004. Kinetic studies of Thermobifida fusca Cel9A active site mutant enzymes. *Biochemistry* **43**:9655–63.

Chapter 2. Structure of a *Thermomonospora fusca* lytic polysaccharide monooxygenase and mutagenesis of key residues

2.1 Introduction

Cellulosic biomass is a promising source of renewable chemicals and fuels, i.e., "second generation" biofuels. Biomass feedstocks can undergo enzymatic deconstruction to their component sugars, which can be used for a variety of bioprocesses. The economic feasibility of cellulosic biofuels is limited by substrate recalcitrance, the physical properties of plant tissue and cell walls that limit the efficiency of sugar release. Cellulolytic bacteria and fungi overcome biomass recalcitrance by secreting complex enzyme mixtures, which can be optimized for industrial application. Novel components that catalyze biomass saccharification are being investigated to optimize commercial enzyme cocktails, thus improving the economic feasibility of second generation fuels and renewable chemicals.

Cellulose is a semicrystalline matrix of anhydro-β-D-glucose linked by β-1,4-glycosidic bonds forming polysaccharide chains (Lynd et al., 2002). Cellulose resists depolymerization by hydrolytic cellulases because of its insolubility in water, highly crystalline structure, and surface complexity. Glycoside hydrolases (comprising 135 CAZy families) perform most of the saccharification of biomass in microbial secretomes and commercial mixtures. Cellulose degrading enzymes are found primarily in GH families 5, 6, 7, 8, 9, 12, 44, 45, 48, 51, 74, and 124. Lytic polysaccharide monooxygenases (LPMOs) are auxiliary activity enzymes that also attack crystalline cellulose, as well as other polysaccharides, using an oxidative mechanism. LPMOs are synergistic with hydrolytic cellulases, and significantly improve digestion by industrial cellulase cocktails (Cannella and Jørgensen, 2014). LPMO genes are abundant in both fungal and bacterial genomes, and multiple genes are often present in fungal genomes (Levasseur

et al., 2013). LPMOs are compact globular enzymes that lack a substrate binding cleft or tunnel, and instead have a copper atom containing active site located near a planar binding surface. In the presence of a reducing agent and oxygen, LPMOs attack the surface of crystalline cellulose (Hemsworth et al., 2015). In contrast to processive cellulases, such as those found in GH families 6, 7, and 48, this mechanism avoids the search for an exposed cellulose chain and the slow process of positioning a cellulose chain end into a tunnel active site.

Oxidative activity on crystalline cellulose by LPMOs was first demonstrated for the LPMOs of *Serratia marcescens* (Forsberg et al., 2011). The LPMO oxidative cleavage mechanism involves creation of an oxygen radical from molecular oxygen, likely a Cu(II)-oxyl species, which abstracts hydrogen and hydroxylates the substrate (Kim et al., 2014). Cellulose and chitin active LPMOs can target either side of the β-glycosidic bond. LPMOs that create oxidized products at glucose C1 are classified as type I; those that attack the C4 position are classified as type II, and those that have both activities are classified as type III (Vu et al., 2014). *Thermobifida fusca*, a model cellulolytic bacterium, secretes two LPMOs, the type I (TfAA10B) and type III (TfAA10A) (Arfi et al., 2014; Forsberg et al., 2014; Moser et al., 2008). The result of oxidative cleavage is an altered cellulose substrate that is more easily degraded by hydrolases. For example, TfAA10A stimulates the activity of the processive exocellulase, TfCel48A (Kostylev and Wilson, 2013).

The structures of several bacterial AA10 LPMOs have been solved, revealing common structural attributes that affect activity. The first LPMO structure to be solved was the chitinolytic LPMO SmCBP21 and more recently several bacterial LPMOs acting on cellulose have been investigated (Li et al., 2012; Vaaje-Kolstad et al., 2005). These structures share a common immunoglobulin-like β-sandwich core fold, a flat binding surface, and a conserved *N*-

terminal histidine. These conserved structural features are similar for LPMOs active on a range of insoluble substrates including chitin, indicating a common strategy for binding and active site positioning. The planar binding surface contains conserved polar residues, which function in binding to planar carbohydrate substrates. The structurally conserved copper coordination site positions the copper atom in proximity to the scissile carbohydrate bond and these residues are essential for tuning the copper redox state (Forsberg et al., 2014). The domain architecture of LPMOs is also shared across species, often with the LPMO catalytic domain (always at the *N* terminus due to the absolutely conserved *N*-terminal histidine involved in copper chelation) alone or attached to a binding domain.

In this work, we report an additional bacterial AA10 LPMO structure, AA10A from the model bacterium, *Thermobifida fusca*, obtained using X-ray diffraction. The activity of AA10A (formerly E7) has received considerable attention recently (Arfi et al., 2014; Forsberg et al., 2014). To explore the mechanism used by LPMOs to bind to and perform oxidative cleavage of crystalline substrates, we characterized TfAA10A surface residue mutants and TfAA10B domain truncations. The results obtained indicate that both polar and aromatic residues on the surface play critical roles for binding and activity; and surprisingly that the CBM2 domain contributes significantly to AA10B binding and activity.

2.2 Materials and Methods

Enzyme Mutagenesis.

TfAA10A and TfAA10B (formerly E7 and E8) were cloned in pET26b+ to replace the endogenous signal peptide with the PelB leader sequence. Mutants of TfAA10A and TfAA10B were created using the QuickChange II XL Site Directed Mutagenesis Kit (Agilent), following established protocols. The TfAA10B domain deletion construct was made by introducing a

HindIII cut site to replace the X1 domain with a two residue (LE) linker sequence. All construct sequences were validated, expressed, and purified using established protocols (Moser et al., 2008). Concentration of purified proteins was determined at OD280 using a calculated extinction coefficient, then they were stored at -80°C.

Crystallization.

TFAA10A crystals were first obtained with sitting drop vapor diffusion using a 96-well plate with Crystal Screen HT from Hampton Research (Aliso Viejo, CA). Fifty µL of well solution was added to the reservoirs and drops were made with 0.2 µL of well solution and 0.2 µL of protein solution using a Phoenix crystallization robot (Art Robbins Instruments, Sunnyvale, CA). The best crystals were grown at 20°C with 0.1 M Sodium acetate trihydrate pH 4.6, 20% v/v 2-Propanol and 0.2 M Calcium chloride dihydrate as the well solution. The protein solution contained 8.5 mg/mL of protein in 20 mM HEPES pH 7.5, 100 mM NaCl, 5% glycerol and 5% ethylene glycol.

Data Collection and Processing.

Both the native and KI soaked TFAA10A crystals were flash frozen in a nitrogen gas stream at 100 K before data collection. The crystallization solution with 12.5% (v/v) ethylene glycol and glycerol was used for freezing the crystal. The potassium iodide was introduced to the crystal by adding 0.5 M KI into the well solution and soaking the crystal in a 5 µL drop for 5 s before flash freezing. Data was collected in an in-house Bruker X8 MicroStar X-ray generator with Helios mirrors and Bruker Platinum 135 CCD detector. Data were indexed and processed with the Bruker Suite of programs version 2011.2-0 (Bruker AXS, Madison, WI).

Structure Solution and Refinement.

Intensities were converted into structure factors and 5% of the reflections were flagged

for R_{free} calculations using programs SCALEPACK2MTZ, ctruncate, MTZDUMP, Unique, CAD, FREERFLAG and MTZUTILS from the CCP4 package of programs (Winn et al., 2011). The structure of TFAA10A was solved using SIRAS with Crank2 (Dauter et al., 2000; Skubák and Pannu, 2013). Buccaneer (Cowtan, 2006) was used to auto build the resulting partial model. Refinement and manual correction was performed using REFMAC5 (Murshudov et al., 2011) version 5.7.0029 and Coot (Emsley et al., 2010) version 0.6.2. The MOLPROBITY method (Chen et al., 2010) was used to analyze the Ramachandran plot and root mean square deviations (rmsd) of bond lengths and angles were calculated from ideal values of Engh and Huber stereochemical parameters (Engh and Huber, 1991). Wilson B-factor was calculated using ctruncate version 1.17.7. The data collection and refinement statistics are shown in Table I.

Table 1 X-ray data collection and refinement statistics. Statistics for the highest resolution bin are in parenthesis.

Data collection	
Space group	P 32 2 1
Unit cell, Å, °	a= 77.23, b= 77.23, c = 66.753
	β = 122.35
Wavelength, Å	1.54178
Temperature (K)	100
Resolution, Å	25.0-1.99 (2.09–1.99)
Unique reflections	28968 (3560)
R_{int} †	0.155 (0.622)

Average redundancy	11.9 (3.4)
$\langle I \rangle / \langle \sigma(I) \rangle$	10.6 (1.4)
Completeness, %	97.9 (90.1)
Refinement	
Resolution, Å	25-2.0 (2.05-2.0)
R/R_{free}	0.167 (0.285)/ 0.233 (0.318)
Protein atoms	3119
Water molecules	345
Other atoms	42
RMSD from ideal bond length, Å[#]	0.017
RMSD from ideal bond angles, ° [#]	1.700
Wilson B-factor	4.8
Average B-factor for protein atoms, $Å^2$	20.6
Average B-factor for water molecules, $Å^2$	28.9
Ramachandran plot statistics, %*	
Allowed	100%
Favored	97.3%
Outliers	0

† Rint = $\sum| I - \langle I \rangle | / \sum|I|$ where I is the intensity of an individual reflection and $\langle I \rangle$ is the mean intensity of a group of equivalents and the sums are calculated over all reflections with more than one equivalent measured

(Engh and Huber, 1991) * (Chen et al., 2010)

Structure Analysis.

Programs Coot and PyMOL (http://www.pymol.org) were used for comparing and analyzing structures. This structure has been deposited to the protein data bank (PDB; www.rcsb.org) with entry code 4GBO.

Substrates and Reducing Agent.

Bacterial cellulose (BC), a gift from Monsanto, was washed and prepared as described previously (Kostylev et al., 2014). The concentration was determined by dry weight using a vacuum oven, and it was stored at 4°C in MilliQ water with 0.02% sodium azide to prevent microbial contamination. All activity assays contained reduced glutathione (Sigma) as a reducing agent to enable LPMO activity. Glutathione was stored dry at -20°C away from light, and a concentrated stock was prepared fresh for each assay by adjusting to pH 5.5 with sodium acetate.

Binding Assays.

Binding affinity assays combined 1 μM of each LPMO variant with 2.5 mg/mL bacterial cellulose and 10 mM sodium acetate to a total volume of 160 μL. All samples were assembled in triplicate with representative buffer, substrate, and enzyme negative controls. Samples were sealed in 96-well Protein LoBind microplates (Eppendorf), and incubated at 50°C with continuous horizontal shaking (160 RPM) for 16 h to ensure a binding equilibrium was reached. After incubation, samples were immediately centrifuged at 4,000 RPM (3313xRCF) for 5 min at room temperature to precipitate the substrate. A 40 μL fraction of the supernatant was carefully removed and combined with 160 μL Quick Start Bradford reagent (Bio-Rad) in a 96-well microplate (Costar). Samples and standards were measured using a Synergy 4 plate reader (Biotek Instruments), and collected as the A_{595}/A_{450} ratio to extend the sensitivity of the Bradford assay (Zor and Selinger, 1996). Unbound protein was quantified using a BSA standard curve,

and extent of binding determined by comparison with enzyme only negative controls.

Activity assays.

Cellulose digestion assays were assembled using similar conditions as binding assays, with the exception of 5.0 mg/mL bacterial cellulose, 1 μM $CuSO_4$, and 2 mM reduced glutathione to facilitate activity. All assays were run in triplicate, with representative controls and standards adjacent to sample wells. Plates containing kinetic time course samples were automatically removed at the designated intervals, transferred to a preheated PCR machine (MJResearch Inc.), heated to 100°C for 5 min, and stored at room temperature until sample processing. Assay supernatant was separated from the insoluble substrate via centrifugation at 4,000 RPM (3313xRCF) for 5 min before careful removal of supernatant for secondary hydrolysis. Secondary hydrolysis was performed to simplify quantification using cdCel5A (50uM) and Novo188 β-D-glucosidase (0.02 CBU, Novozymes) that had been buffer exchanged to remove most background signal.

HPLC Quantification.

Samples from secondary hydrolysis were filtered through a 96 well 0.45 μm Supor® filter (Pall) and were quantified using a Shimadzu Ultramate HPLC equipped with RID and UV detectors. An Aminex HPX-87H column (Bio-Rad) with a standard guard column was operated with isocratic flow at 0.6 mL/min with 0.005 M H_2SO_4 as eluent. Fifty μL of each sample for HPLC detection were injected using a refrigerated autoinjector. Both neutral and oxidized sugars were detected with the RI detector, while oxidized sugars were detected at 200 nm in the UV channel (Cannella et al., 2012a). Monosaccharide standards, glucose and gluconolactone (Sigma), were quantified during each run.

Data Processing.

OriginPro 2016 (OriginLab Corp.) was used to process the raw data. A Gaussian fit was applied to chromatogram peaks after removal of buffer background to yield values of area under the curve. These values were compared to a linear standard curve to determine concentration. The fraction of oxidized product was determined by a standard curve of varying ratios, yielding the concentration of each product. The extent of digestion was determined by comparison of the sum of monosaccharides released as initial anhydrous G1 based on dry weight determination. The kinetic time course data were plotted as % digestion vs time, and nonlinear kinetic parameters were determined based on the two parameter model using existing protocols (Kostylev and Wilson, 2013).

2.3 Results and Discussion

The Crystal Structure of *T. fusca* TFAA10A. The structure of TFAA10A was refined to a resolution of 2.0 Å with R and R_{free} of 0.167 and 0.233, respectively. There are two molecules in the asymmetric unit, a representative structure is shown in Figure 2.1. It has an Ig-like β-sandwich fold with a copper ion bound on the active site. The copper ion is only partially present with incomplete coordination likely due to low occupancy partial conformations at the active site that cannot be modeled properly.

Structural Comparison.

The structure of TfAA10A shares many features with other recently crystallized AA10 LPMOs. The compact secondary structure of the core immunoglobulin-like fold is typical of AA10 LPMOs. Eight similar clusters with 40% sequence identity clustering were obtained from the protein data bank (PDB; www.rcsb.org) using the jFATCAT-rigid algorithm (Prlić et al., 2010; Ye and Godzik, 2003). From these proteins, the most similar was clearly the *Streptomyces coelicolor* lytic polysaccharide monooxygenase (PDB code 4OY6) (Forsberg et al., 2014), which

has a sequence identity of 70% and Cα root mean square deviations of 0.66 Å, showing that the overall backbone is the same. The other seven structures had sequence identities below 30%.

The presence of copper with high occupancy in the AA10A active site is in agreement with similar structures and EPR results supporting copper as the essential metal (Forsberg et al., 2014). The Jahn-Teller copper coordination of the AA10A copper is nearly identical to other structures, supporting a common mechanism of copper coordination. The N47 residue adjacent to the copper active site is highly conserved, with atomic thermal displacement parameters indicating high mobility of the side chain. The adjacent aromatic residue present on the AA10A binding surface, W46, does not have high conservation. It has been suggested that this aromatic residue performs a substrate targeting role rather than increasing total binding affinity, based on models of the *Phanerochaete chrysosporium* GH61D AA9 LPMO (Wu et al., 2013).

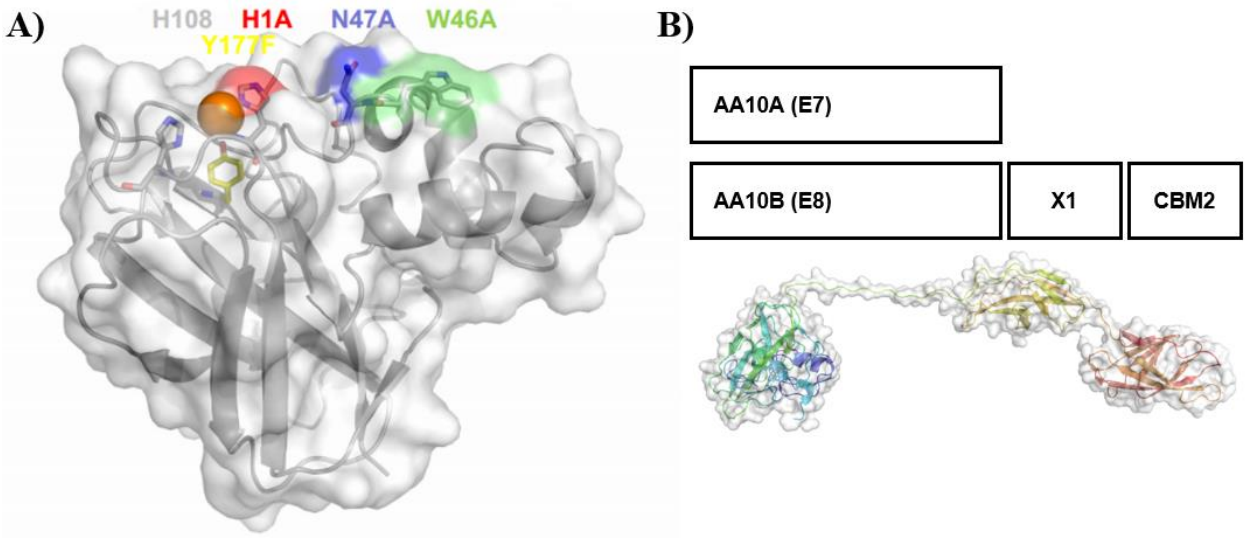


Figure 2.1. A) Crystal structure of *T. fusca* AA10A. Surface diagram positioned to emphasize the planar binding surface. Highly conserved residues surrounding the active site copper are represented by sticks. Position of surface mutant residues are shown in red (H1), yellow (Y177), blue (N47), and green (W46). B) Domain arrangement of *T. fusca* LPMOs, multiple TfAA10B domains shown as modelled structure (HHPRED using ProteinModelPortal.com).

The sequence of AA10A is similar to the catalytic domain of AA10B, sharing 33% residue identity, with the exception of an additional stretch of seven amino acids present in

AA10B. The comparison of AA10B and fungal AA9 proteins indicate that AA10B has a more extended 'wing' on its planar binding surface. An additional similarity between AA10B and fungal AA9 LPMOs is the axial position of the copper coordination sphere, which is tyrosine. In most bacterial LPMOs and in AA10A, a phenylalanine occupies this position.

LPMO Kinetics Quantification.

Multiple small molecule reducing agents have been found to enable LPMO activity. Ascorbate has several disadvantages that complicate activity quantification. Its major oxidized form, dehydroascorbate, degrades into a complex mixture of products when boiled, which co-elute with oxidized monosacharides. Ascorbate is spontaneously oxidized by copper, making it a less suitable reducing agent (Loose et al., 2014). As a source of reducing power, glutathione is capable of single and double electron transfer (Harman et al., 1986). GSH appears to have the same protective role in kinetics assays, as it does in cells, helping to maintain enzyme activity in the presence of oxygen radicals. Glutathione does not interfere with substrate binding and provides electrons to LPMOs with similar efficiency as ascorbate (data not shown). The majority of glutathione forms a stable structure (GSSG) after oxidation that doesn't produce degradation products after boiling.

LPMO reaction kinetics include all the complexity of cellulase kinetics acting on changing recalcitrant substrates, but with additional challenges for assay design and product detection. The wide product distribution created by pseudo-random cleavage makes the kinetics of LPMOs difficult to measure accurately. Excess β-D-glucosidase (0.02 CBU) was used in conjunction with the catalytic domain of TfCel5A (50 nM), an endocellulase capable of degrading longer oligosaccharides. Secondary hydrolysis was complete after overnight incubation based on the absence of neutral or oxidized oligosaccharide products. The β-D-

glucosidase used had no effect on glucose, as all standards were incubated under similar conditions of secondary hydrolysis to confirm the absence of any background lactone signal. Lactonase, which catalyzes the hydrolysis of gluconolactone to gluconic acid, is present in Novozymes 188 (Witteveen et al., 1992). With the presence of lactonase, it was assumed that all soluble lactones were hydrolyzed despite the slower rate of spontaneous lactone hydrolysis at lower pH. However, if C1 oxidation products remained unhydrolyzed, quantification would remain unchanged, as gluconolactone and gluconic acid produced identical standard curves (data not shown).

HPLC determination of oxidized sugars using the Aminex HPX-87H was able to produce the ratio of C1 oxidized aldonic acids in samples based on previous work using commercial mixtures (Cannella et al., 2012b). This detection approach enables quantification using standard saccharide HPLC equipment, but suffers from lower resolution and product detection limits compared to other methods. The gluconolactone standard curve is very linear (R^2=0.99) and serves as an effective standard for quantification of oxidized glucose products. This detection approach is only useful in cases where obtaining the product distribution is not necessary, as it requires complete secondary hydrolysis. This determination of oxidized products from secondary hydrolysis of reactions containing LPMOs relies on detection of the carbonyl group. UV detection cannot directly measure the conversion to gemdiol forms, as alcohol groups do not absorb at 200 nm. A different approach will be required to accurately quantify the net oxidized products of type III LPMOs. The proportion of 4-keto AA10A products which become hydrolyzed to the gemdiol form cannot be quantified using this method, but the total sugar being produced can be evaluated. Due to this characteristic, the number of oxidative cleavage events may be underestimated for AA10A and its mutants.

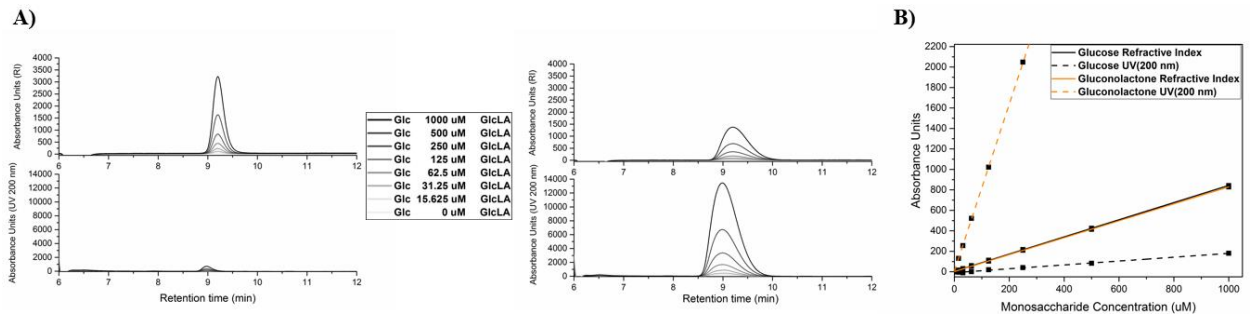


Figure 2.2. HPLC chromatograms of the full range of glucose gluconolactone ratios A) that demonstrate sample neutral/oxidized product ratio using B). Linear standard curves neutral and C1 oxidized monosaccharides for quantification of LPMO products.

Catalase was included in LPMO reactions to preserve activity, based on the results of Scott et al., 2015. The creation of peroxide side products in solution is a predicted mechanism through which LPMOs are inactivated over time (Kjaergaard et al., 2014). LPMO side product peroxide radicals would be expected to interact with the closest available bonds, destroying enzyme activity when biomass is substrate (Scott et al., 2015) or consuming the soluble reducing agent. Catalase may play a role in detoxifying hydrogen peroxide as it is produced. The inhibition of LPMO activity by catalase seen by Vaaje-Kolstaad et al., 2010 was not observed, possibly due to the use of bovine catalase compared with catalase from a fungal origin. Furthermore, catalase does not stimulate activity of LPMOs or hydrolytic cellulases (data not shown). In reaction conditions where reducing agents are not present in excess (i.e., when CDH is employed), the addition of catalase will be essential to protect LPMO activity.

TfAA10A Surface Mutants Binding/Activity on Crystalline Cellulose.

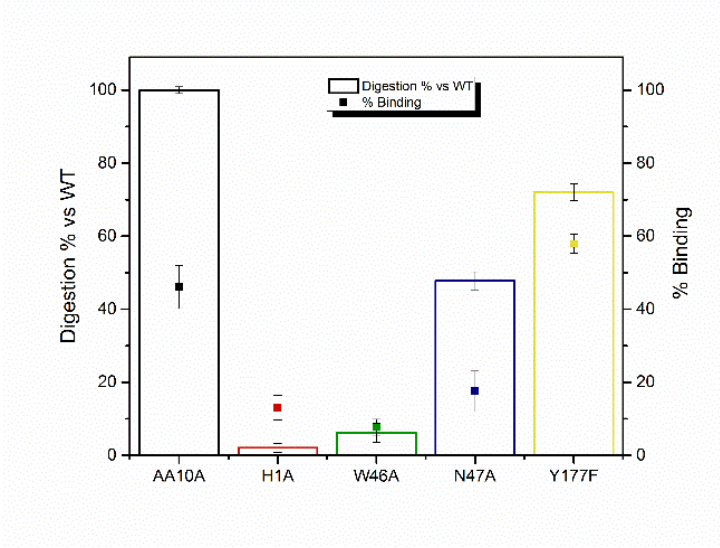


Figure 2.3. Binding and activity of *T. fusca* AA10A binding surface mutants. Digestion (bars) of 0.5 µM TfAA10A incubated 2 h on 5.0 mg/mL BC, with total monosaccharide release compared with WT value. Extent of binding (squares) as fraction of 1.0 µM enzyme lost from solution after binding equilibrium established after 16 h. Binding was measured in the absence of reducing agent. Samples were measured in triplicate, with error bars representing the replicate standard deviation.

Binding is essential for LPMO activity and thus to understand the mechanistic basis of LPMO activity, we must compare mutants with altered binding properties. Mutants of LPMO surface residues and domain architecture constructs help to reveal the mechanism of binding. Multiple residues on the AA10A surface were mutated (Figure 2.1), and constructs removing domains from AA10B were compared in binding to crystalline cellulose (BC).

Mutation of conserved residues on the substrate binding surface had a significant effect on binding. Compared to WT AA10A, AA10A binding surface mutants showed decreased binding affinity when measured at 16 h after equilibrium had been established. The significant decrease in binding due to removal of the binding surface tryptophan, W46, and asparagine N47 indicate that both residues play a critical role in binding to crystalline cellulose. This observation is in agreement with previous results that showed that mutation of the aromatic tyrosine of SmCBP21 (in the same position as AA10A W46) significantly decreases binding to chitin (Vaaje-Kolstad et al., 2005). Removal of the *N*-terminal H1 residue eliminates binding of copper

to the LPMO active site. This change greatly diminishes binding, supporting the significant role of copper in mediating LPMO-substrate interaction. The change of AA10A Y177 to phenylalanine probably does not abolish copper binding, but changes the coordination sphere of the copper to resemble TfAA10B and fungal AA9 LPMOs. This alteration to the copper position relative to the substrate had no effect on binding, and may induce tighter binding to crystalline cellulose.

The very high nonspecific binding in the presence of high conductivity buffers indicates that there is an important role of pH and ionic strength for *T. fusca* LPMO binding (not shown). The redox state of the active site copper may not be important, as external reducing agents are not necessary for binding, but removal of the copper via H1A mutation significantly impairs binding.

Furthermore, AA10A surface mutants have significantly altered activity compared to WT AA10A. The oxidative activity of both AA10A and AA10B were shown by mass spectrometry, confirming previous reports showing AA10A as a type III and AA10B as a type I LPMO (data not shown). The H1A mutation eliminates essentially all oxidative activity relative to WT. This result is expected based on the important role of the H1 residue in providing the correct coordinating shell for the Cu atom within the active site. Also, the activity of W46A on PASC was significantly decreased relative to WT; to a similar degree to the loss of binding. The activity of the N47A mutant on PASC was much decreased relative to WT AA10A, but significantly less than the W46 mutant. This result indicates the importance of this conserved polar residue for binding, but once bound, the active site copper may need to be positioned correctly by W46 to enable substrate attack.

The Y177F mutation to match the residue in the axial position of the copper coordination

sphere in AA10B showed less change relative to WT AA10A. The mutation to Y177F had 28% less activity compared to WT and 41% less binding, indicating that the difference in the axial position between bacterial and fungal LPMOs plays an important role in activity and binding. The presence of an additional hydroxyl group from tyrosine is important for positioning the copper in TfAA10A, but does not completely inhibit the activation of the copper to generate super-oxo species for catalytic attack. This observation is in agreement with previous electron spin resonance results exploring LPMO axial position occupancy (Forsberg et al., 2014).

The trend of AA10A surface mutant activity largely matches the trend of substrate binding, with the exception of the N47A mutant. Moreover, the "disconnect" between binding and activity for the N47A mutant may be relevant to understanding the structure-function relationship for the LPMO AA family. Binding and activity are not always directly coupled, based on results of the N47A mutant, where binding was eliminated but some activity remained. The weak binding at equilibrium indicates that while the binding was not as stable, it existed long enough to position the LPMO active site over the substrate bond. The W46A mutant had a larger effect on activity than the N47A mutant, which supports a role for LPMO surface aromatics performing a critical role in active site positioning; as well as lowering the binding energy (Wu et al., 2013).

TfAA10B Domain Mutants Binding/Activity on Crystalline Cellulose

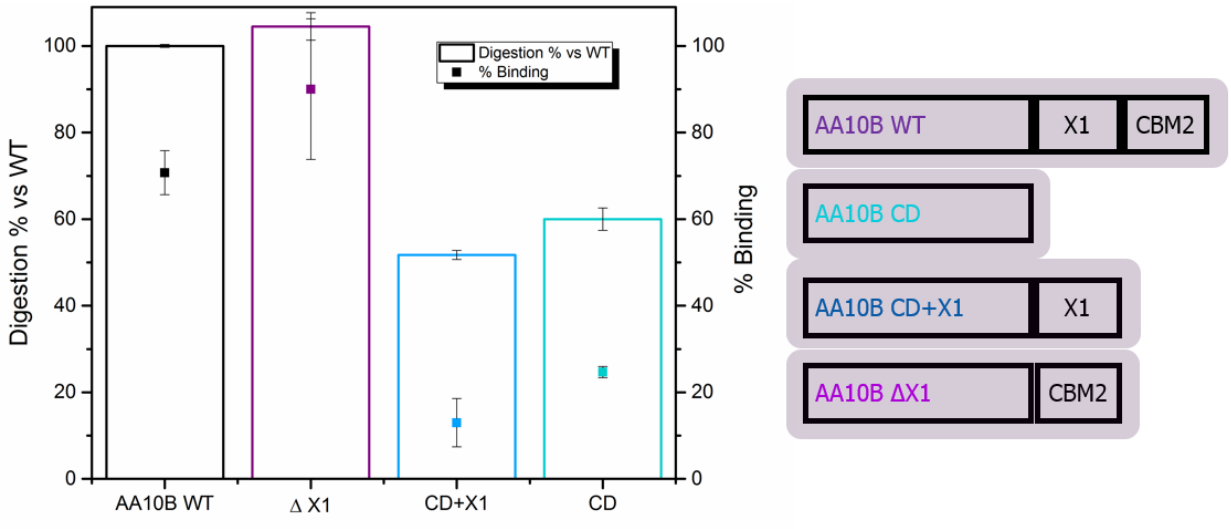
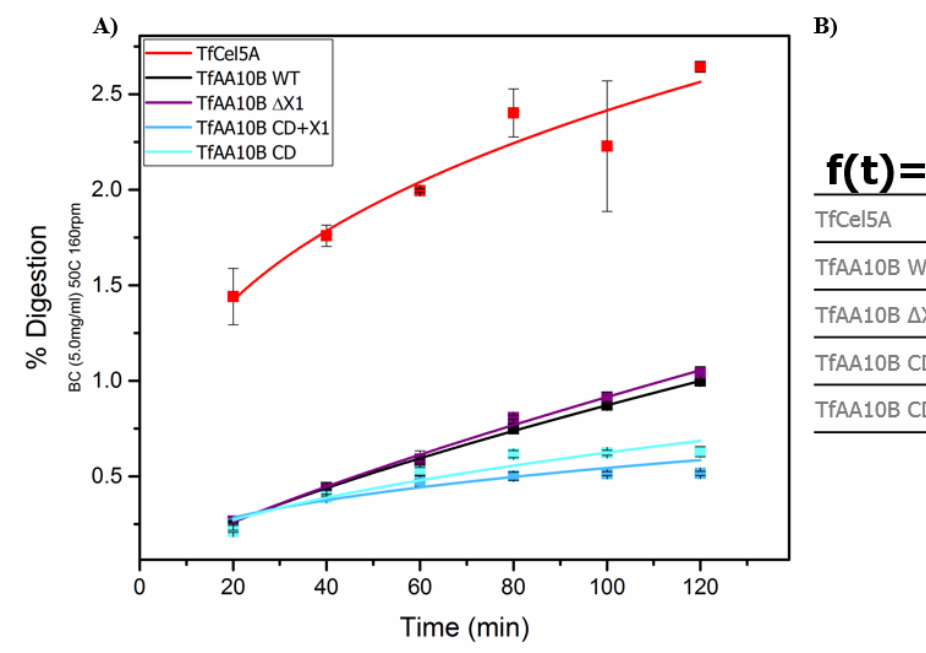


Figure 2.4. Binding and activity of TfAA10B domain truncation mutants. Digestion relative to WT measured as 0.5 µM LPMO on 5.0 mg/mL BC with 2 mM reduced glutathione after 2. As above, binding measured as 1.0 µM LPMO after 16 h incubation in the absence of reducing agent. Samples measured in triplicate, with error bars representing one standard deviation.

The domain mutants of AA10B do not suggest a clear role of the X1 (formerly FNIII-like) domain of AA10B present between the CBM and the catalytic domain. The most significant difference appears to be the removal of the CBM2 domain, as in both mutants, binding to BC is significantly reduced relative to WT. Though similar in size, the CD10B mutant has much less binding affinity than WTAA10A. This is due to the presence of the additional CBM2 domain of AA10B, which provides additional binding to regions of highly crystalline cellulose at the pH optimum for activity.

TfAA10B Domain Mutants Activity on Crystalline Cellulose.



B)

$f(t)=At^b$	*b*	A	R^2
TfCel5A	0.33	0.53	0.92
TfAA10B WT	0.75	0.03	0.99
TfAA10B ΔX1	0.78	0.03	0.99
TfAA10B CD+X1	0.40	0.08	0.89
TfAA10B CD	0.52	0.06	0.89

Figure 2.5. Activity of *T. fusca* AA10B truncation mutants compared with endocellulase TfCel5A. Time course reactions A) measured as 0.5 µM LPMO on 5.0 mg/mL BC with 2 mM reduced glutathione. B) The *A* and *b* parameters resulting from two-parameter nonlinear kinetic fit (Kostylev and Wilson, 2013). All samples measured in triplicate, with standard deviation represented by error bars.

Based on the results presented in Figures 2.4 and 2.5, domain removal has a significant effect on TfAA10B activity. *T. fusca* LPMOs have pH optima similar to those of the secreted hydrolytic cellulases acting on BC, e.g., from pH 5.5 to 6.0 (data not shown). As expected for type I LPMOs, secondary hydrolysis of the activity assay supernatants generated G1 and G1Ox products (Confirmed by MS/MS). The kinetics of the AA10B domain mutants were measured using the nonlinear fit described by Kostylev and Wilson, 2013. The activity decreases over time similar to the nonlinear kinetics of hydrolytic cellulases. Compared to the hydrolytic endocellulase, TfCel5A, TfAA10B releases fewer soluble products. Furthermore, the b parameter of TfAA10B is significantly lower than that of TfCel5A, suggesting less available substrate during the initial phase of digestion relative to Cel5A, which has mostly amorphous regions to attack.

The estimate of initial rate from the linear 'pseudo-steady state' portion of the time course

used by Loose et al. shows a turnover for oxidative cleavage of approximately 1 min^{-1}. The activity of LPMO domain constructs shows that whereas the two LPMO catalytic domains vary in binding ability, the activity of the *T. fusca* type I LPMO benefits significantly from the CBM2 domain. The 10B truncation mutant lacking the intermediate X1 domain shows activity on BC that is unchanged relative to WT AA10B, but that activity was reduced for both mutants lacking the CBM2 domain. This outcome suggests that similar to hydrolytic cellulases, the CBM2 is important for increasing the local concentration of the catalytic domain on the substrate surface in order to perform the catalytic attack. Due to the slower turnover rate reported previously (~1 min^{-1}), this enrichment on the substrate surface appears essential for generating sufficient oxidative cleavages for meaningful product release. This result is in agreement with previous work integrating *T. fusca* LPMO domains into cellulosome scaffolds (Arfi et al., 2014). The removal of the X1 domain had little effect, and the presence of the CBM2 domain was key to providing LPMO activity.

Role of X1/FN3 domain.

The domain mutants lacking the X1 domain showed no difference relative to WT 10B. The removal of the X1 domain, comparing WT to ΔX1, shows no effect. Similarly, the addition of the X1 domain to the CD does not enhance binding. The binding of AA10B is mediated mainly by the CBM2 domain and to some extent by the CD. This binding result is mirrored in activity results, where this domain has no effect. Similarly, the removal of the X1 domain had little effect on TfAA10B incorporated into scaffolds (Arfi et al., 2014).

Although X1 domains are abundant in both LPMO and cellulase genes, the X1 domain is currently a domain of unknown function (Beckham et al., 2011). In some hydrolytic cellulases, such as TfCel48A and in chitinases, found in subfamily A and B of Family 18, where they can be

present in multiple copies (Saito et al., 2003). X1 domains are abundant, found in extremophile amylopullulanase (Lin et al., 2008). Whereas X1 domains have little effect in LPMOs under tested conditions, their removal from a processive endocellulase significantly reduced activity on multiple substrates (Zhou et al., 2004). Also, X1 domain removal from CbhA caused activity reduction to 50% (Kataeva et al., 2002). The structure of the *Ct*CbhA X1 domains do not indicate a clear role and were shown to not destabilize cellulose structure (Brunecky et al., 2012). Only two *T. fusca* cellulases, Cel5A and Cel6B, lack an X1 domain between their CD and CBM.

There have been multiple hypotheses describing the role of X1 domains, including linker protection and extension. The trend of location of X1 domains between the catalytic domain and CBMs in many LPMO genes suggests a role related to mediating binding, possibly replacing glycosylated fungal linkers or protecting long unstructured regions from proteolysis. The position of X1 domains between the CD and CBM2 domains suggest that it may play a role similar to the CBM to assist in direct binding to cellulose, which would replace glycosylated linkers as found in TrCel7A (Payne et al., 2013). The most likely role of X1 domains in bacterial systems is to provide resistance to proteolysis, a structural feature protecting the otherwise unstructured linker between the CD and CBM domains.

Alternatively, unfolding of the X1 domain, similar to the role of the domain in mammalian titin, may give the CD more access to substrate further from the bound CBM. Forced unfolding of X1 domain via SMD has been explored (Craig et al., 2001). Absence of disulfide bonds, which are present in the adjacent CBM2, result in less stability. This role is likely shared with the X1 domains found in other *T. fusca* hydrolyses, suggesting similar optimization of domain arrangements of bacterial LPMOs and cellulases. The X1 domain may play a more significant role in processive enzymes or in cases where substrate access for the CD is impaired.

This supports the concept that the search and engagement to a free chain end is a limiting step for highly active exocellulases. The effect of X1 domain deletion in LPMOs may not be evident under current experimental conditions, due to the abundance of substrate.

2.4 Conclusion

In this work, the two LPMOs of *Thermobifida fusca* were explored through deletion of the multiple domains of TfAA10B and mutagenesis of surface residues of TfAA10A. The crystal structure of TfAA10A was solved, showing significant similarities with other Type III bacterial LPMOs. The activity of the LPMO mutants was measured on crystalline cellulose and characterized using short time course kinetics (Kostylev and Wilson, 2013; Shu et al., 2014). These results support existing models of LPMO binding and activity, showing that polar surface residues were likely selected for enhanced binding and the importance of aromatic residues for positioning the active site copper ion (Wu et al., 2013). The role of the X1 domain has not been conclusively established, but its presence between the CD and CBM domains suggests that it gives increased substrate access on crowded surface (Brunecky et al., 2012) and also acts as a proteolysis resistant linker (Payne et al., 2013). The role of LPMOs and the variability of abundance in genomes is not fully explored. LPMOs likely perform initial attacks into crystalline cellulose to allow larger processive cellulases to bind and attack, but the precise nature of their synergistic behavior remains to be definitively characterized.

REFERENCES

Arfi Y, Shamshoum M, Rogachev I, Peleg Y, Bayer E a. 2014. Integration of bacterial lytic polysaccharide monooxygenases into designer cellulosomes promotes enhanced cellulose degradation. *Proc. Natl. Acad. Sci. U. S. A.* **111**:9109–14.

Beckham GT, Bomble YJ, Bayer E a., Himmel ME, Crowley MF. 2011. Applications of computational science for understanding enzymatic deconstruction of cellulose. *Curr. Opin. Biotechnol.* **22**:231–238.

Brunecky R, Alahuhta M, Bomble YJ, Xu Q, Baker JO, Ding SY, Himmel ME, Lunin V V. 2012. Structure and function of the Clostridium thermocellum cellobiohydrolase A X1-module repeat: Enhancement through stabilization of the CbhA complex. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68**:292–299.

Cannella D, Hsieh CWC, Felby C, Jørgensen H. 2012a. Production and effect of aldonic acids during enzymatic hydrolysis of lignocellulose at high dry matter content. *Biotechnol. Biofuels* **5**:26.

Cannella D, Hsieh CC, Felby C, Jørgensen H. 2012b. Production and effect of aldonic acids during enzymatic hydrolysis of lignocellulose at high dry matter content. *Biotechnol. Biofuels* **5**:26.

Cannella D, Jørgensen H. 2014. Do new cellulolytic enzyme preparations affect the industrial strategies for high solids lignocellulosic ethanol production? *Biotechnol. Bioeng.* **111**:59–68.

Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. 2010. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**:12–21.

Cowtan K. 2006. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**:1002–1011.

Craig D, Krammer a, Schulten K, Vogel V. 2001. Comparison of the early stages of forced unfolding for fibronectin type III modules. *Proc. Natl. Acad. Sci. U. S. A.* **98**:5590–5595.

Dauter Z, Dauter M, Rajashankar KR. 2000. Novel approach to phasing proteins: Derivatization by short cryo-soaking with halides. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **56**:232–237.

Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**:486–501.

Engh RA, Huber R. 1991. Accurate bond and angle parameters for X???ray protein structure refinement. *Acta Crystallogr. Sect. A* **47**:392–400.

Forsberg Z, Mackenzie AK, Sørlie M, Røhr ÅK, Helland R, Arvai AS, Vaaje-Kolstad G, Eijsink VGH. 2014. Structural and functional characterization of a conserved pair of bacterial cellulose-oxidizing lytic polysaccharide monooxygenases. *Proc. Natl. Acad. Sci. U. S. A.* **111**:8446–51.

Forsberg Z, Vaaje-Kolstad G, Westereng B, Bunæs AC, Stenstrøm Y, MacKenzie A, Sørlie M, Horn SJ, Eijsink VGH. 2011. Cleavage of cellulose by a CBM33 protein. *Protein Sci.* **20**:1479–83.

Harman LS, Carver DK, Schreiber J, Mason RP. 1986. One- and Two-electron Oxidation of Reduced Glutathione by Peroxidases. *J. Biol. Chem.* **261**:1642–1648.

Hemsworth GR, Johnston EM, Davies GJ, Walton PH. 2015. Lytic Polysaccharide Monooxygenases in Biomass Conversion. *Trends Biotechnol.* **33**:747–761.

Kataeva I a., Seidel RD, Shah A, West LT, Li XL, Ljungdahl LG. 2002. The fibronectin type 3-like repeat from the Clostridium thermocellum cellobiohydrolase CbHa promotes hydrolysis of cellulose by modifying its surface. *Appl. Environ. Microbiol.* **68**:4292–4300.

Kim S, Ståhlberg J, Sandgren M, Paton RS, Beckham GT. 2014. Quantum mechanical calculations suggest that lytic polysaccharide monooxygenases use a copper-oxyl, oxygen-rebound mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **111**:149–154.

Kjaergaard CH, Qayyum MF, Wong SD, Xu F, Hemsworth GR, Walton DJ, Young NA, Davies GJ, Walton PH, Johansen KS, Hodgson KO, Hedman B, Solomon EI. 2014. Spectroscopic and computational insight into the activation of O2 by the mononuclear Cu center in polysaccharide monooxygenases. *Proc. Natl. Acad. Sci.* **111**:8797–8802.

Kostylev M, Alahuhta M, Chen M, Brunecky R, Himmel ME, Lunin V V, Brady J, Wilson DB. 2014. Cel48A from Thermobifida fusca: structure and site directed mutagenesis of key residues. *Biotechnol. Bioeng.* **111**:664–73.

Kostylev M, Wilson D. 2013. Two-parameter kinetic model based on a time-dependent activity

coefficient accurately describes enzymatic cellulose digestion. *Biochemistry* **52**:5656–5664.

Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. 2013. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol. Biofuels* **6**:41.

Li X, Beeson WT, Phillips CM, Marletta MA, Cate JHD. 2012. Structural basis for substrate targeting and catalysis by fungal polysaccharide monooxygenases. *Structure* **20**:1051–1061.

Lin HY, Chuang HH, Lin FP. 2008. Biochemical characterization of engineered amylopullulanase from Thermoanaerobacter ethanolicus 39E-implicating the non-necessity of its 100 C-terminal amino acid residues. *Extremophiles* **12**:641–650.

Loose JSM, Forsberg Z, Fraaije MW, Eijsink VGH, Vaaje-Kolstad G. 2014. A rapid quantitative activity assay shows that the Vibrio cholerae colonization factor GbpA is an active lytic polysaccharide monooxygenase. *FEBS Lett.* **588**:3435–40.

Lynd LR, Weimer PJ, Willem HVZ, Pretorius IS. 2002. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol. Mol. Biol. Rev.* **66**:506–577.

Moser F, Irwin D, Chen S, Wilson DB. 2008. Regulation and characterization of Thermobifida fusca carbohydrate-binding module proteins E7 and E8. *Biotechnol. Bioeng.* **100**:1066–77.

Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**:355–367.

Payne CM, Resch MG, Chen L, Crowley MF, Himmel ME, Taylor LE, Sandgren M, Ståhlberg J, Stals I, Tan Z, Beckham GT. 2013. Glycosylated linkers in multimodular lignocellulose-degrading enzymes dynamically bind to cellulose. *Proc. Natl. Acad. Sci. U. S. A.* **110**:14646–51.

Prlić A, Bliven S, Rose PW, Bluhm WF, Bizon C, Godzik A, Bourne PE. 2010. Pre-calculated protein structure alignments at the RCSB PDB website. *Bioinformatics* **26**:2983–2985.

Saito a., Fujii T, Miyashita K. 2003. Distribution and evolution of chitinase genes in Streptomyces species: Involvement of gene-duplication and domain-deletion. *Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol.* **84**:7–16.

Scott BR, Huang HZ, Frickman J, Halvorsen R, Johansen KS. 2015. Catalase improves saccharification of lignocellulose by reducing lytic polysaccharide monooxygenase-associated enzyme inactivation. *Biotechnol. Lett.*

Shu Z, Wang Y, An L, Yao L. 2014. The Slowdown of the Endoglucanase Trichoderma reesei Cel5A-Catalyzed Cellulose Hydrolysis Is Related to Its Initial Activity. *Biochemistry* **53**:7650–7658.

Skubák P, Pannu NS. 2013. Automatic protein structure solution from weak X-ray data. *Nat. Commun.* **4**:2777.

Vaaje-Kolstad G, Houston DR, Riemen a. HK, Eijsink VGH, Van Aalten DMF. 2005. Crystal structure and binding properties of the Serratia marcescens chitin-binding protein CBP21. *J. Biol. Chem.* **280**:11313–11319.

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**:235–242.

Witteveen CFB, Veenhuis M, Visser J. 1992. Localization of Glucose-Oxidase and Catalase Activities in Aspergillus-Niger. *Appl. Environ. Microbiol.* **58**:1190–1194.

Wu M, Beckham GT, Larsson AM, Ishida T, Kim S, Payne CM, Himmel ME, Crowley MF, Horn SJ, Westereng B, Igarashi K, Samejima M, Ståhlberg J, Eijsink VGH, Sandgren M. 2013. Crystal structure and computational characterization of the lytic polysaccharide monooxygenase GH61D from the basidiomycota fungus Phanerochaete chrysosporium. *J. Biol. Chem.* **288**:12828–12839.

Ye Y, Godzik A. 2003. Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics* **19**.

Zhou W, Irwin DC, Escovar-Kousen J, Wilson DB. 2004. Kinetic studies of Thermobifida fusca Cel9A active site mutant enzymes. *Biochemistry* **43**:9655–63.

Zor T, Selinger Z. 1996. Linearization of the Bradford protein assay increases its sensitivity: theoretical and experimental studies. *Anal. Biochem.* **236**:302–8.

Chapter 3. Fourier transform infrared spectroscopy (FTIR) to monitor cellulose crystallinity changes

3.1Fourier transform infrared spectroscopy (FTIR) and cellulose crystallinity

The mechanistic basis of cellulase nonlinear kinetics remains unclear, but new techniques can determine the role of changing cellulose crystallinity. Cellulases modify the cellulose substrate during digestion, and the impact of these structural changes on cellulase kinetics requires further investigation (Kostylev and Wilson, 2013; Väljamäe et al., 2003; Xu and Ding, 2007). Substrate modifications are the dominant cause of decreasing cellulase activity; when new substrate is added to an exhausted reaction the product release returns to the high initial rate, but if new cellulase is added to spent substrate the reaction rate remains low (Zhang et al., 1999). Substrate recalcitrance is the combination of physical factors that contribute to reduced cellulase activity. Cellulose crystallinity, the degree of order in cellulose chain packing, is thought to play a major role in the activity of most cellulases. Two hypotheses of the role of crystallinity in substrate recalcitrance have been established. Increasing crystallinity has been persistent as a hypothesis due to its concise logic; crystalline regions are more difficult to degrade, so as cellulases preferentially target the less ordered chains the crystalline regions become enriched and slow the reaction (Mansfield et al., 1999; Park et al., 2010). Alternatively, increasing substrate heterogeneity may provide the mechanistic basis of decreasing cellulase activity. Accessible cellulase active sites decrease as substrate heterogeneity increases, leading to decreased surface area, collapsed pore structure, and surface obstacles that inhibit movement (Zhang et al., 1999). While past investigation has presented conflicting results, Fourier transform infrared spectroscopy (FTIR) can determine whether increasing cellulose crystallinity is correlated with nonlinear cellulase kinetics.

As discussed in previous chapters, Lytic Polysaccharide Monooxygenases (LPMOs) attack crystalline cellulose to enhance cellulase activity, but cellulose crystallinity has not been established as the mechanism behind LPMO synergism. Fourier transform infrared spectroscopy (FTIR) can determine if LPMOs decrease bulk cellulose crystallinity to stimulate cellulase activity. Two main hypotheses have been considered to explain LPMO synergism, a bulk substrate amorphogenesis model and a random cleavage model (Vaaje-Kolstad et al., 2010). The bulk substrate amorphogenesis model proposes that LPMO activity creates amorphogenesis (structural swelling or dispersion) of crystalline cellulose, which loosens the bulk substrate for cellulase (Arantes and Saddler, 2010). The random cleavage model is the default mechanism, that LPMOs create new active sites for cellulase attack through random endo-like oxidative cleavage on otherwise inaccessible chains in the cellulose crystal. By measuring residual substrate crystallinity, FTIR can resolve the debate between models of LPMO synergistic activity.

The cellulase field has worked to understand the close relationship between cellulose recalcitrance and cellulase kinetics. While recalcitrance is thought to involve crystallinity, the relative contributions of crystallinity and other structural properties have not been resolved. Due to the technical challenges of characterizing insoluble cellulose, recalcitrance is most often observed indirectly as decreasing soluble sugar release during cellulase reactions (Kostylev and Wilson, 2013). Cellulose crystallinity measures are now more accessible using FTIR, X-ray diffraction (XRD), and nuclear magnetic resonance (NMR) (Park et al., 2010). Results from studies using multiple techniques have varied due to structural complexity and sample preparation, with some results correlating crystallinity and cellulase kinetics and other studies showing little or no effect. While crystallinity can be measured with some consistency, more

subtle physical properties including accessible volume, pore size distribution and surface chain disorder remain difficult to detect. Recent advances in atomic force microscopy (AFM) enable cellulose surfaces to be observed directly with low resolution, or indirectly using single enzyme particles (Igarashi et al., 2009; Jeoh et al., 2013).

Physical changes connect substrate recalcitrance with nonlinear cellulase kinetics, and increasing crystallinity has been suspected as the property diminishing cellulase activity. Fourier transform infrared spectroscopy (FTIR) enables measurement of crystallinity to determine whether cellulases slow down as they create increasingly crystalline cellulose. Correlation of crystallinity with kinetic drop-off would provide a mechanistic basis of nonlinear cellulase reaction kinetics, as well as explaining the role of LPMOs in disrupting crystalline cellulose regions. FTIR measurements can establish a correlation between changing cellulose crystallinity and the underlying physical basis of nonlinear cellulase kinetics.

3.2. Previous results exploring cellulose hydrolysis using Fourier transform infrared spectroscopy (FTIR)

The long held connection between cellulose crystallinity and cellulase kinetics has been investigated by others to explain nonlinear kinetics. While the correlation between cellulase digestion and cellulose crystallinity has not been clearly established, recent positive results using a new FTIR measurement technique warrant further investigation. Research investigating the crystallinity of residual cellulose has explored multiple variables; utilizing different detection methods, cellulases with different mechanisms, and cellulose model substrates. In some instances no effect was observed, while in others cellulase activity significantly affected cellulose crystallinity.

In early work incorporating FTIR to study cellulose structure, Boisset et al. investigated

cellulose structure and crystallinity changes after deconstruction by cellulases (Boisset et al., 1999). *Clostridium thermocellulm* cellulosomes, large protein scaffolds containing many highly synergistic cellulases, were used to digest both bacterial cellulose (BC) and *Valonia ventricosa* cellulose crystals. The residual cellulose was then characterized using X-ray diffraction (XRD), FTIR, and transmission electron microscopy (TEM). XRD measurements demonstrated that despite being attacked by cellulases, the crystallinity of this residual cellulose remained unchanged. Cellulose crystal packing (arrangement of the cellulose unit cell) was measured using FTIR, and also did not change with increasing digestion. However, visual analysis using TEM showed significant physical changes to the BC: breakage of the cellulose microfiber ribbons, fraying and tapering of the ribbons, and finally a loss of the ribbon structure entirely. Using a complex cellulase mixture, cellulose crystallinity appeared unchanged despite visual observation of significant substrate morphology changes.

Other studies investigated a complex mixture of cellulases, a commercial *T. reesei* cocktail, compared with purified *Tr*Cel7A exocellulase alone using nuclear magnetic resonance (NMR) and XRD (Hall et al., 2010). The crystallinity of Avicel® remained unchanged over the course of digestion, even in samples nearly completely (~90%) degraded. Cellulose with a range of decreasing crystallinity was made using phosphoric acid, which showed that initial hydrolysis rates correlate inversely with crystallinity. The absence of change in crystallinity led to the conclusion that cellulase kinetics slowdown is not a function of substrate crystallinity, but instead results from some other surface modification which accumulates during digestion.

Using purified cellulases rather than complex mixtures, Mansfield and Meder investigated cellulase digestion of Sigmacell®, a highly crystalline model cellulose (Mansfield and Meder, 2003). Changes in crystallinity caused by *Cellulomonas fimi* endocellulases,

exocellulases, and their synergistic mixtures were measured using both FTIR and solid state NMR (26). The results showed that endocellulases increased the overall crystallinity, while exocellulases reduced it. The magnitude of both changes was approximately 5%, close to measurement error, and both approaches produced similar values. These results match expectations based on the mechanisms of the enzymes; endocellulases primarily attack amorphous regions, and exocellulases processively cleave an extracted crystalline chain. The apparent crystallinity changes were modest despite high degree of digestion of up to 30-50% for individual enzymes. A nonstandard determination of crystallinity may have influenced these measurements, and more common FTIR crystallinity indexes will be discussed below.

A similar design of applying purified cellulases to highly crystalline Avicel® and BC was used to investigate the effect of digestion on crystallinity for both native and modified cellulose (Chen et al., 2007). Mercerization involves the use of concentrated base to induce structural rearrangement, converting native cellulose to a more digestible form. Multiple purified cellulases from *Thermobifida fusca* were tested on each substrate: an exocellulase (Cel6B), endocellulase (Cel6A), and a processive endocellulase (Cel9A). XRD results showed no change to crystallinity in either of the native cellulose substrates, but significant crystallinity changes for both mercerized substrates. Although similar with respect to initial crystallinity, mercerized cellulose had undergone extensive structural changes and could not be compared directly with either native bacterial cellulose or cellulose within plant biomass.

More recently, Corgie et al. developed methods using high throughput screening FTIR (HTS-FTIR) to measure thin film BC samples (Corgié et al., 2011). Despite using similar *T. fusca* cellulases as Chen et al. (*Tf*Cel6B, *Tf*Cel5A, and *Tf*Cel9A), Corgie et al. observed very different changes in cellulose crystallinity. Residual BC from each of the cellulase conditions

that were tested resulted in significant changes, as measured using multiple FTIR crystallinity indexes. Using similar substrate and cellulase conditions, the lower magnitude XRD results of Chen et al. are in disagreement with the substantial effects seen using HTS-FTIR. Possibly contributing to the different observations, the HTS-FTIR methods involved minimal sample treatment, sources of overlapping signals, and normalization to a volatile internal standard. This presents an opportunity to validate and improve the HTS-FTIR methodology to obtain more reliable and meaningful data. Cellulose crystallinity results based on improvements to the existing HTS-FTIR methods will be discussed in a later chapter.

Although it is a commonly held hypothesis, the link between changes in crystallinity and cellulase digestion has not been conclusively established by existing work. Measurements of changes in crystallinity appear to differ significantly depending on the sampling approach and conditions used. Multiple reports have demonstrated some change to bacterial cellulose crystallinity using free cellulase systems, justifying further investigation. The results of the HTS-FTIR approach significantly increased substrate crystallinity after cellulase hydrolysis. This work using HTS-FTIR to measure changes in cellulose crystallinity will be replicated and expanded in order to solidify a correlation between increasing cellulose crystallinity and declining rates of cellulase hydrolysis.

3.3. Fourier transform infrared spectroscopy (FTIR) measures bond vibration

FTIR enables faster and lower cost sampling compared with other approaches, which is advantageous for investigating changes occurring gradually over time such as cellulose hydrolysis. Different spectroscopy approaches have different advantages and disadvantages, with FTIR best suited for measuring many cellulose samples. Molecular bonds absorb energy from radiation over a wide range of wavelengths, and molecular spectroscopy makes use of this

absorbance to identify and quantify chemical compounds. Using high energy radiation, X-ray spectroscopy can measure insoluble substrates using dispersion or absorption, but is limited by low throughput and destructive sampling. Spectroscopy in the ultraviolet and visible ranges (UV/Vis) is nondestructive, but relies on electronic excitation of uncommon molecular bonds. Infrared (IR) spectroscopy uses lower energy radiation and instead excites vibrational modes, which are present in all bonds with a dipole moment. At the low end of radiation energy, microwave radiation can measure rotational states of bonds, but is mostly limited to gaseous samples. With unique advantages provided by the underlying technology, FTIR is the spectroscopic approach best suited for measuring cellulose crystallinity in this context.

UV/Vis spectroscopy is the most widely used spectroscopic technique, and the approach of excitation for chemical analysis is conceptually very similar to FTIR. The advantage of UV/Vis spectroscopy is its simplicity and accurate measurement of individual bonds using electron excitation, but it is limited to fewer chemical moieties and cannot measure cellulose structure directly. In UV/Vis spectroscopy, energy provided by absorbed radiation causes electrons within the valence shell to transition to higher energy molecular orbitals (Robinson, James W., Skelly, Eileen, Frame, 2014). These higher molecular orbitals are only accessible when available orbitals or unbound electrons are able to accommodate the higher energy state, requiring the presence of particular chemical moieties. The characteristic UV/Vis absorbance spectra, the aggregate excitation of these bonds, enables reproducible identification and quantification of molecules. Shorter wavelength radiation, visible (800 nm to 400 nm) and ultraviolet light (400 nm to below 200 nm), imparts enough energy to excite electrons within valence shells. With longer wavelengths between 800 nm to 10,000 μm (10 cm), infrared radiation (IR) is lower in energy and not capable of electronic excitation. Instead, IR radiation

excites vibrational levels within molecules. While UV/Vis detects the presence of particular bonds, IR spectroscopy measures bond identity, movement, and interactions in the sample.

The other characteristic difference between UV/Vis and FTIR spectroscopy is how they expose the sample to light, either one frequency at a time or many frequencies simultaneously. UV/Vis spectroscopy is an example of frequency-domain spectroscopy, which measures the intensity of absorbance of many frequencies sequentially to create a spectrum. This sequential frequency measurement is time consuming and constrains detection to a narrow range of wavelengths. In contrast, FTIR is a time-domain spectroscopy approach which measures many frequencies simultaneously by sampling the sum of overlapping signals. The absorbance at each frequency in the signal can then be determined by computational processing. The Fourier transform algorithm converts overlapping time-domain signals of different amplitude and frequency into frequency-domain signals. This transformation results in an absorbance value for each frequency in the initial signal, similar to the spectrum produced by UV/Vis spectroscopy. The use of a much wider wavelength spectrum, exciting all the vibrational modes in a sample simultaneously, is enabled by this time-domain decomposition approach.

At IR frequencies, the peaks and troughs of electromagnetic waves cannot be measured over time directly. Instead, FTIR relies on a combination of hardware and software techniques to enable time-domain spectroscopy. Infrared radiation is too high in frequency for existing phototransducers (which convert IR radiation into current) to match pace with. This hardware limitation is solved by a Michelson interferometer, which modulates high frequency signals to measurable time-domain signals without a loss of fidelity. The modulation of high frequency waves for measurement using a Michelson interferometer is a key concept of FTIR spectroscopy. Within a Michaelson interferometer, the signal is split and recombined by a rapidly moving

mirror to create constructive and destructive interference. An interferogram is produced from measuring this interference. This interferogram is a set of discretely sampled time domain signal measurements that is directly proportional to the original spectrum, which can be processed into a frequency domain spectrum. The Fast Fourier Transform algorithm performs this processing to produce the goal of time-domain spectroscopy, a wide spectrum of sample absorbance as a function of wavenumber (Skoog et al., 2007).

3.4. Fourier transform infrared spectroscopy (FTIR) terminology

Transmittance is the proportion of incident light passing through the target sample. Sample interaction is presented most commonly as absorbance, which is calculated from the initial sample transmittance by:

-log (transmittance)

Absorbance is plotted as a spectrum with respect to wavenumber, a characteristic property of propagating waves defined as the spatial frequency of the wave in units of reciprocal centimeters 1/cm or cm^{-1} (Thompson and Taylor, 2008).

The relationship between wavenumber and wavelength is defined as:

$\nu = 1 / \lambda$.

Where: ν = wavenumber and λ = wavelength

Regions of the electromagnetic spectrum relevant to FTIR are near-, mid- and far-IR (Griffiths and De Haseth, 2007). Near-IR has the shortest wavelength in the 0.8–2.5 μm range, corresponding to wavenumbers of 14000–4000 cm^{-1}. The mid-IR region is the most commonly used region for FTIR. It spans the wavelength range between 2.5–25 μm, or the 4000–400 cm^{-1} wavenumber range. The mid-IR region is useful for most biological molecules due to the

absorbance by vibrational modes common in biological molecules, e.g. carbon and heteroatom single and double bonds (Baker et al., 2014). The far-IR region refers to wavelengths in the 25–1000 μm range, corresponding to wavenumbers of 400–10 cm^{-1}. Wavelengths longer than 50 μm are relevant to biological processes involving much lower frequency modes such as protein folding and DNA macro-scale structural changes, due to the low spatial resolution (Miller et al., 2003).

3.5. Fourier transform infrared spectroscopy (FTIR) measures chemical bond interactions

Energy from infrared radiation is absorbed by chemical bonds to produce changes different from those measured by other spectroscopy approaches. In order to understand how the raw signal of FTIR spectroscopy, transmittance, allows the identification of particular bonds and bond structural arrangements, several concepts must be defined. These include chemical bond absorbance, bond vibrational modes, and neighborhood effects.

A chemical bond is dynamic, the nuclei of bonded atoms constantly move in relation to each other while tethered by shared electrons. Bonds are considered to be IR active when energy absorbed from infrared radiation results in atomic displacement of the bonded nuclei. Only bonds with a dipole moment, nuclei with unequal separated charges moving to create a field, are able to interact with the electric field of IR radiation (Baker et al., 2014). IR active bonds may absorb energy when there is a match between two frequencies, the frequency of the IR electromagnetic field and the frequency of the bond dipole moment field. This resonance of matching frequencies transfers energy to the chemical bond, resulting in vibrational excitation in the bond. A diverse range of chemical interactions may be IR active, as both covalent and non-covalent bonds can absorb IR radiation and change vibrational states.

As an example of IR active bonds absorbing energy from infrared radiation, compare nitrogen and oxygen, atmospheric gasses without dipole moments, to water and carbon dioxide. As homonuclear diatomic molecules, N_2 and O_2 chemical bonds have no charge separation, resulting in a lack of IR absorption at any frequency (Chalmers, 2002). In contrast, due to substantial dipole moments, H_2O and CO_2 readily absorb infrared radiation in the mid-IR range. The characteristic FTIR spectrum of atmospheric gasses is shown in Figure 1.1. Both H_2O and CO_2 have peaks in this FTIR spectrum, between 2000-1200 cm^{-1} and 4000-3500 cm^{-1} for H_2O and at 700 cm^{-1} and 2400 cm^{-1} for CO_2. As would be expected due to a lack of dipole moment, neither N_2 nor O_2 are detected by FTIR.

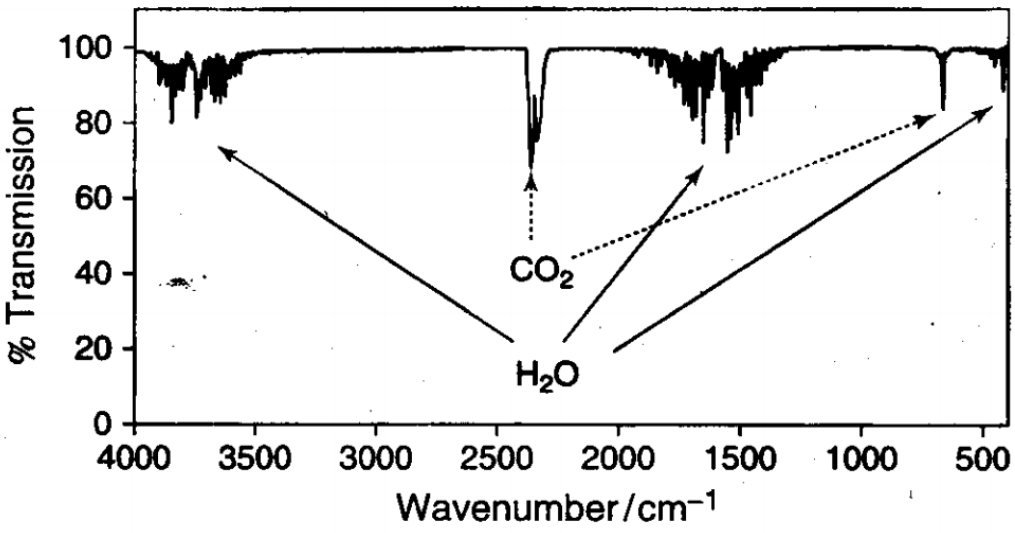


Figure 1.1. FTIR signals resulting from water and carbon dioxide, IR-active atmospheric contaminants (reproduced from Chalmers, 2002).

The IR spectrum for CO_2 and H_2O has multiple peaks, as molecules can vibrate in multiple ways. Vibrational normal modes describe the perturbation of bonded atoms in terms of a harmonic oscillator, the most common include stretching, bending, and wagging (Eichhorn et al., 2005). The vibrational modes of each chemical bond are determined by physical properties,

including atomic mass and bond distance. Badger's Rule, that the strength of a chemical bond is correlated with the frequency of its vibrational modes, illustrates a clear connection between the chemical and spectroscopic properties of a bond (Badger, 1935). In order to be excited to a higher vibrational state, a bond must absorb a discrete amount of energy equal to the vibrational mode transition energy. Bonds with multiple vibrational modes can simultaneously absorb energy from multiple frequencies of infrared radiation, resulting in vibrational excitation that produces a unique absorbance spectrum (Faix, 1992). FTIR measures the IR frequencies that delivered the exact energy needed to excite bond vibrational modes within a molecule.

FTIR can also measure chemical structure, because chemical bonds affect the vibrational modes of nearby bonds. Neighborhood effects are local non-covalent interactions that influence vibrational modes, which include hydrogen bonding, Van der Waals forces, and polar electrostatic interactions. Chemical bonds in a three-dimensional matrix, such as cellulose, have complex neighborhood effects that influence the resulting FTIR spectrum. In solution, the neighborhood effects of intermolecular complexes are more variable, such as the hydrogen bonding interactions of γ -butyrolactone, but remain measurable (Hesse and Suhm, 2009). The structure of biomacromolecules such as cellulose is measurable because of the local noncovalent interactions between adjacent chemical bonds.

3.6. Macromolecule Fourier transform infrared spectroscopy (FTIR) spectral complexity

To make use of IR absorbance to characterize crystallinity, spectra of macromolecules must be carefully interpreted. The complexity of cellulose spectra prevents the measurement of structural crystallinity by quantifying individual peaks. Within the working range of the instrument, chemical bond concentrations are quantifiable since they obey the Beer-Lambert law

(Robinson, James W., Skelly, Eileen, Frame, 2014): signal strength changes in proportion to concentration. While individual peaks are easily quantified, the FTIR spectrum of macromolecules is the additive sum of numerous bonds and their complex interactions. FTIR peaks often overlap due to the multiple chemical moieties, each with multiple vibrational modes, and the neighborhood effects of the resulting combination. Thus, macromolecular FTIR spectra are commonly a mixture of many overlapping peaks difficult to assign as a particular vibrational mode, and fewer separated peaks that are more easily identified. To account for this, these complex spectra are summarized in order to interpret structural features from complex overlapping peaks. Cellulose in particular has regions of spectrum that provide a consistent reference point for identification, as well as a subset of distinct and well characterized peaks which provide a means for structural characterization (Baker et al., 2014).

As a macromolecule composed of ordered polysaccharide chains, exploring the structural features of cellulose using FTIR is complicated by the additive absorbance of the component bonds of anhydrous glucose. Due to the overlap of multiple frequencies, it is difficult to isolate the contribution of particular chemical bond vibrational modes, but the cellulose spectrum has characteristic regions that expose qualitative structural details. The cellulose H-bonding region is centered at 3350 cm^{-1}, and is dominated by intrachain hydrogen bonding between the 3OH...O5 of joined pyranose rings. This H-bonding region is a combination of multiple smaller peaks from the more subtle interchain hydrogen bonds, which overlap in the spectrum to produce a broad band relating to cellulose structure. The assignment of peaks to particular bond vibrational modes began with such interchain interactions, as it was observed that the number of peaks in the FTIR spectrum greatly exceeds the number of OH groups for both cellulose (Michell, 1990) and soluble monosaccharides (Michell, 1988). More thorough assignment of these FTIR peaks was

performed using correlation with X-ray measurements, for example assigning hydrogen bond peaks by correlating frequency with bond length (Michell, 1988). The other characteristic region in cellulose is the fingerprint region, where the many carbon -carbon single bonds produce a complex set of peaks due to the many bending vibrations in crystalline cellulose. The fingerprint region spans approximately 1500-895 cm^{-1}, and provides peaks that can be used to describe cellulose structure. The fingerprint region contains single-bond moiety movements such as C-O and C-H bending, as well as the β-glycosidic bond signal at 896 cm^{-1}.

Returning to the use of FTIR to measure cellulose crystallinity, measuring crystallinity presents challenges due to the macromolecular structure of cellulose. As established above, crystallinity of cellulose cannot be measured based on individual peaks or characteristic regions due to the influence of overlapping signals. Crystallinity can be observed on a broad scale in the H-bonding region, where the disorder of adjacent chains results in fewer interchain interactions. Crystalline cellulose is more structurally ordered, resulting in a much more defined and intensely absorbing hydrogen bonding region compared with amorphous cellulose. Although these regions contain information related to crystallinity, there are too many peaks in this region to reliably assign each contribution and accurately characterize the cellulose structure. Additionally, when cellulose is degraded the resulting spectra may change as neighborhood effects influence peak position, shape, and height. In cases of extensive digestion, removal of a chemical moiety can lead to disappearance of the lower amplitude peak into an overlapping peak. Further complicating analysis, peaks relating cellulase digestion and crystallinity are in the complex hydrogen bonding and cellulose fingerprint regions. While not every peak can be reliably assigned, individual peaks can be consistently assigned based on their relative isolation and characteristic wavenumbers (Carrillo et al., 2004). Some peaks directly correlate with crystalline

structure, but variable sample concentrations prevent using a single peak to describe crystallinity. Due to the complexity of chemical interactions within cellulose, crystallinity cannot be assigned based on peaks derived from deconvoluted regions. The solution lies in using a combination of signals: peaks describing bonds which do not change with changing crystallinity, and peaks reflecting the organization of the cellulose lattice.

3.7. Fourier transform infrared spectroscopy (FTIR) crystallinity indexes

The challenge of using FTIR to characterize cellulose with identical chemical composition but differences in structural crystallinity is addressed using ratios of characteristic bonds. A crystallinity index (CI) is the ratio of a peak related to crystallinity to a peak not representing crystallinity within the same spectrum. Peaks representing crystallinity can be less distinct, as they represent vibrational modes of less constrained bonds. These peaks representing bond freedom, such as bending and H-bonding, are used in conjunction with peaks of more restrained modes to characterize cellulose crystallinity more accurately than using either alone. Multiple crystallinity indexes have been used to describe different cellulose structural features, and are an established approach to describe crystallinity changes in cellulose samples (Park et al., 2010). The crystallinity indexes most commonly employed include the Lateral Order Index (LOI), Hydrogen Bonding Intensity (HBI), and Total Crystallinity Index (TCI). These ratios describe changes to cellulose structure in a consistent way, accounting for sample properties such as differences in concentration, peaks that are low in amplitude (e.g. the glycosidic bond at 896 cm^{-1}), or peaks present only as a shoulder to other peaks. The use of such ratios is well established for FTIR spectroscopy since signal intensity within a single spectrum is internally consistent and less susceptible to variance from sample heterogeneity and preparation (Corgié et

al., 2011). Crystallinity indexes have been used to investigate changes to crystallinity after hydrolysis of crystalline cellulose (Hayashi et al., 1998), and are supported by alternative approaches for measuring cellulose crystallinity such as X-ray diffraction (XRD) and nuclear magnetic resonance (NMR) (Colom et al., 2003).

The Lateral Order Index (LOI) described by Hurtubise and Krassig (Hurtubise and Krassig, 1960), and expanded by Nelson and O'Conner (Nelson et al., 1964), is the ratio of 1427 cm^{-1}/895 cm^{-1}. Cellulose signal at 1429 cm^{-1} is attributed to the scissoring motion of CH_2 bonds. The 1430 cm^{-1} is associated with the amount of crystalline structure of cellulose. The β-glycosidic bond signal at 896 cm^{-1} is attributed to the vibrational mode of C-1 within the β-glycosidic bond between anhydrous glucose (Hayashi et al., 1998). The 895 cm^{-1} peak is associated with the more amorphous region in cellulose (Poletto et al., 2013). The ratio of the amorphous regions to the crystalline cellulose at 1430 cm^{-1} indicates the degree of order within the cellulose crystal.

The Hydrogen Bonding Intensity (HBI), described by Nada et al., is the ratio of 3350 cm^{-1}/1337 cm^{-1} (Nada et al., 2000). This corresponds to the H-bonding of the 3OH...O5 related to the degree of C-OH group in-plane bending. This hydrogen bond intensity value gives an indication of cellulose chain mobility and bond distance. Since chain mobility is a key component of crystallinity, the HBI value is closely related to the crystal system of the cellulose structure and the degree of intermolecular regularity within the crystal.

The Total Crystallinity Index (TCI), described by Nelson and O'Connor (Nelson et al., 1964), is the ratio of 1373 cm^{-1}/2900 cm^{-1}. This ratio describes the amount of bending of C-H, a characteristic peak of cellulose I to the amount of C-H and CH2 stretching. This ratio relates the amount of amorphous cellulose to the more crystalline regions. The bending of C-H represents

the amount of amorphous and non-lattice type of cellulose, and is not affected by interactions with water. The denominator is the C-H and CH2 stretching, which is largely unaffected by changes in crystallinity. Therefore, the change to C-H bending in amorphous cellulose indicates the overall degree of order of the cellulose polymer. In aggregate, these three indexes provide a summary of the total crystallinity within a cellulose sample (Poletto et al., 2013).

3.8. Model cellulose and cellulase rationale

The ultimate goal of investigating cellulose crystallinity using Fourier transform infrared spectroscopy (FTIR) is to apply this approach to optimizing cellulase activity on real plant biomass. Model cellulose provides a representative crystalline microfiber substrate that is useful for FTIR measurement, as current crystallinity indexes developed using pure model cellulose can't be applied to biomass spectra. The physical properties of cellulose in plants are dynamic; plant cellulose crystallinity, length, and diameter change dramatically during development (Seagull et al., 2000). Most higher plants such as cotton produce cellulose microfibers between 5-10 nm in diameter (Brett, 2000), but some such as *Arabidopsis* produce microfibers of 2-5 nm in diameter. Cellulose structure differs between locations within the plant, as cellulose in primary cell walls is significantly smaller in diameter relative to secondary cell walls (Anderson et al., 2010). Within biomass, cellulose is one of many components; even the simplest biomass is composed of numerous structural macromolecules. Adding to this complexity, biomass processing and pretreatment extensively modify covalent and noncovalent bonds in ways that are not well understood. Even if these properties of plant biomass were held constant, digestion modifies both cellulose and non-cellulose signals, changing the local environment and affecting the interaction of all the different components in unpredictable ways. These factors (cellulose

diversity, other macromolecules, and changing bonds) combine to produce biomass FTIR spectra that are currently too complex for reliable peak assignment models. FTIR may become be a powerful tool to measure recalcitrance changes in industrially relevant plant biomass as more sophisticated models are developed.

The choice of model cellulose is a critical decision for FTIR investigation. The ideal model cellulose would have structural properties comparable to those of plant cellulose, including crystalline allomorph (the arrangement of crystalline unit cell and associated hydrogen bonds), degree of crystallinity, and microfiber diameter. Cellulose I is the crystalline allomorph found both in plants and in two commonly used model celluloses, Avicel® and bacterial cellulose (BC). Avicel® is a recalcitrant cellulose made from processed biomass, advantageous because of its high crystallinity, but its large particle size and residual non-cellulose material make it a suboptimal substrate for FTIR. Bacterial cellulose (BC), synthesized by *Acetinobacter xylinum* biofilms, is made up of ribbon-like microfibrils with narrow cross-sections. BC has high crystallinity and surface area, both positive traits for use as a useful model cellulose. Other cellulose allomorphs can be created chemically from cellulose I, including amorphous phosphoric acid swollen cellulose (PASC) (Zhang et al., 2006) and cellulose II produced by mercerization (Park et al., 2010) , both of which decrease the crystallinity and increase digestibility. PASC is a highly homogeneous amorphous cellulose made by dissolving and reprecipitating Avicel® into a disordered structure. Amorphous cellulose has random chain packing, resulting in broad and overlapping peaks that cannot be reliably deconvoluted for crystallinity index measurement. Cellulose II is an easily digested allomorph generated through mercerization, but is so extensively structurally changed that it is difficult to generalize to more native cellulose I. Therefore, for the experimental study of enzymatic changes in cellulose

crystallinity described in the following chapter, bacterial cellulose was chosen as the model substrate based on its multiple advantageous properties. Despite its more ribbon-like structure, BC is the same cellulose I allomorph and has higher crystallinity than most plant cellulose (~75% compared with ~45% for cotton) (Boisset et al., 1999). The largest structural difference, the ribbon-like structure, provides some advantages. The structural morphology of BC gives it a higher surface area for cellulase binding and attack, and enables consistent distribution and homogeneous drying. These features make BC the optimal model cellulose to investigate changes in cellulose crystallinity caused by cellulases.

Cellulases from the *T. fusca* model system were chosen based on their distinct mechanisms, extensive characterization, and prior investigation of their effect on cellulose crystallinity. During their cleavage of cellulose, cellulases reach a digestibility limit that indicates an accumulation of inhibitory substrate changes. For example, exocellulases may remove more easily attacked regions to leave recalcitrant cellulose with increased crystallinity. The exocellulase Cel48A from *T. fusca* is effective in synergistic mixtures, but alone can only digest a small fraction of crystalline cellulose. In contrast, the processive endocellulase Cel9A is the most effective cellulase on crystalline cellulose. Cel9A reaches a higher extent of digestion than exocellulases, and must change the substrate in a different way or overcome the recalcitrance that is produced. This difference in efficacy makes Cel9A a useful counterpoint to Cel48A to observe changes in crystallinity. Cellulase mixtures, including all industrial fungal cocktails, interact synergistically to digest cellulose rapidly. *T. fusca* ER1 crude supernatant is a complex cocktail of many different enzymes, including numerous synergistic cellulases and auxiliary activity enzymes, that breaks down highly crystalline cellulose without experiencing the digestibility limits characteristic of individual cellulases (Kostylev and Wilson, 2013).

Synergistic cooperation likely involves changes to cellulose structure that maximize the activity of the synergistic cellulases. These model cellulases were chosen to reproduce and expand on the significant results from Corgie et al., and to understand how their diverse mechanisms affect cellulose crystallinity.

The FTIR measurement conditions significantly affect the quality of the resulting data, and major choices including detection mode, working range, and internal standards must be carefully considered. Cellulose has bonds with vibrational modes that absorb IR radiation in the same range as most other biomacromoleulces, so FTIR measurement is limited to the mid-IR wavenumbers between 4000 and 400 cm^{-1}. While FTIR can measure cellulose as thin films or pressed KBr cakes, cellulose thin films are superior since they avoid the particle size and packing issues that cause poor signal quality (due to diffuse reflection) in KBr cakes (Chalmers, 2002). FTIR measurements can be taken using transmission or attenuated total reflectance (ATR) modes, but transmission mode is superior for cellulose measurements due to the ATR requirement of homogenous refractive index and the possibility of specular reflection. FTIR measurements deviate from the absorbance/concentration relationship defined by the Beer-Lambert law when the cellulose cake is too thick, the 1.0AU critical limit (corresponding to 10% of initial incident light transmittance) can be used after empirical confirmation of linearity (Griffiths and De Haseth, 2007). Since cellulose samples prepared by drying into thin films have high inherent variance, the highest quality FTIR data is obtained by measuring multiple technical replicates and multiple biological replicates for each condition. Due to high variance, cellulose samples benefit from the inclusion of an internal standard to normalize the resulting FTIR spectra. The two internal standards used previously, sodium acetate and sodium azide, have significant disadvantages which outweigh their benefits as standards. Sodium acetate produces

an FTIR spectrum that overlaps with cellulose significant, due to the similar chemical composition. Sodium azide requires an adequately exhausted sample drying chamber to prevent toxic exposure.

3.9. Conclusion

Fourier transform infrared spectroscopy (FTIR) can be used to investigate the contribution of crystallinity to cellulose recalcitrance as the substrate changes during digestion. Increasing cellulose crystallinity has been a persistent hypothesis to explain the mechanistic basis of recalcitrance, and specifically changing nonlinear kinetics. Other investigations have not conclusively shown that cellulases change crystallinity during digestion, or that crystallinity is a causative factor in the decrease of cellulase activity. Further investigation is warranted in order to validate and expand on recent results showing a correlation between cellulase activity and increasing cellulose crystallinity (Corgié et al., 2011). As will be discussed in the next chapter, HTS-FTIR was used to carefully demonstrate that digestion by purified cellulases does not change the crystallinity of a recalcitrant cellulose substrate.

REFERENCES

Anderson CT, Carroll A, Akhmetova L, Somerville C. 2010. Real-time imaging of cellulose reorientation during cell wall expansion in Arabidopsis roots. *Plant Physiol.* **152**:787–796.

Arantes V, Saddler JN. 2010. Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis. *Biotechnol. Biofuels* **3**:2–11.

Badger RM. 1935. The Relation between Internuclear Distances and the Force Constants of Diatomic Molecules [3]. *Phys. Rev.* **48**:284–285.

Baker MJ, Trevisan J, Bassan P, Bhargava R, Butler HJ, Dorling KM, Fielden PR, Fogarty SW, Fullwood NJ, Heys K a, Hughes C, Lasch P, Martin-Hirsch PL, Obinaju B, Sockalingum GD, Sulé-Suso J, Strong RJ, Walsh MJ, Wood BR, Gardner P, Martin FL. 2014. Using Fourier transform IR spectroscopy to analyze biological materials. *Nat. Protoc.* **9**:1771–91.

Boisset C, Chanzy H, Henrissat B, Lamed R, Shoham Y, Bayer EA. 1999. thermocellum cellulosome : structural and morphological aspects **835**:829–835.

Brett CT. 2000. Cellulose microfibrils in plants: biosynthesis, deposition, and integration into the cell wall. *Int. Rev. Cytol.* **199**:161–199.

Carrillo F, Colom X, Suñol J., Saurina J. 2004. Structural FTIR analysis and thermal characterisation of lyocell and viscose-type fibres. *Eur. Polym. J.* **40**:2229–2234.

Chalmers JM. 2002. Mid-infrared Spectroscopy: Anomalies, Artifacts and Common Errors. *Handb. Vib. Spectrosc.*:2327–2347.

Chen Y, Stipanovic AJ, Winter WT, Wilson DB, Kim Y-J. 2007. Effect of digestion by pure cellulases on crystallinity and average chain length for bacterial and microcrystalline celluloses. *Cellulose* **14**:283–293.

Colom X, Carrillo F, Nogués F, Garriga P. 2003. Structural analysis of photodegraded wood by means of FTIR spectroscopy. *Polym. Degrad. Stab.* **80**:543–549.

Corgié SC, Smith HM, Walker LP. 2011. Enzymatic transformations of cellulose assessed by quantitative high-throughput fourier transform infrared spectroscopy (QHT-FTIR). *Biotechnol. Bioeng.* **108**:1509–20.

Eichhorn SJ, Young RJ, Davies GR. 2005. Modeling crystal and molecular deformation in regenerated cellulose fibers. *Biomacromolecules* **6**:507–513.

Faix O. 1992. 4.1 Fourier Transform Infrared Spectroscopy. In: Lin, SY, Dence, CW, editors. *Methods Lignin Chem.*

Griffiths PR, De Haseth JA. 2007. Fourier Transform Infrared Spectrometry. Wiley. Chemical Analysis: A Series of Monographs on Analytical Chemistry and Its Applications.

Hall M, Bansal P, Lee JH, Realff MJ, Bommarius AS. 2010. Cellulose crystallinity - A key predictor of the enzymatic hydrolysis rate. *FEBS J.* **277**:1571–1582.

Hayashi N, Sugiyama J, Okano T. 1998. Selective degradation of the cellulose I s component in Cladophora cellulose with Trichoderma viride cellulase **305**:109–116.

Hesse S, Suhm M a. 2009. On the low volatility of cyclic esters: an infrared spectroscopy comparison between dimers of gamma-butyrolactone and methyl propionate. *Phys. Chem. Chem. Phys.* **11**:11157–11170.

Hurtubise FG, Krassig H. 1960. Classification of Fine Structural Characteristics in Cellulose by Infrared Spectroscopy. *Anal. Chem.* **32**:177–181.

Igarashi K, Koivula a., Wada M, Kimura S, Penttila M, Samejima M. 2009. High Speed Atomic Force Microscopy Visualizes Processive Movement of Trichoderma reesei Cellobiohydrolase I on Crystalline Cellulose. *J. Biol. Chem.* **284**:36186–36190.

Jeoh T, Santa-Maria MC, O'Dell PJ. 2013. Assessing cellulose microfibrillar structure changes due to cellulose action. *Carbohydr. Polym.* **97**:581–586.

Kostylev M, Wilson D. 2013. Two-parameter kinetic model based on a time-dependent activity coefficient accurately describes enzymatic cellulose digestion. *Biochemistry* **52**:5656–5664.

Mansfield S, Meder R. 2003. Cellulose hydrolysis–the role of monocomponent cellulases in crystalline cellulose degradation. *Cellulose* **10**:159–170.

Mansfield S, Mooney C, Saddler J. 1999. Substrate and Enzyme Characteristics that Limit Cellulose Hydrolysis. *Biotechnol. Prog.* **15**:804–816.

Michell A. 1988. Second derivative Ft-ir spectra of celluloses I and II and related mono-and oligo-saccharides. *Carbohydr. Res.* **173**:185–195.

Michell A. 1990. Second-derivative FT-IR spectra of native celluloses. *Carbohydr. Res.* **197**:53–60.

Miller LM, Smith GD, Carr GL. 2003. Synchrotron-based biological microspectroscopy: From the mid-infrared through the far-infrared regimes. *J. Biol. Phys.* **29**:219–230.

Nada A-AMA, Kamel S, El-Sakhawy M. 2000. Thermal behaviour and infrared spectroscopy of cellulose carbamates. *Polym. Degrad. Stab.* **70**:347–355.

Nelson ML, Fibers P, Connor RTO, Regional S. 1964. Relation of Certain Infrared Bands to Cellulose Crystallinity and Crystal Lattice Type . Part 11 . A New Infrared Ratio for Estimation of Crystallinity in Celluloses I and 11 * **8**:1325–1341.

Park S, Baker JO, Himmel ME, Parilla P a, Johnson DK. 2010. Cellulose crystallinity index: measurement techniques and their impact on interpreting cellulase performance. *Biotechnol. Biofuels* **3**:10.

Poletto M, Pistor V, Zattera AJ. 2013. Structural Characteristics and Thermal Properties of Native Cellulose.

Robinson, James W., Skelly, Eileen, Frame GM. 2014. Visible and Ultraviolet Molecular Spectroscopy. In: . *Undergrad. Instrum. Anal.*

Seagull RW, Oliveri V, Murphy K, Binder a, Kothari S. 2000. Cotton fiber growth and development 2. Changes in cell diameter and wall birefringence. *J. Cott. Sci* **4**:97–104.

Skoog DA, Holler FJ, Crouch SR. 2007. Principles of Instrumental Analysis 6th ed. Thomson Brooks/Cole. International student edition.

Thompson A, Taylor BN. 2008. Guide for the Use of the International System of Units (SI). *NIST Spec. Publ. 811*. Gaithersburg, MD. Vol. 811 1-85 p.

Vaaje-Kolstad G, Westereng B, Horn SJ, Liu Z, Zhai H, Sørlie M, Eijsink VGH. 2010. An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science* **330**:219–222.

Väljamäe P, Kipper K, Pettersson G, Johansson G. 2003. Synergistic cellulose hydrolysis can be described in terms of fractal-like kinetics. *Biotechnol. Bioeng.* **84**:254–257.

Xu F, Ding H. 2007. A new kinetic model for heterogeneous (or spatially confined) enzymatic catalysis: Contributions from the fractal and jamming (overcrowding) effects. *Appl. Catal. A Gen.* **317**:70–81.

Zhang S, Wolfgang DE, Wilson DB. 1999. Substrate heterogeneity causes the nonlinear kinetics of insoluble cellulose hydrolysis. *Biotechnol. Bioeng.* **66**:35–41.

Zhang Y, Cui J, Lynd L, Kuang L. 2006. A transition from cellulose swelling to cellulose dissolution by o-phosphoric acid: evidence from enzymatic hydrolysis and supramolecular structure. *Biomacromolecules* **7**:644–648.

Chapter 4. Investigating correlation between cellulose digestion and crystallinity using Fourier transform infrared spectroscopy (FTIR)

4.1. Introduction

As was described in previous chapters, cellulases exhibit a well characterized decline in rate when degrading recalcitrant cellulose. Recalcitrance is a physical property of cellulose (and other insoluble polysaccharides such as chitin) which has an unknown structural basis. One model of cellulose recalcitrance attributes the decline of cellulase kinetics to an increase of cellulose crystallinity during the course of digestion. This long held model has not been conclusively demonstrated, but recent results support a role for crystallinity in cellulose recalcitrance. Cellulose recalcitrance may also be caused by other physical properties of the cellulose substrate, for example the obstructing surface features described in the surface erosion model (Väljamäe et al., 1998). Fourier transform infrared spectroscopy (FTIR) is a nondestructive technique to measure cellulose crystallinity, and is employed to clarify the contribution of crystallinity to cellulose recalcitrance and cellulase nonlinear kinetics.

Cellulose is a crystalline lattice of closely packed polysaccharide chains, non-covalently linked by an extensive hydrogen bonding network. Cellulose crystallinity, the degree of organization of the lattice, is correlated with more stable cellulose (Lionetto et al., 2012). As cellulases degrade this organized cellulose lattice, observing changes to cellulose crystallinity would provide insight into the changing cellulase-cellulose interaction and the basis of cellulose recalcitrance. However, measuring changes to the many cellulose non-covalent inter- and intra-molecular bonds presents a technical challenge. FTIR is capable of rapid and noninvasive quantification of both covalent and non-covalent interactions within cellulose, enabling measurement of crystallinity of the material. FTIR is also applicable to more complex and

industrially relevant cellulose applications, and may be used to investigate the effect of biomass type, pretreatment, and enzyme cocktail interactions on sugar release. Characterizing the crystallinity changes resulting from cellulase digestion of a pure cellulose model substrate serves as the basis for future work to overcome cellulose recalcitrance in industrial contexts.

FTIR has been used to measure cellulose crystallinity after cellulase digestion, and is accepted as a powerful approach to explore this relationship. While investigations using multiple techniques to describe cellulose after digestion have generated mixed results, a recent report using FTIR has strongly suggested a correlation between cellulose crystallinity and cellulase digestion. The FTIR technique is sensitive to the cellulose bonds of interest, but is not sufficiently discriminatory of the source of bonds within complex samples such as cellulose hydrolysates. These overlapping chemical bond signals create issues for accurately interpreting the resulting spectra and may have led to erroneous conclusions in prior studies.

This chapter describes experiments using FTIR to measure cellulose crystallinity after digestion by cellulases, in an attempt to correlate cellulase kinetic parameters to crystallinity changes. The goal was to validate and expand on the work of Corgie et al. (Corgié et al., 2011) by using the same measurement approach and cellulase system. Using a model cellulose substrate hydrolyzed using either individual cellulases, incomplete synergistic mixtures, or an intact cellulase secretome, this work builds upon several reported correlations between cellulose crystallinity and digestibility (Boisset et al., 1999; Cao and Tan, 2002; Chen et al., 2007; Corgié et al., 2011).

4.2. Materials and Methods

4.2.1. Substrates and Enzymes.

Cellulose substrates were prepared using methods described in previous chapters, summarized in brief here. Bacterial cellulose (a gift from Monsanto, San Diego, CA) was resuspended overnight at 4 °C, then washed with Milli-Q water with four cycles of centrifugation and resuspension. PASC was prepared according to Zhang et al. (Zhang et al., 2006) by solubilizing Avicel® PH-105 in chilled phosphoric acid at 4 °C, then neutralizing and thoroughly washing using chilled Milli-Q water. The final concentration of each substrate was determined by measuring dry weight in triplicate. Sodium azide was added to all substrates at a final concentration of 0.02% to prevent microbial contamination. The cellulases used, *Thermobifida fuasca* Cel48A, Cel9A, and ER1 crude supernatant, were prepared using previously described methods (Irwin et al., 2003; Kostylev and Wilson, 2011; Li et al., 2010).

4.2.2. Hydrolysis Assays Assembly.

Enzyme and substrate stocks (master mixes) were prepared separately at twofold concentration and combined in Eppendorf protein LoBind 96-well format microplates. Cellulose substrates, ranging in final concentration from 1.25-5.0 mg/ml were distributed by manually dispensing the substrate master mix using wide aperture pipette tips. An equal volume of enzyme master mix, final concentration of 10-1000 nM, was added using an Eppendorf EpMotion 5075Vac liquid handling robot, bringing each well to a final volume of 160 µl and a final buffer concentration of 50 mM sodium acetate (pH 5.5). Plates were sealed (AB-0745 seals, Fisher Scientific), then vortexed briefly on each corner to ensure substrate suspension at the start of incubation. Plates were incubated in a 48-plate tower incubator (StoreX STX40, Liconic US Inc, Woburn, MA) at 50 °C with constant horizontal shaking (150 rpm). Samples were removed at

the appropriate time by a Zymark plate handling system (Twister II, PerkinElmer Inc., Waltham, MA), and reactions were stopped by heating to 100 °C for 5 minutes within a PCR thermocycler heating block (PTC-100, MJResearch inc., Waltham, MA).

4.2.3. Extent of Digestion Quantification.

Samples were centrifuged for 5 minutes at 4,000 rpm (RCF = 3313) to separate residual substrate, then 40 µl of supernatant was transferred to a new plate. For complete secondary hydrolysis of short oligosaccharides to glucose, 120 µl of excess Novo188 β-glucosidase (buffer exchanged with 5 mM NaOAc, pH 5.5, 10% glycerol, 0.22 CBU/ml) (Novozymes, Davis, CA) was added, maintaining the final buffer concentration of 50 mM NaOAc (pH 5.5). A cellobiose standard curve was added to each plate to confirm complete hydrolysis, and plates were incubated for 16 hours as described above.

The extent of digestion was determined by quantifying released glucose using the bismuth-catalyzed p-hydroxybenzoic acid (PAHBAH) method of Lever (Lever, 1977), modified for microscale analysis and validated using existing HPLC quantification methods (Kostylev et al., 2014). The PAHBAH reaction of 120 µl PAHBAH reagent added to 40 µl of hydrolysate was heated using a PCR thermocycler at 70 °C for 10 minutes, cooled to 4 °C for 5 minutes, and finally brought to room temperature as recommended by King et al. (King et al., 2009). The reaction was transferred to flat bottom polystyrene plates (Costar 9017, Corning Inc., Corning, NY), and measured at 410 nm using a microplate spectrophotometer (Synergy 2, BioTek Instruments Inc., Winooski, VT). PAHBAH signal was compared to the hydrolyzed G2 standard curve, quantified as units of anhydrous glucose (162 g mol^{-1}) (Olsen et al., 2014), and extent of digestion was calculated using the following formula:

$$\% \, Digestion = \left(\frac{x \, umol \, glucose}{\left(\frac{x \, mg \, BC \, initial}{\frac{0.162 \, mg}{1 \, umol}} \right)} \right) * 100 \quad (4.1)$$

4.2.4. Cellulose Processing and Plating.

The residual cellulose was washed and the concentration normalized prior to plating. First, the cellulose within the assay hydrolysate was thoroughly resuspended and transferred to a 96-well filter plate (Pall 380 µl 0.45 µm filter plate, SUPOR membrane). The liquid fraction was separated by vacuum filtration and stored for validation by HPLC. The residual cellulose was resuspended in Milli-Q water and washed three times on the filter to remove all soluble compounds. In order to normalize the cellulose concentrations, residual cellulose was resuspended in an appropriate volume of Milli-Q water to account for the fraction of cellulose digested during hydrolysis. For comparison of pipette tip aperture, either narrow or truncated tips were used to resuspend the residual cellulose.

Silicon FTIR HTS plates (Bruker Optics Ltd, Milton, Ontario) were washed by the recommended protocol and thoroughly dried prior to use. The cellulose samples were thoroughly resuspended and 50 µl was transferred to the plate manually using a multichannel pipette or using the EpMotion liquid handling system. Each sample biological replicate was plated in triplicate on the FTIR plate. On each plate, a substrate standard curve of known concentrations was added in triplicate. After samples were distributed onto the silicon plate, variance in surface tension caused occasional incomplete well coverage, which required manual spreading of the sample to evenly cover the well surface. Sample plates were placed in a freshly recharged desiccator for gradual drying at least 16 hours at room temperature. Drying was completed by heating in a vacuum oven at 70 mbar and 70 °C for at least 8 hours. The plates were allowed to

return to 25 °C prior to measurement.

4.2.5. FTIR Measurement and Data Processing.

All FTIR spectra were collected using a Bruker Vertex 70 FTIR spectrometer (Bruker Optics, Germany) using a HTS-XT module. Prior to detection, calibration of the equipment was performed to ensure P and Q validation according to manufacturer instructions. Humidity was monitored and maintained below specification levels. A single well on each silicon 96-well plate was designated as the background position. Spectra were collected in transmission mode for the 7000 cm^{-1} and 400 cm^{-1} wavenumber range, with the spectral resolution set at 2 cm^{-1}. The final signal was the averaged interferogram of 32 scans per well. Two factor zero filling was applied during data collection, and atmospheric compensation was used to remove atmospheric water and carbon dioxide signal.

Data processing was performed with the OPUS software package using a macro script for consistent processing. The raw spectra were converted from transmittance to absorbance and cut to remove wave numbers above 4000 cm^{-1}. Rubber band baseline correction was applied, and the resulting spectra were analyzed without normalization. Second derivative peak height was used for quantitation based on the consistency of signal amplitude after concentration normalization. The intensity values of second derivative peak heights for each sample were exported to a CSV table, and were used to calculate the three crystallinity indexes described in the previous chapter: hydrogen bonding intensity (HBI), lateral order index (LOI), and total crystallinity index (TCI). The resulting values for each sample were imported into OriginPro 2015 (OriginLab, Northampton, MA) for visualization and graphing. The time course reaction kinetics were analyzed by applying the two-parameter power law fit to % digestion data according to Kostylev and Wilson (Kostylev and Wilson, 2013).

4.3. Results

4.3.1. Method improvements.

The processed bacterial cellulose (BC) FTIR spectrum between 4000 cm^{-1} and 400 cm^{-1} contains several characteristic regions serving as landmarks. These regions contain the peaks from individual chemical vibrational modes relevant for measuring crystallinity. A representative standard curve of untreated BC spectra is shown in Figure 4.1.A, showing a BC concentration range from 0.2 - 1.7 mg/ml in 0.3 mg/ml increments. The spectra closely match previously reported cellulose I FTIR spectra, indicating no rearrangement of the cellulose structure during sample processing (Corgié et al., 2011). At the left of this cellulose FTIR spectrum is the hydrogen bonding region, which extends from 3800 cm^{-1} to 3000 cm^{-1} and contains signals from inter and intra-molecular interactions between hydrogen and oxygen in the case of cellulose. A key interaction is measured in the 3350 cm^{-1} peak, which results from the hydrogen bonding interaction between the 3OH-O5 adjacent to the β-glycosidic bond of the cellulose Iα structure. Adjacent to this is the C-H bonding region, which is much narrower between 3000 cm^{-1} and 2800 cm^{-1}, and corresponds to the stretching of C-H moieties. In the middle of the Mid-IR wavenumber range, signals from other heteroatoms such as nitrogen as well as double bonds are absent from spectra of pure cellulose. The most complex region of cellulose FTIR spectra is the fingerprint region from 1430 cm^{-1} to approximately 850 $cm^{-1,}$ which contains signals from the numerous sp3 single bond vibrational modes. Finally, between 850 cm^{-1} and 400 cm^{-1} is the region containing multiple bands corresponding to heavy-atom bending and rotation (Michell, 1990).

Crystallinity indexes calculated from ratios of individual peak heights depend on the consistency of peak resolution, peak wavenumber and peak shape over a large concentration

range. Peak consistency is especially relevant as critical bonds may decrease in height significantly relative to others due to hydrolysis of cellulose. The standard curve of cellulose shown in Figure 4.1.A, represents the entire detection range for HTS-FTIR. In this standard curve, the peak shapes of characteristic cellulose signal regions do not change with respect to concentration. A standard curve demonstrating the linearity of individual peaks critical for determining cellulose crystallinity is shown in Figure 4.1.B. These peaks are consistent in wavenumber position, shape, and have a linear relationship with concentration. Not shown is the smaller 896cm^{-1} peak which is used for calculation of the LOI crystallinity index.

The concentration of cellulose samples used is limited by the maximum signal intensity in the case of BC, as there was no observed tendency to aggregate or flake at higher concentrations. Based on the signal resulting from concentrations shown in Figure 4.1.B, the working range for BC was determined to be 1.0 mg/ml to remain below the Bruker recommended signal limit of 1.0AU for the maximum peak absorbance.

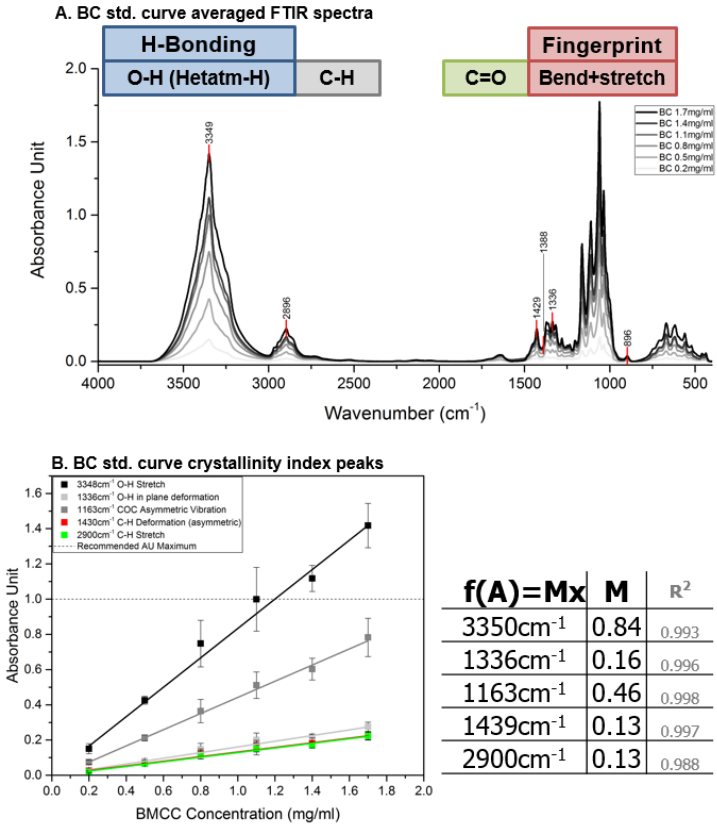


Figure 4.1. BC standard curve measured via transmittance HTS-FTIR. A) Averaged processed FTIR spectra indicating key characteristic peaks used for crystallinity index calculation. B) Second derivative peak height values of characteristic CI peaks (error bars representing standard deviation), with slopes of corresponding peak height values shown with corresponding goodness-of-fit (r^2) values of the linear regressions.

The samples presented in Figure 4.1 were processed using narrow mouth tips, one of several method improvements to improve data quality. The effect of narrow tip aperture was compared with resuspension using wide aperture tips, summarized in Figure 4.2. Samples resuspended using wide aperture pipette tips (4.2.A) or narrow aperture tips (4.2.B) were processed and the crystallinity indexes compared. The coefficient of variance (CV) of narrow tip resuspension, 3.42%, is much lower than that of wide aperture tip resuspension, 14.34%. The white spots observed in wide aperture resuspension samples are regions of very dense BC.

While there is a high consistency for the TCI and HBI crystallinity indexes over the working concentration range of the transmission HTS-FTIR measurement approach, the effect of path length on LOI can be seen where LOI decreases with increasing BC concentration. Due to

this change in LOI, resuspension to similar final concentrations (adjusting for % digestion after buffer exchange) was found to be an essential for meaningful data.

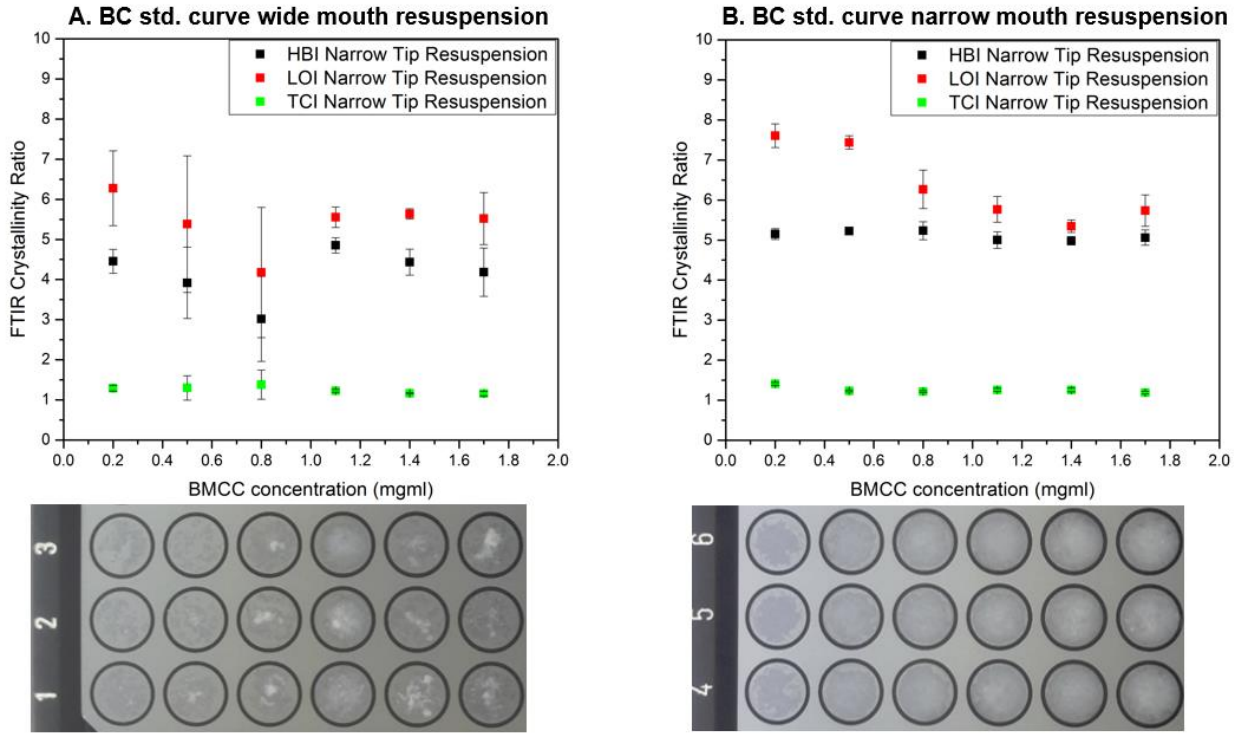


Figure 4.2. Effect of tip aperture on BC standard curve crystallinity index values and variance. A) Resuspension using wide-aperture tips. B) Resuspension using narrow-aperture tips.

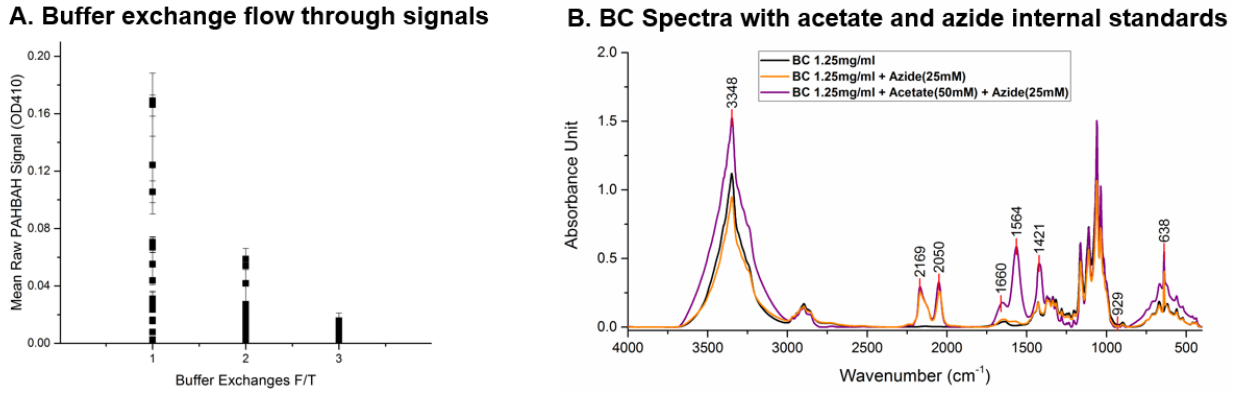


Figure 4.3. A) Decreasing PAHBAH reducing end signals of solvent after sequential washes. Buffer exchange of hydrolysis assay samples using Milli-Q water, as solvent results in low oligosaccharide in the residual substrate. B) Signals contributed by optional internal standards. Acetate contributes signal at 3350, 1660, 1560, 1421 and 929cm^{-1} most significantly. Azide contributes signal at 2169, 2050, and 638cm^{-1}.

A third improvement, washing of BC samples using Milli-Q water, is sufficient to remove reducing end signals. This reduction of soluble signals indicates nearly complete

removal of soluble oligosaccharide products from the hydrolysis sample. In Figure 4.3.A., three washing steps are sufficient to remove soluble sugars from the residual BC, which serve as a representative signal for the various hydrosylate components which produce overlapping spectra, including sugars, glycerol, acetate, and proteins. Washing is validated in the resulting FTIR spectra, where characteristic bands from acetate (1660 cm^{-1} and 1560 cm^{-1}) or protein (1700-1500 cm^{-1}) are absent. A minor but significant modification to sample preparation, the optimized two stage plate drying procedure is successful in eliminating water from the residual cellulose cake. Signals from residual water were monitored for in the characteristic signal regions and not observed. Free water will appear at 1630 cm^{-1}, which could indicate either incomplete drying or potentially trapped water within micropores/interstial spaces, and will significantly affect hydrogen bonding if present (Corgié et al., 2011).

The coefficient of variance is a critical indicator of data reliability. The CV of sample crystallinity index ratios is low, as seen above, while the CV of individual peak raw data is high owing to the variability of sample drying and resulting cake thickness. This suggests that the variance of cake thickness well to well is quite high, but the use of crystallinity ratios provides a certain amount of intra-sample normalization.

Another way to account for the variability from sample drying is to include an internal standard for normalization. The internal standard used by Corgie et al., acetate, has several disadvantages. Acetate signals are very strong in the same regions as cellulose, even at low concentrations can be seen in the 1700-1400 cm^{-1} as well as the 3800-3000 cm^{-1} range (figure 4.4.B). Acetate was volatilized at moderate temperature and vacuum conditions (70 °C and 50 mmbar), and spread from samples containing acetate to neighboring wells and plates within the same space during the drying (not shown). An alternative internal standard, sodium azide, was

investigated and deemed to be feasible in cases where the pH is controlled to avoid volatilization of hydrazoic acid during the drying process. The azide signal is present in the 2200-2000 cm^{-1} wave number region and is sufficiently separated from BC signals (Figure 4.3.B).

4.3.2. Initial cellulose hydrolysis testing.

Initial hydrolysis assays were conducted prior to developing the method improvements described above, and resulting FTIR spectra are highly similar to those presented by Corgie et al. The FTIR spectra of BC hydrosylate after digestion using Cel48A, shown in Figure 4.4.B, can be compared to the spectra of digestion by Cel6B reported by Corgie et al. in in Figure 4.4.A. In both cases very substantial changes are observed, with a decrease in peak height greater than 50% for BC hydrolyzed by either exocellulase. This decrease occurs in multiple characteristic regions, including the hydrogen bonding and cellulose fingerprint regions and the β-glucosidic bond. This large change occurs despite the low extent of digestion, ~14.5% for Cel6B and ~3% for Cel48A. The disproportionate decrease of FTIR spectra amplitude in relation to extent of digestion was observed for all hydrolysis samples processed using the methods of Corgie et al.

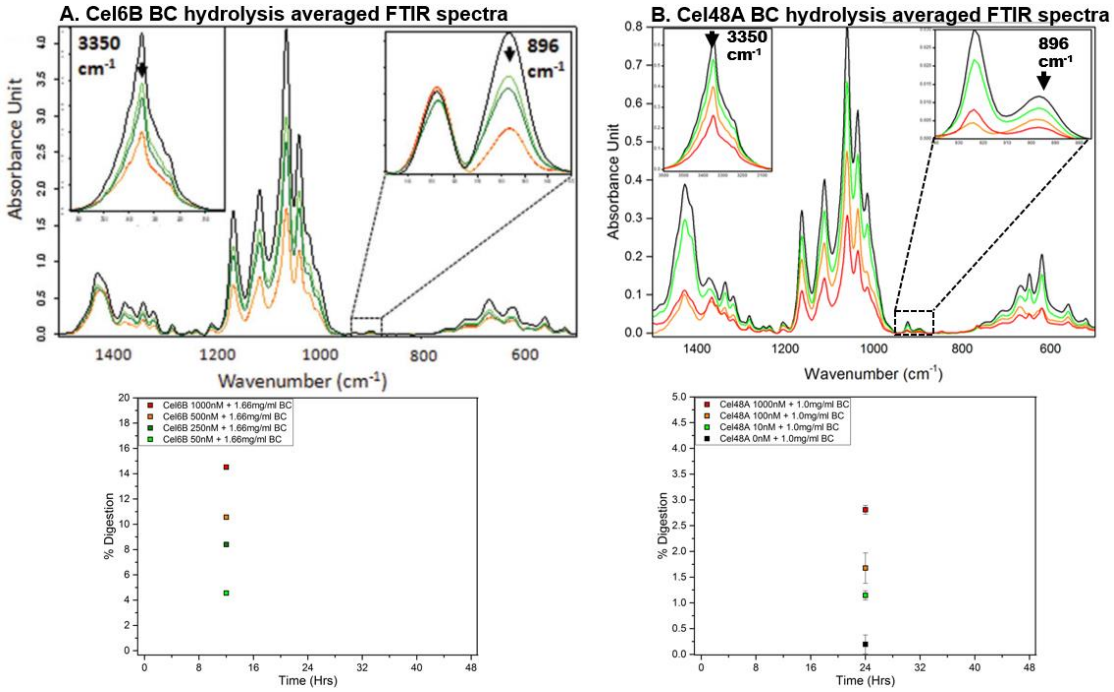


Figure 4.4. Characteristic FTIR spectra regions of BC showing similar changes to those observed by Corgie et al. A) Cel6B + BC hydrolysis at 12 hours, reproduced from Corgie et al., 2011(Corgié et al., 2011) B) Cel48A + BC hydrolysis at 24 hours, using similar sample preparation methods.

4.3.3. Revised Methods Hydrolysis FTIR Samples.

Measuring a time course of Cel48A hydrolysis showed changes to multiple crystallinity indexes which appeared to correlate with the kinetic falloff of hydrolysis, as shown in Figure 4.5, (colored points) and 4.7 (colored points).

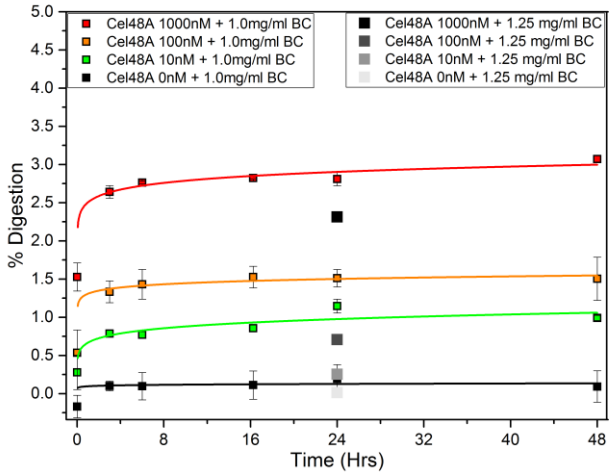


Figure 4.5. Cel48A + BC hydrolysis time course using methods from Corgie et al. (colors), and Cel48A + BC hydrolysis end point using updated methods (greyscale). Parameters for fitted curves not shown.

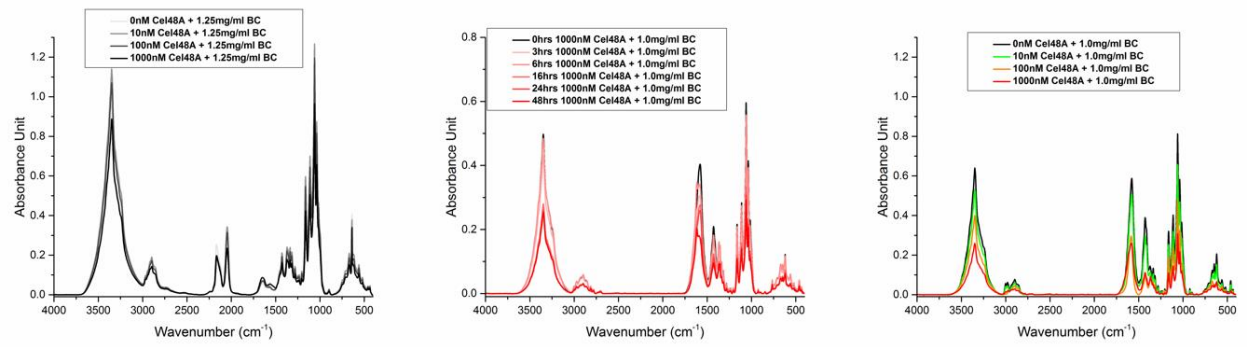


Figure 4.6. Averaged spectra for Cel48A time course, end point showing change to FTIR spectra due to hydrolysis. Greyscale, Cel48A endpoint showing consistency due to concentration normalization.

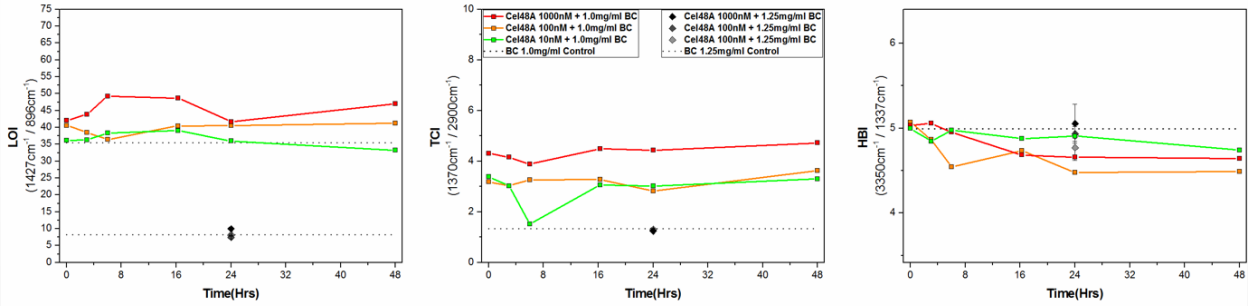


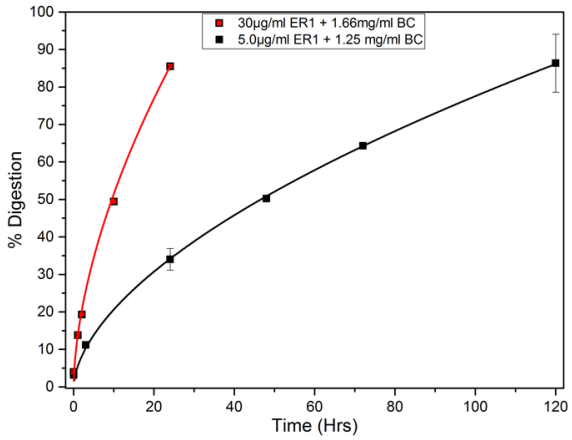
Figure 4.7. Crystallinity indexes of Cel48A time course, using methods from Corgie et al.(colors), compared with BC control, (dotted line), and Cel48A endpoint with revised methods(greyscale). Using existing methods, changes to crystallinity indexes follow trend of % digestion over time for a Cel48A hydrolysis time course.

Comparing the Cel48A hydrolysis time course samples processed using the methods of Corgie et al. with Cel48A end point samples using updated methods shows a difference in the resulting FTIR crystallinity indexes. After removal of soluble hydrolysate and normalization of BC concentration according to extent of digestion, the overlap of processed FTIR spectra indicate similar BC concentrations. With these changes, the crystallinity indexes of Cel48A hydrolysis did not change with increasing digestion.

4.3.5 ER1 time course.

To confirm that low extent of hydrolysis was not the cause of the lack of crystallinity index change after Cel48A digestion, BC samples from hydrolysis using *T. fusca* ER1 crude supernatant (a complete high activity cellulase system) were measured over a wide range of digestion.

ER1 crude hydrolysis time course samples with a very high extent of digestion, (Figure 4.8) resulted in kinetic parameters agreeing with previously reported values (Kostylev and Wilson, 2013). FTIR spectra of these samples using previous methods showed significant change to all three measures of crystallinity over time (red, Figure 4.9.A).



$f(t)=At^b$	A	b	R^2
30µg/ml ER1	13.6	0.57	0.99
5.0µg/ml ER1	5.51	0.58	0.99

Figure 4.8. Hydrolysis time course by *T. fusca* ER1 crude supernatant. Shown in black is a lower loading over longer time used for revised methods. In red, higher loading using older methods.

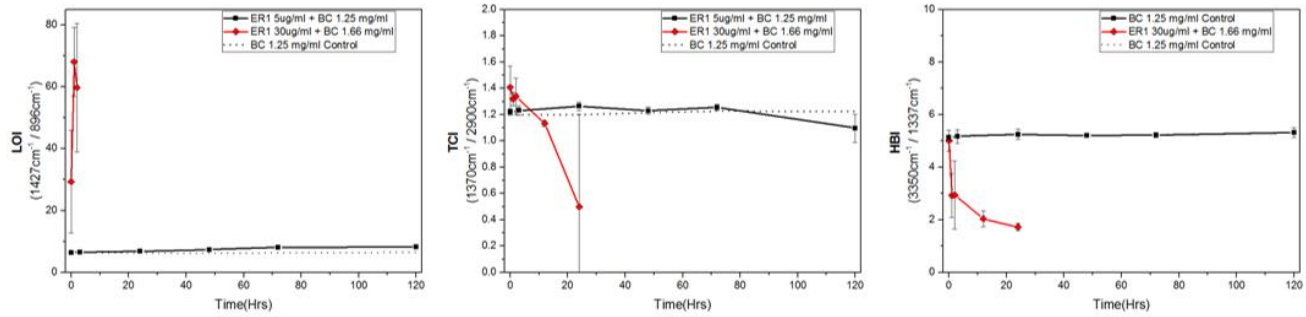


Figure 4.9. Crystallinity indexes of ER1 crude supernatant + BC hydrolysis Corgie et al. methods (red), and new methods (greyscale).

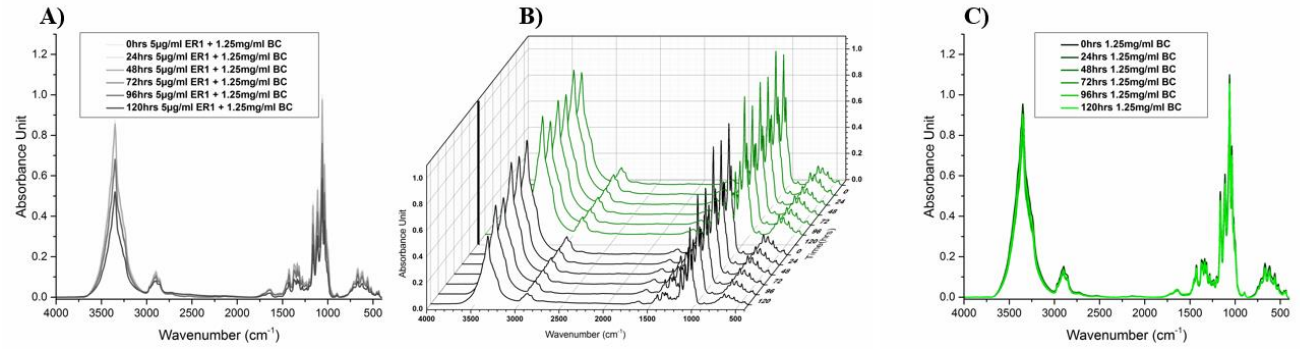


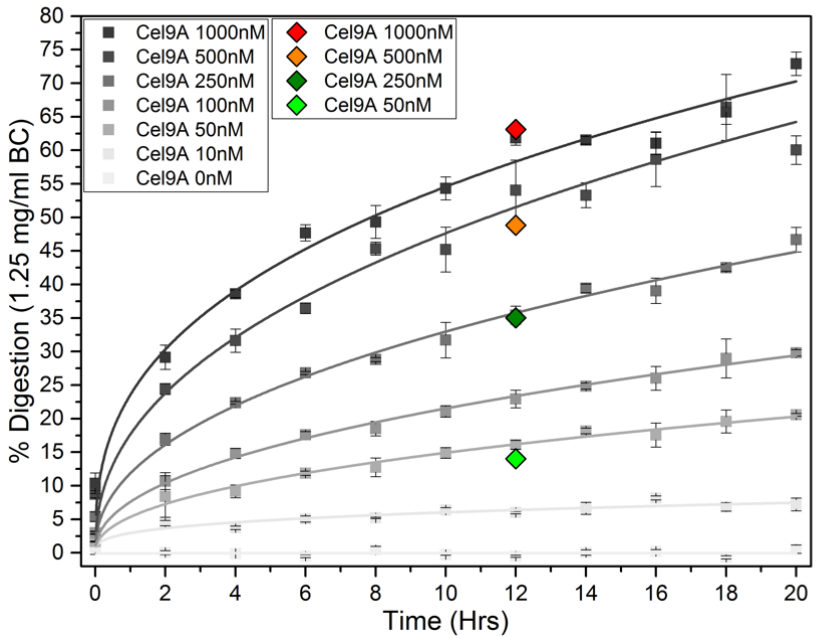
Figure 4.10. Change to FTIR spectra of ER1 crude hydrolysis time course over time. A) Resuspended residual BC after hydrolysis by ER1 crude supernatant, showing decreasing signal over time. B) Comparison of FTIR spectra showing decreasing signal. C) BC only control showing consistent FTIR spectra over time.

Despite having similar extent of digestion, FTIR spectra collected from ER1 time course residual BC using new methods show no changes to the HBI or TCI crystallinity indexes (Figure 4.9). The lateral order index (LOI) appears to negatively correlate with extent of digestion. This change was the result of misestimating % digestion prior to sample concentration normalization, shown in the FTIR spectra in Figure 4.10 and validated by sulfuric acid-phenol quantification of BC concentration (not shown). The residual BC deposited on the FTIR plate had decreasing concentration, resulting in a significant change in cellulose cake thickness and a corresponding change to LOI. This is as would be expected based on the BC standard curve results shown in Figure 4.1.B, where the LOI crystallinity index has a negative correlation with BC concentration.

4.3.6. Digestion by T. fusca Cel9A.

The most significant correlation between cellulase digestion and increasing crystallinity observed by Corgie et al. resulted from digestion by the processive endocellulase Cel9A. To reevaluate the effect of Cel9A digestion on BC crystallinity, the conditions for the Cel9A time course hydrolysis assay were chosen to match the extent of digestion observed by Corgie et al., as seen in the color-filled diamonds in Figure 4.11. A range of eight Cel9A concentrations from 10 to 1000 nM, and three BC substrate concentrations from 1.25 to 5.0 mg/ml were tested. The close agreement with prior extent of digestion results indicates that similar enzymes, substrates, and incubation conditions were employed.

A)



B)

$f(t)=At^b$	*A*	*b*	R^2
Cel9A 1000nM	23.5	0.36	0.96
Cel9A 500nM	17.6	0.43	0.95
Cel9A 250nM	11.8	0.44	0.98
Cel9A 100nM	7.6	0.45	0.98
Cel9A 50nM	5.3	0.44	0.97
Cel9A 10nM	2.9	0.31	0.77

Figure 4.11. % digestion Cel9A time course. A) Data points shown in diamond colors are those reported by Corgie et al., showing high agreement B) Values of the two-parameter nonlinear fit of Cel9A time course digestion values processed according to Kostylev and Wilson (Kostylev and Wilson, 2013).

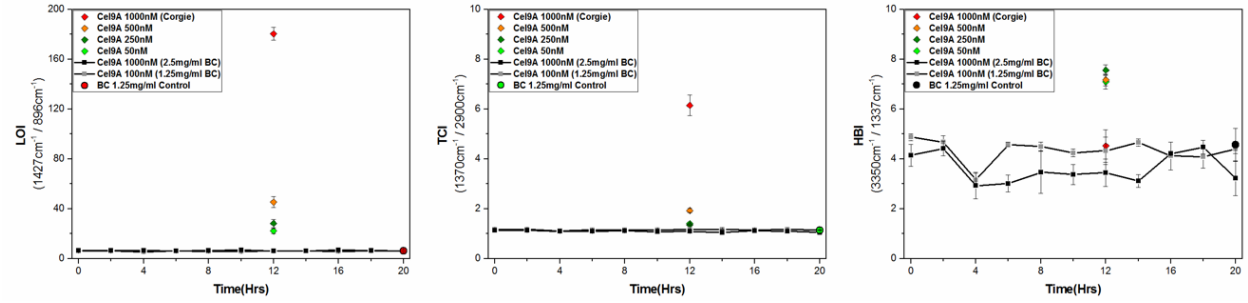


Figure 4.12. FTIR LOI, TCI, and HBI crystallinity index values of two concentrations of Cel9A plotted with respect to time for Cel9A hydrolysis. Values with diamond symbols are those reported by Corgie et al., values in large circles are those for BC only controls.

Crystallinity index changes Cel9A BC are shown in Figure 4.12. Two representative data sets are shown, with similar results of other concentration combinations omitted for clarity. In all conditions, the crystallinity indexes of the residual substrate were not significantly different compared with the BC only negative control. Samples reported by Corgie et al. showed a strong correlation between digestion and crystallinity index ratio values (colored points).

4.3.7. Oxidative cleavage by T. fusca AA10B.

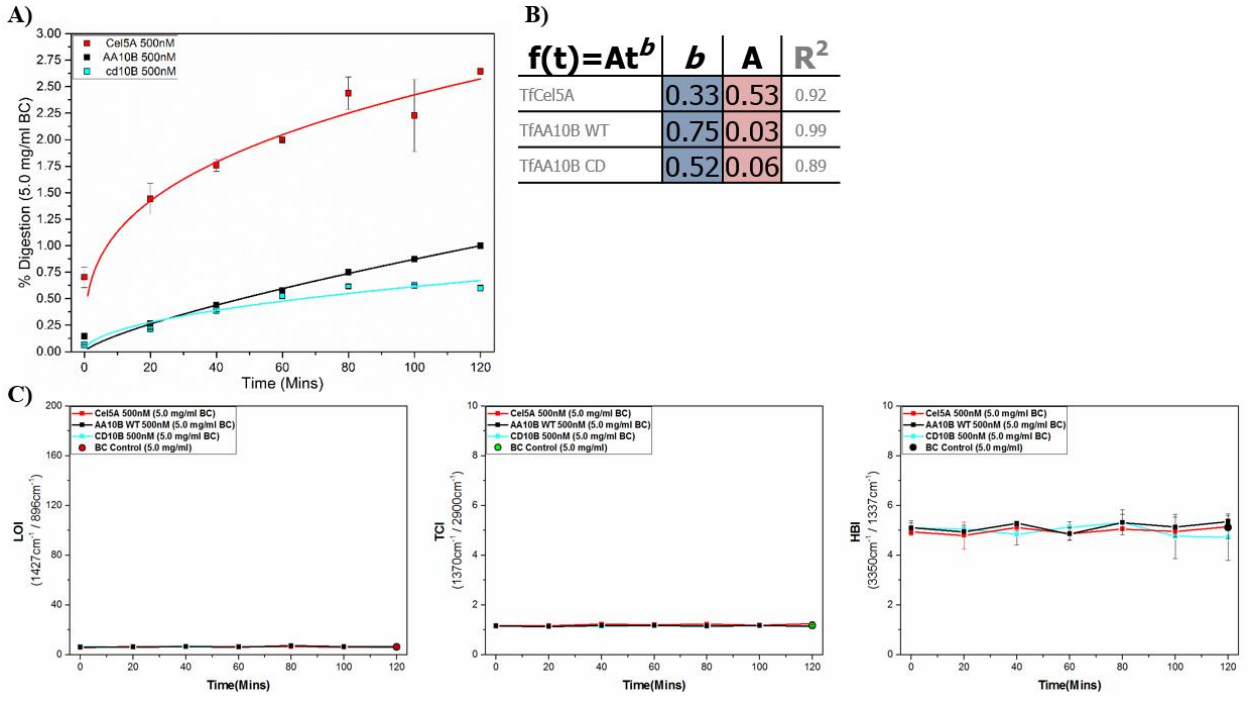


Figure 4.13. A) Extent of BC (5mg/ml) digestion by 0.5uM TfCel5A (red), TfAA10B WT (black), and TfAA10B CD only (light blue). B) Kinetic parameters of the three enzymes showing differences in b parameter. C) FTIR LOI, TCI, and HBI crystallinity index values of BC after digestion. Points represent the average of three values, with error bars corresponding to standard deviation.

The effect of lytic polysaccharide oxidative cleavage (LPMO) oxidative cleavage on bacterial cellulose structure was compared with hydrolytic cleavage by the classical endocellulase Cel5A. Two forms of the *T. fusca* type I LPMO AA10B were tested, full length wild type (WT) and the LPMO catalytic domain (CD) form. LPMOs were assayed in the presence of a suitable reducing agent (2mM reduced glutathione) as described in chapter 2. As seen in Figure 4.13, digestion of BC by either Cel5A or AA10B resulted in no significant change to the three crystallinity indexes as compared to the substrate only negative control.

4.4. FTIR Discussion

4.4.1. Method improvements.

Cellulose samples with varying extents of digestion can only be compared meaningfully using the HTS-FTIR technique when many variables are carefully controlled. Changes to the HTS-FTIR sample preparation methods described by Corgie et al. were necessary in order to obtain clean raw spectra and reliable measures of cellulose crystallinity. The sample preparation methods were improved to remove extraneous overlapping signals, provide accurate extent of digestion, and correct errors caused by varying concentrations. Though the residual cellulose was handled more extensively, these improved preparation methods did not affect the cellulose structure or alter the resulting crystallinity index values.

To more accurately measure cellulose crystallinity, samples were washed to remove unrelated sources of signal. Unprocessed hydrolysate contains several compounds at sufficient concentration to contribute to cellulose FTIR spectra: proteins (cellulases), cryoprotectants (glycerol), buffers (acetate), and hydrolysis products (free mono/oligosaccharides). These compounds share chemical bond compositions with cellulose, and thus have peaks which may overlap with those from cellulose vibrational modes. The polar compounds will interact non-covalently with cellulose when dried into the cellulose matrix, inflating the hydrogen bonding region centered at 3350 cm^{-1}. As seen in Figure 4.3, buffer exchange using Milli-Q water was effective at removing soluble hydrolysate components to produce a clean spectrum of the residual cellulose.

Acetate buffer signal at 1650 cm^{-1} was used as an internal standard in previous work, but this was removed due to signal overlap and poor reliability. The rapid drying of samples within the vacuum oven chamber volatilizes acetic acid, which transfers between wells and results in

peak height that is not representative of cellulose concentration. Thus, acetate was removed during buffer exchange and the internal standard normalization step during data processing was omitted. Sodium azide was found to be more stable during drying, and was tested for use as an internal standard. The azide signal at 2069cm^{-1} is a linear v3 mode signal, which is uncommon in carbohydrates and provides a unique and useful peak (Gray and Waddington, 1957). However, the peak shape of azide at 2170 cm^{-1} was observed to change with varying concentration (not shown), and requires further validation for use as a standard in future work.

HTS-FTIR sample processing requires a rapid % digestion assay to avoid the erroneous LOI measurement caused by differing sample concentrations described above. The PAHBAH assay provided a sensitive reducing end detection method that can accurately measure the extent of cellulose digestion, but cannot differentiate between different reducing sugars or determine the product distribution of oligosaccharide mixtures. PAHBAH requires secondary hydrolysis to homogenize the hydrolysate product distribution and obtain a true measure of the number of glucose units in solution. This secondary hydrolysis and PAHBAH method worked as intended, providing high quality digestion measurements in close agreement with values measured using HPLC.

The lateral order index (LOI) is sensitive to sample concentration in a way that the other two indexes are not, as this ratio is based on a vibrational mode with low absorbance efficiency (β-glycosidic bond at 895 cm^{-1}). Cellulose cakes of different thickness can't be compared, as their different path lengths disproportionately affect small peaks near the lower limit of the detector linear range. Crystallinity index ratios that rely on these small peaks will change with respect to concentration, as one of values will not obey Beer's law. This effect can be seen in the representative standard curve in Figure 4.1, where the LOI index negatively correlates with BC

concentration despite constant chemical and structural composition. This disadvantage of the LOI index appears in the ER1 crude digestion time course results in Figure 4.10. Insufficient β-glucosidase (BglC) concentration resulted in misestimating of % digestion by the PAHBAH method, and the incorrect sample concentration produced LOI values correlated with digestion. This was corrected in all other samples by accurately normalizing the final BC concentration of all samples to within the detector linear range.

To process raw FTIR spectra, two approaches are commonly used: second derivative peak height and peak integration after deconvolution (Faix, 1992). Second derivative peak height is the differentiation of the raw spectrum to provide better resolution of hidden peaks, an approach that has been used effectively to analyze cellulose FTIR spectra (Michell, 1990). Second derivative peak height is superior to integration of area using peak deconvolution when definitive assignment of all component peaks isn't possible, such as for complex cellulose samples with ambiguous overlapping bands. The second derivative approach has several disadvantages that are prevented by maintaining consistent detection parameters: increased signal noise, small values for broader peaks, and derivative height values influenced by spectral resolution (Saarakkala et al., 2010). For quantification of carbohydrate FTIR spectra, second derivative peak height is currently the most appropriate compromise between noise and peak resolution.

Cellulose samples were carefully handled to limit exposure to conditions that could affect crystallinity. High intensity sonication was used by Corgie et al. to improve cellulose homogeneity and digestibility, but it was not used in this work to avoid inadvertent disruption of non-covalent interactions. In testing, sonication did not have a measurable effect on digestibility (Kostylev, unpublished) or on FTIR crystallinity indexes (not shown). Cellulose samples were

never frozen to avoid formation of ice crystals within cellulose that could disrupt inter- or intra-molecular bonds (cellulose can contain 10%-15% w/v trapped water) (Thakur, 2013), but samples stored at -20 °C showed no difference in structure as measured using FTIR (not shown). The effect of heating on cellulose crystallinity was investigated, as heating to 100 °C for 5 minutes is necessary to stop enzymatic hydrolysis via denaturation. Heating BC under these conditions did not alter crystallinity indexes (not shown). Exposure to sonication or extreme temperature may have a more significant impact on more complex biomass substrates, due to the presence of more sensitive non-covalent cellulose-hemicellulose interactions.

4.4.2. Cellulose Hydrolysis and FTIR Crystallinity Ratios.

The use of peak ratios is well established in the FTIR literature, and many useful ratios have been proposed and validated for the investigation of different insoluble substrate properties. The chemical bond vibrational modes which make up these ratios give distinct measures of substrate properties. In the case of chemically homogenous but structurally diverse polysaccharide substrates, such as the model bacterial cellulose, these crystallinity index ratios provide a summary of the physical state and structural proximity of the polysaccharide chemical bonds. Besides crystallinity, additional structural information can be obtained from FTIR spectra including proportion of cellulose I α and β phases within the crystalline lattice. However, consistent deconvolution of overlapping peaks in the 800 cm^{-1} to 400 cm^{-1} region prevented accurate calculation of these values in this work.

Most exocellulases (with the exception of the fungal GH7 family) reach a slowing of kinetic rate at a low extent of digestion, an indicating sign of substrate change. Thus, exocellulases were an intuitive starting point for investigating recalcitrance and crystallinity change. Exocellulases act by binding a free cellulose chain end, directing the free chain into the

active site, then cleaving this single chain multiple times. The two exocellulases in the *T. fusca* system both act processively, but they differ both in the direction of cellulose chain binding (Cel6B attacking the non-reducing and Cel48A the reducing end), and in the extent of digestion achieved on crystalline cellulose (Cel6B slowing at ~10% and Cel48A at ~1%). In the context of substrate crystallinity, exocellulases would be expected to digest a significant portion of surface chains with exposed termini, resulting in cellulose with depleted paracrystalline material and enriched crystalline regions. Past results of Cel6B activity showed crystallinity index changes in line with these expectations: BC digested by Cel6B had significantly increased LOI and TCI, indicating increasing substrate crystallinity and decreased HBI, corresponding to reduced chain mobility and bond distances.

This work focused on Cel48A, the reducing end directed *T. fusca* exocellulase, based on the expectation that the effects observed with Cel6B would be more pronounced. The more rapid drop off of Cel48A activity suggests a more rapid enrichment of substrate crystallinity. Using previous processing methods, similar trends were observed in the raw spectra (Figure 4.4) and crystallinity index value changes (colored squares in Figure 4.7). These Cel48A time course samples showed similar crystallinity changes to Cel6B despite reaching much lower % digestion, as the LOI and TCI appeared to increase and HBI decrease over the course of digestion. In previous work, Cel6B did not significantly modify the HBI of BC at 12 hours, but Cel48A appeared to modify the HBI value at all three concentrations by 16 hours. The LOI and TCI values stabilized rapidly, possibly due to Cel48A rapidly reaching kinetic drop off on highly crystalline material. Though initial results supported previous observations and the expectations of a crystallinity enrichment model, digestion samples processed with updated methods did not show the same trend (grey squares in Figure 4.7). Comparing the extent of digestion with the

resulting crystallinity index changes, Cel48A with similar extent of digestion but with improved processing methods showed no changes with respect to digestion at 24 hours. Similarly, Cel6B at multiple concentrations processed using improved methods resulted in no change to CI value (not shown). This lack of correlation challenges the initial observation of exocellulase hydrolysis inducing BC crystallinity changes.

To test the possibility that the lack of changes to BC crystallinity after digestion by Cel48A was due to the low extent of digestion, a complete mixture of cellulases was used to degrade the substrate more completely. The *T. fusca* ER1 supernatant is a complex cellulase mixture that has no kinetic drop off when digesting pure cellulose substrates, due to extensive synergistic interactions. It is unknown whether ER1 crude supernatant digests all cellulose regions simultaneously, or completely digests less crystalline regions prior to degrading more recalcitrant regions. The lack of kinetic drop off suggests the former, which would result in either no crystallinity change or an initial increase in crystallinity followed by stabilization as an equilibrium between digestible and recalcitrant cellulose is established.

Digestion using a high loading of ER1 crude supernatant rapidly reached a high extent of digestion, approximately 85% within 24 hours (red squares, Figure 4.9). When these samples were prepared using the previous methods, significant changes to crystallinity indexes were observed. A large increase to LOI and a rapid drop of both TCI and HBI suggested cellulose structure changes that would be advantageous for an efficient synergistic cellulase system. The cellulose structural crystallinity represented by TCI and HBI was being reduced during digestion by ER1 crude supernatant, despite the lateral order of the remaining crystalline fiber increasing as indicated by the LOI.

However, when ER1 crude supernatant hydrolysis time course samples at similar extent

of digestion were processed with updated methods, no significant changes were observed in the TCI and HBI crystallinity indexes. The only change to crystallinity correlating with increasing digestion was a change to LOI caused by incorrect sample concentration, as discussed previously. The lack of change to residual cellulose crystallinity suggests a lack of significant change to the residual BC structure, supporting the hypothesis of simultaneous activity on each region of BC.

The processive endocellulase Cel9A provides a final test of the correlation between digestion and crystallinity for individual hydrolytic cellulases. Cel9A is an important component of the *T. fusca* cellulase system, with high activity on both amorphous and crystalline cellulose regions. The Cel9A digestion mechanism is an initial endocellulolytic cleavage of a crystalline chain within the cellulose lattice, followed by processive cleavage. Cel9A produces the highest extent of digestion of any single *T. fusca* cellulase on crystalline cellulose, and showed the most significant change to all three crystallinity ratios in previous investigations.

In the previous work by Corgie et al., Cel9A digestion of BC resulted in crystallinity index changes far surpassing the highly active endocellulase and exocellulase synergistic mixture of Cel5A and Cel6B. The increase of both LOI and TCI, with a decrease to HBI, indicated a significant increase in crystalline cellulose. This was attributed to preferential digestion of exposed amorphous paracrystalline regions prior to engaging highly crystalline substrate. The ability of Cel9A to substantially increase the TCI value, considered an indicator of a 'crystallinity limit' for the individual Cel5A and Cel6B enzymes, was in agreement with Cel9A tolerance to substrate crystallinity over other cellulases.

Cel9A time course samples were prepared with an extent of digestion matching previously reported digestion values by Corgie et al., with values obtained using the PAHBAH

method validated using HPLC quantification (grey squares, Figure 4.11). When the residual cellulose was processed using updated methods, the crystallinity indexes did not change significantly during the course of digestion (grey squares, Figure 4.12). This may be explained as Cel9A not exhibiting a preference towards cellulose regions of different crystallinity, or as the cellulose crystallinity equilibrating by some other mechanism. Cel9A activity is a significant portion of the total ER1 crude activity on crystalline cellulose, and it matches expectations that both showed a similar lack of crystallinity ratio change using updated FTIR processing methods.

Lytic polysaccharide monooxygenases (LPMOs) perform oxidative cleavage to degrade crystalline polysaccharides, but the physical changes to the substrate that enable synergism with hydrolytic cellulases are not understood. The amorphogenesis model of LPMO activity was based upon a physical disruption of cellulose crystallinity resulting in a more digestible material (Arantes and Saddler, 2010). Using the same residual cellulose preparation methods as with hydrolytic cellulases, no change to cellulose crystallinity was induced by LPMO attack. This observation does not support the amorphogenesis model of LPMO attack, where oxidative cleavages would introduce substrate disruption to produce a physically altered substrate.

In comparison, the classical endocellulase Cel5A predominately attacks more amorphous and solvent exposed cellulose chains. While Cel5A produced a significant change to crystallinity in previous results, here there was no change to cellulose structure observed. The similar absence of crystallinity index change by either cellulose cleavage mechanism suggests cellulose crystallinity unchanged by either enzyme.

The extent of LPMO digestion as measured by soluble products was low, approximately 1% digestion, due to their random binding and intra-chain attack. It is likely that due to the substrate concentration and time period chosen, many oxidative cleavage products remained on

the substrate surface that were not represented by the detection of soluble products. It is unlikely that higher extent of digestion would result in significant change to crystallinity. As synergistic components of native or industrial enzyme mixtures, LPMO oxidative products are typically low in concentration (often less than 2% of total products) despite producing a large synergistic effect (Cannella et al., 2012). In order to effectively induce substrate amorphogenesis, LPMOs would cause disruption before the substrate becomes highly digested.

4.4.3. Interpretation of raw signals

The results presented above indicate that cellulase digestion does not cause significant changes to cellulose crystallinity indexes when the samples are buffer exchanged, concentration normalized, gradually dried, and measured without normalization to an internal standard. However, large and reproducible changes to both the raw spectra and the resulting crystallinity indexes are observed when the previous processing methods are used, and these changes should be explained. The most likely causes of the changes in these samples are the conversion of ordered cellulose to soluble mono- and oligosaccharide products, and the corresponding decreased path length of the dried cellulose sample. Soluble products would be expected to significantly affect the resulting crystallinity indexes due to the changed structural interaction of these products with the bulk cellulose. As described in the previous chapter, all three crystallinity indexes depend on the relationship between peaks representing bond mobility or non-covalent interactions with those that are unaffected by lattice structure. Free oligosaccharides pack more randomly between the ordered microfibrils during drying, resulting in reduced non-covalent interactions while the signal from pyranose ring covalent bonds remained unchanged. The conversion of ordered cellulose to soluble products could produce the diminished spectra and changes to crystallinity indexes observed, despite the crystallinity of the residual cellulose

remaining unchanged.

The action of cellulases are expected to reduce the samples path length, as the soluble oligosaccharide products are much smaller and contribute little to the packing of much larger cellulose microfibers. However, the signal reduction observed in samples with low extent of digestion, such as the Cel48A time course shown in Figure 4.7. is disproportionate to the amount of soluble products removed. This may be explained as digestion inducing more compact packing of the cellulose microfibers, resulting in a disproportionately thin cellulose cake. It is possible that the less ordered and bulky surface chains were removed in these samples, resulting in a shorter path length without significant removal of mass. This change to packing thickness may be entirely uncorrelated with crystallinity if reorganization of surface chains results in the same degree of disorder on the individual microfiber scale (a likely scenario since highly disordered chains would have minimal inter-chain interactions). This type of 'microfiber pruning' would not significantly affect signals from the minimal contact between microfibers, but would lead to a closer microfiber packing and a reduced sample path length. A shorter path length is expected to reduce the amplitude of the raw spectra, but should not affect crystallinity indexes when the cellulose cake thickness is sufficiently within the detector linear range. In contrast to soluble products, shorter path length would result in significant changes to the lateral order index (LOI), as seen in Figure 4.1 and discussed previously.

Another possibility to explain significant crystallinity index changes for samples with high extent of digestion is the reduced coverage of the sample well by the residual cellulose. Cellulose samples that fail to cover the well surface would have more stray light which does not pass through thick cellulose. Stray light causes the detector to receive more background light, leading to different effects on bands with different absorptivity (Chalmers, 2002). Stray light also

artificially limits the working range of the detector relative to strongly absorbing bands, since no transmittance can be below this minimum value (Griffiths and De Haseth, 2007). This non-interacting stray light is a disadvantage for transmission mode FTIR of thin film polymer samples such as pure cellulose cakes, but normalization of sample concentrations will remove this source of error. Normalization to an internal standard may improve comparison between homogenous samples of differing thickness, but would not correct those errors related to the detector (insufficient linear range in the case of inaccurate LOI, and excess background in the case of stray light). When soluble products are removed and the concentration of cellulose is normalized, the resulting cellulose samples are directly comparable. These improved methods eliminate both the reduced spectral height and changes to crystallinity indexes resulting from cellulase digestion.

4.4.3. Comparison with previous work

Previous exploration of the effect of cellulase activity on cellulose crystallinity largely agree with the results presented here. While no consensus has formed, digestion of pure cellulose substrates by individual or synergistic cellulase systems has typically resulted in undetectable changes to substrate crystallinity. Boisset et al. did not observe changes to BC crystallinity by cellulsomes using XRD, or Iα/Iβ ratios using FTIR (Boisset et al., 1999). Mansfield and Meder (Mansfield and Meder, 2003) showed changes to Sigmacell® crystallinity by two *Cellulomonas fimi* cellulases, but the small effect may be explained by other factors. The 5% increase of crystallinity by *Cf*Cel48A may be explained by the use of a nonstandard crystallinity index, the ratio of 1282/1202 cm^{-1}. The residual cellulose was washed, but the concentration was not normalized. Additionally, spectra were normalized to an unreported peak, and integration after peak deconvolution was employed for quantification. Consistent deconvolution of changing

cellulose samples is difficult, especially when hidden peaks dynamically change during hydrolysis. The combination of these factors may have caused the small change to cellulose crystallinity observed.

In a different study, multiple *T. fusca* cellulases had little effect on the crystallinity of BC as measured by X-ray diffraction. An increase in crystallinity of approximately 2% was caused by Cel9A digestion, a larger effect than in the results presented here. It is possible that this change was within measurement error due to the low throughput of XRD technique. In contrast, the crystallinity of two forms of mercerized cellulose (the cellulose II allomorph) increased significantly after digestion. Based on the similar initial crystallinity of the two allomorphs and the much higher digestion of the cellulose-II form, these results support the interpretation that crystallinity alone is not likely to be directly correlated with cellulose recalcitrance (Chen et al., 2007).

The demonstration by Hall et al. of the influence of crystallinity on initial rate and extent of digestion support this conclusion that changing cellulase kinetics are not due to increasing substrate crystallinity (Hall et al., 2010). Avicel® crystallinity was negatively correlated with initial hydrolysis rate, but crystallinity did not change over digestion or influence the extent of digestion. This result was similar for digestion by an individual *Tr*Cel7A exocellulase and a commercial *T. reesei* cellulase cocktail. These results agree with the results presented here using individual cellulases and a complex *T. fusca* mixture, suggesting that the amorphous regions are not preferentially digested to produce highly crystalline cellulose.

Despite initial results using the sample preparation methods of Corgie et al., cellulose crystallinity indexes do not change with cellulase digestion when improved methods are used. Cellulase hydrolysis does not induce changes to bacterial cellulose crystallinity. This lack of

correlation between crystallinity and recalcitrance is in agreement with the results of other studies, and thus an alternate model of cellulose recalcitrance is needed in order to explain the decline of cellulase rates. Initial substrate crystallinity plays a role in hydrolysis, but appears to be limited to determining the initial rate of reaction. In contrast to the model of cellulose recalcitrance directly correlated with increased cellulose crystallinity, the surface erosion model described in the previous chapter predicts the results observed here. The surface erosion model proposes that the kinetic drop off and extent of digestion is determined by the accumulation of surface blockages which prevent productive cellulase engagement.

This is consistent with observations of constant crystallinity during digestion if reordering occurs as surface chains are removed. Rather than a static lattice which becomes more crystalline, the chains newly exposed to the solvent interface must adopt the same amount of disorder as those that were removed. This equilibrium of crystallinity is supported by the results observed here, where cleavage of crystalline cellulose to any extent does not change the crystallinity of the residual material. The constant crystallinity, despite increasing recalcitrance, indicates that the crystallinity is less critical than the complexity of the cellulose surface, and in the context of the surface erosion model this provides an explanation of the kinetic fall off observed.

4.5. Conclusion

4.5.1. Project objectives.

The objective of this project was to gain an insight into the substrate changes causing nonlinear cellulase kinetics, and the role of lytic polysaccharide monooxygenase (LPMO) activity to enable degradation of crystalline cellulose. The HTS-FTIR platform was a clear choice to test the correlation between cellulase digestion and cellulose crystallinity based on a

foundation of prior research and compelling results. The results here show an absence of increasing cellulose structural crystallinity during cellulose digestion, supporting an alternative model of both cellulose recalcitrance and the role of cellulose-active LPMOs in synergistic mixtures. With the sample preparation method improvements described here, HTS-FTIR may be valuable for future work in this space as the models of industrially relevant biomass substrates become more mature. The lack of crystallinity increases caused by cellulases does not provide a model of cellulose recalcitrance, the absence informs a more accurate cellulase-cellulose model for designing a more efficient biomass deconstruction system.

The cellulose crystallinity index changes observed in past work were likely caused by the additional components in the hydrolysate samples introducing confounding signals. These components, glycerol, proteins, and monosaccharides, overlap with some regions of the cellulose spectrum essential for crystallinity ratios, while leaving other critical peaks unmodified. The release of different amounts of mono- and oligosaccharides into solution by the different extents of digestion, compounded by the change in path length of the residual cellulose, would be expected to change the spectrum in the manner that was observed. Introducing the removal of non-cellulose components prevented these sources of signal from affecting the FTIR spectrum and ratios calculated based on it.

4.5.2. FTIR fundamental constraints.

FTIR has great potential for investigating structural changes to complex biomacromolecules, but there are fundamental constraints of the HTS-FTIR approach that must be understood and accommodated. For samples involving hydrolysis or soluble compounds, a thorough buffer exchange is essential to remove contaminating signal. The detector linear range is a critical constraint for determining sample loading, as data quality is affected by both

oversaturation of peaks of interest and nonlinear signal from small peaks. Sample heterogeneity and variable drying is a limitation of insoluble cellulose or biomass when used for FTIR. Initial testing suggested that an internal control such as sodium azide (with caution regarding potential toxicity) would allow for accurate normalization as it dries into the insoluble sample matrix.

4.5.3. FTIR future applications.

As described above, alternative techniques can be used to measure cellulose crystallinity changes. The NMR and XRD approaches are considered more robust for investigating cellulose crystallinity, but are limited by their much lower throughputs compared with HTS-FTIR (Mansfield and Meder, 2003). These techniques have produced similar results addressing the role of crystallinity in cellulase kinetics, and are in agreement with the results presented here.

While FTIR does not provide the basis of pure cellulose recalcitrance, it may serve as an essential tool for understanding biomass deconstruction. Industrially relevant biomass contains complex interactions between cellulose and the numerous structural plant polymers including hemicellulose, lignin, and pectin. The structure of these interactions is essential to biomass recalcitrance, and changes significantly due to pretreatment and enzymatic hydrolysis. These interactions within biomass are often non-covalent, and as such are more readily detectable by FTIR compared with other techniques. While the complexity of biomass FTIR peak assignment presents a major challenge, progress is being made to develop accurate models, such as the use of multivariate analysis and Principle Component Analysis, to characterize changes in biomass caused by pretreatment (Sills and Gossett, 2012).

4.5.4. Project conclusion.

The results presented here do not support the hypothesis that nonlinear cellulase kinetics on recalcitrant cellulose is due to increased cellulose crystallinity, as measured by the three most

common FTIR crystallinity indexes. The substrate recalcitrance which inhibits cellulase activity, and which LPMOs help to overcome, appears to be a change to cellulose structure that does not involve crystallinity. These combined results suggest that during digestion, the removal of cellulose chains does not affect the substrate crystallinity. Chains from the more highly ordered crystalline microfiber core become surface chains as digestion proceeds. As they become exposed, these ordered chains may become more disordered through contact with solvent, resulting in continual equilibration of crystallinity to the starting value as digestion progresses. According to a surface erosion model, increasing cellulose recalcitrance is not caused by increasingly ordered chains, but through the accumulation of surface disorder with the same proportion of crystallinity as the starting material.

The surface erosion model is the mechanistic explanation that is most in line with these results, and attributes increasing surface complexity as the basis of increasing cellulose recalcitrance. The results presented here successfully reproduced prior data using similar enzymes and equipment, but methodological changes produce data that do not support the cellulose crystallinity model of recalcitrance. Current results alter the interpretation of previously presented data, and provide a valuable direction for design of future recalcitrance investigation. The conclusions presented are in agreement with the current understanding of cellulose recalcitrance within plant biomass, where recalcitrance is correlated with increasing cellulose complexity rather than cellulose crystallinity (Hu et al., 2015). Moving toward the goal of biomass deconstruction for fuel and chemicals will require a more nuanced understanding of the subtle changes occurring to the structure of cellulose, and how these changes influence cellulase activity within the context of real biomass substrates.

REFERENCES

Arantes V, Saddler JN. 2010. Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis. *Biotechnol. Biofuels* **3**:2–11.

Boisset C, Chanzy H, Henrissat B, Lamed R, Shoham Y, Bayer EA. 1999. thermocellum cellulosome : structural and morphological aspects **83**5:829–835.

Cannella D, Hsieh CWC, Felby C, Jørgensen H. 2012. Production and effect of aldonic acids during enzymatic hydrolysis of lignocellulose at high dry matter content. *Biotechnol. Biofuels* **5**:26.

Cao Y, Tan H. 2002. Effects of cellulase on the modification of cellulose. *Carbohydr. Res.* **337**:1291–1296.

Chalmers JM. 2002. Mid-infrared Spectroscopy: Anomalies, Artifacts and Common Errors. *Handb. Vib. Spectrosc.*:2327–2347.

Chen Y, Stipanovic AJ, Winter WT, Wilson DB, Kim Y-J. 2007. Effect of digestion by pure cellulases on crystallinity and average chain length for bacterial and microcrystalline celluloses. *Cellulose* **14**:283–293.

Corgié SC, Smith HM, Walker LP. 2011. Enzymatic transformations of cellulose assessed by quantitative high-throughput fourier transform infrared spectroscopy (QHT-FTIR). *Biotechnol. Bioeng.* **108**:1509–20.

Faix O. 1992. 4.1 Fourier Transform Infrared Spectroscopy. In: Lin, SY, Dence, CW, editors. *Methods Lignin Chem.*

Gray P, Waddington TC. 1957. Fundamental vibration frequencies and force constants in the azide ion. *Trans. Faraday Soc.* **53**:901.

Griffiths PR, De Haseth JA. 2007. Fourier Transform Infrared Spectrometry. Wiley. Chemical Analysis: A Series of Monographs on Analytical Chemistry and Its Applications.

Hall M, Bansal P, Lee JH, Realff MJ, Bommarius AS. 2010. Cellulose crystallinity - A key predictor of the enzymatic hydrolysis rate. *FEBS J.* **277**:1571–1582.

Hu J, Gourlay K, Arantes V, Van Dyk JS, Pribowo A, Saddler JN. 2015. The Accessible Cellulose Surface Influences Cellulase Synergism during the Hydrolysis of Lignocellulosic Substrates. *ChemSusChem* **8**:901–907.

Irwin D, Leathers TD, Greene R V, Wilson DB. 2003. Corn fiber hydrolysis by Thermobifida fusca extracellular enzymes. *Appl. Microbiol. Biotechnol.* **61**:352–8.

King BC, Donnelly MK, Bergstrom GC, Walker LP, Gibson DM. 2009. An optimized microplate assay system for quantitative evaluation of plant cell wall-degrading enzyme activity of fungal culture extracts. *Biotechnol. Bioeng.* **102**:1033–44.

Kostylev M, Alahuhta M, Chen M, Brunecky R, Himmel ME, Lunin V V., Brady J, Wilson DB. 2014. Cel48A from Thermobifida fusca: Structure and site directed mutagenesis of key residues. *Biotechnol. Bioeng.* **111**:664–673.

Kostylev M, Wilson D. 2013. Two-parameter kinetic model based on a time-dependent activity coefficient accurately describes enzymatic cellulose digestion. *Biochemistry* **52**:5656–5664.

Kostylev M, Wilson DB. 2011. Determination of the Catalytic Base in Family 48 Glycosyl Hydrolases. *Appl. Environ. Microbiol.* **77**:6274–6.

Lever M. 1977. Carbohydrate determination with 4-hydroxybenzoic acid hydrazide (PAHBAH): effect of bismuth on the reaction. *Anal. Biochem.* **81**:21–7.

Li Y, Irwin DC, Wilson DB. 2010. Increased crystalline cellulose activity via combinations of amino acid changes in the family 9 catalytic domain and family 3c cellulose binding module of thermobifida fusca ce19a. *Appl. Environ. Microbiol.* **76**:2582–2588.

Lionetto F, Del Sole R, Cannoletta D, Vasapollo G, Maffezzoli A. 2012. Monitoring Wood Degradation during Weathering by Cellulose Crystallinity. *Materials (Basel).* **5**:1910–1922.

Mansfield S, Meder R. 2003. Cellulose hydrolysis–the role of monocomponent cellulases in crystalline cellulose degradation. *Cellulose* **10**:159–170.

Michell A. 1990. Second-derivative FT-IR spectra of native celluloses. *Carbohydr. Res.* **197**:53–60.

Olsen SN, Borch K, Cruys-Bagger N, Westh P. 2014. The Role of Product Inhibition as a Yield-

Determining Factor in Enzymatic High-Solid Hydrolysis of Pretreated Corn Stover. *Appl. Biochem. Biotechnol.* **174**:146–155.

Saarakkala S, Rieppo L, Rieppo J, Jurvelin JS. 2010. Fourier Transform Infrared (FTIR) Microspectroscopy of Immature , Mature and Degenerated Articular Cartilage. *Education*:403–414.

Sills DL, Gossett JM. 2012. Using FTIR to predict saccharification from enzymatic hydrolysis of alkali-pretreated biomasses. *Biotechnol. Bioeng.* **109**:353–362.

Thakur VK. 2013. Green Composites from Natural Resources. Ed. V K Thakur. Taylor & Francis 419 p.

Väljamäe P, Sild V, Pettersson G, Johansson G. 1998. The initial kinetics of hydrolysis by cellobiohydrolases I and II is consistent with a cellulose surface-erosion model. *Eur. J. Biochem.* **253**:469–475.

Zhang Y, Cui J, Lynd L, Kuang L. 2006. A transition from cellulose swelling to cellulose dissolution by o-phosphoric acid: evidence from enzymatic hydrolysis and supramolecular structure. *Biomacromolecules* **7**:644–648.

Chapter 5. Conclusion

5.1 Summary of research

This thesis describes the exploration of two lytic polysaccharide monooxygenases (LPMOs), auxiliary activity enzymes from the *Thermobifida fusca* secretome, in order to better understand the role of this diverse and industrially relevant enzyme family. LPMOs have become essential to the improvement of industrial cellulolytic enzyme cocktails, the key component of processes that produce low-cost renewable fuels and chemicals from renewable biomass (Cannella and Jørgensen, 2014). The initial aim of this work was to clarify the structure-function relationship of key LPMO surface residues, and to resolve the amorphogenesis model of LPMO synergism. These aims were accomplished through production of the crystal structure of TfAA10A, characterization of mutants forms of AA10A and AA10B, and using Fourier transform infrared spectroscopy (FTIR) to observe cellulose crystallinity changes after digestion.

5.2 LPMO structure and activity

The structure of *T. fusca* AA10A reported in chapter 2 showed similarities with other bacterial AA10 LPMOs, including a compact β-sandwich fold, conserved copper-chelating active site, and conserved surface residues. Site directed mutagenesis of surface residues showed significantly altered activity. Mutation of W46 resulted in elimination of both binding and activity, while N47 severely reduced binding while having a moderate effect on activity. Domain mutants of TfAA10B showed the importance of the CBM2 binding module for binding at the activity pH optimum, and the minimal role of the X1 domain under conditions of substrate abundance.

The quantification of insoluble products remains a challenge to LPMO kinetic investigation. The secondary hydrolysis method improves existing techniques by eliminating the

need for strong anion-exchange (HPAEC) chromatography. Attempts were made to measure substrate-bound lactones using FTIR, but the technique lacked the sensitivity of detection (not shown). In existing reports, LPMO activity has been reported qualitatively to show product distributions and relative activity through endpoint or time course data. Time course reactions of purified LPMOs have been conducted by several groups, but the kinetic rates reported were of single points and did not reflect the changing substrate in the reaction (Forsberg et al., 2014; Loose et al., 2014). Measuring kinetic rates using soluble substrate-directed PMOs will be more straightforward compared to insoluble substrate-directed LPMOs. Due to the gross similarities of LPMO active sites and binding surfaces, the insights gained from soluble substrate acting LPMOs should be informative for insoluble acting LPMOs.

5.3 LPMO reaction conditions

LPMO activity is driven by delivery of reducing power to the active site copper, making the mechanism of electron transfer an important outstanding question. Small molecules remain the most common reducing agents for LPMO assays, but enzyme-mediated electron transfer may prove to be more effective and allow for tighter LPMO control in native systems. Glutathione occupies a middle ground of redox potential, and could be present from plant cell lysis in native contexts (Foyer and Noctor, 2011). Glutathione can transfer either one or two electrons from the sulfhydryl group, and is typically forms the stable GSSG oxidized form (Noctor et al., 2012). In contrast, ascorbate can form many products after oxidation, complicating product detection using HPLC.

The source of electrons for LPMO activation has not been conclusively established, but several pathways are capable of delivering electrons to the LPMO active site copper. Cellobiose dehydrogenase (CDH) efficiently delivers electrons to multiple families of LPMO, but are not

found in bacterial genomes (Levasseur et al., 2013). Bacteria may use a quinone-based redox enzyme to provide similar activation (Hemsworth et al., 2015). CDH-like feruloyl esterases may perform this role using lignin as the substrate (Arfi et al., 2014).

While current models of LPMO catalytic cycle suggest an order to redox events, it is not known whether LPMOs become active in solution or after binding their substrate (Kim et al., 2014). The LPMO side reaction producing H_2O_2 suggests that activation can occur in the absence of substrate (Kjaergaard et al., 2014). Preliminary tests showed less LPMO inactivation as substrate concentration increased, suggesting that the inactivation mechanism was counteracted by catalytic domain binding (data not shown). Catalase is a solution to LPMO production of H_2O_2 due to high dismutation capacity over a broad pH and temperature range. Catalase was effective at rescuing industrial cocktail, and will likely be a component of future cocktails containing redox active enzymes (Scott et al., 2015). The *T. fusca* ER1 crude supernatant had H_2O_2 dismutation activity, while the crude supernatant of *H. jecorina* did not (data not shown). The presence of two catalases in the *T. fusca* genome, one preceded by a secretion signal, suggests a need for H_2O_2 mediation in native secretomes containing LPMOs (Lončar and Fraaije, 2015).

5.4 Crystallinity and recalcitrance model

The decline of cellulose hydrolysis over time has been attributed to increasing substrate recalcitrance rather than other factors such as inhibition or inactivation (Zhang et al., 1999). Crystallinity was once thought to be the key contributor to recalcitrance, but the influence of crystallinity has recently been debated (Hall et al., 2010; Hu et al., 2015; Park et al., 2010). The understanding of recalcitrance is shifting to taking multiple physical properties into account, especially substrate accessibility and surface disorder, in addition to crystallinity. Based on

results presented in Chapter 4, cellulose crystallinity does not change over the course of bacterial cellulose digestion by hydrolytic cellulases or oxidative LPMOs. This provides additional support for a surface erosion model of recalcitrance that emphasizes physical properties other than crystallinity (Hu et al., 2015; Väljamäe et al., 1998). A current model for the role of LPMO activity in cellulase systems is a contribution to exocellulase synergism through creating new breaks in cellulose chains, enabling processive enzymes to engage with these new chain termini (Vermaas et al., 2015).

The surface erosion model depends on successive removal of cellulose surface layers, analogous to peeling an onion (Hu et al., 2015). This is supported by both experimental and computational work exploring the interaction between exocellulases and their synergistic partners. Exocellulases are able to extract and process along a crystalline cellulose chain after binding to an exposed end, and are responsible for the majority of activity of cellulase mixtures on crystalline cellulose (Kipper et al., 2005). Computational modeling suggests that exocellulases stall on a disordered cellulose surface. The optimal degree of synergistic effect (DSE) occurs when endocellulases are present to liberate stalled exocellulases (Shang and Chu, 2014). This model was supported by experimental results, where exocellulases are released from stalling by endocellulase activity to rebind and continue activity (Hu et al., 2015). These results support an increase of substrate surface complexity as a physical basis of cellulose recalcitrance, rather than cellulose crystallinity alone.

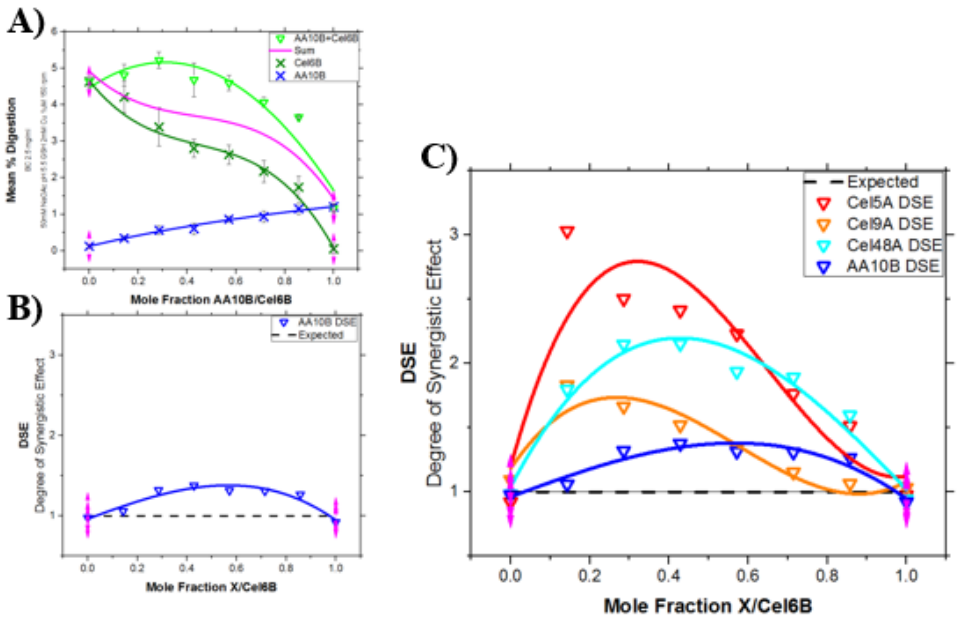


Figure 5.1. Degree of synergistic effect between Cel6B, cellulases, and LPMO A) Extent of digestion by a range of concentrations of TfCel6B (Green X), TfAA10B (Blue X), or their combination (Light green Triangle) in comparison with the expected sum of activity (Magenta line). B) The degree of synergistic effect (DSE) resulting from A. C) DSE of multiple synergistic partners with Cel6B, endocellulase Cel5A (Red), processive endocellulase Cel9A (Orange), exocellulase Cel48A (Light Blue), and LPMO AA10B (Dark Blue).

Experimental results based on the computational simulation of Shang and Chu give context to explore the mechanism of LPMO synergism. Figure 14 shows the results of digestion by a range of ratios between the *T. fusca* Cel6B exocellulase and multiple synergistic partners (end point methods as described in Chapter 2). LPMOs follow a different trend of synergism with the exocellulase Cel6B compared with endo- and exocellulases. The DSE of both endocellulases is shifted left, a result described by Shang and Chu as freeing stalled exocellulases from the cellulose surface. As an exocellulase attacking the opposite reducing end of cellulose chains, Cel48A has an optimal DSE in the center of the ratios. The AA10B LPMO has an optimal ratio shifted slightly towards more LPMO, indicating that while it is able to increase Cel6B engagement it does not free stalled enzymes. In the context of Shang and Chu, this suggests a model of LPMO-exocellulase synergism that involves initial complexation but not release of stalled enzymes, shown in Figure 15.

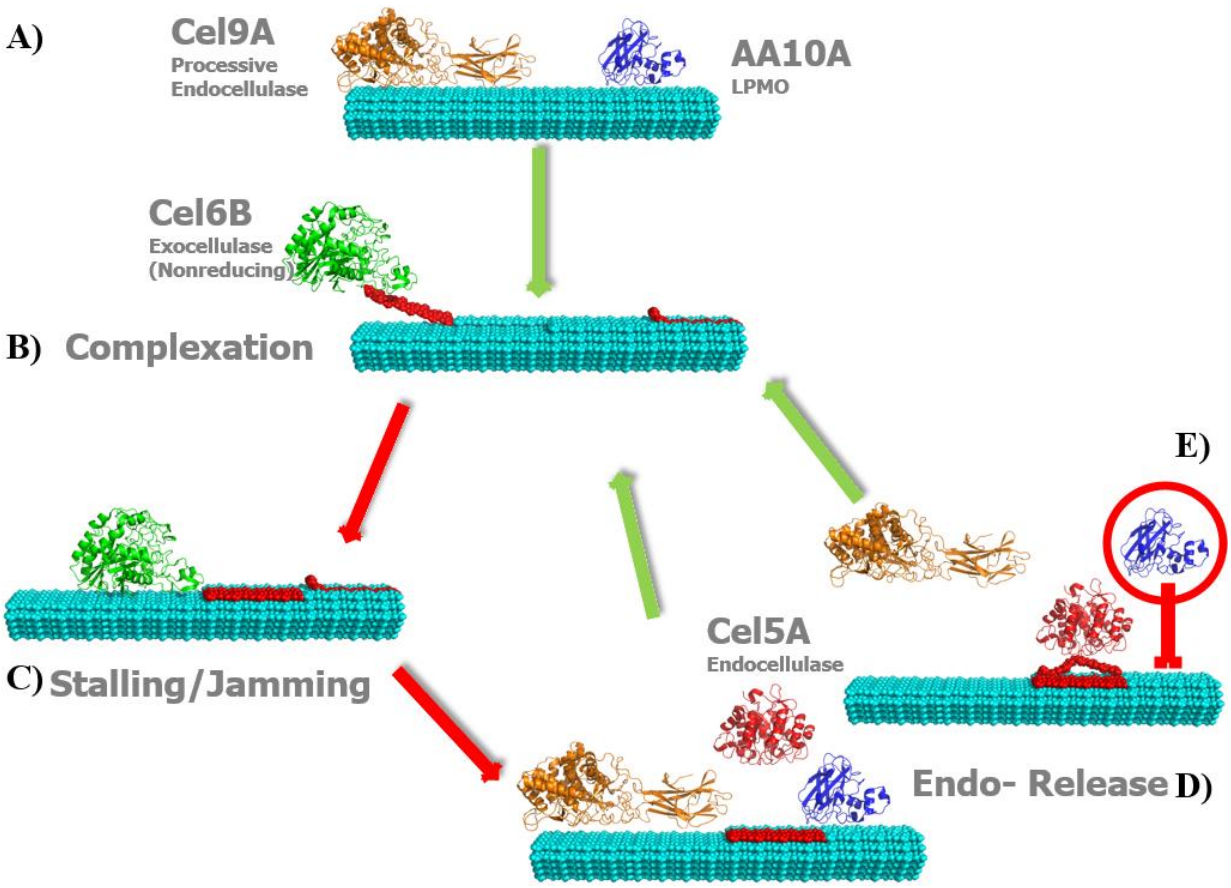


Figure 5.2. A model contrasting exo-endo synergism with exo-LPMO synergism. A) The crystalline cellulose substrate is first cleaved by a suitable enzyme, such as a processive endocellulase (Cel9A, orange), or LPMO (AA10A, Blue) in the *T. fusca* secretome. B) Exocellulases (Cel6B, Green) engage with the exposed chain terminus and processively cleave the substrate. C) Surface irregularities cause exocellulases to stall and remain unproductively bound. D) endocellulases (Cel5A, Red) remove blockages to release unproductively bound exocellulases. E) LPMOs create initial cleavages that enable exocellulase complexation, but cannot free stalled enzymes.

5.5. The role of LPMOs in cellulolytic systems

Reflecting on the structure of substrates that LPMO-containing secretomes evolved to attack, the role of LPMOs may be expanded beyond the surface erosion model. Cellulases and auxiliary activity enzymes break down cellulose when they are able to access it, but cellulose within plant cell walls is typically embedded within a complex matrix of hemicellulose, lignin, and proteins (Mansfield et al., 1999). LPMOs are small relative to cellulases and can attack intact cellulose, suggesting a role during the initial phases of biomass degradation. There is little evidence for LPMO amorphogenesis, but initial LPMO cleavages can create the starting

substrates for cellulase attack (Vermaas et al., 2015).

A single model of plant cell wall architectures is not possible, but common organizational themes exist. Secondary xylem fibers, the dominant structural plant cells, are relevant for paper and bioenergy applications due to their cellulose-rich "xylan-type" secondary cell walls (Gorshkova et al., 2012). The xylan-type secondary cell wall is a thick mesh of overlapping cellulose microfibers cross-linked by abundant xyloglucans (Mikshina et al., 2013). This mesh is surrounded by lignin and pectin, which bind extensively and reduce pore sizes (Lin et al., 2015). The structural organization of xylan-type secondary cell walls is highly conserved across plant species, suggesting that cellulose is initially inaccessible to cellulases in the majority of biomass substrates (Gorshkova et al., 2012).

Cellulases have difficulty accessing cellulose within plant cell walls, which must be exposed (Mansfield et al., 1999). Hemicellulases degrade bound xylan, pectinases remove pectin linkages, and laccases depolymerize lignin to expose buried cellulose. With their smaller size, LPMOs could access cellulose before the much larger cellulases to cleave the intact substrate. LPMOs can be activated by removing electrons from lignin, abundant in the early stages of digestion (Cannella and Jørgensen, 2014). Additionally, uncontrolled LPMO H_2O_2 production would be useful to weaken complex substrates if cellulases are sterically excluded.

LPMOs synergize efficiently with cellulases through random oxidative cleavage of the cellulose surface, despite performing few cleavages and not affecting the structural crystallinity of cellulose. LPMO traits suggest a role distinct from cellulases; early expression during biomass deconstruction, unregulated surface exposed redox active sites, and compact catalytic domains. It can be hypothesized that small LPMOs can access buried cellulose microfibrils to perform the initial oxidative cleavage that enable exocellulase complexation, preparing crystalline surfaces

for more efficient cellulase attack. The introduction of oxidized ends may also modify cellulose surface charges, disrupting cellulose-hemicellulose hydrogen bonds in a positive feedback loop to increase cellulose access. Initial cleavages of buried cellulose provides an explanation to the presence of LPMOs in secretomes containing cellulases effective on crystalline cellulose. The processive endocellulase Cel9A does not synergize with LPMOs, and the exocellulase Cel7A is only weakly synergistic (data not shown), but both are much larger and would be excluded from initial cellulose access. LPMOs may play a key role in microbial secretomes by increasing cellulose surface area for attack by much larger cellulases, performing initial cleavages and possibly remodeling complex cell walls to facilitate deconstruction.

5.6 Conclusion

The wide diversity and potent synergism of lytic polysaccharide monooxygenases (LPMOs) make this enzyme family promising avenue for industrial cellulase cocktail improvement. The results presented here show the importance of conserved surface residues for *Thermobifida fusca* AA10A activity, the large contribution of the CBM2 domain for AA10B binding, and the absence of changes to cellulose crystallinity after digestion. LPMO activity has been reported to alter cellulose crystallinity (Eibinger et al., 2014; Tanghe et al., 2015), but results do not support LPMOs as a sole C_1 factor. It is more likely that synergistic interactions, of which LPMO oxidative cleavage is one component, all contribute to degrade difficult substrates. The contributions of multiple physical properties, including crystallinity and surface heterogeneity, must be further characterized in order to fully understand how cellulolytic microorganisms overcome cellulose recalcitrance.

REFERENCES

Arfi Y, Shamshoum M, Rogachev I, Peleg Y, Bayer E a. 2014. Integration of bacterial lytic polysaccharide monooxygenases into designer cellulosomes promotes enhanced cellulose degradation. *Proc. Natl. Acad. Sci. U. S. A.* **111**:9109–14.

Cannella D, Jørgensen H. 2014. Do new cellulolytic enzyme preparations affect the industrial strategies for high solids lignocellulosic ethanol production? *Biotechnol. Bioeng.* **111**:59–68.

Eibinger M, Ganner T, Bubner P, Ros S, Kracher D, Haltrich D, Ludwig R, Plank H, Nidetzky B. 2014. Cellulose Surface Degradation by a Lytic Polysaccharide Monooxygenase and Its Effect on Cellulase Hydrolytic **289**:35929–35938.

Forsberg Z, Mackenzie AK, Sørlie M, Røhr ÅK, Helland R, Arvai AS, Vaaje-Kolstad G, Eijsink VGH. 2014. Structural and functional characterization of a conserved pair of bacterial cellulose-oxidizing lytic polysaccharide monooxygenases. *Proc. Natl. Acad. Sci. U. S. A.* **111**:8446–51.

Foyer CH, Noctor G. 2011. Ascorbate and glutathione: the heart of the redox hub. *Plant Physiol.* **155**:2–18.

Gorshkova T, Brutch N, Chabbert B, Deyholos M, Hayashi T, Lev-Yadun S, Mellerowicz EJ, Morvan C, Neutelings G, Pilate G. 2012. Plant Fiber Formation: State of the Art, Recent and Expected Progress, and Open Questions. *CRC Crit. Rev. Plant Sci.* **31**:201–228.

Hall M, Bansal P, Lee JH, Realff MJ, Bommarius AS. 2010. Cellulose crystallinity - A key predictor of the enzymatic hydrolysis rate. *FEBS J.* **277**:1571–1582.

Hemsworth GR, Johnston EM, Davies GJ, Walton PH. 2015. Lytic Polysaccharide Monooxygenases in Biomass Conversion. *Trends Biotechnol.* **33**:747–761.

Hu J, Gourlay K, Arantes V, Van Dyk JS, Pribowo A, Saddler JN. 2015. The Accessible Cellulose Surface Influences Cellulase Synergism during the Hydrolysis of Lignocellulosic Substrates. *ChemSusChem* **8**:901–907.

Kim S, Ståhlberg J, Sandgren M, Paton RS, Beckham GT. 2014. Quantum mechanical calculations suggest that lytic polysaccharide monooxygenases use a copper-oxyl, oxygen-rebound mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **111**:149–154.

Kipper K, Väljamäe P, Johansson G. 2005. Processive action of cellobiohydrolase Cel7A from Trichoderma reesei is revealed as “burst” kinetics on fluorescent polymeric model substrates. *Biochem. J.* **385**:527–35.

Kjaergaard CH, Qayyum MF, Wong SD, Xu F, Hemsworth GR, Walton DJ, Young NA, Davies GJ, Walton PH, Johansen KS, Hodgson KO, Hedman B, Solomon EI. 2014. Spectroscopic and computational insight into the activation of O2 by the mononuclear Cu center in polysaccharide monooxygenases. *Proc. Natl. Acad. Sci.* **111**:8797–8802.

Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. 2013. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol. Biofuels* **6**:41.

Lin D, Lopez-Sanchez P, Gidley MJ. 2015. Binding of arabinan or galactan during cellulose synthesis is extensive and reversible. *Carbohydr. Polym.* **126**:108–121.

Lončar N, Fraaije MW. 2015. Not so monofunctional-a case of thermostable Thermobifida fusca catalase with peroxidase activity. *Appl. Microbiol. Biotechnol.* **99**:2225–2232.

Loose JSM, Forsberg Z, Fraaije MW, Eijsink VGH, Vaaje-Kolstad G. 2014. A rapid quantitative activity assay shows that the Vibrio cholerae colonization factor GbpA is an active lytic polysaccharide monooxygenase. *FEBS Lett.* **588**:3435–40.

Mansfield S, Mooney C, Saddler J. 1999. Substrate and Enzyme Characteristics that Limit Cellulose Hydrolysis. *Biotechnol. Prog.* **15**:804–816.

Mikshina P, Chernova T, Chemikosova S, Ibragimova N, Mokshina N, Gorshkova T. 2013. Cellulosic Fibers : Role of Matrix Polysaccharides in Structure and Function. In: . *Cellul. - Fundam. Asp.*, pp. 91–112.

Noctor G, Mhamdi A, Chaouch S, Han Y, Neukermans J, Marquez-Garcia B, Queval G, Foyer CH. 2012. Glutathione in plants: An integrated overview. *Plant, Cell Environ.* **35**:454–484.

Park S, Baker JO, Himmel ME, Parilla P a, Johnson DK. 2010. Cellulose crystallinity index: measurement techniques and their impact on interpreting cellulase performance. *Biotechnol. Biofuels* **3**:10.

Scott BR, Huang HZ, Frickman J, Halvorsen R, Johansen KS. 2015. Catalase improves saccharification of lignocellulose by reducing lytic polysaccharide monooxygenase-

associated enzyme inactivation. *Biotechnol. Lett.*

Shang BZ, Chu J-W. 2014. Kinetic Modeling at Single-Molecule Resolution Elucidates the Mechanisms of Cellulase Synergy. *ACS Catal.* **4**:2216–2225.

Tanghe M, Danneels B, Camattari A, Glieder A, Vandenberghe I, Devreese B, Stals I, Desmet T. 2015. Recombinant Expression of Trichoderma reesei Cel61A in Pichia pastoris: Optimizing Yield and N-terminal Processing. *Mol. Biotechnol.* **57**:1010–1017.

Väljamäe P, Sild V, Pettersson G, Johansson G. 1998. The initial kinetics of hydrolysis by cellobiohydrolases I and II is consistent with a cellulose surface-erosion model. *Eur. J. Biochem.* **253**:469–475.

Vermaas J V, Crowley MF, Beckham GT, Payne CM. 2015. Effects of lytic polysaccharide monooxygenase oxidation on cellulose structure and binding of oxidized cellulose oligomers to cellulases. *J. Phys. Chem. B* **119**:6129–43.

Zhang S, Wolfgang DE, Wilson DB. 1999. Substrate heterogeneity causes the nonlinear kinetics of insoluble cellulose hydrolysis. *Biotechnol. Bioeng.* **66**:35–41.