Cloning and characterization of new cellulases from *Cellulomonas fimi* and *Cellulomonas flavigena*

by

Marc Aristaeus de Asis

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Science

in

Biology

Waterloo, Ontario, Canada, 2013

Author's declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Lignocellulose is one of the most abundant carbon sources in nature. This naturally-occuring substance is an underutilized source of bioenergy. A major bottleneck in biofuel processing is the enzymatic hydrolysis of lignocellulose into its ultimate fermentable product, glucose. *Cellulomonas fimi* is a well-studied soil organism known for its capabilities to efficiently hydrolyze cellulose. Recently sequenced genomes of *Cellulomonas fimi* and *Cellulomonas flavigena* have allowed analysis to reveal previously unidentified cellulases from several glycoside hydrolase (GH) families. This study also includes the expression of secreted cellulases from families GH 5, 6, and 9 at the protein level by the native organism after growth in media supplemented with carboxymethylcellulose or soluble xylan. In order to find enzymes with novel qualities, the cloning and expression of these newly identified cellulases from *C. fimi* and *C. flavigena* were done. One of these enzymes is Celf_1230 (Cel6C), a putative cellobiohydrolase from the glycoside hydrolase family 6. Using substituted cellulose derivatives as substrates, we have characterized Celf_1230 to be a thermostable enzyme with endoglucanase activity.

Acknowledgments

I would like to extend my deepest gratitude to my supervisor Dr. Warren Wakarchuk for all his support and guidance throughout the entire process. I am thankful that he gave me an opportunity to prove myself and allow me to learn and grow as a scientist. He widened my perspective and introduced me to the many facets of carbohydrate research and to notable people in the field.

I am also indebted to Drs. Todd Holyoak and David Rose for their generosity in providing space and resources for me to continue my research during my stay at the University of Waterloo. They also provided valuable insight and technical expertise. The support they provided was beyond what was asked of them and I am grateful for that.

I would like to thank my committee members, Drs. Trevor Charles and Bernard Glick for their opinions and guidance throughout the course of my graduate degree.

I would also like to thank the following people for their contributions to the work: Drs. Marie Couturier and Emma Master of the University of Toronto for their expertise and generosity with the biomass substrates and the ToF-SIMS assay; Dr. Anthony Clarke of the University of Guelph for his generous gift of the Celf_3184 plasmid; Helen Stubbs of the Department of Chemistry for her help and advice on the DSC; The NRC organization and the individuals (Dr. John Kelly, Denis Brochu, Theresa Lindhout, Melissa Schur, and many more) for their work on the mass-spectrometry and their advice during my summer stay there.

Lastly, the journey would not have been complete without the support of my friends and family. Their encouragement allowed me to pursue my education wholeheartedly.

Table of contents

List of Figures

List of Tables

Chapter 1

Summary

Lignocellulose is an abundant organic compound that is currently a highly underutilized carbon source that has potential for biofuel production. Enzymes are necessary for the saccharification of cellulose into oligosaccharides, cellobiose, and ultimately into glucose for downstream fermentation. In nature, cellulose decomposition is achieved through the action of cellulase enzymes from microorganisms like fungi and bacteria. These organisms are responsible for the recycling of carbon that is trapped in the terrestrial biosphere. Thus, the glycoside hydrolases of interest, which are necessary for the complete degradation of this substrate, are: endo-β-1,4-glucanases (endoglucanases), exo-β-1,4-glucanases (cellobiohydrolases), β-glucosidases, and the corresponding enzymes that act on hemicellulosic substrates. *Cellulomonas* spp. potentially possess the key enzymes to improve industrial production of bioproducts. Analysis of the recently sequenced genomes of actinobacteria *Cellulomonas fimi* and *Cellulomonas flavigena* have revealed previously unidentified cellulases of several glycoside hydrolase (GH) families. This thesis focuses on expression and characterization of recombinant glycosidases from families GH 5, 6, and 9. This permitted me to investigate uncharacterized endo-β1,4-glucanases and cellobiohydrolases. Moreover, GH 5 and 9 are large families with diverse substrate specificities that are of interest for biomass utilization.

In Chapter 2, biomass is introduced and the composition of lignocellulose is discussed in detail. Additionally, cellulosic substrates used in research are explained and compared. Additionally, glycoside hydrolases, specifically those that act on cellulosic substrates, are discussed in this chapter. This details discussing the mechanisms utilized by these enzymes. Cellulase-producing organisms are briefly compared in Chapter 2. Furthermore, this chapter introduces the traits of the *Cellulomonas* spp. for a complete profile. Chapter 3 presents the complete materials and methods for the entire thesis. Chapter 4 presents the results of the thesis. This includes

bioinformatics, proteomic and RT-PCR data and procedures used to elucidate the expression of cellulases from *C. fimi* and *C. flavigena*. Information gathered from this, along with genomic analysis, provided an initial screen for targets to be cloned, expressed, and purified for detailed characterization of enzymatic activity. Among the dozens of genes that were cloned, there was one successful and interesting cellulase Celf_1230. The expression, purification, and characterization of this enzyme is the subject of the second half of Chapter 4. The last chapter will focus on the future perspectives that can be pursued to complement the work that is presented here.

Chapter 2

Introduction

2.1 Biofuels

The decreasing global reserves of fossil fuels and the exponential increase in demand have necessitated the search for alternative sources of energy. In particular, biofuels are one of the popular options for renewable energy for several reasons. First, biofuels are readily introduced into our current energy infrastructure due to the similarity to fossil fuels. In fact, contemporary fuel is already supplemented with as much as 5-10% bioethanol in certain geographic regions [1,2]. Also, slight modifications to combustion engines can allow this machinery to run on 85% biofuels today [2]. Secondly, its liquid property allows energy to be stored and transported for long durations and distances. This is in contrast to the limitations of the energy grid of other alternative solutions (i.e. hydro, wind, and solar energy). Biofuels can be subcategorized into two broad categories. Contemporary biofuels are either sugarcane- or starch-based; while second-generation are based on lignocellulosic starting material. The intrinsic difference between the two feedstocks is their characterization as alpha- and beta-carbohydrates, respectively.

Contemporary (first generation) biofuels poses two problems: First, existing supplies of starch raw materials are insufficient to meet increasing demand[3]. Secondly, they are a controversial resource for bioconversion due to rising food costs[3]. However, relative to starch, lignocellulose is difficult to process due to its composition of lignin, hemicellulose, and crystalline cellulose. For this reason, contemporary biofuels continue to be the choice of biofuel producers.

Production of lignocellulose biofuel (next generation) requires three fundamental steps: Pretreatment of lignocellulose; hydrolysis/saccharification of long-chain carbohydrates; and fermentation of monosaccharides into ethanol [4]. One of the major bottlenecks in lignocellulose processing are enzyme production and saccharification, which is the hydrolysis of cellulose and

hemicellulose into their respective fermentable sugars[4,5] As a result of the current pretreatment methods, the industrial bioconversion of cellulose into ethanol is currently uneconomical and unfeasible due to the lack of inexpensive efficient biocatalysts that can perform at harsh conditions [4,5].

2.2 Starch

Starch, a source of first-generation biofuel, is a carbohydrate macromolecule composed of amylopectin and amylose [6]. Amylose is a linear polymer made of α-1,4 linked D-glucopyranosyl monomers. On the other hand, amylopectin is a polysaccharide composed of α-1,4-glucan chains that branches with α-1,6 linkages in a cluster-like design [6]. Amylase (EC 3.2.1.1) hydrolyzes amylose & amylopectin by either acting at random sites or from the non-reducing end to ultimately produce maltose, which is a disaccharide of glucose molecules in an α-1,4 configuration [7]. Amylase has huge economic and industrial impact, which makes up about 30% of the world's enzyme production [8].

2.3 Lignocellulose

Plants contain 99% of the carbon found in living organisms, approximately 27×10^{10} tons [9]. Lignocellulose is a highly abundant organic compound from plant biomass that is underutilized in the rising efforts towards renewable green energy [1,4,5,10,11]. There are rich sources of lignocellulose from agricultural residues and industrial waste, which makes it an excellent alternative compared to biofuel from food crops [5]. Lignocellulose is made up of three significant substances **(Figure 1)**: Cellulose makes up the core and acts as the framework; Hemicellulose surrounds the cellulose as a matrix component; and Lignin is a resilient encrusting substance that

gives wood its hardy characteristic. It can be extrapolated that about 40% of the plant-carbon is bound in cellulose, 30% in lignin, 26% in other polysaccharides [9].

2.4 Plant cell wall

The fundamental organization of bundled cellulose is the elementary fibril. A cross-sectional view of an elementary fibril shows that it could contain roughly 40 cellulose chains. Aggregates of elementary fibrils are covered by hemicellulose and lignin [12]. These micelles then form highly ordered cellulose microfibrils and its width can vary between different materials [13,14]. These microfibrils are indefinite in length and are linked laterally into lamellae that form different cell wall layers [13]. The conceptual organization of the cell wall contains a thin primary outer layer and three secondary layers [13]. These layers are differentiated from one another based on the order and orientations of the microfibrils.

2.5 Lignin

Lignin accounts for roughly 30% of the organic carbon found in the biosphere, which makes it the second most abundant biopolymer in existence [15]. Here, the importance of lignin lies within understanding the pretreatment methods and innovations for processing efficiency. However, due to its carbon content, lignin itself is actually a source of potential energy and useful chemical components.

The lignin architecture is fundamental for the structural integrity of the cell wall and thus the rigidity of the wood [15]. Likewise, the structure and constitution of lignin makes it highly resistant to degradation and enzymatic action. It is a three-dimensional racemic aromatic polymer with an irregular and disordered branching structure [9,15]. Its configuration is a random combination of phenylpropane [hydroxycinnamyl alcohol] units (p-coumaryl, coniferyl, and sinapyl alcohols)

that differ in their degree of methoxylation [15,16]. However, studies in the past decade show that there are other monolignols involved aside from the three aforementioned phenylpropanoids [15].

2.6 Hemicellulose

Hemicellulose is a heterogeneous polysaccharide that has a complex structure made up of polymers of different sugars. Depending on the plant material, hemicellulose will have varying content of pentoses (xylose and arabinose), hexoses (glucose, mannose, and galactose), sugar-acids, and deoxy-sugars [16]. For hardwoods, the primary hemicellulose component is xylan (polymer of xylose), while softwoods contain a heteropolymer such as glucomannan and arabinogalactans [9,16,17]. Hemicelluloses differ from celluloses by containing shorter chains and significant branching of these chains [9]. In particular, xylan is comprised of a homopolymer backbone of xylose that is interlaced with an irregular pattern of 4-O-methylglucuronic acid sidechains as well as traces of rhamnose and galacturonic acid [9]. These sugar-acid molecules are suggested to be the reason for alkali resistance of hemicellulose [9,17,18]. Furthermore, hemicellulose serves as the link that connects the lignin polymer and the crystalline cellulose to create a more rigid complex [19].

Among the three main components of lignocellulose, hemicellulose is the most thermo-chemical sensitive component [20]. During pre-treatment, the sidegroups of hemicellulose react first then this is followed by reaction with the backbone.

2.7 Pectin

This polymer is a minor component in the lignin-hemicellulose network that encapsulates the microfibril structure [21]. The pectin polysaccharide is a matrix component for plant cell wall [21]. Its structure is made up of alternating blocks of copolymers (branched or unbranched). The

branched block has a main chain that is comprised of galacturonan subunits and interlaced & bent by rhamnose units with various sidechains (arabinose and/or galactose). The second copolymer is unbranched and is made up of a galacturonan backbone [21].

2.8 Cellulose

At the center of each microfibril is a core of densely-packed crystalline cellulose, a linear polysaccharide of β-1,4-linked D-glucose subunits[9]. The dry portion of plant material can be composed of 40-45% of cellulose [17]. Furthermore, the cellulose chain has a reducing and non-reducing end. The OH-group at the C1 position has reducing properties due to the pyranose ring formation creating an aldehyde hydrate group at one end [9]. The non-reducing end of the cellulose chain is an alcoholic hydroxyl group at the C4 position.

In order to further appreciate the structure of the cellulose polymer, it is crucial to understand its biosynthesis. Cellulose synthase (a specific glycosyl transferase) acts in a processive manner by adding two -D-glucoses from UDP-glucose to the growing chain [22]. The two-by-two addition of D-glucose molecules removes the need for a cellulose primer; and maintains the two-fold screw axis shaped by the repeating cellobiose (disaccharide) units[22]. Fundamentally, the repeating unit of cellulose is a cellobiose unit. The adjacent glucose molecules are joined by elimination of one water molecule between their hydroxylic groups at carbon 1 (C1) and carbon 4 (C4) of the first and second unit respectively[9]. The β-linkage requires the second glucose unit to rotate 180° around the C1-C4 axis **(Figure 2)**. This allows the linear chain to maintain a single plane arrangement. An α-linkage would lead to a helical formation as observed in amylose.

Moreover, the cellulose fibers have a strong tendency to form intra- and intermolecular hydrogen bonds, which intensifies its supramolecular structure[17]. The cellulose synthase polymerizes

adjacent glucan chains in close proximity that encourages crystallization [22]. A structure called the terminal complex is found at the growing end of the microfibril. The variations of the terminal complex are responsible for producing different minisheets of crystalline cellulose [22]. Segments of the native cellulose core are amorphous regions that lack hydrogen bonding and are more susceptible to enzymatic hydrolysis.

Crystalline cellulose has been investigated by X-ray diffraction to solve for its structure. Native cellulose (cellulose I) have their chains organized in a parallel orientation [17]. In contrast, regenerated cellulose (cellulose II) has a staggered and antiparallel orientation. Cellulose I is converted into cellulose II by alkali swelling, which introduces more hydrogen bonding and deforms the lattice [17]. Cellulose III and IV are formed when I and II are exposed to heat and chemical treatments [17].

2.8.1 Cellulose as a substrate

Pretreatment of natural substrates required the delignification of the plant material. The product achieved after delignification is called holocellulose. Ideally, this process should completely remove all lignin material without chemically modifying or hydrolyzing the polysaccharide inside [9]. For research purposes, the two most common methods for delignification are chlorination (then alcohol extraction) and acidification [9]. Multiple investigators have modified these processes to obtain hemicellulose and cellulose that are less unaffected. Factors that were considered include reaction temperature, time, chemical amount, and pH of solution [9]. Lignocellulose pretreatment is a vast subject matter on its own and will not be further discussed here. Detailed descriptions are given in the following review articles (e.g. Fan et al., 1982; Hendriks & Zeeman, 2009; Mosier et al., 2005; Sun & Cheng, 2002).

Producing pure cellulose for research requires intensive treatments involving partial hydrolysis, dissolution, and precipitation, which consequently leads to very short molecular chains [9]. Thus, when isolating cellulose from natural substrates, it is common to achieve a crude preparation called α-cellulose. This compound refers to the material that remains insoluble after treating wood with strong sodium hydroxide solution. β-cellulose refers to the soluble fraction that can be precipitated by neutralizing the solution. The matter that remains soluble in the neutralized fraction is called γ-cellulose [9].

It is worth noting that alkali alone is unable to solubilize native cellulose, as only depolymerized cellulose fragments with a low degree of polymerization are alkali soluble [17]. Alkali celluloses, those that are swollen by sodium hydroxide, are important intermediates due to its increased reactivity. This allows other reagents to penetrate the substrate and permit downstream modifications and derivations [24].

Commercial grades of cellulose derivatives are distinguished based on their degree of substitution (DS) and degree of polymerization (DP). These two factors have a profound effect on the viscosity of solubilized substrates. An increase in DS increases the solubility of the modified cellulose. To be considered as cellulose, the glucan chains must have at least several hundred residues; although a very low DP of 6-8 will already render the Beta-1,4-glucan as insoluble [22]. Native cellulose can have DP of 10,000 glucose residues [25] but carboxymethylcellulose would be roughly 3200 (Sigma C5013).

2.8.2 Cellulose derivatives

Cellulose derivatives have garnered interest for both industrial use and research purposes. Derivatives such as regenerated cellulose and cellulose ethers are used as filters, coatings, films,

and as additives (i.e. food, building material, and pharmaceuticals) [24]. Each D-glucopyranose within a polysaccharide chain retains three reactive hydroxyl groups (OH-2, OH-3 and OH-6) [17]. Cellulose can be esterified with either organic (i.e. acetic acid) or inorganic acids (e.g. nitric, sulfuric, and phosphoric acids). It is essential that the acid has the ability to induce swelling of the polysaccharide chain [17]. Nitrocellulose and cellulose acetates are two common materials for filtration systems. Cellulose esters are typically insoluble in water but soluble in various organic solvents or in acidic/alkaline solutions.

Cellulose ethers are, in general, water soluble substrates such as hydroxyethylcellulose and carboxymethylcellulose [17]. These compounds are prepared by treating alkali-swollen cellulose with various reagents like alkyl or aryl halides. Synthesis and molecular descriptions of these compounds will not be discussed further in detail.

Cellulose (and hemicellulose) derivatives can be further modified to crosslink dye compounds. These substrates are then used to assess glycoside hydrolases for enzymatic activity. However, most cellulases are inhibited by the modifications on these substrates hence there is a decrease in available hydrolysis sites. This necessitates the need for a more "natural" substrate to assess cellulolytic activity. Cotton, which consists of 95-99% crystalline cellulose can be used to assess exocellulases [9]. Unmodified swollen cellulose is a prime example of a pretreated natural substrate. Swelling of avicel (crystalline cellulose) can be economically performed by investigators. The most popular method is the preparation of phosphoric acid-swollen cellulose (PASC). Details of this process will be discussed further under experimental procedures (Section 6.1.3).

2.9 Glycoside hydrolases

2.9.1 Annotation by the CAZy database

Carbohydrate-active enzymes are those that create, modify, or cleave glycosidic bonds. The CAZy [***C***arbohydrate ***A***ctive en***Zy***mes] database (http://www.cazy.org/) classifies known CAZymes according to family based on amino acid sequence similarity[26,27]. There are only five[26] enzyme classes that are presently annotated by CAZy: glycoside hydrolases (GH) (EC 3.2.1.-); glycosyltransferases (GT) (EC 2.4.x.y); polysaccharide lyases (PL) (EC 4.2.2.-); carbohydrate esterases (CE) (mainly EC 3.1.1.-), and recently, auxiliary activities (AA) (mainly EC 1.1.3.- and 1.11.1.-). Additionally, CAZy also annotates associated modules. Presently, this category only covers carbohydrate-binding modules (CBMs).

Glycosyltransferases catalyze the formation of glycosidic bonds, while glycoside hydrolases degrade or rearrange glycosidic bonds. To date, there are 132 glycoside hydrolase (GH) families. However, some families have since been deleted (i.e. GH-21, 40, 41, 60, 69) due to reclassification. The list of families is systematically updated based on NCBI releases and using the database's internal BLAST. Positive hits are listed either automatically or manually under the family with sequence similarity. Enzyme activity/biochemistry and structure are extracted from the literature after they are experimentally demonstrated and publicly available.

Members of these families will have similar mechanism, protein fold, and catalytic residues responsible for hydrolysis but then vary between different families. Moreover, hydrolase families can be further classified into larger groups called clans that are based on significant similarity of their tertiary structure, which is better conserved than their sequences [28]. Thus, predicted protein

folds of the new cellulases, compared to folds of well-studied enzymes, should provide clues to its potential action on cellulosic substrates.

This thesis focuses on families GH 5, 6, and 9 since these are the major families that include cellulases. GH5[29] employs a retaining mechanism while GH families 6[30] and 9[31] uses a single-displacement reaction thereby causing an inversion of stereochemistry configuration at the anomeric carbon (discussed further below). The GH 5 family belongs to the clan GH-A where it shares a $(\beta/\alpha)_8$ barrel fold with members of that clan[32]. GH 9 has an $(\alpha/\alpha)_6$ barrel and is classified as part of the GH-G clan[33]. On the other hand, GH 6 does not belong to any clan and members of this family have varying folds in respect to their activity.

2.9.2 Mechanism

In terms of mechanism, all glycoside hydrolases act by a general acid/nucleophile mechanism where two amino acids in the active site take part in either a single- or double-displacement reaction[34] **(Figure 3).** There are two types of mechanisms for glycoside hydrolases: inverting or retaining, depending on the anomeric hydroxyl group configuration after the reaction [35]. The inverting mechanism relies on two amino acids in the catalytic pocket. Both of these amino acids have a –COOH side group (i.e. aspartic acid and glutamic acid) where one would serve as a general acid while the other as a general base [36]. The catalytic acid residue donates a proton to the anomeric carbon of the substrate, "while the catalytic base removes a proton from a water molecule to increase its nucleophilicity" [35]. The reaction proceeds through an oxocarbenium ion-like transition state [37].

The retaining glycoside hydrolases generally follow the classical Koshland retaining mechanism. Initially, one of the carboxyl-containing amino acids serves as a general acid by protonating the

oxygen of the glycosidic bond. The other catalytic residue serves as a nucleophile to the anomeric carbon which forms a glycosyl-enzyme intermediate [35]. Subsequently, the carboxylate deprotonates a water molecule and increases its nucleophilicity to disrupt the glycosyl-enzyme intermediate [35,36]. Other glycosyl hydrolases that degrade non-cellulosic substratres have been observed to follow a non-Koshland retaining mechanism. These enzymes employ either neighboring group participations, alternative nucleophiles, NAD as a cofactor, or substrate-assisted mechanisms[35]. These mechanisms are beyond the scope of this work and will not be discussed further here.

2.9.3 Classes of Glycoside Hydrolases

Glycoside hydrolases can also be classified based on their substrate specificity. In particular, cellulases are GHs that act on cellulosic substrates. Based on their mode of action, cellulases can then be sub-categorized into endoglucanases, exoglucanases/cellobiohydrolases, and β-glucosidases. Endoglucanases **(Figure 4)**, whose ability is to hydrolyze β-1,4 glycosidic linkages, are enzymes that act randomly on internal sites in the chain thereby increasing the amount of shorter chains available[4]. Exoglucanases **(Figure 4)** then degrades the molecule further by processively removing cellobiose either from the reducing end or non-reducing end (CBHI and CBHII, respectively)[4]. After removing the disaccharide product, several subsites on the enzyme allow it to remain bound to the polysaccharide substrate[38]. β-glucosidases hydrolyzes cellobiose into glucose in preparation of downstream fermentation. Substantial hydrolysis of native cellulose typically requires the synergistic activity of these three types of cellulases. [39]

Some glycoside hydrolases have a multi-modular organization that contains one or more catalytic domains (CD) and one or more non-catalytic domains[38,40]. One of the non-catalytic

modules is the carbohydrate-binding module (CBM), which plays a role in substrate binding, while other modules (fibronectin type III and immunoglobulin-like domains) have unknown functions[38,40–42]. Non-modular cellulases, those that do not contain a CBM, have decreased activity on insoluble substrates but maintains their ability on soluble substrates[43]. A similar observation is seen on modified modular enzymes when the CBM is genetically or proteolytically excised[33,44,45].

2.10 Cellulase-producing organisms

2.10.1 Overview

Cellulase-producing organisms span the phylogenetic tree across all three kingdoms. Most of the lignocellulose degradation is accomplished by fungi and bacteria [39].

Animal parasites such as insects (e.g. boring beetles, termites, and ants), crustaceans (i.e. gribble) and mollusks (i.e. shipworms) are also able to feed on wood [9]. Herbivores and soil macro-invertebrates rely on symbiotic gut bacteria to utilize cellulosic material [46]. As exemplars, ruminants process their diet in their fore-stomach that allows slow transit, anaerobic conditions, and thus the formation of short-chain fatty acids for host utilization [46]. The rumen holds a complex microbial community consisting of anaerobic fungi, protozoa, bacteria, and methanogenic archaea [46]. This community contains numerous organisms whose growth conditions have not been elucidated and thus remain elusive as research strains. As a result, there is considerable amount of interest in metagenomic analysis to discover novel glycoside hydrolases from various sources [47,48].

2.10.2 Eukaryotic sources

Fungal cellulases are currently popular in industrial applications due to its potency, ease of purification, and sheer yield of enzyme production. The fungal model organism is *Trichoderma reesei*, but substantial attention has also been given to *Aspergillus* spp. and *Humicola* spp. among other industrially relevant cellulase producers [49]. Cellulolytic fungi can be subdivided based on the wood condition after degradation (i.e. brown-rot, white-rot, and soft-rot) [9]. These groups vary in their ability to degrade lignin and polysaccharides. For instance, brown-rot fungi are able to hydrolyze polysaccharides of lignocellulose with minimal degradation of the lignin coating [9].

2.10.3 Prokaryotic sources

Although fungal cellulases are the preferred enzymes for industrial applications, bacterial cellulases still receive strong interest in the quest for novel enzymes. Bacterial enzymes represent greater diversity compared to their fungal counterparts. It was common to view the cellulolytic systems of the two as similar. However, it is evident that cellulolytic fungi are better equipped to penetrate natural woody substrates [9]. Furthermore, cell-free supernatants of cellulolytic bacteria have a decreased activity on crystalline cellulose even though the cultures are able to feed on the substrates [50]. However, current bioprocessing practices rely on heat and acid for the pretreatment of lignocellulose. Thus considerable attention has been focused on discovering thermophilic and acidophilic cellulases [3]. That being said, there have been relatively few thermophiles and acidophiles whose genomes have been fully sequenced. This can be attributed to the difficulty in cultivating these organisms under laboratory conditions [3]. Additionally, recombinant strategies to produce these thermostable enzymes in a mesophilic host bacteria can be challenging [51,52]. It has been assumed that bacterial communities act synergistically and symbiotically to degrade

varying lignocellulosic substrates. As an exemplar, it was surprising that *Clostridium thermocellum* can efficiently degrade xylan but is unable to utilize xylose as a carbon source [53,54]. Within this community are cellulolytic species of aerobic actinobacteria.

The model organisms for cellulolytic actinobacteria have been *Thermobifida fusca* and *Cellulomonas fimi* [55]. The following section further describes *Cellulomonas*, which is the focus of this thesis.

2.10.3.1 *Cellulomonas*

Cellulomonas spp. are Gram-positive and coryneform rod bacteria of the Actinobacteria phylum (Gram-positive bacteria with high G+C ratios)[56]. For *Cellulomonas*, its G+C ratio has been determined to be in the range of 71-76 mol% [57,58]. They are best known for their capability of decomposing lignocellulosic substrates[10]. Their main habitat is the soil where they were initially isolated [57]. More specifically, *Cellulomonas fimi* and *C. flavigena* were isolated from a landfill of domestic refuse [57]. There are currently 10 species within this genus, all of which have cellulolytic activities [57]. They are differentiated phenotypically by peptidoglycan determination, cell wall sugars, and utilization of various substrates [57]. *Cellulomonas* spp. has received considerable attention relative to other cellulolytic bacterial species [59]. Although *C. flavigena* is considered the type species for the *Cellulomonas* genus, most of the biochemical studies on hydrolases were done with *C. fimi*[57].

All species of *Cellulomonas* can grow at aerobic or microaerophilic conditions while *C. uda* and *C. fermentans* can also grow under strict anaerobic conditions [57]. With *C. fimi*, it was observed that glucose uptake was doubled under aerobic conditions relative to N_2 or H_2 growth conditions [57].

Under laboratory conditions, *C. fimi* and *C. flavigena* are typically grown under aerobic conditions in low-salt media (containing yeast extract or peptone at neutral pH) at a moderate temperature (30°C).

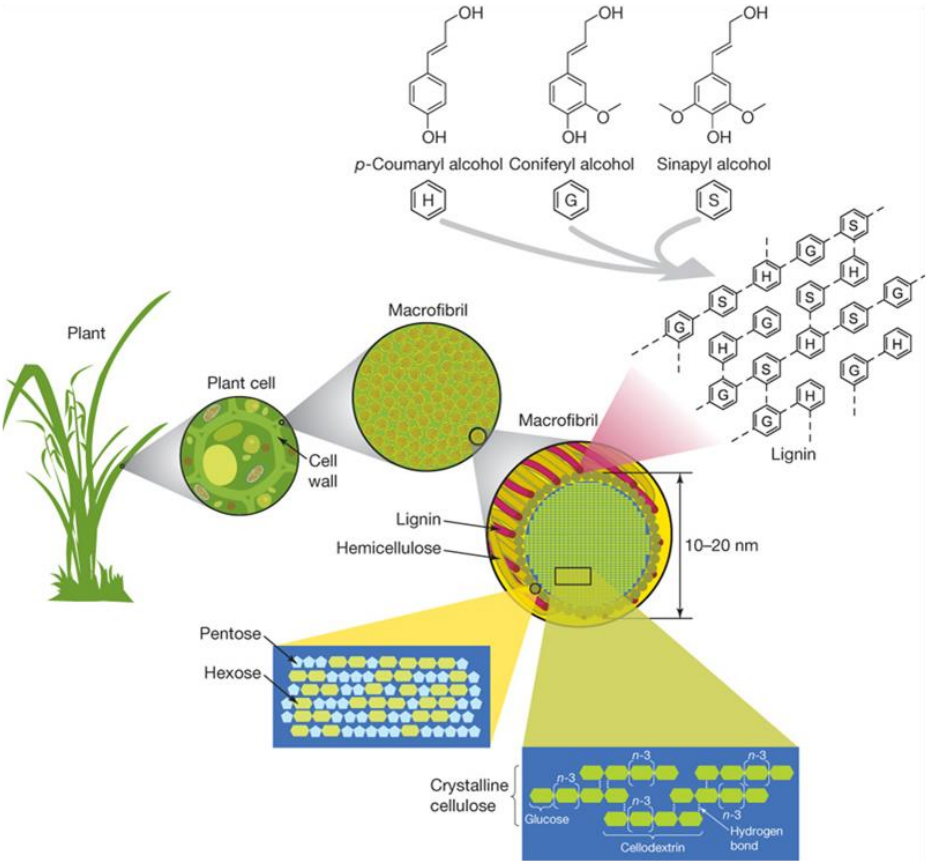
Recently, widespread genome sequencing has allowed the identification of putative glycoside hydrolases in numerous organisms, both eukaryotic and prokaryotic. Currently, *Cellulomonas fimi* (NC_015514.1, NCBI) and *C. flavigena* (NC_014151.1)[56] are the only two species of their genus to have their genomes fully sequenced. This has provided reason to revisit the two *Cellulomonas* spp. to identify and characterize hydrolases that were previously overlooked due to technical limitations.

C. fimi's CAZome[1] consists of 109 glycoside hydrolases, while *C. flavigena*[2] has 89 of these enzymes. Among these, only 14 have been characterized in *C. fimi* ATCC 484, and only 5 enzymes in *C. flavigena* CDBB-531 (a different strain to our group's strain). Characterized enzymes of interest from *C. fimi* are: 4 endo-1,4-glucanases (A, B, C, & D), 2 cellobiohydrolases (A & B), 3 xylanases (Cex, C, & D), and a β-glucosidase. The few enzymes that are studied from *C. flavigena* are: endo-1,4-glucanase B, cellobiohydrolase A, and a xylanase (fragment).

The overall goal of this project is the purification, biochemical and molecular analysis of newly identified cellulases in *Cellulomonas fimi* and *C. flavigena*. This will hopefully identify enzymes with novel or superior qualities that are beneficial for industrial applications.

[1] http://www.cazy.org/b1651.html Accessed on: **2013-07-23**
[2] http://www.cazy.org/b1248.html Accessed on: **2013-07-23**

Structure of lignocellulose.



EM Rubin *Nature* **454**, 841-845 (2008) doi:10.1038/nature07190

nature

Figure 2.1. Structure of lignocellulose. Reprinted with permission from Rubin EM. Genomics of Cellulosic Biofuels. *Nature*. Copyright (2008) Nature Publishing Group.

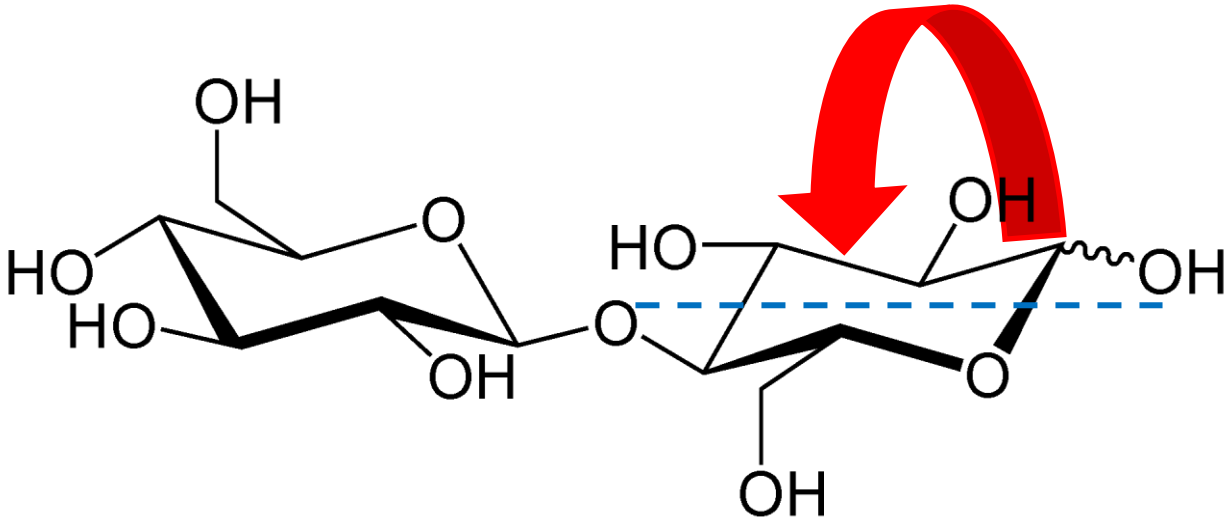


Figure 2.2. Cellobiose. Two glucose molecules with a two-fold screw axis along the C1-C4 plane.

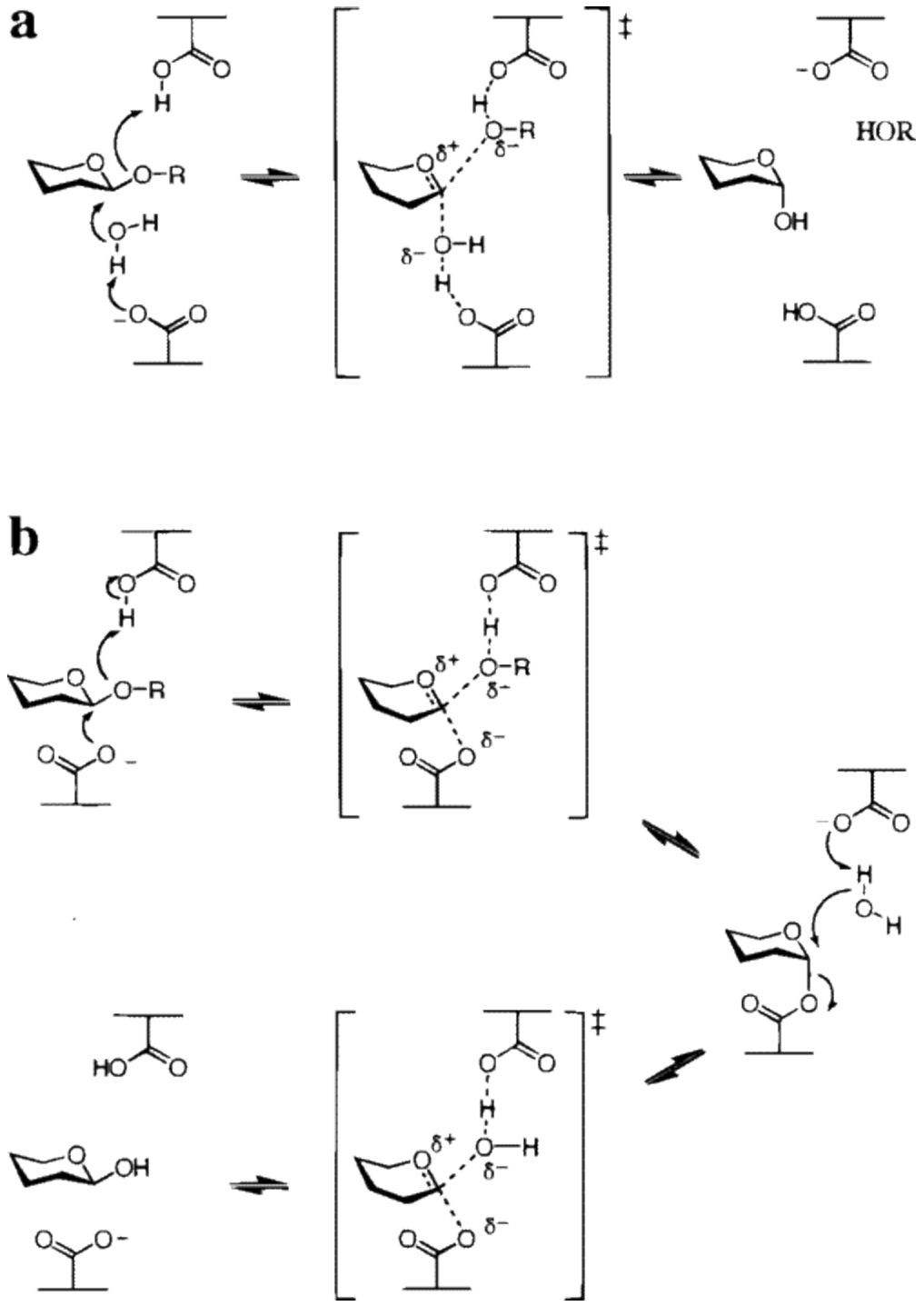


Figure 2.3. General mechanisms for (a) inverting and (b) retaining glycosidases. Reprinted with permission from Zechel DL and Withers SG. Glycosidase Mechanisms: Anatomy of a Finely Tuned Catalyst. *Accounts of Chemical Research*. Copyright (2000) American Chemical Society.

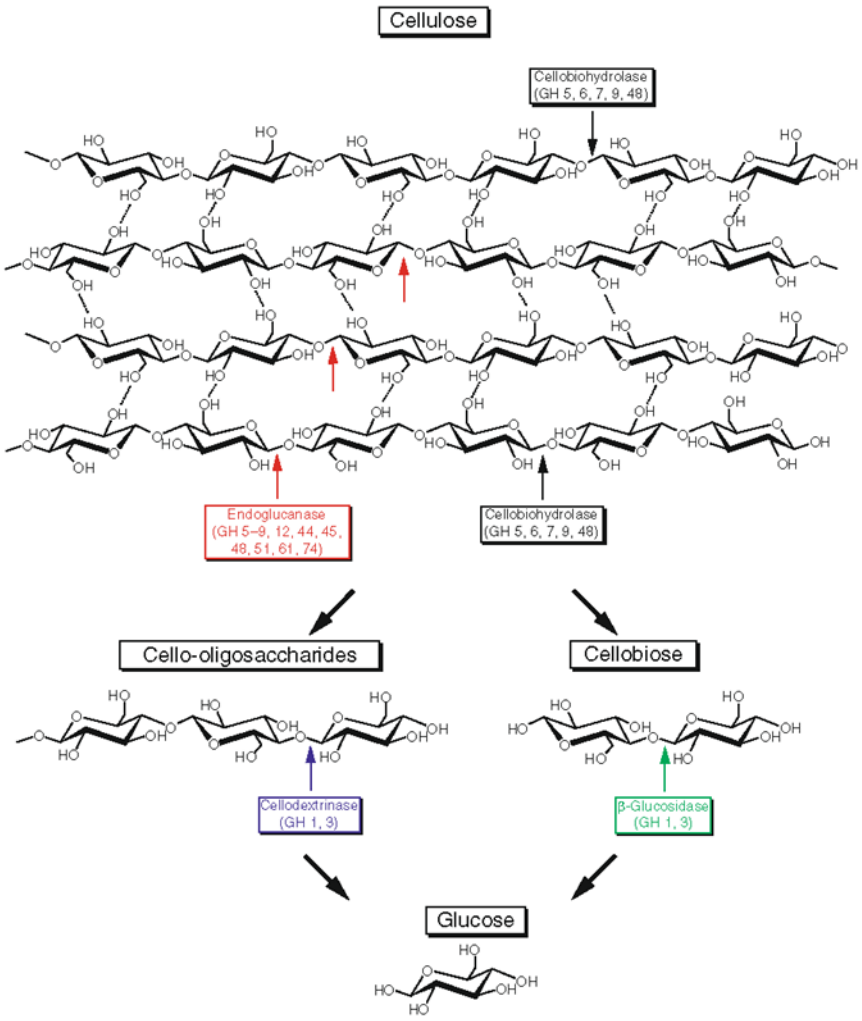


Figure 2.4. Enzyme classes of cellulases. Reprinted with permission from Yeoman CJ et al. Thermostable Enzymes as Biocatalysts in the Biofuel Industry. *Advances in applied microbiology.* Copyright (2010) Elsevier.

Chapter 3

Materials and Methods

3.1 Strains

Cellulomonas fimi ATCC 484 and *Cellulomonas flavigena* ATCC 482, both of which are type strains of their respective species, were source for genomic DNA for cloning purposes. AD202[3] is an OmpT protease-deficient E. coli strain (K-12 derivative) used for expression of recombinant proteins. Shuffle Express™ (New England Biolabs)[4] was used for protein expression that requires disulfide bond formation. Shuffle Express™ is an E. coli BL21 derivative that constitutively expresses cytoplasmic disulfide bond isomerase DsbC, which also serves as a chaperone for protein folding. In addition, it has gene deletions for glutaredoxin reductase (*Δgor*) and thioredoxin reductase (*ΔtrxB*) to prevent the reduction of cysteines in the cytoplasm. Occasionally, various cloning strains such as DH5α or XL10 were used for troubleshooting purposes.

3.2 Supernatant proteomics by LC-MS analysis

"Overnight (16 hrs)" bacterial cultures (1 Litre) of *C. fimi* and *C. flavigena* were grown in low-salt LB media (1.0% Tryptone; 0.5% Yeast extract; 0.1% NaCl (w/v)). This was used to inoculate fresh media supplemented with either carboxymethylcellulose (CMC) or oat-spelt xylan at 0.2% w/v. These were incubated at 30°C with shaking for 24 h. After a centrifugation step (8,000 x g for 10 min) to remove cellular matter and debris, the supernatant was pooled and collected. The supernatant proteins were concentrated by TCA precipitation (20% w/v, on ice 30 min, then collected by centrifugation 20,000 x g for 30 min). The proteins were then solubilized for 1D SDS-PAGE analysis. Bands were visualized after staining with Biosafe Coomassie blue (BioRad). For GelLC, each of the sample-containing lanes were cut into 25 equal slices, using a

[3] *F-, [araD139]$_{B/r}$, Δ(argF-lac)169, ompT1000::kan, λ-, flhD5301, Δ(fruK-yeiR)725(fruA25), relA1, rpsL150*(strR), *rbsR22, Δ(fimB-fimE)632(::IS1), deoC1*. http://cgsc.biology.yale.edu/Strain.php?ID=30158

[4] *fhuA2 [lon] ompT ahpC gal* λatt::pNEB3-r1-*cDsbC* (SpecR, *lacI^q) ΔtrxB sulA11* R(*mcr-73::miniTn10*--Tet^S)*2 [dcm] R(zgb-210::Tn10* --Tet^S) *endA1 Δgor Δ(mcrC-mrr)114::IS10*

gel cutter, from about 150 kDa to just above the dark smear at the bottom of the gel. Tryptic fragments were then examined by LC-MS analysis. The protein ID was performed using MASCOT, and proteins with a score over 50 are shown in Figures 4.4-4.7. Processing and analysis were performed by the Biological Mass Spectrometry and Glycomics Facility of the National Research Council-Institute of Biological Sciences (NRC-IBS).

3.3 Cloning and expression of putative cellulases from families GH 5, 6, 9

3.3.1 Genomic DNA extraction

C. fimi and *C. flavigena* gDNA was harvested from saturated 40 hour cultures grown in low-salt LB. The procedure follows the standard protocol for Gram-positive bacteria found in DNeasy Blood & Tissue kit (Qiagen).

3.3.2 Primer design and Polymerase Chain Reaction

The list of primers used in this study is found in Table 1. Primers were supplied by either Eurofins MWG Operon or Sigma-Aldrich. Constructs were designed to include an EcoRI-NdeI-His6[5] in the 5' end, and a Stop codon-SalI-HindIII site on the 3' end. The primers had complementary sequence between 20-28 base pairs. Primer design was optimized for % GC ratios and primer length, and also against hairpins and dimerization. Primer melting temperature (Tm) could not be optimized due to the intrinsic high GC ratio of the *Cellulomonas* genome. Thus, PCR amplification of the genes of interest had various annealing temperatures and conditions. For cloning purposes, Phusion High-Fidelity polymerase and GC buffer (New England Biolabs) were used and DMSO was frequently used as an additive to achieve successful amplification.

[5] Hexamer of histidine residues to serve as a tag for protein purification

3.3.3 Digestion, Ligation, and Transformation

Initially, the restriction enzymes used were NdeI and SalI but due to unforeseen internal restriction sites, NdeI and HindIII were used for cloning purposes instead. A 10-fold (enzyme loading) overdigestion was regularly followed as recommended by the manufacturer. Furthermore, an additional 3-fold increase was used when dealing with supercoiled plasmids. The reaction was allowed to proceed for 1-4 hours prior to enzyme inactivation at 65°C. Restriction enzymes were purchased from New England Biolabs.

After digestion, insert DNA was purified by MinElute PCR Purification Kit (Qiagen) while the cut vector backbone was gel extracted before being applied to a MinElute spin column. Ligation reactions were typically prepared in a 3:1 (Insert : Vector) molar ratio with approximately 30 fmol of vector. The reaction mixture consists of linear vector, insert DNA, ligase buffer, and T4 DNA ligase supplied by Fermentas. Transformations into bacterial strains were performed by either heat-shock or electroporation. Ligated DNA samples were purified and eluted with water to remove salts for electroporation. This step was not necessary for chemical transformation. Electrocompetent cells were thawed on ice (no longer than 20 min) and ice-cold DNA sample was added to the cells. The DNA-cell mixture was then transferred to an ice-cold electroporation cuvette (2 mm gap size). An electric charge was delivered to the cuvette using a Micropulser (Bio-Rad) set at Ec2 (2.5 kV, 1 pulse). SOC (2% w/v Tryptone, 0.5% w/v Yeast extract, 10 mM NaCl, 2.5 mM KCl, 10mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM D-glucose) media was immediately added and mixed with the DNA-cell suspension. The cells were allowed to recuperate at 37°C with shaking prior to plating on antibiotic-containing plates.

3.3.4 Preparation of Competent Cells

Chemical competent

An overnight culture of an *E. coli* strain is prepared in 5 mL LB-Miller broth at 37°C. This is used to inoculate a 100 mL flask of fresh LB-Miller broth at 1:100 dilution. The culture is incubated again at 37°C until it reaches an early-log growth phase indicated by an OD600 between 0.2-0.4. The cells are chilled on ice for 10 min then pelleted by centrifugation (5,000 g for 5 min). The supernatant media is removed and the cells are washed with sterile ice-cold 0.1 M $CaCl_2$ and left on ice for at least 30 minutes. The cells are subjected to a second centrifugation step (5,000 g for 5 min) and then resuspended in ≤3 mL of 0.1 M $CaCl_2$. Glycerol is added to a final concentration of 16% (v/v) to allow long term storage in the freezer. Cells are divided into 50 µL aliquots for each transformation experiment.

Electrocompetent

Similarly, an overnight culture of an *E. coli* strain is prepared in 10 mL LB-Miller or Super Optimal Broth (SOB, Beckton-Dickinson). This is used to inoculate a 500 mL flask of LB-Miller or SOB broth at 1:100 dilution. The culture is incubated at 37°C until it reaches midexponential phase indicated by an OD600 of 0.5-0.7. The cultures are placed on ice for at least 15 minutes and it is crucial for the cells to stay ice-cold from this point on. The cells are centrifuged (5000 g for 10 min) and then washed with 200 mL of chilled sterile 10% (v/v) glycerol. This step is repeated at least 3 times to remove the salts from the media. The cells are eventually resuspended in 2 mL of 10% (v/v) glycerol, which leads to an OD_{600} of 200-250. The high density of cells is crucial to successful electroporation. The cells are divided into 50 µL aliquots and stored at -80°C.

3.3.5 Cloning verification

To verify successful ligation of PCR products into digested plasmid vectors, colony PCR was performed. Colonies were picked off the plates using sterile wooden toothpicks and swirled into sterile water. This dilute cell suspension was used as the template for the PCR reaction with either Taq or Phusion polymerase. Cycling conditions were adjusted to accommodate the change in enzyme. Occasionally, the purified plasmid is further subjected to double digest and diagnostically run on an agarose gel for a more conclusive analysis.

3.3.6 Gene sequencing

Plasmids were prepared from clones using standard protocol from Qiaprep Spin MiniPrep kit (Qiagen). To accommodate for low yields, the "low-medium copy" plasmid protocol was followed. Purified plasmids containing inserts of the appropriate size were sent for gene sequencing using specific sequencing primers (Table 3) by either the NRC (Ottawa) or The Centre for Applied Genomics (TCAG, Toronto).

3.3.7 Small-scale expression of recombinant proteins

E. coli expression strain AD202 was used for all expression trials unless otherwise stated. Strains containing cellulase constructs were grown overnight in LB media (1.0% w/v Tryptone, 0.5% w/v Yeast extract, 1.0% w/v NaCl) with 150 µg/mL Ampicillin (Amp150) as a selective marker. A fresh 20 or 50 mL LB-Amp150 was inoculated 1:100 and grown at 37°C with shaking in beveled flasks. These were grown to log-phase (OD_{600} 0.5-0.7) and the shaking was discontinued while the incubator was changed to the desired induction temperature. After the incubator has achieved its new setting, the flasks were allowed to equilibriate with its new temperature prior to adding sterile isopropyl β-D-1-thiogalactopyranoside (IPTG). The final concentration of IPTG

was either 0, 0.1, or 0.5 mM unless otherwise specified. Induction and protein expression was allowed to go on for 3-15 hours. The cells were then harvested by centrifugation (8000 g for 10 min at 4°C). The cell pellet was then frozen at -20°C until needed. The cell pellet was then thawed and resuspended in binding buffer (20 mM $NaPO_4$, 20 mM Imidazole, 500 mM NaCl, pH 7.4 @ 25°C) to a final concentration of 10% w/v. The cell suspension was then sonicated on ice for cell lysis. A 20 µL fraction was set aside to represent the whole cell extract (WCE). The rest was centrifuged for 15 min at 18,000 g at 4°C to remove cellular debris. The supernatant fraction containing the soluble proteins was set aside for SDS-PAGE analysis.

3.4 SDS-PAGE analysis

The whole cell extract and the supernatant fractions were diluted 1:4 in water to prevent overloading. The diluted samples were then added to the 6x sample buffer (0.375 M Tris pH 6.8, 10% w/v SDS, 50% v/v glycerol, 10% w/v β-mercaptoethanol, 0.025% w/v bromophenol blue) and heated for 5 min at 95°C before being loaded onto 0.75 mm-thick 10% acrylamide gels. Standard discontinuous buffer system by Laemmli[60] was employed to separate proteins for analysis. The proteins were visualized by InVision in-gel His stain (Invitrogen) followed by Coomassie R-250 stain. Step-by-step procedure for the His-stain followed the manufacturer's recommendation with slight modifications to the volumes used. Occasionally, the His-stain is skipped and only Coomassie blue staining was employed for visualization. Gels were incubated with the stain and subjected to microwave treatment to accelerate the process. Once sufficiently stained, the background stain was removed using a destain solution (10% acetic acid, 50% EtOH) and also heated by microwave treatment until satisfactory.

3.5 Congo Red analysis

Agar plates containing 1.0% (w/v) CMC and 50 mM KPO4 (pH 7 at 25°C) was used to screen and qualitatively detect endoglucanase activity. Filter-paper discs were soaked with enzyme-containing solutions (of various purity) were placed on top of the agar and allowed to incubate for a few hours or overnight at 37°C. After the incubation, the filter paper discs are removed and the petri dish is flooded with congo red solution (0.1% w/v) (Figure S5). After soaking for at least 30 minutes, the solution is poured off the plate and quickly rinsed with deionized water to remove residual solution. The petri dish is then flooded with 1 M NaCl for 5 minutes twice. Optionally, the color contrast can be increased by soaking the plate in 5% (v/v) acetic acid solution because the reagent is sensitive to pH. The congo red reagent was purchased from J.T. Baker (F788). The sodium chloride solution interferes with the hydrogen bonds that formed between the cellulose and the azo & amine groups of congo red. It is possible to completely wash off the congo red stain but it preferentially decreases the stain at sites of hydrolysis. At those sites, the congo red molecule is more readily liberated off shorter chains of cellulose.

Alternatively, the agar plates can contain LB-Miller, Cellulose, and IPTG to induce growth of strains that contain plasmid constructs. After the colonies have grown to sufficient size, the colonies are smeared away using a gloved finger and then washed away with deionized water. The congo red staining follows the aforementioned procedure from this point forward. Colonies that expressed active cellulases are able to produce zones of hydrolysis without an added lysis step. This method works for both secreted and unsecreted recombinant cellulases.

3.6 Expression and purification of recombinant proteins

3.6.1 Celf_1230 (from here on known also as Cel6C)

Celf_1230 (aa51-415) was codon-optimized for *E. coli* expression and subsequently gene synthesized by GeneArt® (Life Technologies). The construct featured EcoRI-NdeI restriction sites at the 5' end and a tandem stop codon followed by HindIII-XbaI restriction sites at the 3' end. This construct proved difficult to express and purify, and thus required modifications. The original construct was modified by PCR and DNA techniques to extend the N- and C-terminus that contained aa31-433. This resulting construct was subcloned into pVEK-06 (see supplementary information) and renamed as Cfi-28, which features an N-terminal His-tag fusion protein. The nucleotide sequence was verified prior to downstream experiments. The primers used to lengthen the original construct can be found in Table 1. Cfi-28 was transformed into Shuffle Express™ (New England Biolabs) for protein production. Another construct was made to extend only the C-terminus of the original construct and produce residues aa51-433. This plasmid is renamed as Cfi-29, which produces soluble active protein but was not thoroughly investigated.

E. coli Shuffle Express™ strain containing Cfi-28 was inoculated into 20 mL of 2YT broth (Sigma) containing Amp150 and grown overnight at 30°C with shaking. The saturated bacterial culture was then diluted into a fresh 1 L of 2YT broth containing Amp150. This was incubated at 30°C for a few hours until the OD600 was between 0.6-0.8 indicating mid-log growth phase. The culture was induced with a final concentration of 0.5 mM IPTG and left to express overnight at 30°C. The cells were harvested by centrifugation (8000g for 10 min) and frozen at -20°C and then stored at -80°C until needed. For lysis and purification, the cells were thawed in the

refrigerator and resuspended in 10% w/v of binding buffer (20 mM NaPO4, 20 mM Imidazole, 500 mM NaCl, pH 7.4 at 25°C) containing cOmplete EDTA-free Protease Inhibitor Cocktail (Roche). The cell suspension was then subjected to lysis by using an Emulsiflex C5 (Avestin) with >17,000 psi. The lysate was centrifuged (18,000 g for 20 min) to remove debris. The supernatant fraction of this was subjected to further centrifugation (44,000 rpm for 1 hour) to remove material that clogs the chromatography columns. The clarified cell lysate was then injected into a pre-equilibrated HisTrap column (GE Healthcare) using a Bio-Rad FPLC equipment. Afterwards, the protein bound to the Ni-NTA resin was eluted using a linear gradient from 0-60% of elution buffer (20 mM NaPO4, 500 mM Imidazole, 500 mM NaCl, pH 7.4 at 25°C) over 50 mL. Selected fractions were pooled and placed in an Amicon®-Ultra 10K (Millipore) for concentration and buffer exchange to remove the contaminating imidazole. The concentrated protein sample was aliquoted in 20 µL amounts and frozen with liquid nitrogen for storage.

3.6.2 Celf_3184 (Cel6A)

The Celf_3184 construct was created by Dr. Anthony Clarke and his group at the University of Guelph. This construct featured the full-length protein and was inserted into a pET-30a(+), which fused a C-terminal His-tag. For more details, please see the corresponding reference [64]. Also, the expression and purification of the protein followed their methods with minor changes. For my purposes, the immobilized metal affinity chromatography (IMAC) was sufficiently pure without the need of anion exchange chromatography.

3.7 Assays

3.7.1 HBAH reducing sugar assay

The increase of reducing ends produced by glycoside hydrolases was measured using ρ-hydroxybenzoic acid hydrazide (HBAH) as a reagent developed by Lever[66]. Several changes were made to his original methods and the following procedure contains these modifications. The alkali solution (0.5 M NaOH, 0.02 M EDTA pH 8, and 0.02 M $CaCl_2$) was prepared fresh daily in advance from stock solutions and then kept chilled on ice. The HBAH reagent is weighed and added to the alkali solution to a final concentration of 1.0% w/v. This solution is left in the refrigerator on a stirplate until it is fully dissolved. The solution is then dispensed in 3 mL volumes using a repeat pipetter into pre-chilled glass test tubes. The solution-containing tubes are kept on ice and must be used within a few hours once the reagent is dissolved. To stop a time point, 500 µL of the assay is added to the 3 mL of HBAH solution. These sample-HBAH mixtures are kept on ice bath until all time points are completed. A vigorously boiling-water bath is used to heat the solutions for 6 min then promptly returned to ice to stop the heating process. A 200 µL aliquot of these samples are transferred to a 96-well plate for colorimetric quantification at a wavelength of 410 nm. Aside from the Michaelis-Menten experiments, all HBAH assays were performed on 1.0% CMC (medium viscosity, Sigma C-4888). Standard curves of glucose were prepared to determine the amount of reducing sugars released. One unit of activity (U) is defined as the number of micromoles of glucose equivalents released per minute. The typical amount of enzyme used was approximately 0.667 µg/mL of assay.

3.8 Preparation of PASC (Phosphoric acid-swollen cellulose)

Avicel PH-102 (2.5 g) is slowly added to 50 mL of cold neat H_3PO_4 (85% w/v ortho-phosphoric acid) and left on a stirplate in the refrigerator for 1 hour. The cold temperature is necessary to decrease acid hydrolysis of the cellulose. The suspension is poured into 1 L of ice-cold water and left to stir for 30 min. Pour off the water and repeat rinse with 1 L for 2 more times for 10 min each. Next, wash the swollen cellulose with cold 1% (w/v) of NaHCO3 three times to neutralize the solution. Centrifugation (8000 g for 15 min) was used to sediment the cellulose in between washes. Wash (3x) the cellulose with 500 mL of ice-cold water again to remove the sodium bicarbonate solutes. The slurry is collected by centrifugation and homogenized by an overhead stirrer (Wheaton) (Figure S3). Hardy lumps of avicel are removed and discarded. Sodium azide is added to a final concentration of approximately 5 mM as a preservative and then the suspension is kept at 4°C. The concentration (% w/w) of the substrate is determined comparing the dry weight over the wet weight after placing an aliquot in a 37°C incubator overnight.

3.9 Differential Scanning Calorimetry

In order to determine the melting temperature of the enzymes, a differential scanning calorimetry (VP-DSC) machine from MicroCal was used. The enzymes were diluted to 0.5-1.0 mg/mL in buffer (25 mM MOPS (or $NaPO_4$), 50 mM NaCl, pH 7 at 25°C) and degassed using a vacuum pump. The DSC was equilibrated with an overnight buffer-buffer scans to create a baseline. During a downscan, the buffer is quickly removed from one of the vessels in the DSC and replaced with enzyme-containing sample. After the measurements are complete, the enzyme sample is removed and kept aside to measure the actual protein concentration within the vessel. Baseline determination was done automatically by the Origin program for the DSC machine.

3.10 Thin-layer Chromatography

Determination of hydrolysis products was done by separating the molecules by thin-layer chromatography (TLC). Using a classical silica gel 60 plate on a glass backing (Merck), there was a distinct separation of cellodextrins (G1-G5 molecules) as observed with the standard that had a concentration of 2 mg/mL of each molecule. The hydrolysis of cellopentaose (G5) was done with or without Cel6C in a 60°C water bath for approximately one hour. The sample was then dried onto the silica plate prior to exposing it to the liquid phase solvent. The solvent (Ethyl acetate : MeOH : Water : Acetic acid in a 4:2:1:0.1 ratio by volume) was placed in a glass jar and sealed to maintain a saturated air environment. Finally, the carbohydrates on the TLC plate are visualized by dipping the plate in the staining solution (5% v/v H_2SO_4 in ethanol), air-dried, then charred over a hot plate until bands are visible.

3.11 Dye-conjugated cellulosic substrates

Various substrates have to be used to give a complete picture of cellulolytic activity since cellulases are diverse in nature. Azurine-crosslinked-substrates (Megazyme) (AZCL) like AZCL-Hydroxyethylcellulolose (AZCL-HEC), AZCL-Avicel, and AZCL-Xylan was used to determine substrate specificity in a colorimetric assay. These substrates are insoluble and tend to adhere to tubes and pipets making them problematic to handle. The powdered substrate is directly weighed in microfuge tubes to approximately 10 mg. Buffer (25 mM MOPS, 50 mM NaCl, pH 7 at 25°C) and water was added to a final volume of 1 mL then incubating it in a water bath for preheating. The enzyme or buffer is added and then mixed by inverting the tubes. The reaction is allowed to proceed for 1-4 hours depending on the enzyme activity. The assay is chilled on ice to stop the

enzymatic action then centrifuged (5000 g for 10 min) to pellet the insoluble material and 100-200 µL is moved to a 96-well plate for quantification and read at 590 nm.

3.12 2,4-dinitrophenyl-β-D-cellobioside

The soluble substrate 2,4-dinitrophenyl-β-D-cellobioside (dNPC) was generously provided by Dr. Steve Withers from the University of British Columbia. It has been shown to be an excellent substrate for screening for both types of cellulases, endoglucanases and cellobiohydrolases [63]. This characteristic can be attributed to the modified analogue (2,4-dinitrophenol) which makes it a better leaving group relative to the 4-nitrophenol at neutral pH [67]. The reaction conditions consisted of 1 mM of substrate, buffer (20 mM MOPS, 50 mM NaCl, pH 7 at 25°C). The reaction was initiated upon addition of the enzyme in a temperature controlled plate reader. The molar extinction coefficient of dNPC is 0.9384 $mM^{-1}mm^{-1}$, which was derived experimentally using a nanodrop spectrophotometer. The dNPC substrate can be detected at 300nm, while the 2,4-dinitrophenol product is detected at 360nm. The kinetic assays are done on standard polystyrene flat-bottomed 96-well plates (Greiner Bio-one). The enzyme concentration for Cel6A and Cel6C was 4.61 and 6.22 µg/mL of assay, respectively.

3.13 Determination of the Michaelis constant

The Michaelis constant (K_M) was determined using the HBAH reducing sugar assay and using CMC as the substrate. A range of substrate concentrations was used from 0.05 to 1.4% at pH 7 using 20 mM MOPS with 50 mM NaCl as the buffer. The amount of enzyme used was 0.05 µg per mL of assay to minimize substrate depletion. The K_M was determined by nonlinear regression of the Michaelis-Menten plot based on the substrate inhibition (uncompetitive) equation using the enzyme kinetics module of SigmaPlot (SysStat Software Inc).

3.14 Ball-milled Biomass substrates

Time of Flight-Secondary Ion Mass Spectrometry (ToF-SIMS)

In order to better characterize Cel6C, natural substrates was used to detect hydrolytic activity. With the help of Drs. Emma Master and Marie Couturier of the University of Toronto, I was able to perform analysis on ball-milled biomass substrates (specifically aspen and spruce, which are hardwood and softwood respectively). Surface analysis of these substrates was performed by ToF-SIMS. This technique employs bombardment of barium ions to the sample which liberates secondary ions from the sample. The molecules containing the secondary ions project towards the detector and their time of flight is a derivative of their mass. Unhydrolyzed substrates would have spectra that contain more polysaccharide-known peaks, while hydrolyzed substrates would have a higher proportion of lignin-peaks since the celluloses have been degraded. Analysis of this data was done using Principal Component Analysis (PCA) that allows pattern recognition for complex spectra. For more details on this technique and its application for investigating cellulases, please see the following reference from Goacher, Jeremic, and Master [68]. A reducing sugar assay accompanied the ToF-SIMS result to quickly check for hydrolytic activity on the biomass substrate. The reducing sugar assay used here was the dinitrosalicylic acid (DNS) assay, details of which will not be provided.

3.15 Enzyme concentrations

Enzyme concentrations were determined by A_{280} on a nanodrop spectrophotometer. The theoretical extinction coefficient (M^{-1} cm^{-1}) used for Cel6A and Cel6C are 107,745 and 108,205, respectively. The coefficients used are with the assumption that all pairs of Cys residues form

disulfide bonds. The predicted molecular weights are 48.25 kDa and 45.45 kDa respectively. Both parameters were determined by ExPASy ProtParam[69].

Chapter 4

Results

4.1 Bioinformatics, cloning, and test expressions

4.1.1 Bioinformatics

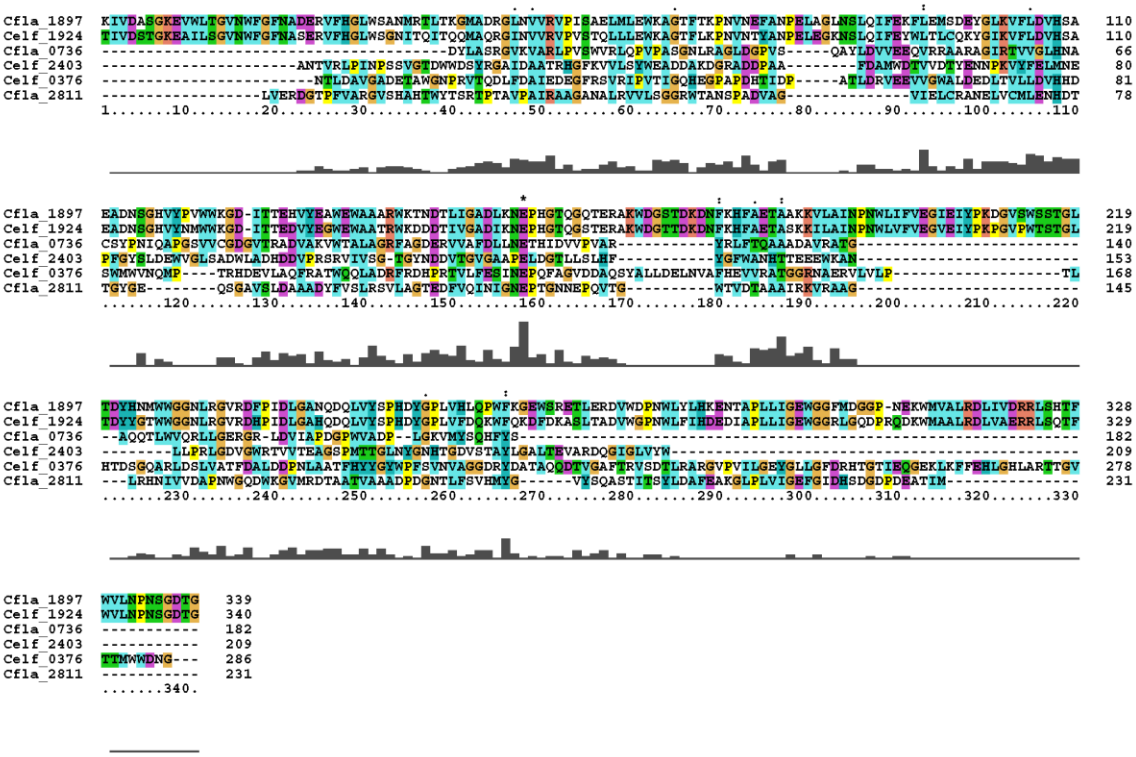


Figure 4.1 Amino acid sequence alignment between the catalytic domains of Family 5 Glycoside Hydrolases of Cellulomonas fimi and Cellulomonas flavigena. Quality curve, depicting the alignment's conservation score, is shown below each row of amino acid sequence. The alignment was performed using ClustalX2 software.

The alignment and comparison of the catalytic domains of GH 5, 6, and 9 putative cellulases from *C. fimi* and *C. flavigena* are shown in Figures 5.1-5.3, respectively. Only Celf_1924 (Cel5A/CenD) has been characterized in detail [61]. Cfla_1897 appears to be a close homolog in *C. flavigena* for Cel5A.

The GH 6 family consists of two distinct groups and two outliers. One group (Celf_3184, Cfla_2912, Cfla_2913) contains putative endoglucanases since Celf_3184, also known as Cel6A,

has been characterized previously [62–64]. Cfla_2912 and Cfla_2913 have closely related amino acid sequence and their position in the *C. flavigena* genome suggest this arose from a gene duplication event. The second group (Celf_1925, Cfla_1986) are cellobiohydrolases since Celf_1925, also known as Cel6B, has been characterized previously [45,65]. Celf_1230 and Celf_0233 are the two outliers with unknown function and substrate specificity.

In the GH 9 family, Celf_0019 (Cel9A/CenB) and Celf_1537 (Cel9B/CenC) are two *C. fimi* enzymes that have been characterized as endoglucanases. There appears to be substantial sequence similarities among the Cel9A-like group and the Cel9B-like group. However, the various differences in amino acid residues could lead to significant kinetic differences. It also begs the question of the physiological importance of the redundancy of Cel9A-like enzymes.

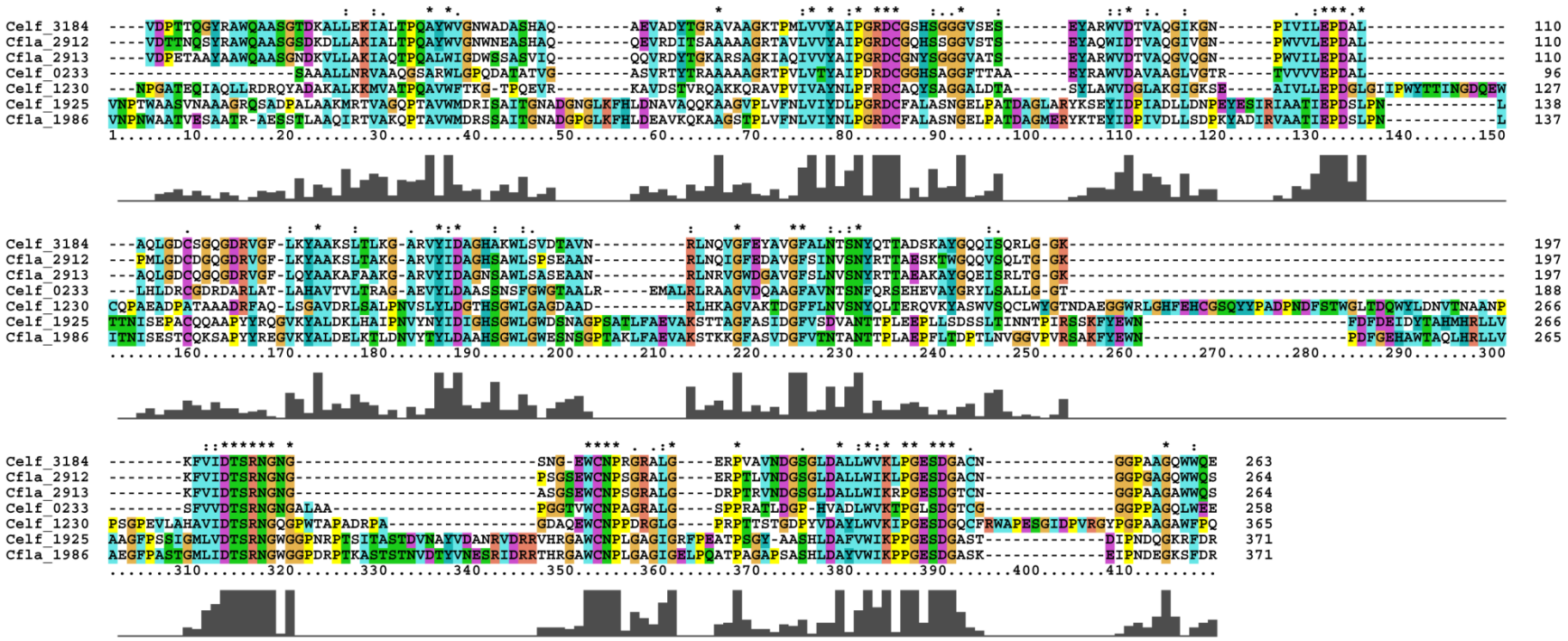


Figure 4.2 Amino acid sequence alignment between the catalytic domains of Family 6 Glycoside Hydrolases of *Cellulomonas fimi* and *Cellulomonas flavigena*. Quality curve is shown below each row of amino acid sequence. The alignment was performed using ClustalX2 software.

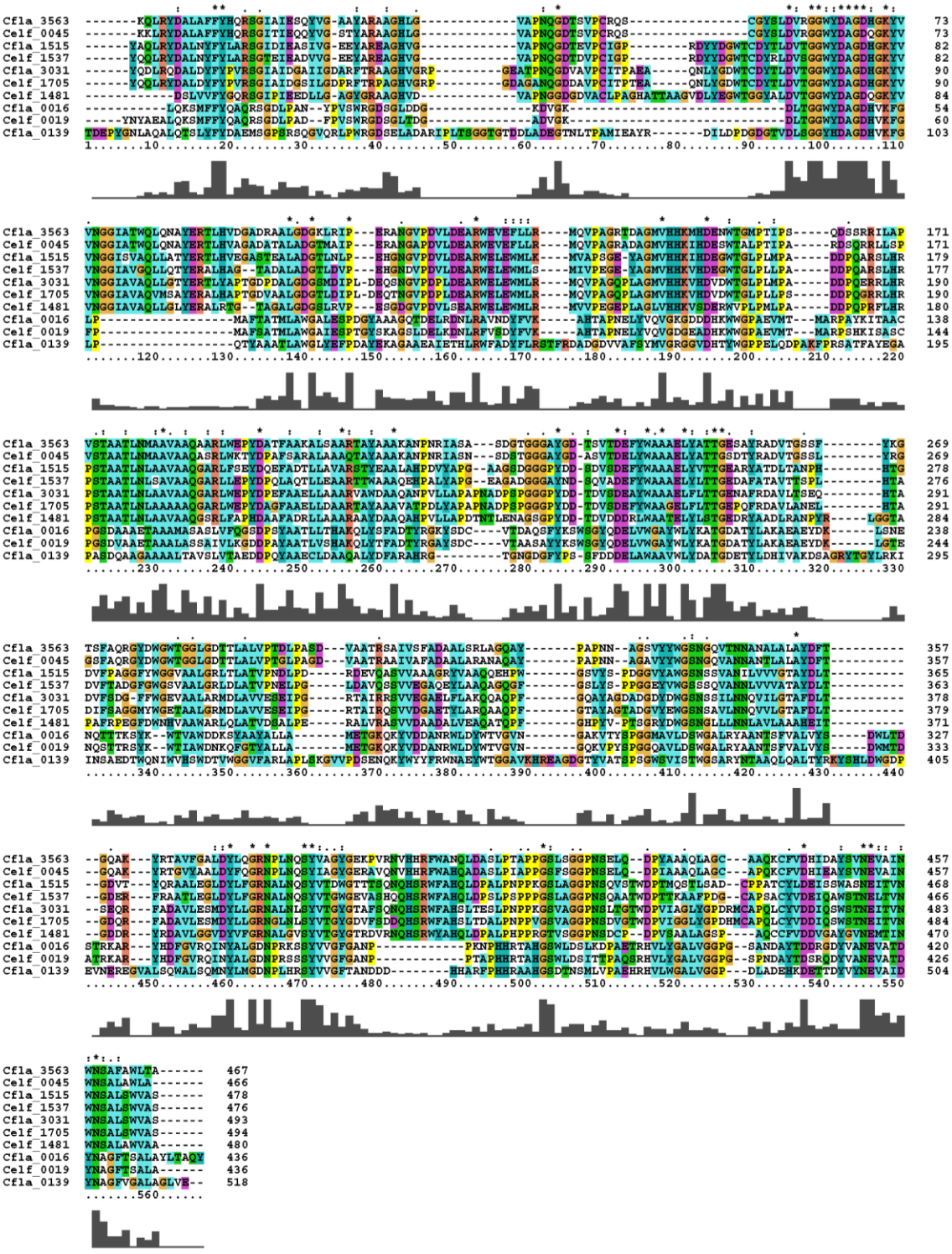


Figure 4.3 Amino acid sequence alignment between the catalytic domains of Family 9 Glycoside Hydrolases of *C. fimi* and *C. flavigena*. Quality curve is shown below each row of amino acid sequence. The alignment was performed using ClustalX2 software.

4.1.2 Proteomics of secreted proteins

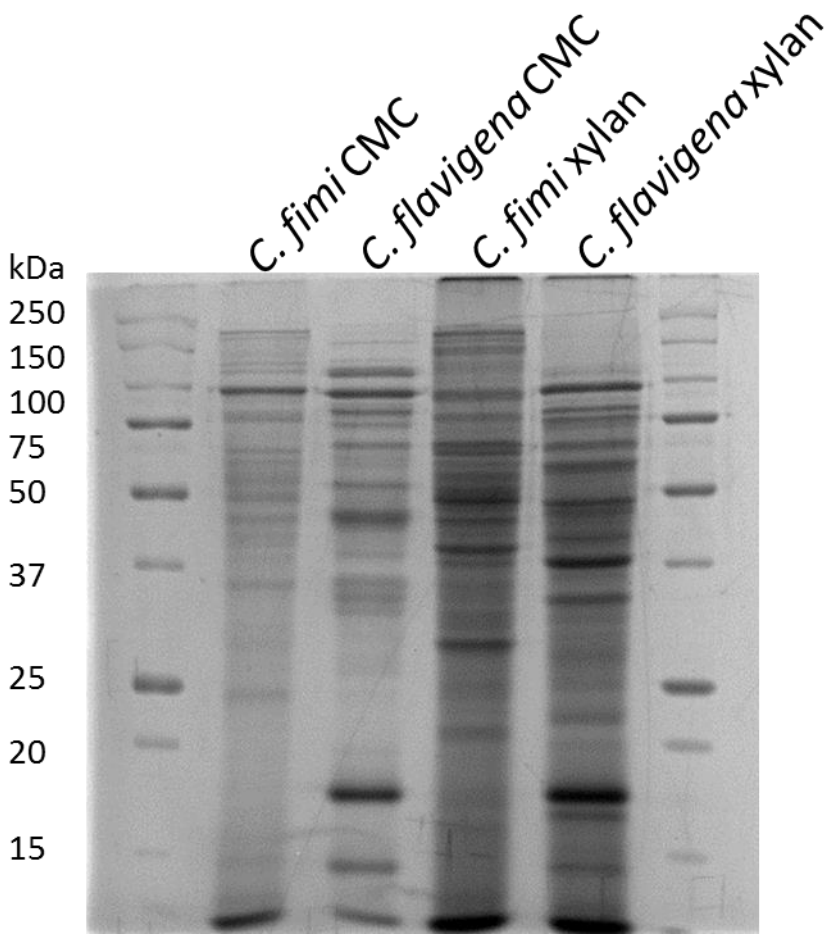


Figure 4.4. SDS-PAGE analysis of the secretome of *C. fimi* and *C. flavigena* under CMC or xylan. Gel was visualized by staining with Biosafe Coomassie blue (Bio-Rad). Preparation of the secretome for proteomic analysis. Supernatants were precipitated with 20% w/v TCA, and the proteins were then solubilized for 1D SDS-PAGE analysis (Figure 5.4)

The following data (Figures 4.4-4.7) are provided by Dr. Wakarchuk and the Biological Mass Spectrometry and Glycomics Facility of the National Research Council-Institute of Biological Sciences (NRC-IBS).

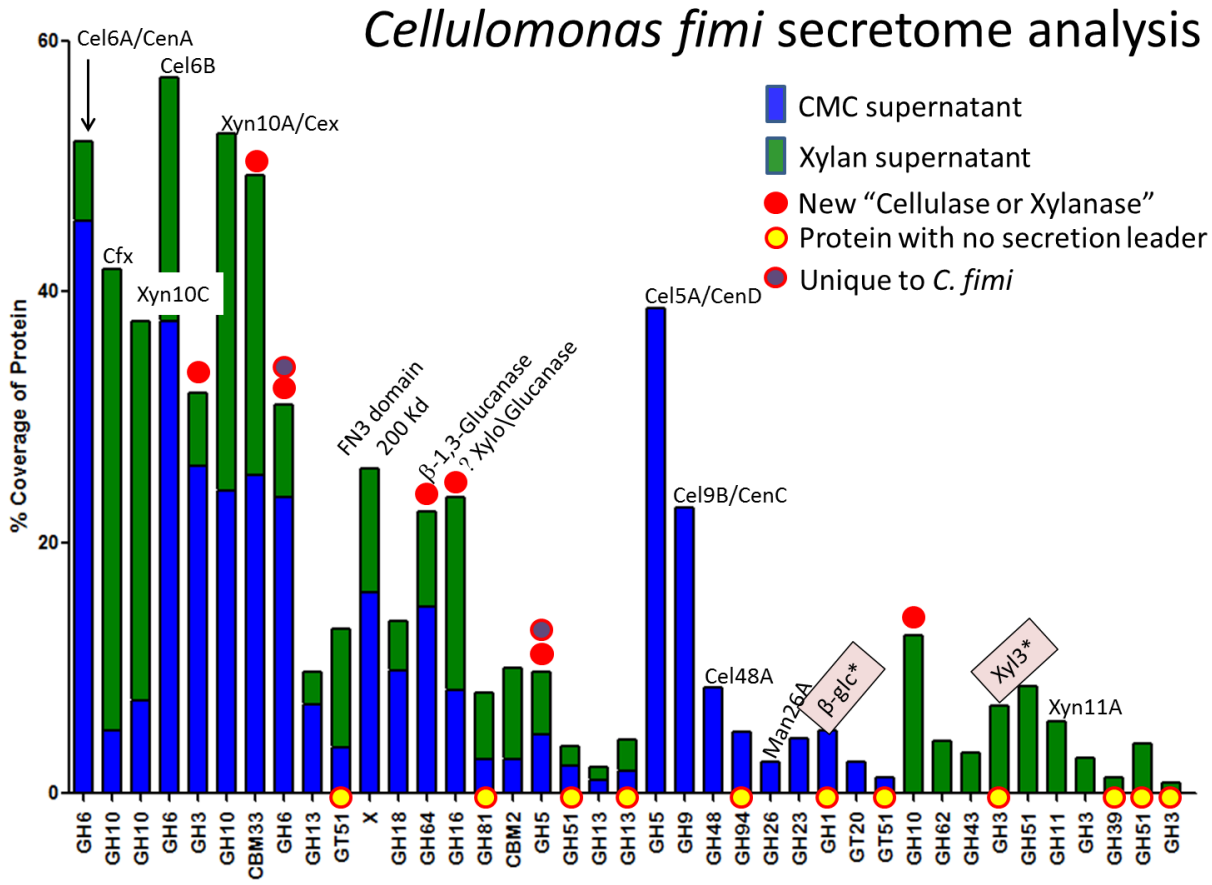


Figure 4.5. Secretome analysis by mass-spectrometry of *C. fimi* cellulases or xylanases.

Proteins with a MASCOT score of less than 50 are not shown. The % coverage of protein is a correlation to the relative amount of protein. Proteins that are unique to *C. fimi* are qualitatively determined by amino acid sequence. Proteins that were identified in both supernatant samples are represented by stacked bars.

Several proteins that had high % coverage are labeled in Figure 5.5 as these proteins matched known characterized proteins from their corresponding literature. Those experiments involved blind cloning from screening recombinant libraries and from work involving purified enzymes from the host organism. Thus, it goes to show that the highly expressed cellulases were more easily detected with those techniques. However, a few "new" glycoside hydrolases were overlooked even though they have a relatively modest expression level.

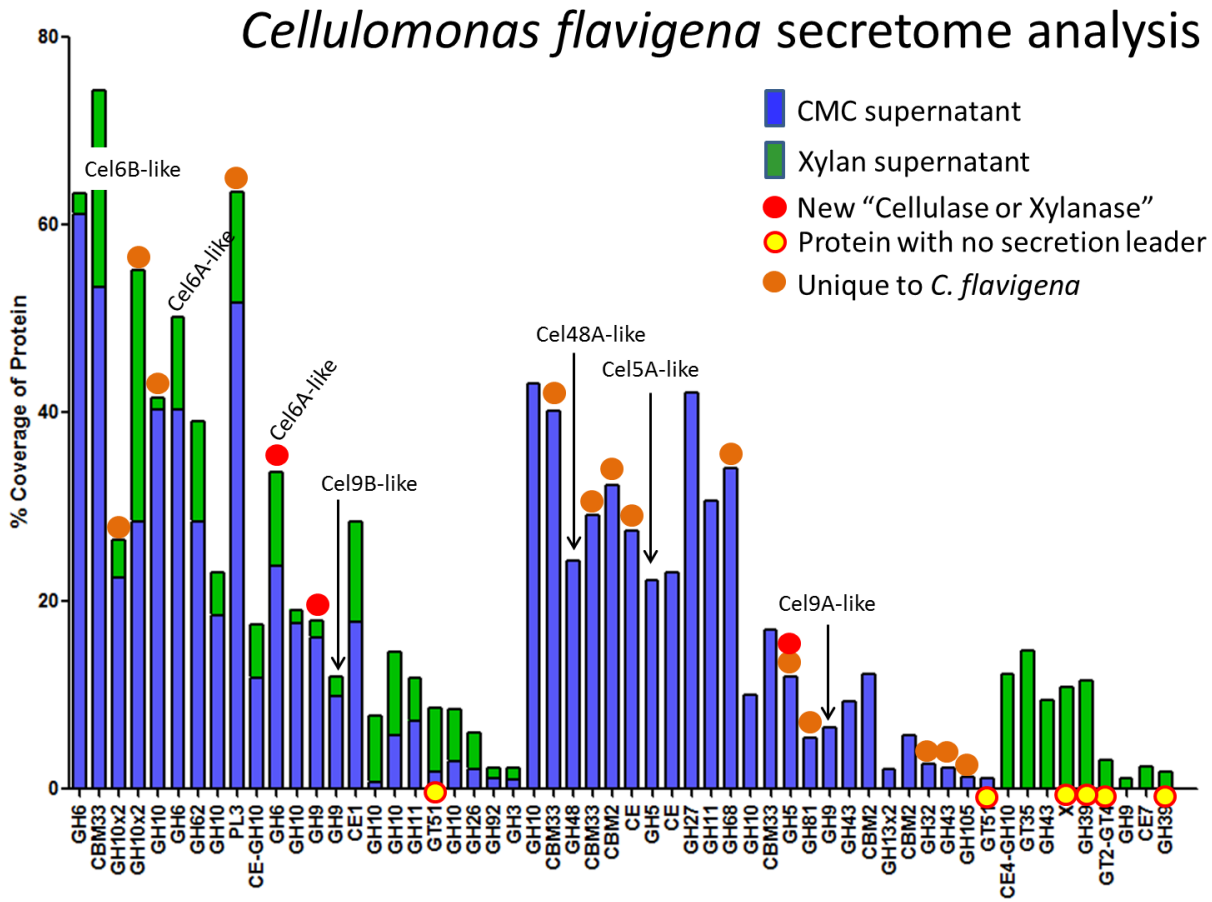


Figure 4.6. Secretome analysis by mass-spectrometry of *C. flavigena* cellulases or xylanases. Proteins with a MASCOT score of less than 50 are not shown. The % coverage of protein is a correlation to the relative amount of protein. Proteins that are unique to *C. flavigena* are qualitatively determined by amino acid sequence. Proteins that were identified in both supernatant samples are represented by stacked bars.

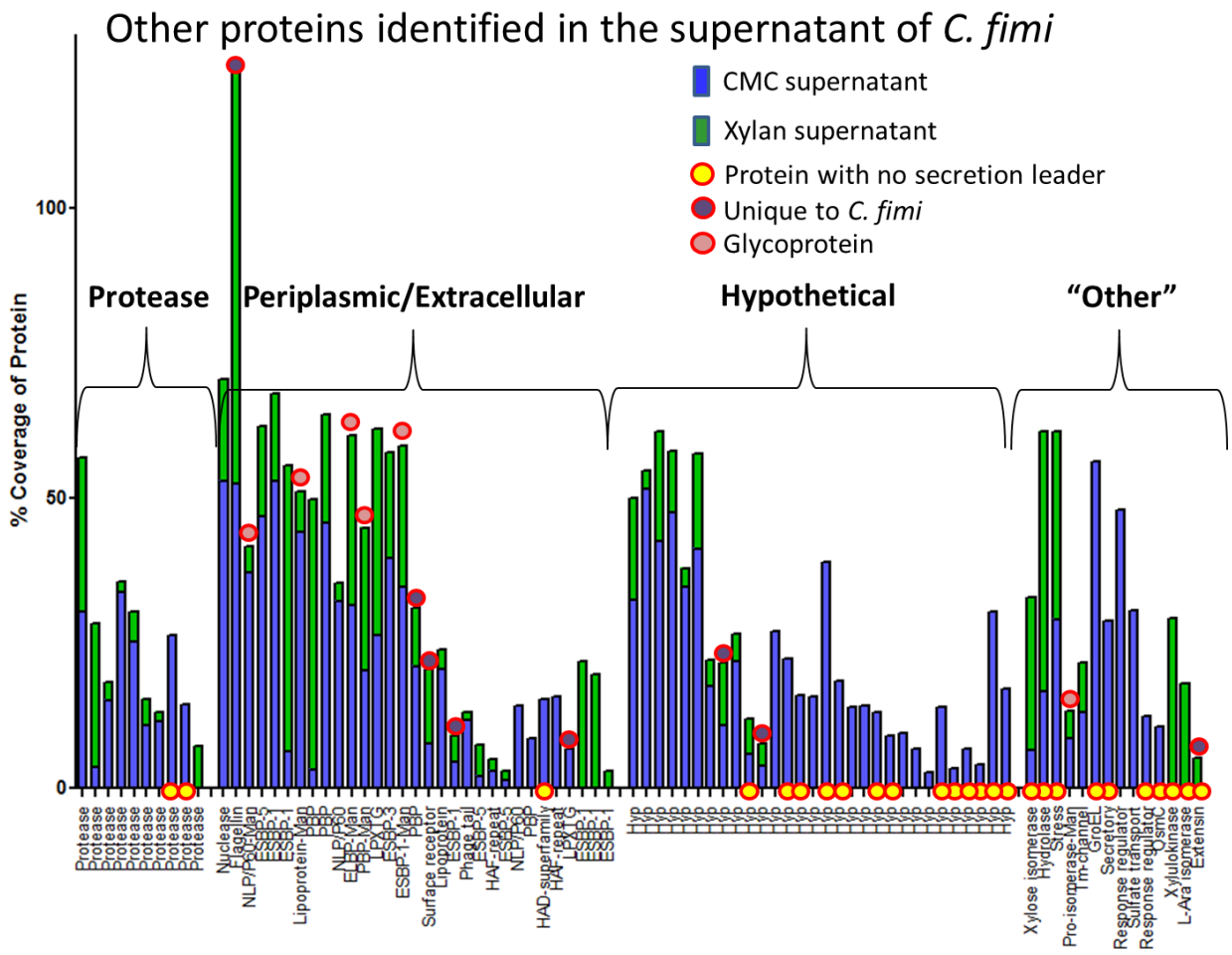
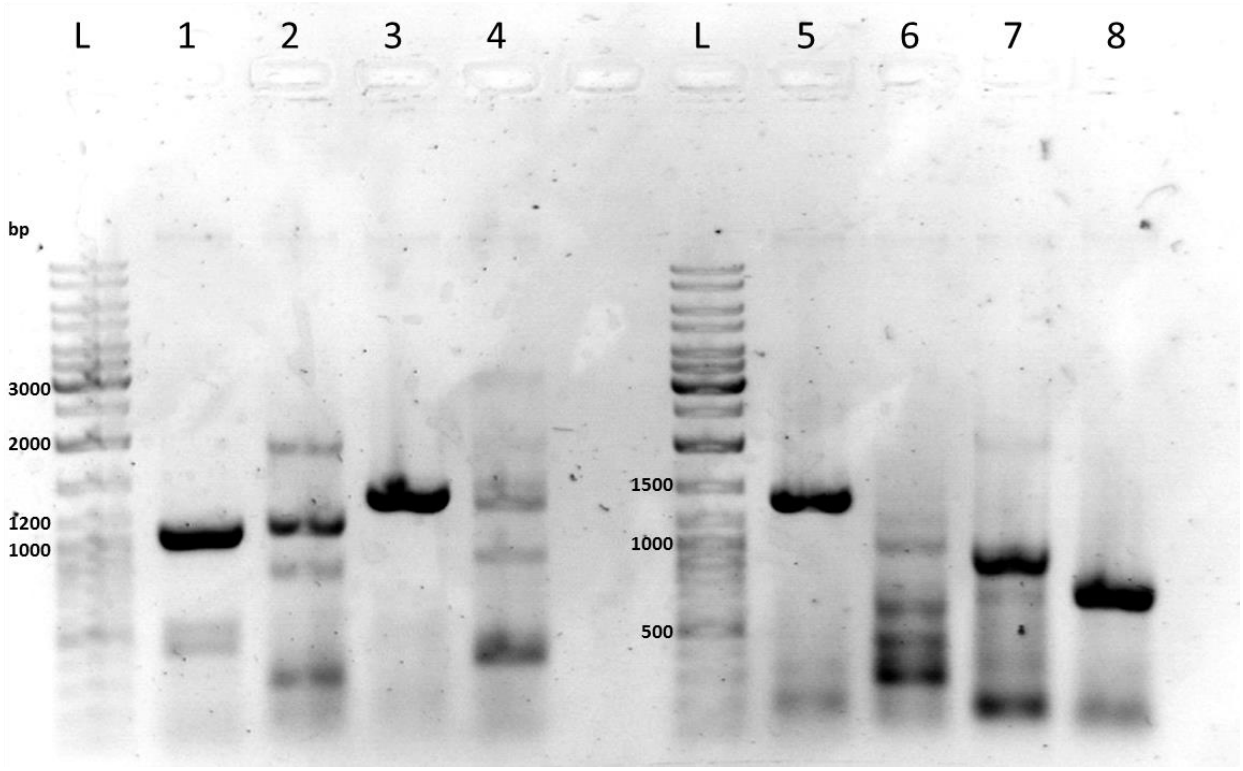


Figure 4.7. Secretome analysis by mass-spectrometry of other *C. fimi proteins*. Proteins with a MASCOT score of less than 50 are not shown. The % coverage of protein is a correlation to the relative amount of protein. Proteins that are unique to *C. fimi* are qualitatively determined by amino acid sequence.

Non-cellulase proteins were also found to be secreted by *C. fimi*. A substantial number of these are hypothetical proteins and more than half of these were only expressed under growth with CMC as a substrate.

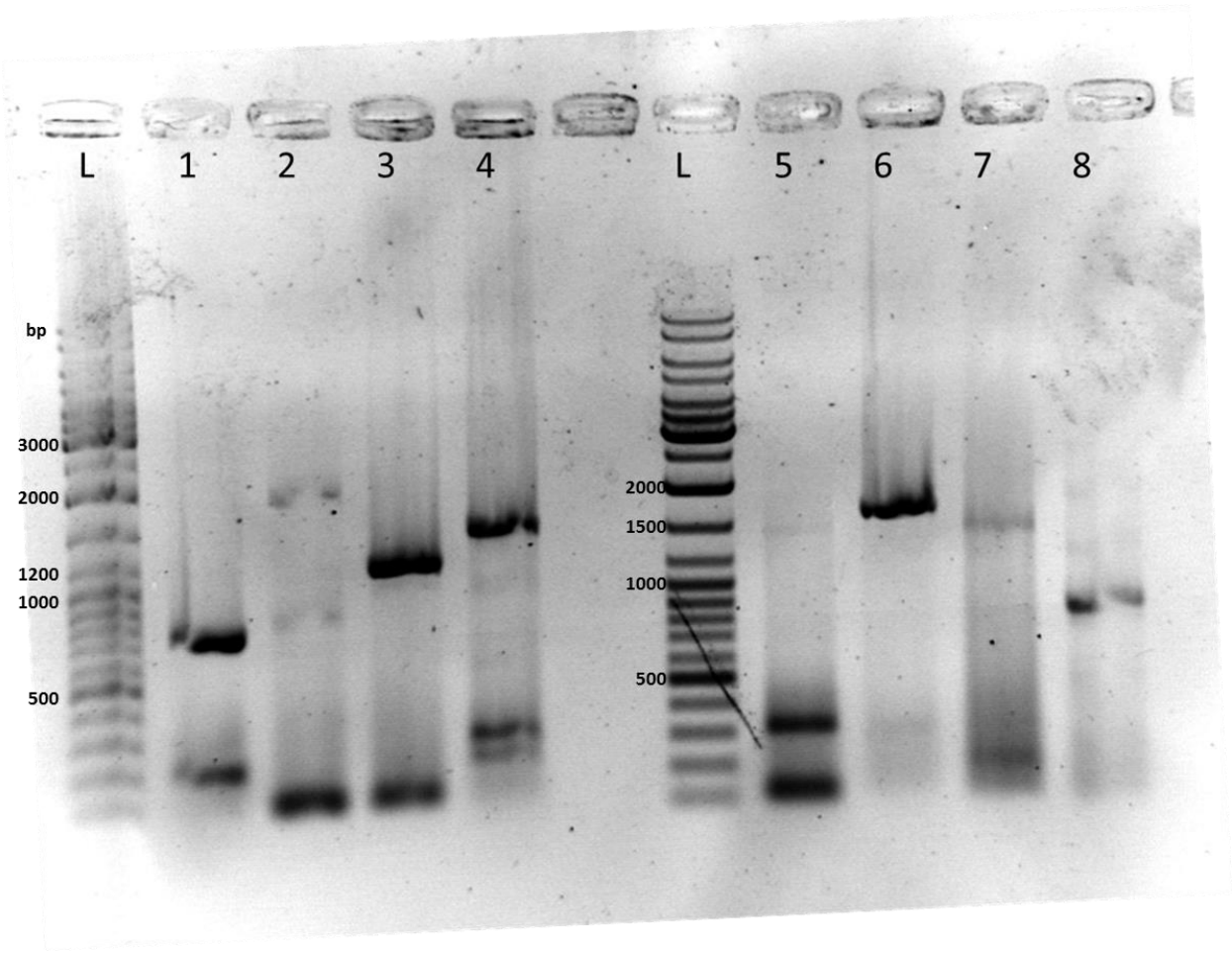
4.1.3 Cloning of catalytic domains of putative cellulases

PCR attempts for downstream cloning



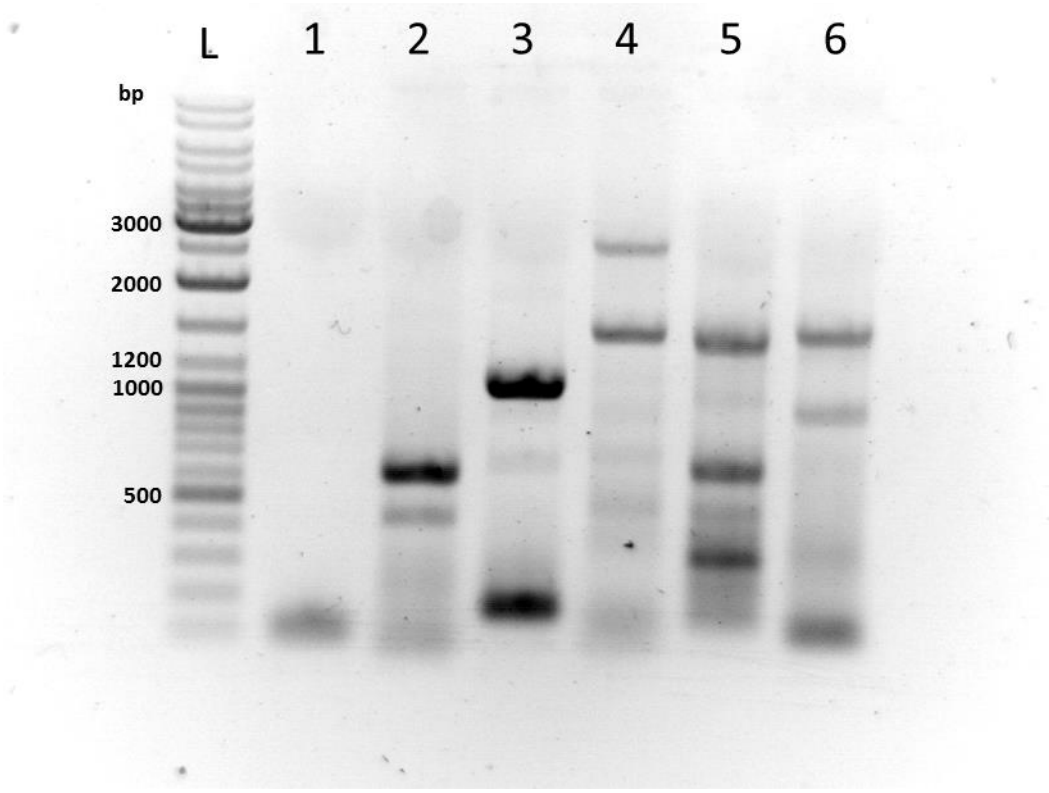
L: O Gene Ruler DNA Ladder Mix (Fermentas SM1173) (Appendix)
1: Celf_1924 (Cel5A) – 1077 bp
2: Celf_1925 (Cel6B) – 1170 bp
3: Celf_0019 (Cel9A) – 1362 bp
4: Celf_1537 (Cel9B) – 1485 bp
5: Celf_0019 – 1362 bp
6: Celf_1705 – 1539 bp
7: Celf_0376 – 915 bp
8: Celf_2403 – 684 bp

Figure 4.8. PCR attempt to check cycling conditions and annealing temperatures for successful cloning. Lanes 4 and 6 were unsuccessful. Other lanes produced prominent bands that correspond to the specific gene target, among other non-specific amplifications.



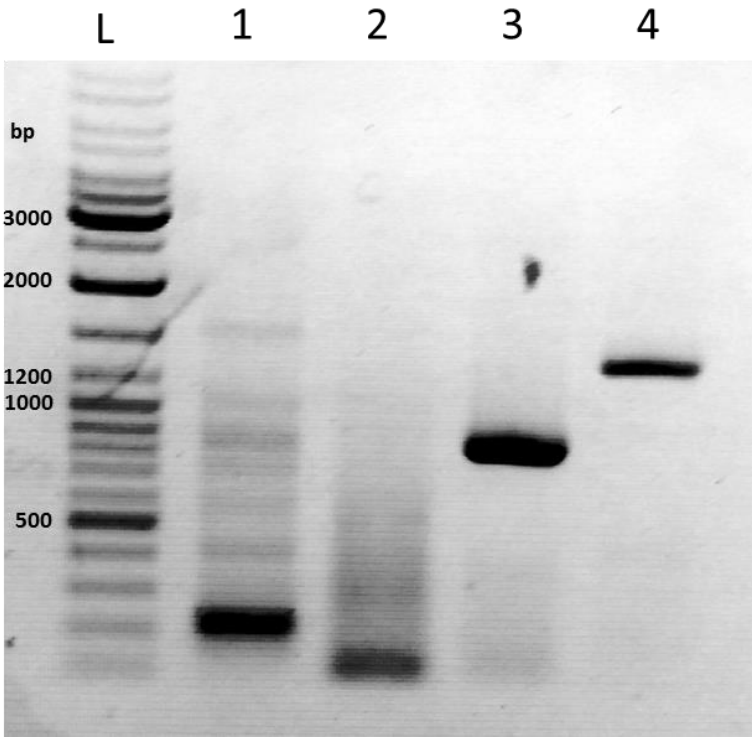
L: O Gene Ruler DNA Ladder Mix
1: Cfla_2811 – 750 bp
2: Cfla_2913 – 849 bp
3: Cfla_1986 – 1170 bp
4: Cfla_1515 – 1491 bp
5: Cfla_3563 – 1458 bp
6: Cfla_0139 – 1581 bp
7: Celf_0045 – 1455 bp
8: Celf_0233 – 831 bp

Figure 4.9. PCR attempt to check cycling conditions and annealing temperatures for successful cloning. Lanes 2, 5, 7, and 8 all produced a weak band of the correct size. Other lanes produced prominent bands that correspond to the specific gene target, among other non-specific amplifications.



L: O Gene Ruler DNA Ladder Mix
1: Celf_1537 (Cel9B) – 1485 bp
2: Cfla_0736 – 603 bp
3: Cfla_1897 – 1074 bp
4: Celf_1481 – 1497 bp
5: Celf_1705 – 1539 bp
6: Cfla_3031 – 1536 bp

Figure 4.10. PCR attempt to check cycling conditions and annealing temperatures for successful cloning. Only lane 1 was unsuccessful. Other lanes produced prominent bands that correspond to the specific gene target, among other non-specific amplifications.



L: O Gene Ruler DNA Ladder Mix
1: Celf_3184 (Cel6A) – 846 bp
2: Celf_1230 – 1152 bp
3: Cfla_2912 – 849 bp
4: Cfla_0016 – 1365 bp

Figure 4.11. PCR attempt to check cycling conditions and annealing temperatures for successful cloning. Lane 1 produced a weak band of the correct size, while lane 2 was unsuccessful. Lanes 3 and 4 were able to produce prominent bands that correspond to the specific gene targets.

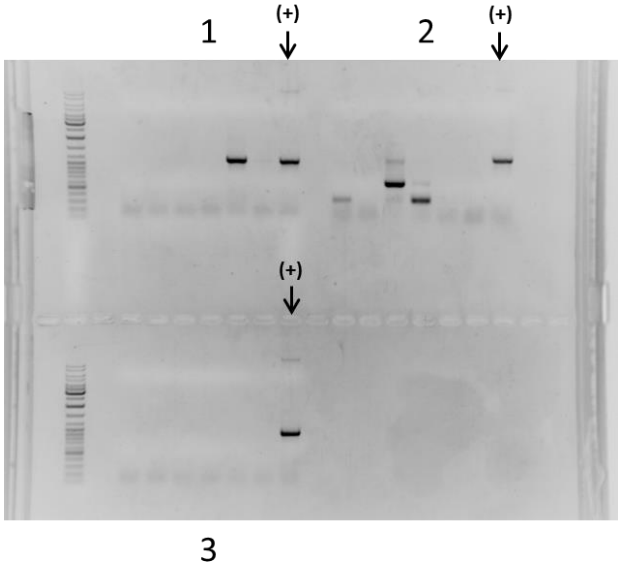


Figure 4.12. Colony PCR of clones made with NdeI and HindIII as restriction enzymes. Quadrants 1-3 correspond to the following gene targets: (1) Celf_1924 (Cel5A), (2) Celf_1925 (Cel6B), (3) Celf_0376. The first 6 lanes in each quadrant refer to 6 different colonies tested for the presence of inserts. The 7th lane (indicated by arrows) of each quadrant was performed with genomic DNA to serve as positive controls.

Not all cloning experiments are shown here.

4.1.4 Test expressions of clones

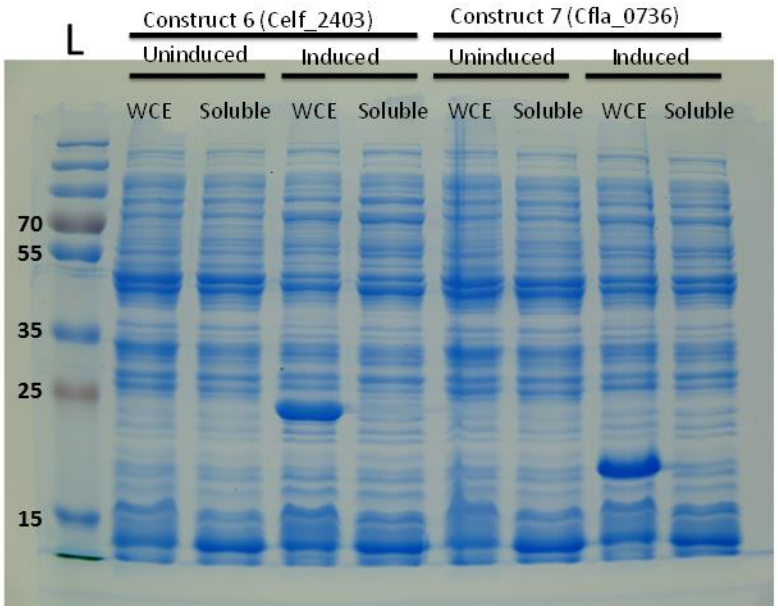


Figure 4.13. SDS-PAGE analysis of expression trial of Celf_2403 and Cfla_0736. WCE: whole cell extract. Expected sizes of Celf_2403 and Cfla_0736 are 23.76 kDa and 20.79 kDa, respectively. The ladder (L) used was Pageruler Plus Prestained from Thermo Scientific. The induction was initiated by the addition of 0.5 mM IPTG at 37°C for 3 hours. Only insoluble recombinant protein is observed here.

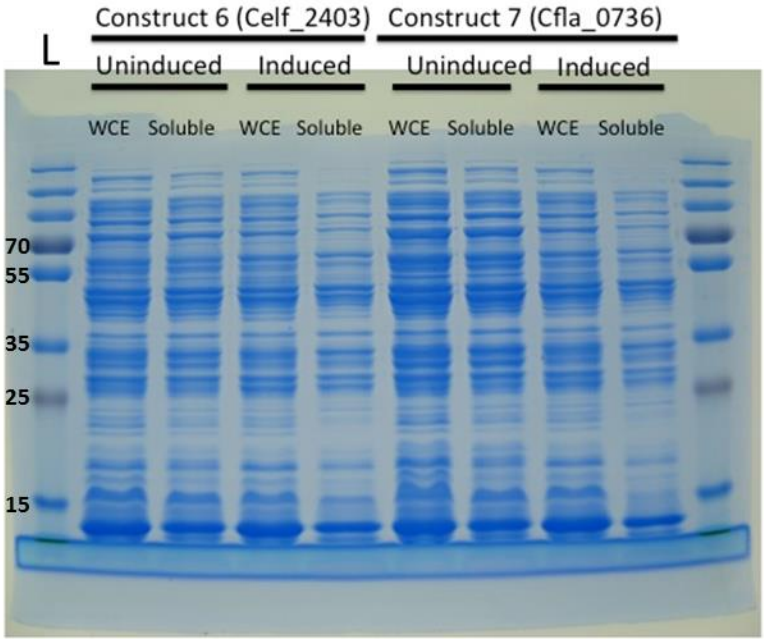


Figure 4.14. SDS-PAGE analysis of expression trial of Celf_2403 and Cfla_0736. WCE: whole cell extract. Expected sizes of Celf_2403 and Cfla_0736 are 23.76 kDa and 20.79 kDa, respectively. The ladder (L) used was Pageruler Plus Prestained from Thermo Scientific. The induction was initiated by the addition of 0.5 mM IPTG at 30°C and left overnight. There is no noticeable overexpression detected.

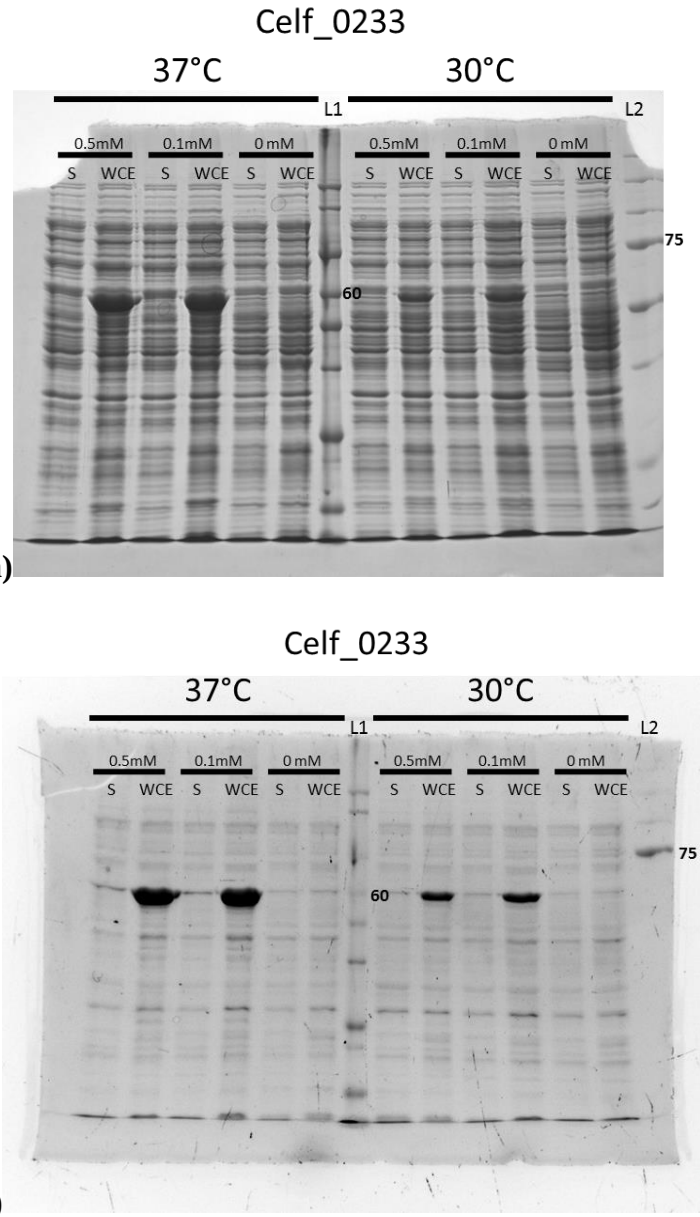


Figure 4.15. SDS-PAGE analysis of expression trials of Celf_0233 for 37°C and 30°C.

Different induction conditions are represented based on induction temperature (°C) or the final concentration of IPTG (mM). The expected size of Celf_0233 is 60 kDa. The protein gel was (a) stained with Coomassie after (b) the same gel had been stained using InVision In-gel His-stain (Invitrogen). The ladders used were: (L1) BenchMark™ His-tagged Standard (Invitrogen) and (L2) Precision Plus (Bio-Rad). WCE: whole cell extract.

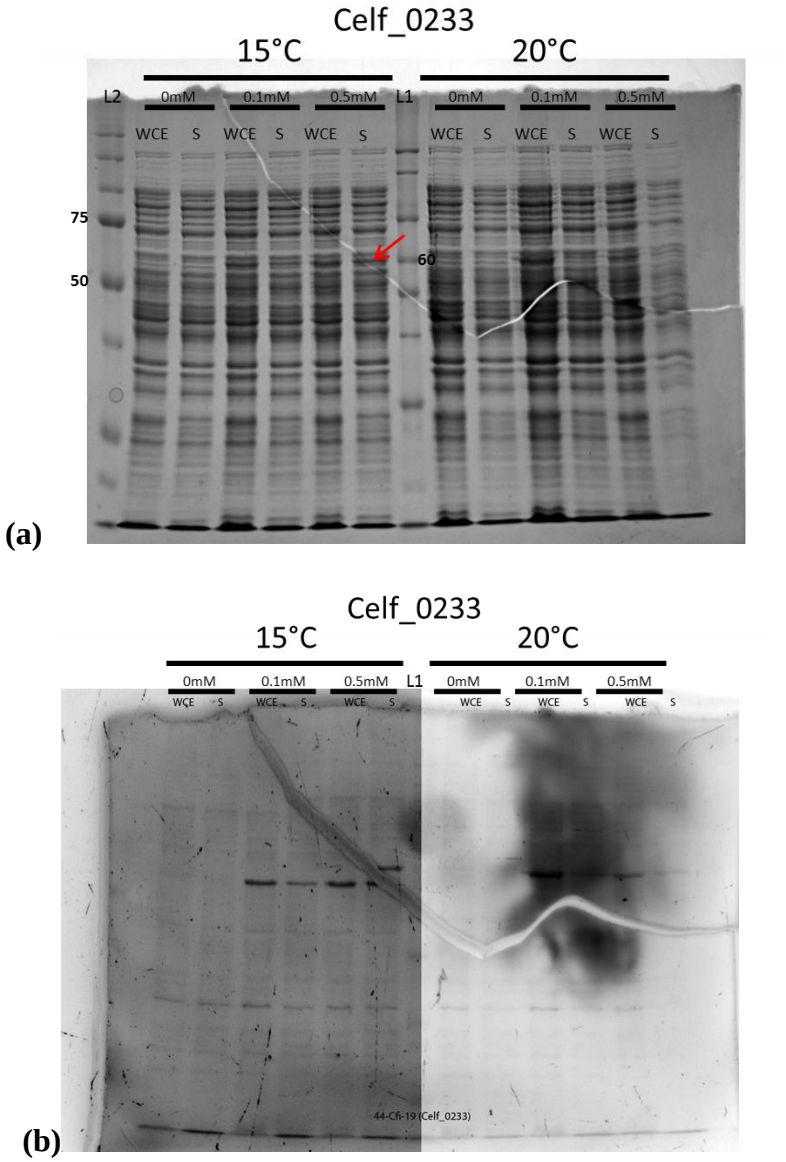


Figure 4.16. SDS-PAGE analysis of expression trials of Celf_0233 for 15°C and 20°C. Different induction conditions are represented based on induction temperature (°C) or the final concentration of IPTG (mM). The expected size of Celf_0233 is 60 kDa. The protein gel was (a) stained with Coomassie after (b) the same gel had been stained using InVision In-gel His-stain (Invitrogen). The ladders used were: (L1) BenchMark™ His-tagged Standard (Invitrogen) and (L2) Precision Plus (Bio-Rad). WCE: whole cell extract. The red arrow indicates the protein

band is at 60 kDa as seen with the coomassie stain since the ladder was not visible during the His-staining.

Celf_0233 in Vek-06 (Cfi-19) was placed in Shuffle Express™ for expression trials (Figures 5.20-5.21). It mostly produces insoluble protein except when induced at 15°C overnight. This produced a little but mostly soluble protein (Figure 5.21 b). However, the expressed protein was not able to bind to a nickel column (not shown).

4.1.5 Summary

Here we show sequence comparisons, secretome analysis, and test expressions of various recombinant cellulases from *C. fimi* and *C. flavigena*. Though most of our expressions proved to be unproductive, work must continue on expressing and characterizing the unique glycoside hydrolases identified by sequence comparisons. It is very likely that the constructs designed to express the catalytic domains of these cellulases were flawed. The boundaries defined for the catalytic domain was too restrictive and relied on BLAST homology determined by NCBI. Thus, the constructs resulted in truncated catalytic domains and could be missing necessary amino acids for proper folding. This design error likely resulted in the insoluble expression that was observed. Future directions include extending these domains to proper boundaries using PCR amplification or re-cloning all the gene targets to make full-length protein constructs. There have been expression trials with a vector (pMalE-Thr) that fuses a maltose-binding protein to the N-terminal of the construct (data not shown). This appeared to produce soluble protein but was inactive on preliminary assays. The assumption from this observation is that the actual cellulase construct continues to produce misfolded protein but is then able to maintain solubility due to the large maltose-binding protein. Thus, this approach was abandoned for this particular project.

Other attempts to express Celf_0233 involved moving it into vectors pCW (no his-tag) or pVEK-07 (C-terminal his-tag)[6]. An attempt to remove the stop codon from the Celf_0233 insert using PCR was unsuccessful. Thus, the Vek-07 construct would have been unusable since the C-terminal tag would not be expressed. The pCW-Celf_0233 construct produced no significant overexpressed protein. The Celf_0233 expression project has also been put on hold and other expression strategies are currently being explored by other members of the group.

Among all the constructs, Celf_1230 in pVEK-06 showed promise in solubility and activity. The work done with this enzyme is the focus of the following section. This enzyme is renamed as Cel6C, in reference to being the third GH 6 enzyme characterized for *Cellulomonas fimi*.

[6] For more information on these vectors, consult supplementary image (S4). The Cfi-28 (pVEK-06 with insert) is derived from pCW.

4.2 Characterization of Celf_1230 (Cel6C)

4.2.1 Bioinformatics of Cel6C

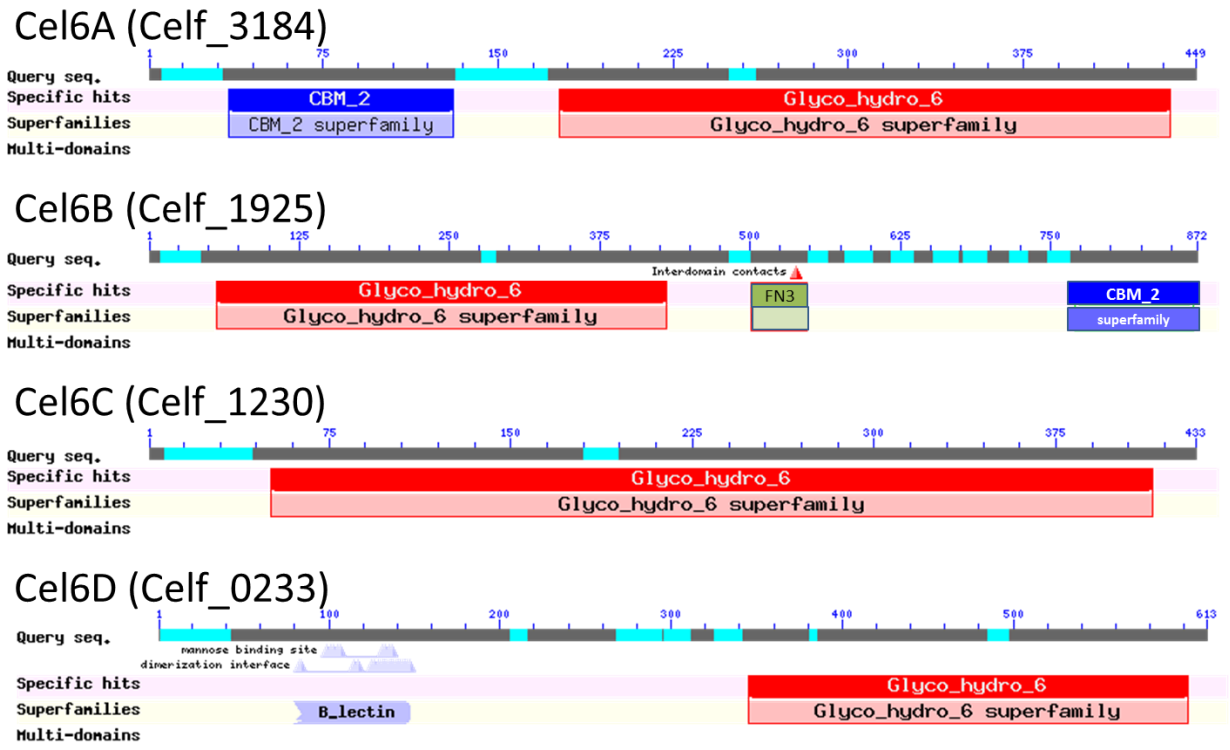


Figure 4.17. Comparisons of the different members of the Family 6 glycoside hydrolase from *C. fimi*. The different domains of the GH6 cellulases are shown. Cel6A has a N-terminal CBM-2 domain while Cel6B has a C-terminal CBM-2 domain. Cel6C has no identifiable CBM domain while Cel6D has a putative mannose-binding B-lectin domain on the N-terminal end.

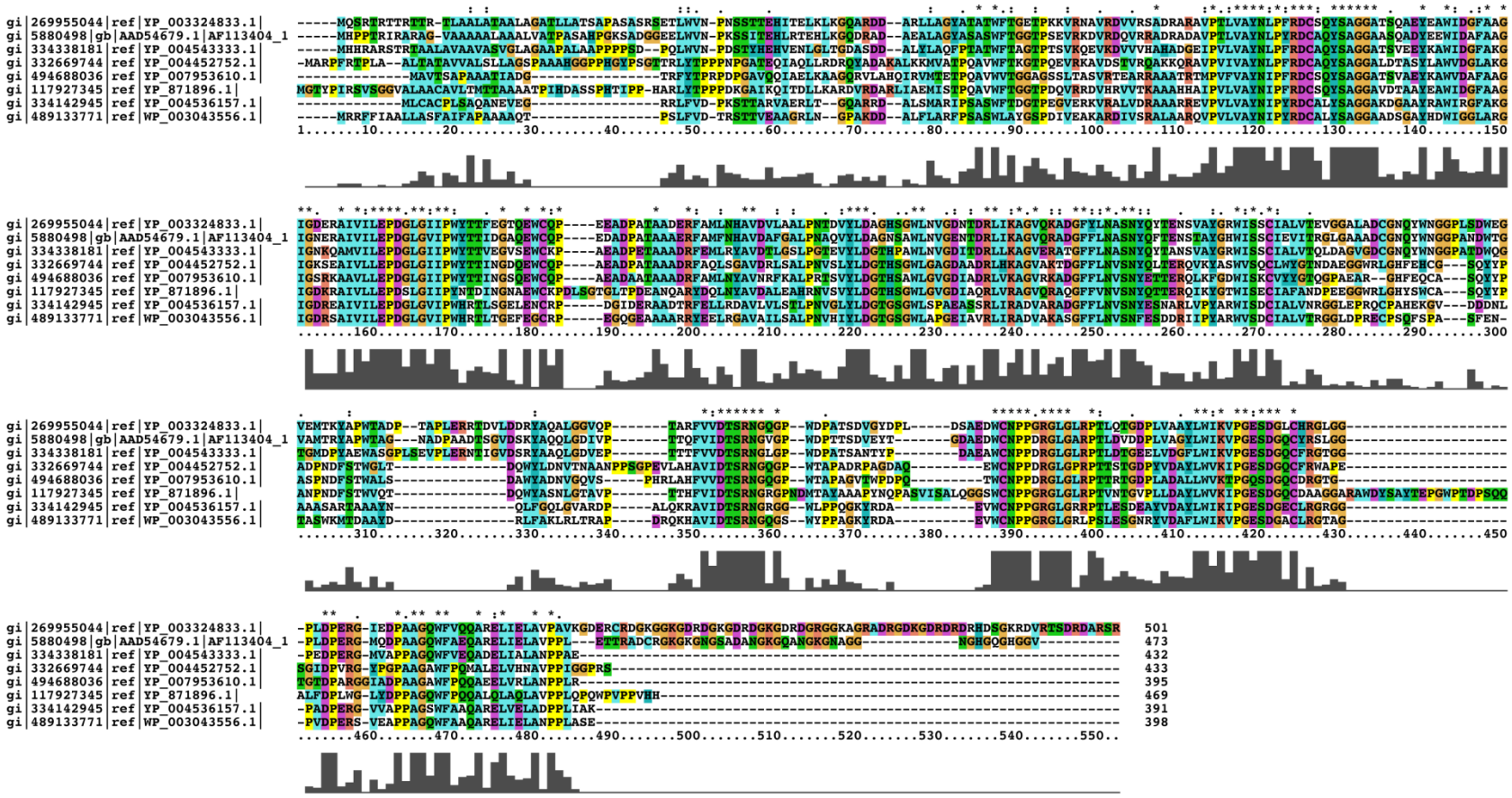


Figure 4.18 Amino acid sequence alignment between the full-length protein of Celf_1230 and its closest homologs determined by BLAST. Celf_1230 can be identified here as YP_004452752.1. For the identity of the other proteins, please refer to their respective accession numbers. Quality curve is shown below each row of amino acid sequence. The alignment was performed using ClustalX2 software.

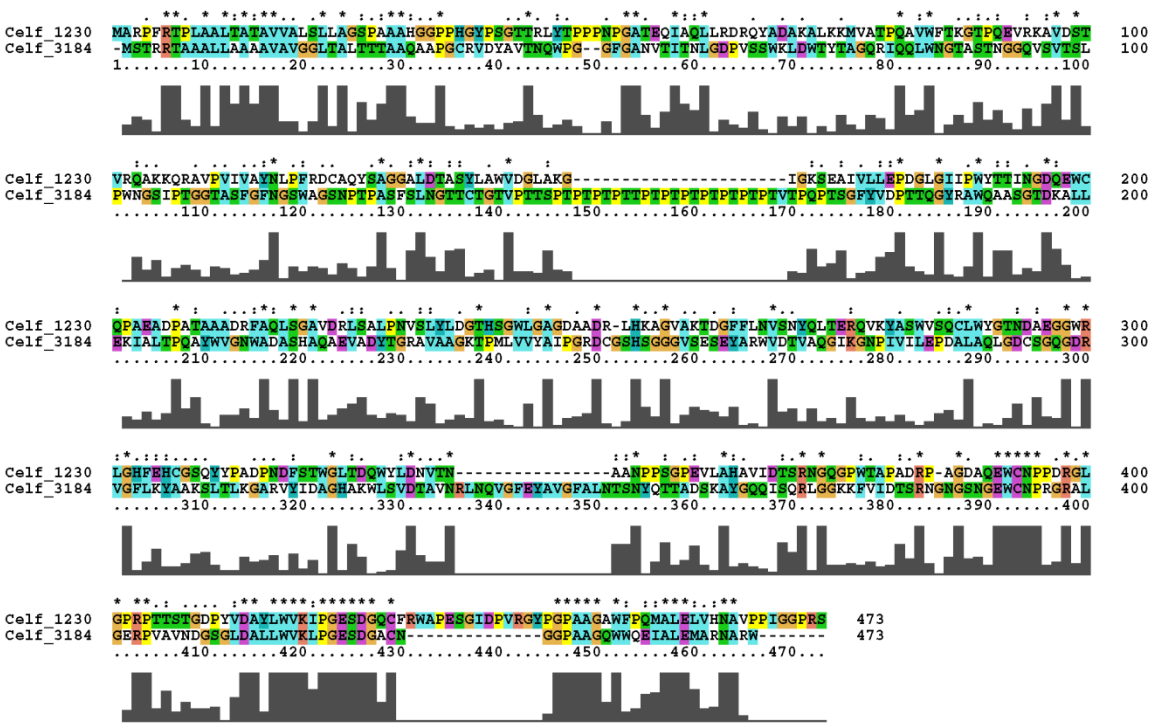


Figure 4.19. Amino acid sequence alignment of full-length Celf_1230 (Cel6C) and Celf_3184 (Cel6A). Quality curve is shown below each row of amino acid sequence. The alignment was performed using ClustalX2 software.

Score	Expect	Method	Identities	Positives	Gaps
110 bits(275)	5e-31	Compositional matrix adjust.	99/357(28%)	132/357(36%)	103/357(28%)

```
Query  69    DAKALKKMVATPQAVWFTK-GTPQEVRKAVDSTVRQAKKQRAVPVIVAYNLPFRDCAQYS  127
             D   L+K+  TPQA W         +   V     +A      P++V Y +P RDC  +S
Sbjct  193   DKALLEKIALTPQAYWVGNWADASHAQAEVADYTGRAVAAGKTPMLVVYAIPGRDCGSHS  252
Query  128   AGGALDTASYLAWVDGLAKGIGKSEAIVLLEPDGLGIIPWYTTINGDQEWCQPAEADPAT  187
              GG +  + Y  WVD +A+GI K   IV+LEPD L                  A+    +
Sbjct  253   GGG-VSESEYARWVDTVAQGI-KGNPIVILEPDAL------------------AQLGDCS  292
Query  188   AAADRFAQLSGAVDRLSALPNVSLYLDGTHSGWLGAGDAADRLHKAGVAKTDGFFLNVSN  247
                DR   L  A   L+ L    +Y+D  H+ WL    A +RL++ G     GF LN SN
Sbjct  293   GQGDRVGFLKYAAKSLT-LKGARVYIDAGHAKWLSVDTAVNRLNQVGFEYAVGFALNTSN  351
Query  248   YQLTERQVKYASWVSQCLWYGTNDAEGGWRLGHFEHCGSQYYPADPNDFSTWGLTDQWYL  307
             YQ T     Y   +SQ             RLG                        ++ +
Sbjct  352   YQTTADSKAYGQQISQ-------------RLGG----------------------KKFVI  376
Query  308   DNVTNAANPPSGPEVLAHAVIDTSRNGQGPWTAPADRPAGDAQEWCNPPDRGLGPRPTTS  367
             D   N                       G W  P  R  G+              RP
Sbjct  377   DTSRNG------------------NGSNGEWCNPRGRALGE--------------RPVAV  404
Query  368   TGDPYVDAYLWVKIPGESDGQCFRWAPESGIDPVRGYPGPAAGAWFPQMALELVHNA  424
                  +DA LWVK+PGESDG C                GPAAG W+ ++ALE+  NA
Sbjct  405   NDGSGLDALLWVKLPGESDGAC--------------NGGPAAGQWWQEIALEMARNA  447
```

Figure 4.20. Aligned two sequence by BLAST (bl2seq) for the full-length amino acid sequences of Cel6A and Cel6C. The query and subject proteins are Cel6C and Cel6A, respectively. The identities found are based on 82% coverage as indicated by the result.

Celf_1230 (YP_004452752.1) was subjected to a BLAST analysis to find homologs (Figure 6.2). None of the homologs produced an identity >60%. Notably, one of the close homolog is an uncharacterized putative cellulase from *Acidothermus cellulolyticus* 11B (YP_871896.1). This cellulase from *A. cellulolyticus* has a 52% identity for 96% coverage of Celf_1230.

A direct comparison of Celf_1230 and Celf_3184 was conducted on their amino acid sequences using ClustalX2 (Figure 6.3)or BLAST (Figure 6.4). For the BLAST method, the subject alignment starts at aa193, which are about 20 amino acids into the catalytic domain of Cel6A. These residues start to have similarity with aa69 for Cel6C, which are about 35 amino acids after its signal peptide. Note that the sequence alignments of ClustalX2 and BLAST differ from one another.

4.2.2 Boundary extensions of Cel6C construct

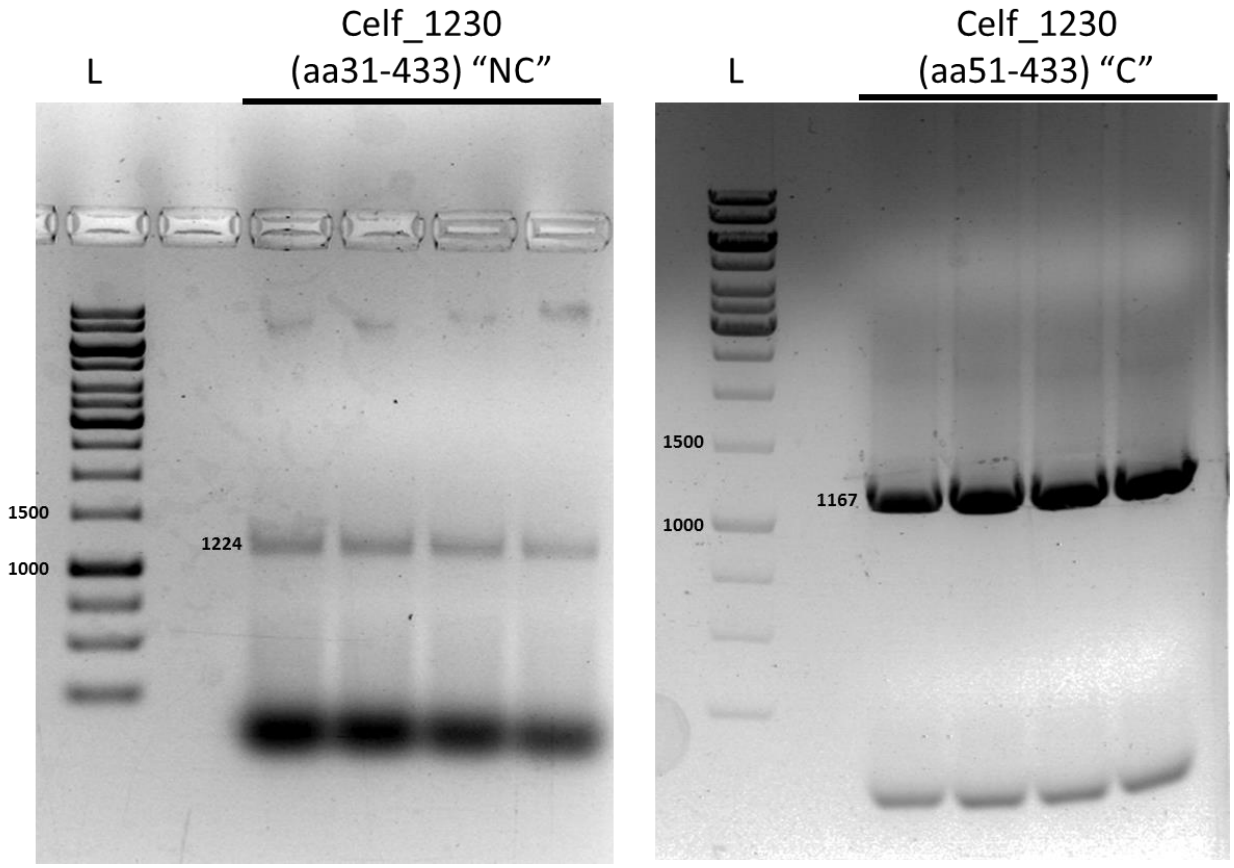


Figure 4.21. PCR-based addition of nucleotides to extend the boundaries of the Celf_1230 construct. A band of the correct size was detected for both PCR experiments. (L) GeneRuler 1 Kb ladder (Fermentas)

Celf_1230-NC (aa31-433) was cloned successfully into pVEK-06 and renamed as Cfi-28. Celf_1230-C (aa51-433) was also successfully cloned into pVEK-06 and renamed as Cfi-29. The enzymatic characterization of Celf_1230 was only done using the Cfi-28 construct. However, preliminary data suggests that Cfi-29 produces relatively equal amounts of protein with similar activity (not shown).

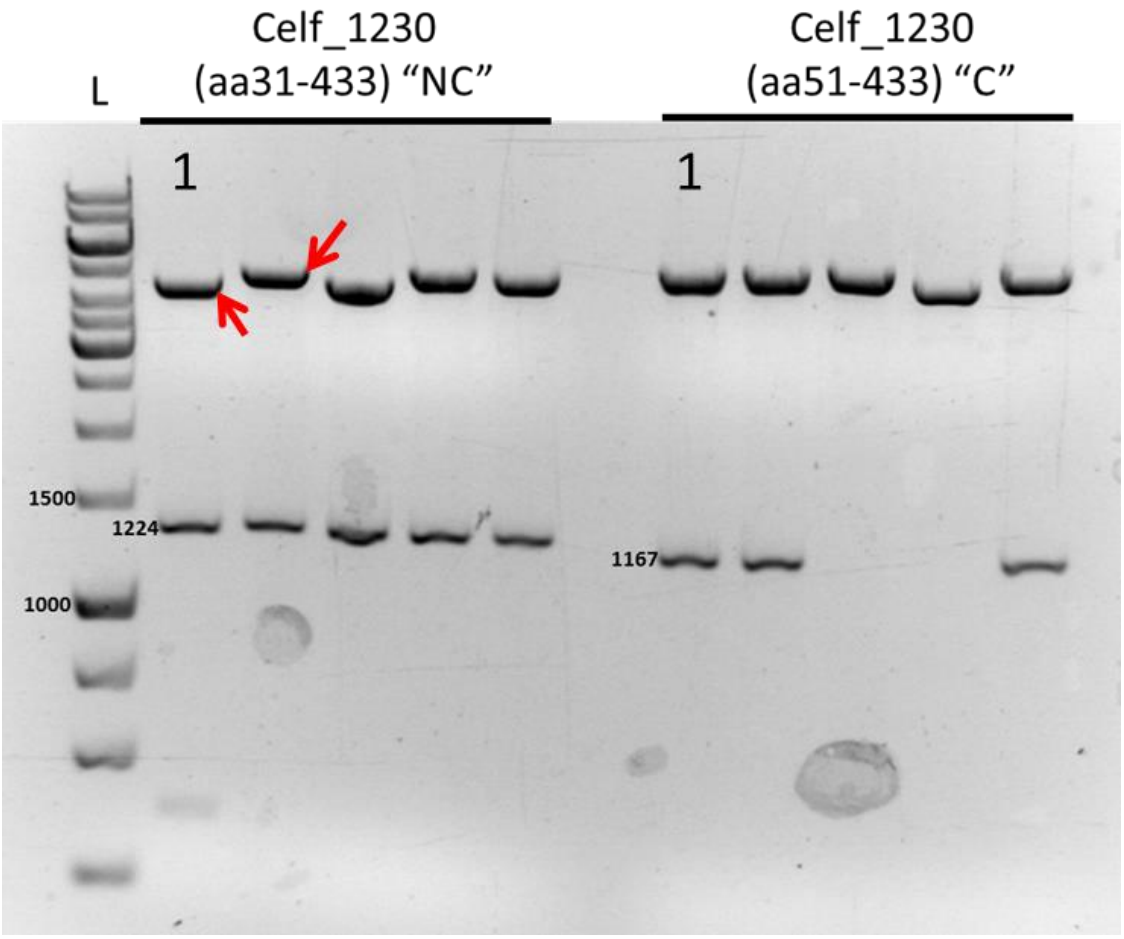


Figure 4.22. Detection of insert-positive plasmids by a double restriction digest (NdeI and HindIII). Note that the vector backbone between colonies 1 and 2 (red arrows) have slightly different sizes. (L) GeneRuler 1Kb ladder (Fermentas)

Colony 1 of both constructs were sent for sequencing and subsequently used for protein expression. The shorter backbone (1) has been attributed to a mutation in the vector that produced a second HindIII site away from the insert (upstream of the promoter region). This appears to have no effect on the expression of Celf_1230 but it prevents the reuse of this backbone for other NdeI-HindIII cloning projects.

```
Alignment: Insert of Cfi-28 (Celf_1230 aa31-433)
Celf_1230 NC1 Read      atgggtggtc cgcctcatgg ttatccgagc ggcaccaccc gtctgtatac accgcctccg aatccgggtg caaccgagca gattgcacag
Celf_1230 NC Template   ATGGGTGGTC CGCCTCATGG TTATCCGAGC GGCACCACCC GTCTGTATAC ACCGCCTCCG AATCCGGGTG CAACCGAGCA GATTGCACAG

Celf_1230 NC1 Read      ctgctgcgtg atcgtcagta tgcagatgca aaagcactga aaaaaatggt tgcaacaccg caggcagttt ggtttaccaa aggtacaccg
Celf_1230 NC Template   CTGCTGCGTG ATCGTCAGTA TGCAGATGCA AAAGCACTGA AAAAAATGGT TGCAACACCG CAGGCAGTTT GGTTTACCAA AGGTACACCG

Celf_1230 NC1 Read      caagaagttc gtaaagcagt tgatagcacc gttcgtcagg caaaaaaaca gcgtgcagtt ccggttattg ttgcatataa tctgccgttt
Celf_1230 NC Template   CAAGAAGTTC GTAAAGCAGT TGATAGCACC GTTCGTCAGG CAAAAAAACA GCGTGCAGTT CCGGTTATTG TTGCATATAA TCTGCCGTTT

Celf_1230 NC1 Read      cgtgattgtg cacagtatag tgccggtggt gcactggata ccgcaagcta tctggcatgg gttgatggtc tggcaaaagg tattggtaaa
Celf_1230 NC Template   CGTGATTGTG CACAGTATAG TGCCGGTGGT GCACTGGATA CCGCAAGCTA TCTGGCATGG GTTGATGGTC TGGCAAAAGG TATTGGTAAA

Celf_1230 NC1 Read      agcgaagcaa ttgttctgct ggaaccggat ggcctgggta ttattccgtg gtatacaaca attaacggcg atcaagaatg gtgtcagcct
Celf_1230 NC Template   AGCGAAGCAA TTGTTCTGCT GGAACCGGAT GGCCTGGGTA TTATTCCGTG GTATACAACA ATTAACGGCG ATCAAGAATG GTGTCAGCCT

Celf_1230 NC1 Read      gcagaagcag atccggcaac cgcagcagca gatcgttttg cccagctgag cggtgcagtt gatcgtctga gcgcactgcc gaatgttagc
Celf_1230 NC Template   GCAGAAGCAG ATCCGGCAAC CGCAGCAGCA GATCGTTTTG CCCAGCTGAG CGGTGCAGTT GATCGTCTGA GCGCACTGCC GAATGTTAGC

Celf_1230 NC1 Read      ctgtatctgg atggcaccca tagcggttgg ctgggtgcgg gtgatgcagc ggatcgtctg cataaagccg gtgttgccaa aaccgatggt
Celf_1230 NC Template   CTGTATCTGG ATGGCACCCA TAGCGGTTGG CTGGGTGCGG GTGATGCAGC GGATCGTCTG CATAAAGCCG GTGTTGCCAA AACCGATGGT

Celf_1230 NC1 Read      ttttttctga atgtgagcaa ctatcagctg accgaacgtc aggttaaata tgcaagctgg gttagccagt gtctgtggta tggcaccaat
Celf_1230 NC Template   TTTTTTCTGA ATGTGAGCAA CTATCAGCTG ACCGAACGTC AGGTTAAATA TGCAAGCTGG GTTAGCCAGT GTCTGTGGTA TGGCACCAAT

Celf_1230 NC1 Read      gatgcagaag gtggttggcg tctgggtcat tttgaacatt gtggtagcca gtattatccg gcagatccga atgattttag cacctggggt
Celf_1230 NC Template   GATGCAGAAG GTGGTTGGCG TCTGGGTCAT TTTGAACATT GTGGTAGCCA GTATTATCCG GCAGATCCGA ATGATTTTAG CACCTGGGGT

Celf_1230 NC1 Read      ctgaccgatc agtggtatct ggacaatgtt accaatgcag caaatccgcc tagcggtccg gaagttctgg cacatgcagt tattgatacc
Celf_1230 NC Template   CTGACCGATC AGTGGTATCT GGACAATGTT ACCAATGCAG CAAATCCGCC TAGCGGTCCG GAAGTTCTGG CACATGCAGT TATTGATACC

Celf_1230 NC1 Read      agccgtaatg gtcagggtcc gtggaccgca ccggctgatc gtcctgccgg tgatgcccaa gagtggtgta atccgcctga tcgtggtctg
Celf_1230 NC Template   AGCCGTAATG GTCAGGGTCC GTGGACCGCA CCGGCTGATC GTCCTGCCGG TGATGCCCAA GAGTGGTGTA ATCCGCCTGA TCGTGGTCTG

Celf_1230 NC1 Read      ggtccgcgtc cgaccaccag taccggtgat ccgtatgttg atgcatatct gtgggttaaa attccgggtg aaagtgatgg tcagtgtttt
Celf_1230 NC Template   GGTCCGCGTC CGACCACCAG TACCGGTGAT CCGTATGTTG ATGCATATCT GTGGGTTAAA ATTCCGGGTG AAAGTGATGG TCAGTGTTTT

Celf_1230 NC1 Read      cgttgggcac cggaaagcgg tattgatccg gttcgtggtt atccgggtcc ggcagcgggt gcatggtttc cgcagatggc actggaactg
Celf_1230 NC Template   CGTTGGGCAC CGGAAAGCGG TATTGATCCG GTTCGTGGTT ATCCGGGTCC GGCAGCGGGT GCATGGTTTC CGCAGATGGC ACTGGAACTG

Celf_1230 NC1 Read      gttcataatg cagttccgcc tattggtggt ccgcgtagct aataa
Celf_1230 NC Template   GTTCATAATG CAGTTCCGCC TATTGGTGGT CCGCGTAGCT AATAA
```

Figure 4.23. Nucleotide sequence verification of Cfi-28. The line "Celf_1230 NC1 Read" indicates the sample sent for sequencing. There are no variations in the two nucleotide sequences.

4.2.3 Expression and purification of recombinant Cel6A and Cel6C

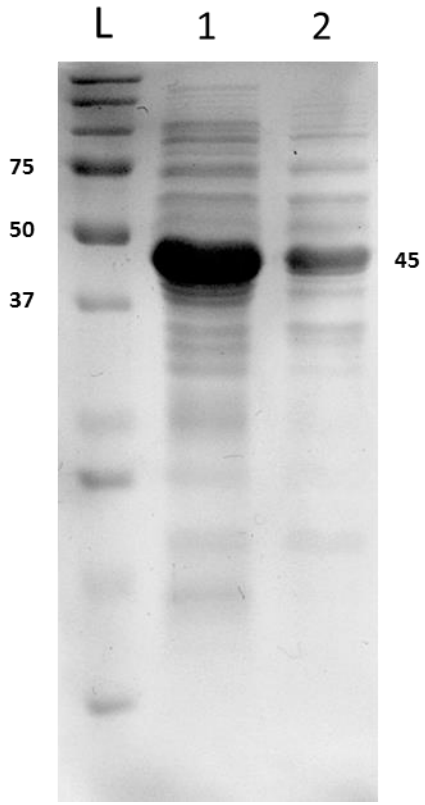


Figure 4.24. SDS-PAGE analysis of the expression of Cfi-28 (Celf_1230 aa31-433). This gel was visualized by Coomassie stain to detect the overexpressed protein, which has an expected size of 45.45 kDa. Lane 1 is the whole cell extract while lane 2 is only the supernatant soluble fraction. The ladder (L) used here was Precision Plus Dual Color (Bio-Rad).

This SDS-PAGE analysis shows that approximately half or more of the recombinant protein remains insoluble. The protein was expressed at 30°C overnight with 0.5 mM IPTG.

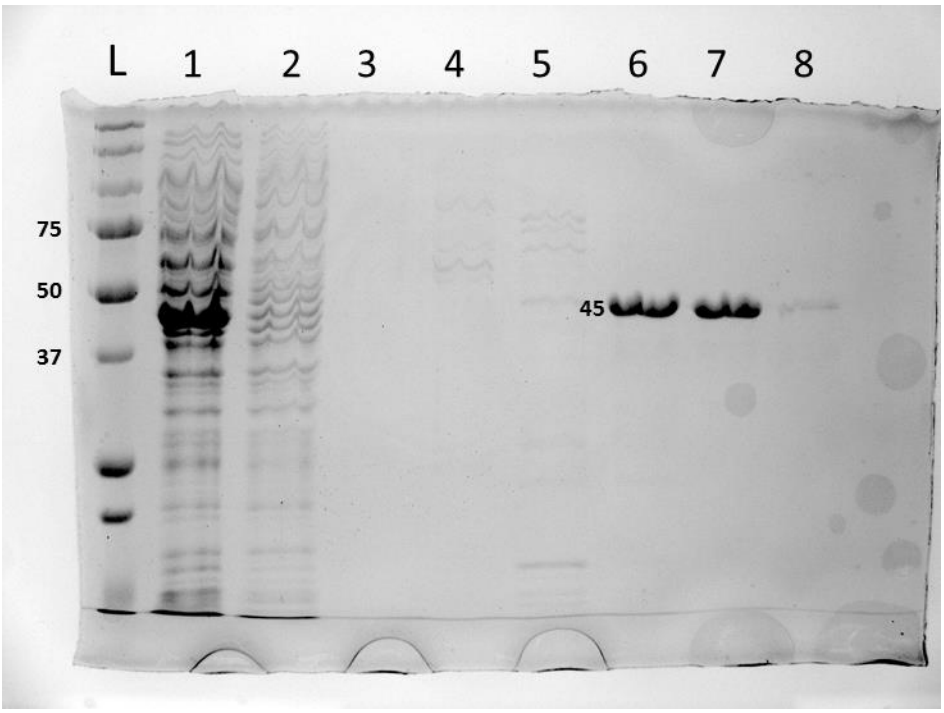


Figure 4.25. SDS-PAGE analysis of the purification of Cfi-28 (Celf_1230 aa31-433) by IMAC. This gel was visualized by Coomassie stain to detect the purified recombinant protein, which has an expected size of 45.45 kDa. Lane 1 is the flowthrough fraction while lane 2 and 3 are the wash fractions. Lanes 4-8 are various fractions taken during the elution gradient. A prominent band of the correct size is detected in both lanes 6 and 7 with minimal contaminating proteins. The ladder (L) used here was Precision Plus Dual Color (Bio-Rad).

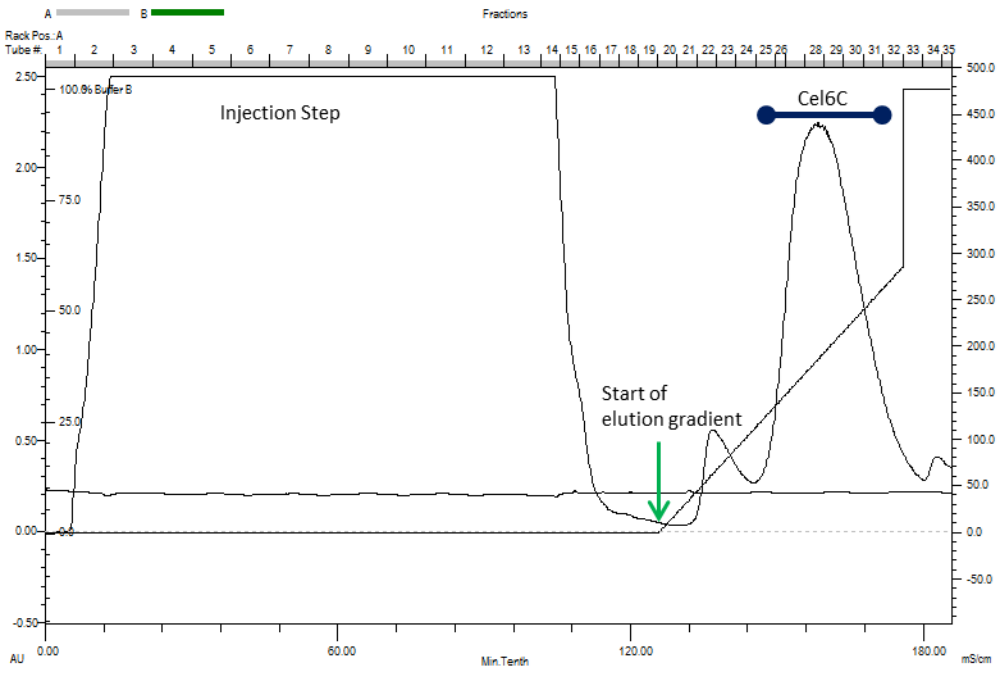


Figure 4.26. Chromatogram of Celf_1230 (aa 31-433) IMAC purification using an FPLC. The largest peak corresponds to the overexpressed recombinant Cel6C/Celf_1230. The protein was detected by a UV-Vis detector set for A_{280nm}. The x-axis is duration of the purification in minutes.

Figure 4.27. SDS-PAGE analysis of purity of Cfi-28 (Celf_1230 aa31-433). This gel was visualized by Coomassie stain to detect the pooled purified recombinant protein. In order to detect contaminating proteins, the gel was substantially overloaded to meet the minimum sensitivity of the reagent to detect protein levels.

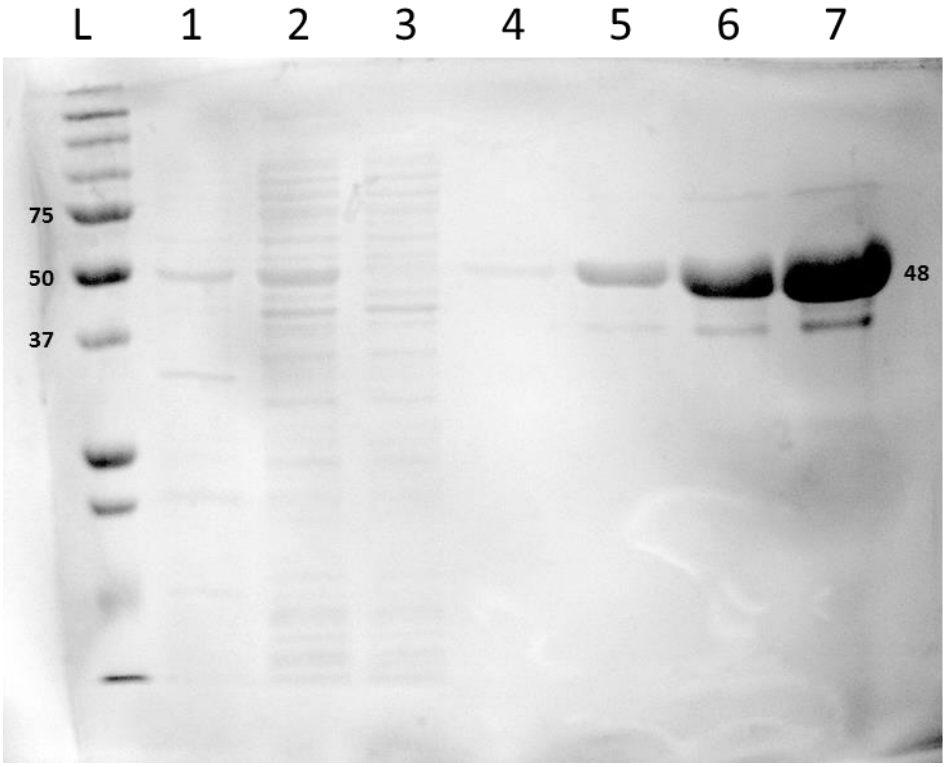


Figure 4.28. SDS-PAGE analysis of batch purification of Cel6A produced from the construct pACDC-003. The following lanes contain: (1) the solubilized pellet, (2) the supernatant soluble fraction, (3) purification flowthrough, and (4) the wash step. Lanes 5-7 are the three collected fractions from the elution steps and a band of the correct size (48 kDa) is detected. The ladder (L) used here was Precision Plus Dual Color (Bio-Rad).

4.2.4 Assays

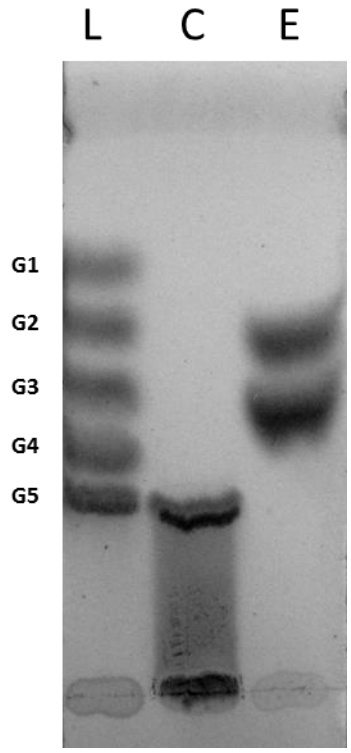


Figure 4.29. Thin-layer chromatography to detect the hydrolysis products produced by Cel6C on cellopentaose. The experimental lane (E) shows cellobiose (G2) and cellotriose (G3) as products when cellopentaose is incubated with Cel6C for approximately an hour at 60°C. The control (C) was treated the same but excluding the enzyme. The standard (L) is made up of various amounts of cellodextrins from glucose (G1) up to cellopentaose (G5).

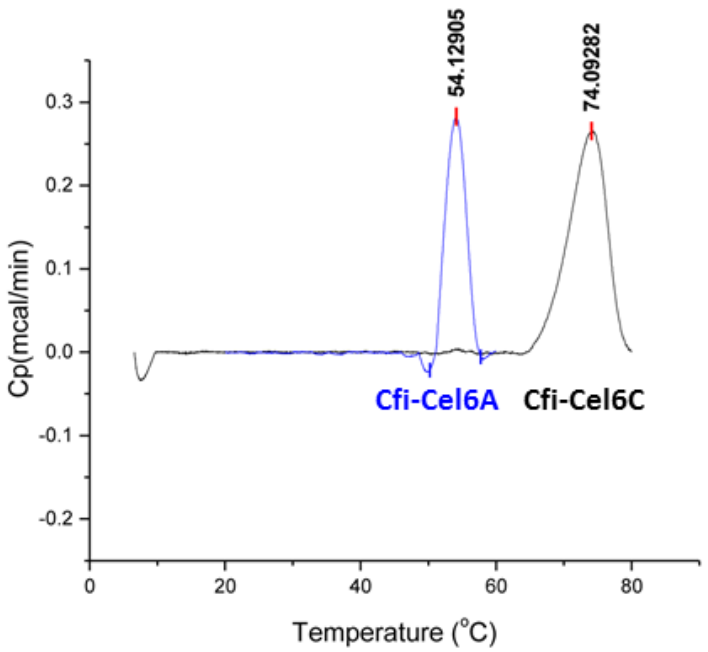


Figure 4.30. Differential scanning calorimetry (DSC) analysis of Cel6A and Cel6C. The two proteins were analyzed separately but their graphs are overlaid here. The peaks for each enzyme correspond to their respective melting temperature.

The DSC was performed in buffer containing 20 mM MOPS and 50 mM NaCl at a pH of 7. A related experiment tested the melting temperature (Tm) of Cel6C in the presence of 20mM NaPO4 buffer instead of MOPS (not shown). The Tm of Cel6C in the phosphate buffer was 73.86, which is not significantly different from the Tm with MOPS (74.09). This concludes that the thermostability is not due to a buffer effect.

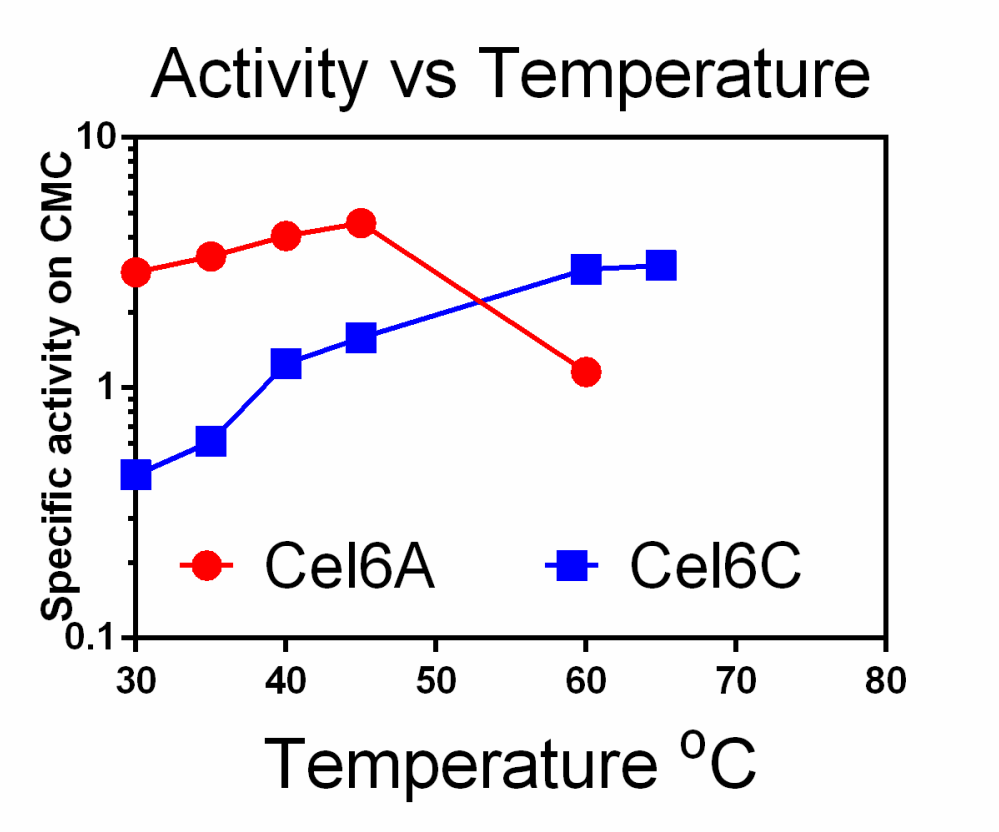


Figure 4.31. Activity vs Temperature profile of Cel6A and Cel6C on CMC as a substrate. This assay was conducted using the HBAH method to detect reducing sugars.

The measurement was done based on initial reaction rates so that substrate depletion was not an issue. This allowed Cel6A to have a detectable amount of activity at an assay temperature of 60°C since it was able to hydrolyze CMC before being completely denatured under these conditions. Specific activity is in µmol/min/mg**.**

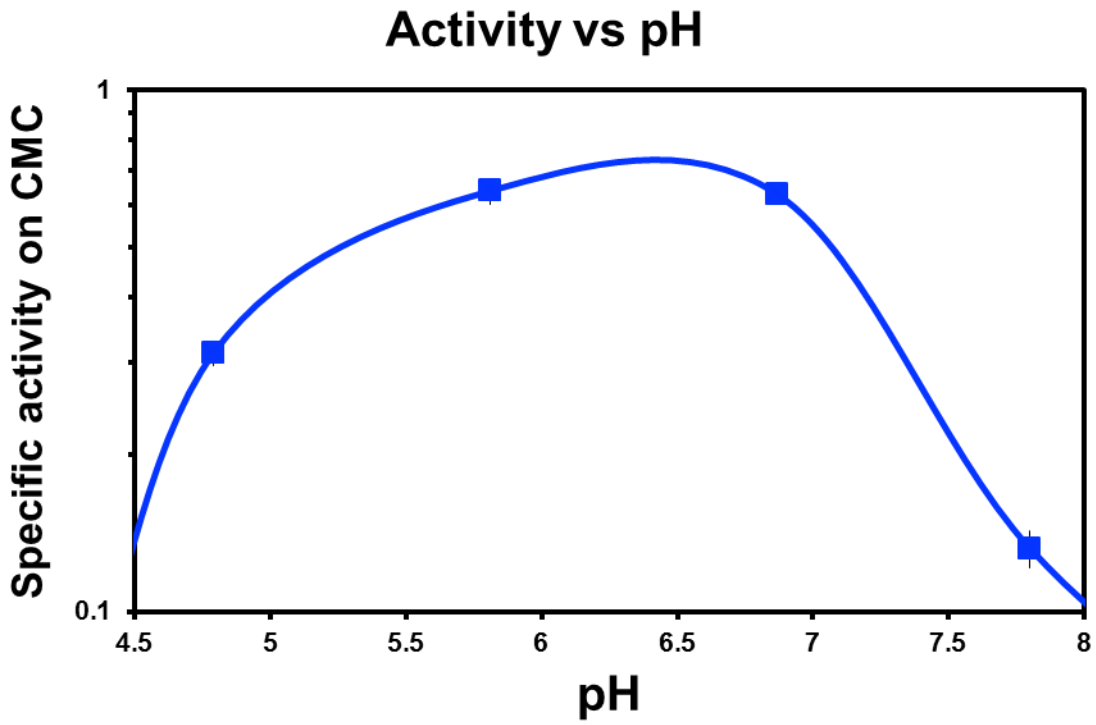


Figure 4.32. Activity vs pH profile of Cel6C on CMC as a substrate. This assay was conducted at an assay temperature of 30°C on 1.00% CMC. Cel6C prefers buffers within a neutral pH but can tolerate mild acidity (pH 5-7).

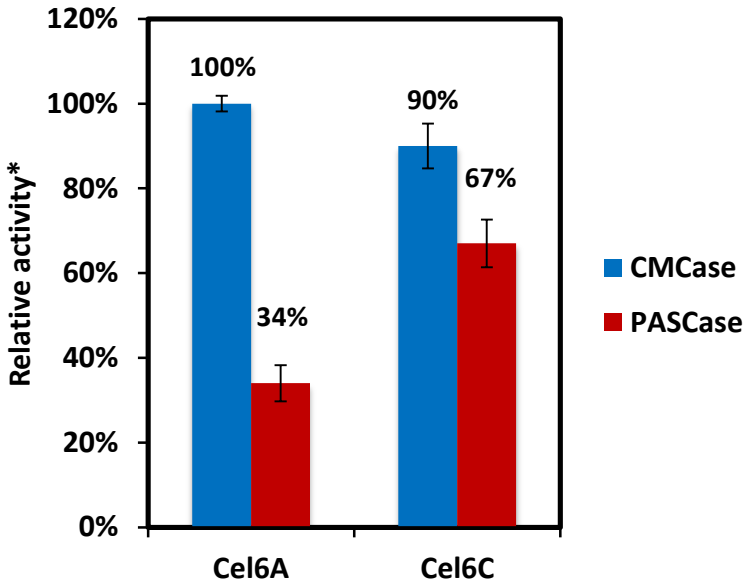


Figure 4.33. The relative activites of Cel6A and Cel6C on CMC or phosphoric acid-swollen cellulose (PASC) as substrates. The relative activities were normalized to Cel6A's activity on CMC.

Cel6C has comparable activity (90%) to Cel6A on CMC as a substrate. Both Cel6A and Cel6C had a significant decrease in activity when the substrate was switch to PASC. However, Cel6C has nearly double the activity of Cel6A on PASC.

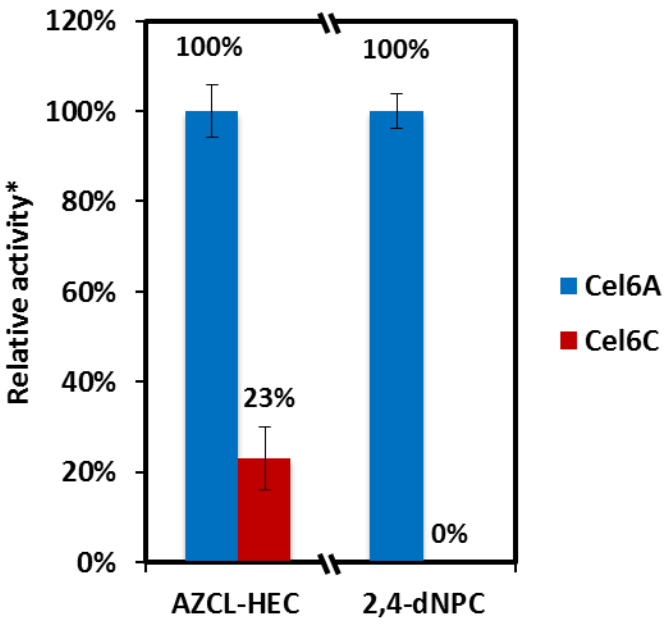
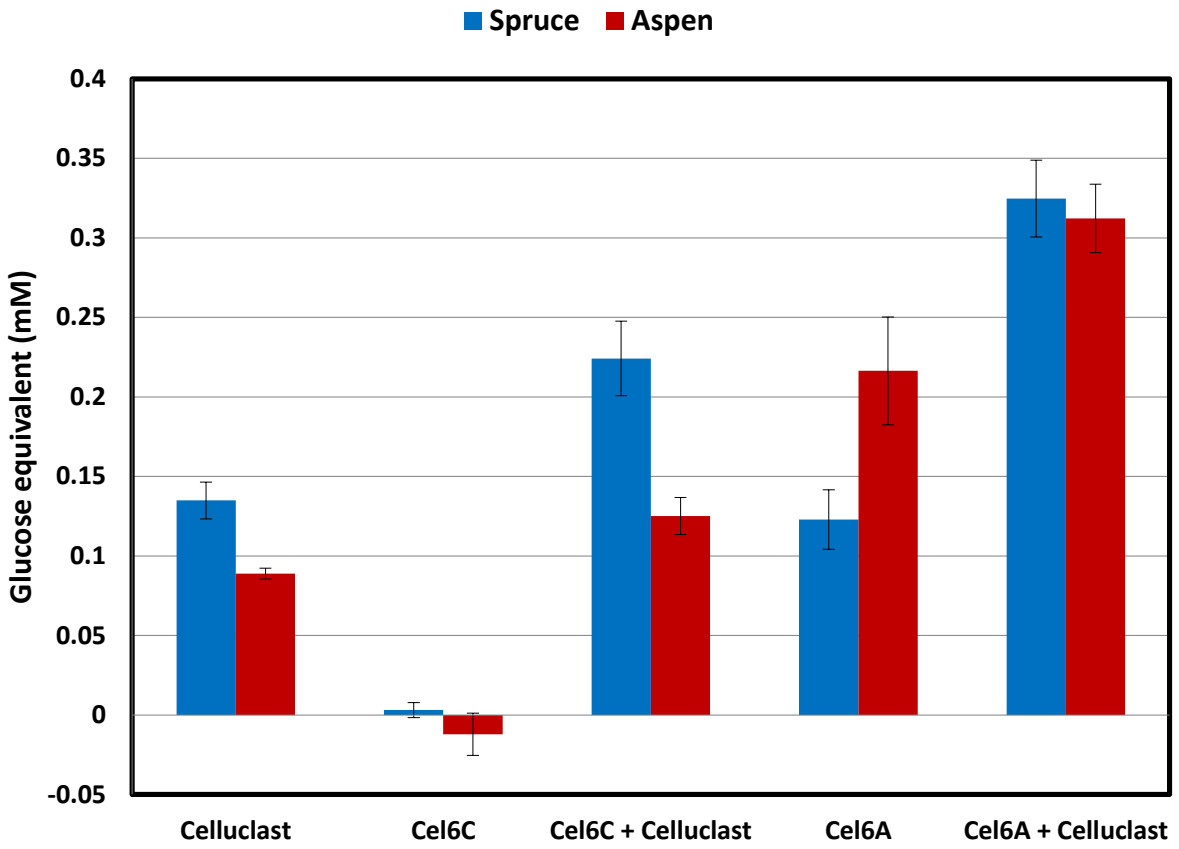


Figure 4.34. The relative activites of Cel6A and Cel6C on modified substrates. The relative activities were normalized to the highest activity for each substrate between the two enzymes.

Cel6C only has 23% activity of Cel6A on Azurine-crosslinked hydroxyethyl cellulose (AZCL-HEC). Cel6C has no activity at all on 2,4-dinitrophenyl cellobioside (2,4-dNPC) but Cel6A had an average Vmax of 14.2 (mili-A_{360nm}/min).

(a)



(b) ToF-SIMS Principal Component Analysis

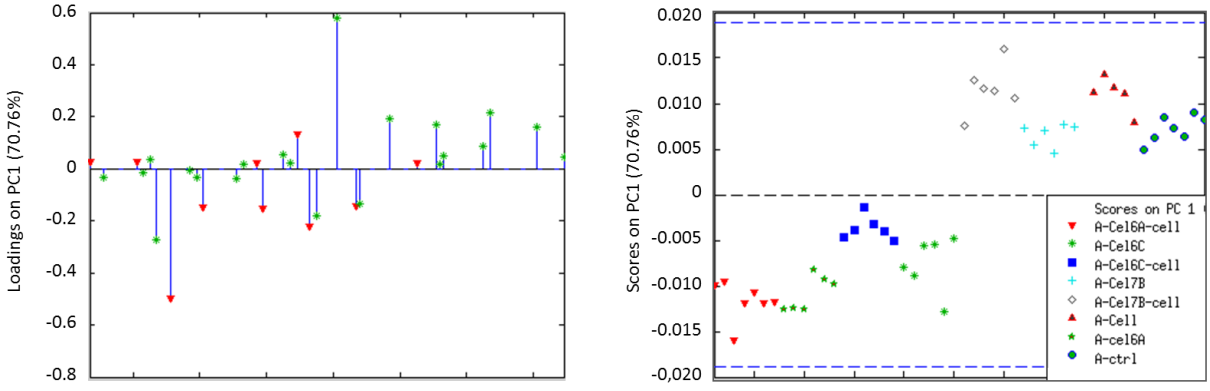


Figure 4.35. Hydrolytic activity on ball-milled biomass substrates. (a) DNS assay showed that Cel6A is able to hydrolyze spruce (softwood) or aspen (hardwood) on its own. Cel6A and Celluclast® (a commercial blend of Trichoderma cellulases) have an additive effect on these

substrates. Cel6C has little to no activity on these substrates on its own. However, with the addition of Cellulclast®, Cel6C shows some hydrolytic activity. (b) (right) ToF-SIMS analysis of the biomass substrates shows stark differences between the *C. fimi* GH6 cellulases (Cel6A and Cel6C) and the rest (Cel7B (*T. reesei* source), Cellulclast® only, and negative control). On the left, there is no clear pattern of separation between the lignin peaks (green stars) and the polysaccharide peaks (red triangles). Thus, the scores and clustering observed on the right graph cannot be attributed to polysaccharide hydrolysis but to some other parameter that is currently unknown.

The end-point assay was conducted by incubating the enzymes with the substrate at 45°C. The enzymatic activity was quantified based on reducing sugar release using the DNS assay.

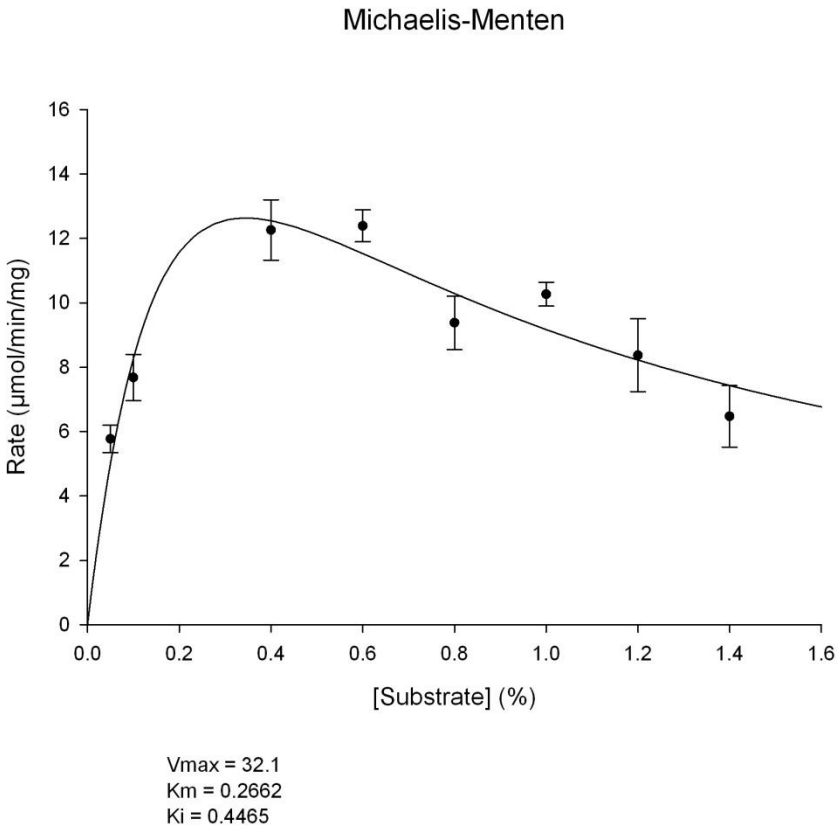


Figure 4.36. Michaelis-menten plot of Cel6C on CMC as a substrate. Cel6C shows substrate inhibition for [CMC] ≥ 0.45%.

The model with the best R-squared value is shown here, which is based on the substrate inhibition equation. The R-squared value here was 0.72447.

Chapter 5

Discussion

5.1 Results and their significance

Here I present evidence that Cel6C is a novel endoglucanase from *C. fimi*. Relative to known endoglucanases of the organism, this enzyme has significant thermostability over its counterparts. In addition, cellulases are typically multi-modular that consists of a signal peptide, a carbohydrate-binding module, and the catalytic domain. In the GH 6 family of *C. fimi*, Cel6A (an endoglucanase) has an N-terminal CBM approximately 100 amino acid residues long. On the other hand, Cel6B (a cellobiohydrolase) has a C-terminal CBM that is about 100 amino acid residues long and possibly longer depending on the actual boundary. Cel6C appears to have no CBM and its catalytic domain is nearly 100 residues longer than Cel6A's catalytic domain. The CBM enhances the enzymatic hydrolysis of cellulose by facilitating the association of the enzyme with the substrate. This is particularly evident when dealing with insoluble and highly inaccessible substrates. Gilkes et al[45] showed that the proteolytic excision of the cellulose-binding domain from CenA (Cel6A) results in an apparent increase in activity on soluble substrates but had a decrease against microcrystalline cellulose. The fact that Cel6C has no identifiable CBM explains why the enzyme has very little activity on pretreated insoluble biomass substrates. However, cellulases that do not have a CBM still require the ability to bind to the substrate during hydrolysis. Further research is required into the amino acid residues responsible for sugar-binding of catalytic domains.

Glycoside hydrolases can have one of three types of active sites: the pocket, the cleft, or the tunnel [38]. Classical endoglucanases have a relatively open cleft at its catalytic site and this allows it to attack at random internal sites of a cellulose chain [38]. Nonetheless, some endoglucanases act in a processive manner, where it cleaves cellulose at an internal point then releases oligosaccharides before detaching from the substrate [70,71]. In simplest terms, a processive

endoglucanase is a hybrid of endo- and exo-acting enzymes. It is normally required to have the synergistic effect of endo- and exoglucanases to solubilize filter paper but processive endoglucanases are able to do this on its own [70]. It has been shown previously that Cel6A acts in a relatively nonprocessive manner on CMC [72]. Overnight incubation of Cel6A or Cel6C with filter paper results in no detectable solubilization (data not shown). This suggests that the known GH6 *C. fimi* endoglucanases are relatively nonprocessive.

Based on results on PASC, modified, and conjugated-substrates, the substrate specificity of Cel6A and Cel6C provides clues to the shape of its active-site. Based on NCBI, both Cel6A and Cel6C amino acid sequence predicts cellobiohydrolase. Furthermore, Phyre2[7] (secondary structure prediction) also incorrectly models Cel6C to be a cellobiohydrolase (not shown) but correctly models Cel6A's catalytic domain to be an endoglucanase. To date, neither Cel6A nor Cel6C from *C. fimi* has had their molecular structure solved experimentally. Since Cel6C has aforementioned interesting qualities, it is only logical to determine its molecular structure. This would provide insight into its thermostability, binding cleft, and substrate-binding ability. An initial high-throughput screen conducted by the Hauptman-Woodward Institute (HWI) resulted only in salt crystals (not shown). Collaboration with Dr. Alisdair Boraston of the University of Victoria (Canada) is now underway to hopefully produce protein crystals for x-ray diffraction.

5.2 Future Directions

Enzyme engineering

Aside from structure determination of Cel6C, future directions involving this enzyme involve constructing a chimeric enzyme. The rationale behind this project is to increase the hydrolytic

[7] Protein Homology/analogY Recognition Engine V 2.0
http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index (Kelley and Sternberg 2009)

activity of Cel6C while maintaining or improving its thermostability. The objective would be to fuse one or more thermostable CBMs from other organisms to either the N- or C-terminal end of Cel6C catalytic domain. There is precedent in this type of work that involves optimizing the number of domains added as well as optimizing the linker regions between the domains [73]. A study showed that the addition of a CBM to endoglucanases, which naturally had no CBM, caused a two- to three- fold increase in activity on insoluble substrates [74]. Furthermore, there are naturally occurring multifunctional glycoside hydrolases that contain several catalytic domains from various families [75,76]. This can serve as precedent for attempting to fuse other catalytic domains onto Cel6C.

To further understand which regions or which residues are important for thermostability, random mutagenesis on Cel6C can be done. Mutations that lead to a significant reduction on thermostability would provide indicators as to which secondary structures or residues are crucial.

Other types of assays

The reducing sugar analysis on ball-milled biomass showed that Cel6C has little to no activity on this insoluble substrate. However, when simultaneously treated with the addition of Celluclast®, there appears to be a synergistic increase in liberated reducing sugars. This provides evidence that Cel6C has novel activity relative to Cel6A on natural substrates. Work will be continued on these biomass substrates to have a better understanding on the difference between these enzymes.

A protein sample of Cel6C has been given to Dr. Jose Moran-Mirabal of the McMaster University (Canada) for a unique assay. His group utilizes single-molecule investigation of biomolecular interactions. Using specially fabricated slides, they are able to maintain high temperatures and observe assays in real time using high resolution fluorescence microscopy.

Hopefully, this collaboration would elucidate the cellulase-cellulose interaction of Cel6C on various substrates.

Additionally, a more comprehensive examination of synergistic effects is needed to understand Cel6C's role in a complete cellulase system. This would require testing Cel6C with cellobiohydrolases, β-glucosidases, lytic polysaccharide monooxygenases, and hemicellulases on real biomass substrates.

5.3 Concluding remarks

Currently, cellulases for industrial applications are generally from fungal sources. This is due to the fact that fungal enzymes tend to be produce in copious amounts, have high catalytic activity, and some of which are thermostable and stable in non-neutral pH levels. Thermostability is a desired trait since the pretreatment results in a prolonged heated environment. Also, the Arrhenius' equation dictates that reaction rates are temperature dependent and positively correlated. However, it is also discussed that it is both economically inefficient and energetically wasteful to produce fuel that requires heat input. Therefore, progress needs to continue to find pretreatment alternatives that use less energy and remain effective. That being said, research into mesophilic cellulose degradation should go hand in hand. In the meantime, the discoveries and explanations that come from Cel6C should benefit the field of cellulases and provide new topics to explore.

Tables

Table 1. Gene targets and primers used

Gene name	Tube number	Primer Sequence (5' to 3')
Celf_1924	1 - Forward	ACTAAGGAATTCCATATGCATCATCACCATCACCACACGATCGTCGACTCGACCGGGAAGGAG
(aka Cel5A/CenD)	1 - Reverse	AAGTCAAAGCTTGTCGACCTAGCCGGTGTCGCCCGAGTTCGGGTT
Celf_1925	2 - Forward	ACTAAGGAATTCCATATGCATCATCACCATCACCACGTGAACCCCACGTGGGCCGCCTC
(aka Cel6B/CbhA/ Cex)	2 - Reverse	AAGTCAAAGCTTGTCGACCTAGCGGTCGAACCGCTTGCCCTGGTCG
Celf_0019	3 - Forward	ACTAAGGAATTCCATATGCATCATCACCATCACCACAACTACGCCGAGGCCCTGCAGAAGTC
(aka Cel9A/CenB)	3 - Reverse	AAGTCAAAGCTTGTCGACCTAAGCGAGGGCGCTGGTGAAGCC
Celf_1537	4 - Forward	ACTAAGGAATTCCATATGCATCATCACCATCACCACTACCAGCAGCTCCGGTACGACG
(aka Cel9B/CenC)	4 - Reverse	AAGTCAAAGCTTGTCGACCTACGAGGCGACCCAGGACAGC
Celf_0376	5 - Forward	ACTAAGGAATTCCATATGCATCATCACCATCACCACAACACCCTCGACGCGGTCGGC
	5 - Reverse	AAGTCAAAGCTTGTCGACCTAGCCGTTGTCCCACCACATCGTCGT
Celf_2403	6 - Forward	ACTAAGGAATTCCATATGCATCATCACCATCACCACGCGAACACCGTGCGCCTGCCGATCAAC
	6 - Reverse	AAGTCAAAGCTTGTCGACCTACCAGTAGACGAGCCCGATGCCCTGG
Cfla_0736	7 - Forward	ACTAAGGAATTCCATATGCATCATCACCATCACCACGACTACCTCGCGAGCCGCGGCGT
	7 - Reverse	AAGTCAAAGCTTGTCGACCTAGGAGTAGAAGTGCTGGGAGTACATGACC
Cfla_1897	8 - Forward	ACTAAGGAATTCCATATGCATCATCACCATCACCACAAGATCGTCGACGCCTCCGGCAAG
	8 - Reverse	AAGTCAAAGCTTGTCGACCTACCCGGTGTCACCCGAGTTCGGGTTGAGC
Cfla_2811	9 - Forward	ACTAAGGAATTCCATATGCATCATCACCATCACCACCTCGTCGAGCGCGACGGCACCCCCTTCG
	9 - Reverse	AAGTCAAAGCTTGTCGACCTACATGATCGTCGCCTCGTCGGGGTCACC
Celf_3184	10 - Forward	ACTAAGGAATTCCATATGCATCATCACCATCACCACGTCGACCCGACGACGCAGGGCTA
(aka Cel6A/CenA)	10 - Reverse	AAGTCAAAGCTTGTCGACCTACTCCTGCCACCACTGGCCGG
Celf_0233	11 - Forward	ACTAAGGAATTCCATATGCATCATCACCATCACCACTCCGCGGCCGCGCTGCTGAACAGGG
	11 - Reverse	AAGTCAAAGCTTGTCGACCTACTCCTCCCACAGCTGCCCCGCCGGG
Celf_1230	12 - Forward	ACTAAGGAATTCCATATGCATCATCACCATCACCACAACCCCGGCGCGACCGAGCAGA
	12 - Reverse	AAGTCAAAGCTTGTCGACCTACTGCGGGAACCACGCGCCCG
Cfla_2913	13 - Forward	ACTAAGGAATTCCATATGCATCATCACCATCACCACGTCGACCCCGAGACGGCGGCCTA
	13 - Reverse	AAGTCAAAGCTTGTCGACCTACGACTGCCACCAGGCGCCGGCC

Cfla_1896	14 - Forward	ACTAAGGAATTCCATATGCATCATCACCATCACCACGTGAACCCGAACTGGGCAGC GACGG
	14 - Reverse	AAGTCAAAGCTTGTCGACCTAGCGGTCGAAGCTCTTGCCCTCGTCGT
Cfla_2912	15 - Forward	ACTAAGGAATTCCATATGCATCATCACCATCACCACGTCGACACCACGAACCAGTC GTACC
	15 - Reverse	AAGTCAAAGCTTGTCGACCTACGACTGCCACCACTGACCGG
Celf_0045	16 - Forward	ACTAAGGAATTCCATATGCATCATCACCATCACCACAAGAAGCTGCGGTACGACGC GCTCGCG
	16 - Reverse	AAGTCAAAGCTTGTCGACCTAGGCGAGCCAGGCCAGCGCCGAGTT
Celf_1481	17 - Forward	ACTAAGGAATTCCATATGCATCATCACCATCACCACGACTCGCTCGTGGTGTTCTAC GG
	17 - Reverse	AAGTCAAAGCTTGTCGACCTACGCCGCGACCCAGGCCAGCGCCGAGTT
Celf_0019*	18 - Forward	ACTAAGGAATTCCATATGCATCATCACCATCACCACAACTACGCCGAGGCCCTGCA GAAGTC
	18 - Reverse	AAGTCAAAGCTTGTCGACCTAAGCGAGGGCGCTGGTGAAGC
Celf_1705	19 - Forward	ACTAAGGAATTCCATATGCATCATCACCATCACCACTACCAGCAGCTGCGGTACGA CG
	19 - Reverse	AAGTCAAAGCTTGTCGACCTACGACGCCACCCAGGACAGCGCCGAGT
Cfla_1515	20 - Forward	ACTAAGGAATTCCATATGCATCATCACCATCACCACTACGCGCAGCTGCGGTACGA CGC
	20 - Reverse	AAGTCAAAGCTTGTCGACCTACGACGCGACCCACGACAGCGCCGAGT
Cfla_3563	21 - Forward	ACTAAGGAATTCCATATGCATCATCACCATCACCACAAGCAGCTCCGGTACGACGC GCTCGC
	21 - Reverse	AAGTCAAAGCTTGTCGACCTACGCCGTCAGCCACGCGAACGCCGA
Cfla_0016	22 - Forward	ACTAAGGAATTCCATATGCATCATCACCATCACCACCTGCAGAAGTCGATGTTCTT
	22 - Reverse	AAGTCAAAGCTTGTCGACCTAGTACTGGGCCGTCAGGTACG
Cfla_0139	23 - Forward	ACTAAGGAATTCCATATGCATCATCACCATCACCACAACCTCGCGCAGGCCCTGCA GACCT
	23 - Reverse	AAGTCAAAGCTTGTCGACCTACGCGAGCGCGCCCACGAACCCGGCGTT
Cfla_3031	24 - Forward	ACTAAGGAATTCCATATGCATCATCACCATCACCACTACCAGGACCTGCGGCAGGA CG
	24 - Reverse	AAGTCAAAGCTTGTCGACCTACGACGCCACCCACGAGAGTGCCGAGT
Celf_1230 NC	NC-Forward	GAATTCCATATGGGTGGTCCGCCTCATGGTTATCCGAGCGGCACCACCCGTCTGTAT ACACCGCCTCCGAATCCGGGTGCAACCGAGCAG
	NC-Reverse	TCTAGAAAGCTTTTATTAGCTACGCGGACCACCAATAGGCGGAACTGCATTATGAA CCAGTTCCAGTGCCATCTGCGGAAACCATGCACCCGC
Celf_1230 C	C-Forward	CATATG-AATCCGGGTGCAACCGAGCAG
	C-Reverse	(same as NC-Reverse)

*Celf_0019 was designed and synthesized twice by mistake.

Theoretical primer Tm is not shown here since it varies greatly with the experimental annealing temperatures that were successful.

Table 2. RT-PCR primers

Gene name	Tube number	Primer Sequence (5' to 3')	Primer Tm (°C)
Celf_1924 (aka Cel5A/CenD)	1 - Forward	AAGGACAACTTCAAGCACTTCGCC	70.2
	1 - Reverse	AAGTCCTTCTGGAACCACTTCTGG	67.4
Celf_1925 (aka Cel6B/CbhA/Cex)	2 - Forward	GCTCGTGTTCAACCTCGTCATCTA	68.3
	2 - Reverse	AGTGGCCGATGTCGATGTAGTTGT	69.5
Celf_0019 (aka Cel9A/CenB)	3 - Forward	GTTCGTCAGCGACTACTTCGTCAA	68.4
	3 - Reverse	TACGTGCCGAACTGCTTGTTGT	68.7
Celf_1537 (aka Cel9B/CenC)	4 - Forward	TACGACGCGCTCAACTACTTCTAC	65.3
	4 - Reverse	TTGACGACGTACTTGCCGTGGT	70.6
Celf_0376	5 - Forward	TGCTGTTCGAGAGCATCAACGA	70.3
	5 - Reverse	AAGAACTTCAGCTTCTCGCCCT	66.4
Celf_2403	6 - Forward	AGAAGCCGTAGAAGTGCAGGGACA	70.8
	6 - Reverse	ACGAGAACAACCCGAAGGTCTACT	66.9
Cfla_0736	7 - Forward	ACCTGCTCAACGAGACCCACAT	69.1
	7 - Reverse	AAGTGCTGGGAGTACATGACCTTG	67.2
Cfla_1897	8 - Forward	AGTTCCTCGAGATGAGTGACGAGT	66.4
	8 - Reverse	GGCGAAGTGCTTGAAGTTGTCCTT	70.2
Cfla_2811	9 - Forward	GTGCAGATCAACATCGGCAACGAA	73.8
	9 - Reverse	TACATGTGGACGGAGAACAACGTG	69.8
Celf_3184 (aka Cel6A/CenA)	10 - Forward	TTCCTCAAGTACGCCGCCAAGT	70.6
	10 - Reverse	AGGTGTCGATGACGAACTTCTTGC	69.4
Celf_0233	11 - Forward	TTTCGCGGTCAACACGTCGAACTT	73.9
	11 - Reverse	TCTTGACCCACAGGTCCGCCA	74.3
Celf_1230	12 - Forward	ACGCCAAGGCCCTGAAGAAGAT	70.8
	12 - Reverse	TGAGCTGGTAGTTCGAGACGTTGA	69
Cfla_2913	13 - Forward	AAGATCGCGCAGATCGTCGTGTA	71.7
	13 - Reverse	ACGTGTCGATGACGAACTTCTTGC	70.3
Cfla_1896	14 - Forward	AGATGTTCGTGATGAGGTTCGGGA	71.5
	14 - Reverse	TCAACCTCGTCATCTACAACCTGC	68.2
Cfla_2912	15 - Forward	TCGACACCACGAACCAGTCGTA	69.9
	15 - Reverse	ACGTGTCGATGACGAACTTCTTGC	70.3
Celf_0045	16 - Forward	AACCCGCTGAACCAGTCGTACAT	69.5
	16 - Revesre	TTCCAGTTGATCGCGACCTCGTT	72.6
Celf_1481	17 - Forward	AACGGCCTGCTGCTCAACAA	70.2
	17 - Reverse	AGTTGATCGTCATCTCGTTCACGC	70.3
Celf_0019*	~~18 - Forward~~	~~AAGTCGATGTTCTTCTACCAGGCG~~	
	~~18 - Reverse~~	~~TTGACGAAGTAGTCGCTGACGAAC~~	
Celf_1705	19 - Forward	TACGACGCGCTCGACTACTTCTA	66.9
	19 - Reverse	TCCAGTCGACGTCATGGACCTTGT	72.8
Cfla_1515	20 - Forward	TACGACGCGCTGAACTACTTCTACCT	68.5

	20 - Reverse	AACTCCTGGTCGTACTCGCTGAA	68
Cfla_3563	21 - Forward	ACAAGATGCACGACGAGAACTGGA	71
	21 - Reverse	TTGTAGAATGACGAGCCCGTGACA	71.4
Cfla_0016	22 - Forward	ACGACTACTTCGTCAAGGCACACA	68.8
	22 - Reverse	TCTCGTTGCTCAGCTTGTCGTACT	68.2
Cfla_0139	23 - Forward	AAGTACTGGTACTACTTCCGCTGG	63.3
	23 - Reverse	TGTCGCCCATGAGGTAGTTCATCT	69.6
Cfla_3031	24 - Forward	CTCGATCCTCAACAACCAGGTGAT	69.2
	24 - Reverse	GTTCCAGTTGACGGTGATCTCGTT	69.2
C. fimi 16s	25 - Forward	GCGGATTAATTCGATGCAACGCGA	75.4
	25 - Reverse	TGCGGGACTTAACCCAACATCTCA	72
C. flavigena 16s	26 - Forward	TTTCAGCAGGGAAGAAGCGAGAGT	70.1
	26 - Reverse	GCCCAATAATTCCGGACAACGCTT	72.6

*This Celf_0019 RT-PCR primer is different from Tube 3 but was not ordered since it serves the same purpose.

Table3. Sequencing primers used

Oligo Name	Sequence (5' to 3')	Length	Direction	Tm (°C)	Comments
PCW-FP	TTG CGC CAT TCG ATG GTG TC	20	FP	58.6	pCW primer, ~400 bp from end
PCW-RP	TTC GTC TTC AAG CAG ATC TG	20	RP	52.3	pCW primer, ~100 bp from end
MB1041	GGT GAT CAA CGC CGC CAG CGG TCG	24	FP	73.1	forward primer for malE, after malE-F2
RCW70	AGG CCC TTT CGT CTT CAA GCA GAT C	25	RP	66.2	pCW primer, 70 bp from end
FCW1	ACAGGATCCATCGATGCTTAGGAG	24	FP	64.6	pCW primer, covers Bam HI site
RCW20	TGTTTGACAGCTTATCATCG	20	RP	56.3	pCW primer, 20 bp from end
malE-F1	GAA GCG TTA TCG CTG ATT TAT AAC	24	FP	52.2	forward primer for malE
malE-F2	CCG TTC GTT GGC GTG CTG AG	20	FP	61.6	forward primer for malE
malE-R1	CTC AGC ACG CCA ACG AAC GG	20	RP	61.6	reverse primer for malE
malE-R2	GTT ATA AAT CAG CGA TAA CGC TTC	24	RP	52.2	reverse primer for malE

Supplementary information

Figure S1. Homogenizer using an overhead stirrer. Image is a representation of the equipment used.

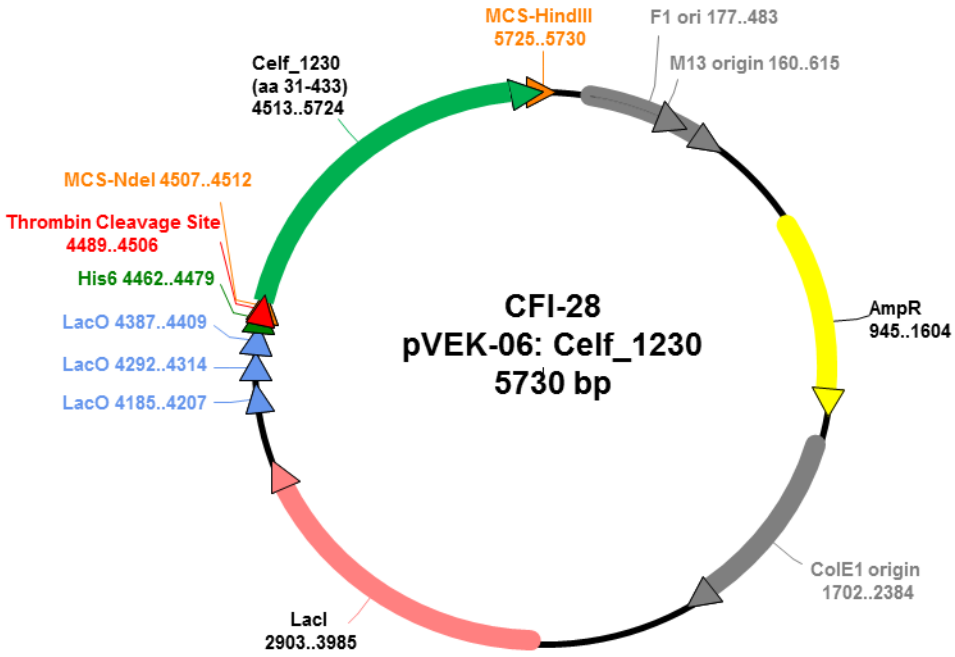


Figure S2. Vector map of Cfi-28. Celf_1230 aa31-433 inserted into the multiple cloning site of pVEK-06.

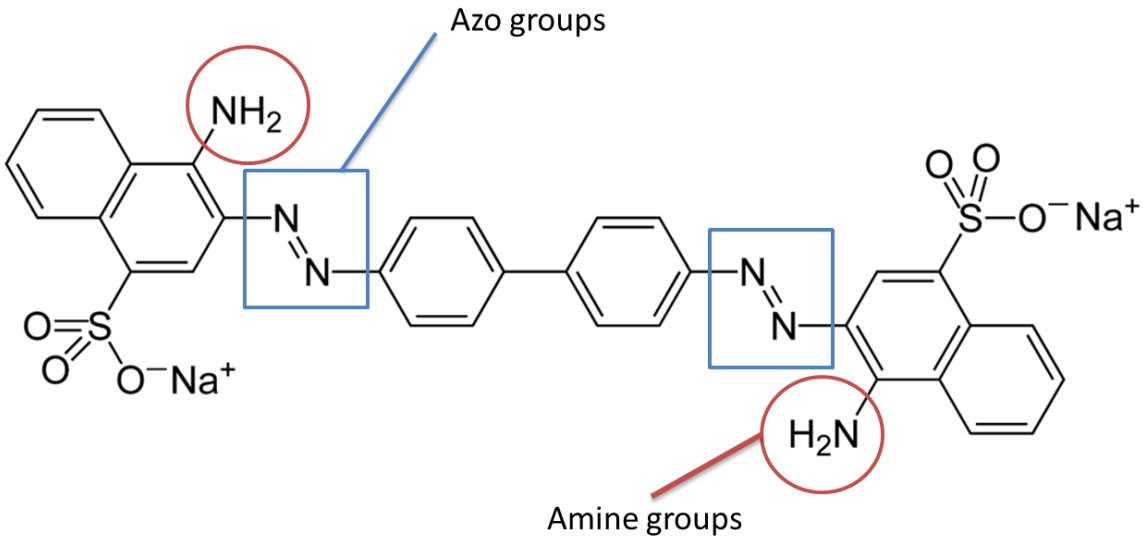


Figure S3. sodium salt of 3,3'-([1,1'-biphenyl]-4,4'-diyl)bis(4-aminonaphthalene-1-sulfonic acid) or Congo Red. The hydroxyl groups of the cellulose chain forms hydrogen bonds with the azo and amine groups of congo red. Adapted with permission from Puchtler et al. On the Binding of Congo Red by Amyloid. Journal of Histochemistry & Cytochemistry. Copyright (1962). SAGE Publications.

S4. Nucleotide sequence of Celf_1230 (aa31-433) that was codon-optimized for *E. coli* expression.

```
ATGGGTGGTCCGCCTCATGGTTATCCGAGCGGCACCACCCGTCTGTATAC          50
ACCGCCTCCGAATCCGGGTGCAACCGAGCAGATTGCACAGCTGCTGCGTG          100
ATCGTCAGTATGCAGATGCAAAAGCACTGAAAAAAATGGTTGCAACACCG          150
CAGGCAGTTTGGTTTACCAAAGGTACACCGCAAGAAGTTCGTAAAGCAGT          200
TGATAGCACCGTTCGTCAGGCAAAAAAACAGCGTGCAGTTCCGGTTATTG          250
TTGCATATAATCTGCCGTTTCGTGATTGTGCACAGTATAGTGCCGGTGGT          300
```

GCACTGGATACCGCAAGCTATCTGGCATGGGTTGATGGTCTGGCAAAAGG 350

TATTGGTAAAAGCGAAGCAATTGTTCTGCTGGAACCGGATGGCCTGGGTA 400

TTATTCCGTGGTATACAACAATTAACGGCGATCAAGAATGGTGTCAGCCT 450

GCAGAAGCAGATCCGGCAACCGCAGCAGCAGATCGTTTTGCCCAGCTGAG 500

CGGTGCAGTTGATCGTCTGAGCGCACTGCCGAATGTTAGCCTGTATCTGG 550

ATGGCACCCATAGCGGTTGGCTGGGTGCGGGTGATGCAGCGGATCGTCTG 600

CATAAAGCCGGTGTTGCCAAAACCGATGGTTTTTTTCTGAATGTGAGCAA 650

CTATCAGCTGACCGAACGTCAGGTTAAATATGCAAGCTGGGTTAGCCAGT 700

GTCTGTGGTATGGCACCAATGATGCAGAAGGTGGTTGGCGTCTGGGTCAT 750

TTTGAACATTGTGGTAGCCAGTATTATCCGGCAGATCCGAATGATTTTAG 800

CACCTGGGGTCTGACCGATCAGTGGTATCTGGACAATGTTACCAATGCAG 850

CAAATCCGCCTAGCGGTCCGGAAGTTCTGGCACATGCAGTTATTGATACC 900

AGCCGTAATGGTCAGGGTCCGTGGACCGCACCGGCTGATCGTCCTGCCGG 950

TGATGCCCAAGAGTGGTGTAATCCGCCTGATCGTGGTCTGGGTCCGCGTC 1000

CGACCACCAGTACCGGTGATCCGTATGTTGATGCATATCTGTGGGTTAAA 1050

ATTCCGGGTGAAAGTGATGGTCAGTGTTTTCGTTGGGCACCGGAAAGCGG 1100

TATTGATCCGGTTCGTGGTTATCCGGGTCCGGCAGCGGGTGCATGGTTTC 1150

CGCAGATGGCACTGGAACTGGTTCATAATGCAGTTCCGCCTATTGGTGGT 1200

CCGCGTAGC 1209

Permissions

Selected figures used in this thesis with permission that was granted for publication by the following permission statements.

Figure 2.1. Structure of lignocellulose.

NATURE PUBLISHING GROUP LICENSE TERMS AND CONDITIONS

Aug 20, 2013

This is a License Agreement between Marc de Asis ("You") and Nature Publishing Group ("Nature Publishing Group") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number 3213400626102
License date Aug 20, 2013
Licensed content publisher Nature Publishing Group
Licensed content publication Nature
Licensed content title Genomics of cellulosic biofuels
Licensed content author Edward M. Rubin
Licensed content date Aug 14, 2008
Volume number 454
Issue number 7206
Type of Use reuse in a thesis/dissertation
Requestor type academic/educational
Format print and electronic
Portion figures/tables/illustrations
Number of figures/tables /illustrations
1
High-res required no
Figures Figure 2. Structure of lignocellulose.
Author of this NPG article no
Your reference number
Title of your thesis / dissertation
CLONING AND CHARACTERIZATION OF NEW CELLULASES FROM CELLULOMONAS FIMI AND CELLULOMONAS FLAVIGENA
Expected completion date Sep 2013
Estimated size (number of pages)
100
Total 0.00 USD
Terms and Conditions

Terms and Conditions for Permissions

Nature Publishing Group hereby grants you a non-exclusive license to reproduce this

material for this purpose, and for no other use,subject to the conditions below:
NPG warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is

original to Nature Publishing Group and does not carry the copyright of another entity (as credited in the published version). If the credit line on any part of the material you have requested indicates that it was reprinted or adapted by NPG with permission from another source, then you should also seek permission from that source to reuse the material.

1. Permission granted free of charge for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to the work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.Where print permission has been granted for a fee, separate permission must be obtained for any additional, electronic re-use (unless, as in the case of a full paper, this has already been accounted for during your initial request in the calculation of a print run).NB: In all cases, web-based use of full-text articles must be authorized separately through the 'Use on a Web Site' option when requesting permission.
2. Permission granted for a first edition does not apply to second and subsequent editions and for editions in other languages (except for signatories to the STM Permissions Guidelines, or where the first edition permission was granted for free).
3. Nature Publishing Group's permission must be acknowledged next to the figure, table or abstract in print. In electronic form, this acknowledgement must be visible at the same time as the figure/table/abstract, and must be hyperlinked to the journal's homepage.
4. The credit line should read:

Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)
For AOP papers, the credit line should read:
Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

Note: For republication from the *British Journal of Cancer,* the following credit lines apply.

Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)For AOP papers, the credit line should read:
Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

5. Adaptations of single figures do not require NPG approval. However, the adaptation should be credited as follows:

Adapted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

Note: For adaptation from the *British Journal of Cancer,* the following credit line applies.

Adapted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

6. Translations of 401 words up to a whole article require NPG approval. Please visit http://www.macmillanmedicalcommunications.com for more information.Translations

of up to a 400 words do not require NPG approval. The translation should be credited as follows:
Translated by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication).
Note: For translation from the *British Journal of Cancer,* the following credit line applies.
Translated by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

7. We are certain that all parties will benefit from this agreement and wish you the best in the use of this material. Thank you.

Figure 2.3. General mechanisms for (a) inverting and (b) retaining glycosidases.

ELSEVIER LICENSE
TERMS AND CONDITIONS

Aug 21, 2013

This is a License Agreement between Marc de Asis ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier Elsevier Limited
The Boulevard,Langford Lane
Kidlington,Oxford,OX5 1GB,UK
Registered Company Number
1982084
Customer name Marc de Asis
Customer address 200 University Ave W
Waterloo, ON N2L3G1
License number 3213910125661
License date Aug 21, 2013
Licensed content publisher Elsevier
Licensed content publication Elsevier Books
Licensed content title Advances in Applied Microbiology, Volume 70
Licensed content author Carl J. Yeoman,Yejun Han,Dylan Dodd,Charles M. Schroeder,Roderick I. Mackie,Isaac K.O. Cann
Licensed content date 2010
Number of pages 55
Start Page 1
End Page 55
Type of Use reuse in a thesis/dissertation
Intended publisher of new work
other
Portion figures/tables/illustrations
Number of figures/tables /illustrations
1
Format both print and electronic
Are you the author of this Elsevier chapter?
No
Will you be translating? No
Order reference number
Title of your thesis/dissertation
CLONING AND CHARACTERIZATION OF NEW CELLULASES FROM CELLULOMONAS FIMI AND CELLULOMONAS FLAVIGENA

Expected completion date Sep 2013
Estimated size (number of pages)
100
Elsevier VAT number GB 494 6272 12
Permissions price 0.00 USD
VAT/Local Sales Tax 0.0 USD / 0.0 GBP
Total 0.00 USD

Terms and Conditions

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the

combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement

and publisher reserves the right to take any and all action to protect its copyright in the materials.
9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.
10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.
12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).
13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.
14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the

amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:
15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.
16. **Website**: The following terms and conditions apply to electronic reserve and author websites:
Electronic reserve: If licensed material is to be posted to website, the web site is to be password-protected and made available only to bona fide students registered on a relevant course if:
This license was made in connection with a course,
This permission is granted for 1 year only. You may obtain a license for future website posting,

All content posted to the web site must maintain the copyright information line on the bottom of each image,
A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com , and
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
17. **Author website** for journals with the following additional clauses:
All content posted to the web site must maintain the copyright information line on the bottom of each image, and the permission granted is limited to the personal version of your paper. You are not allowed to download and post the published electronic version of your article (whether PDF or HTML, proof or final version), nor may you scan the printed edition to create an electronic version. A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx . As part of our normal production process, you will receive an e-mail notice when your article appears on Elsevier's online service ScienceDirect (www.sciencedirect.com). That e-mail will include the article's Digital Object Identifier (DOI). This number provides the electronic link to the published article and should be included in the posting of your personal version. We ask that you wait until you receive this e-mail and have the DOI to do any posting.
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
18. **Author website** for books with the following additional clauses:

Authors are permitted to place a brief summary of their work online only.
A hyper-text must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version.
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
19. **Website** (regular and for author): A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal /xxxxx. or for books to the Elsevier homepage at http://www.elsevier.com
20. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.
21. **Other Conditions**:
v1.6
If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number

RLNK501095583.
Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.
Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006
For suggestions or comments regarding this order, contact RightsLink Customer Support: customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.
Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

Figure 2.4. Enzyme classes of cellulases.

Figure S3. sodium salt of 3,3'-([1,1'-biphenyl]-4,4'-diyl)bis(4-aminonaphthalene-1-sulfonic acid) or Congo Red.

References

1. Sun, Y. & Cheng, J. Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresource Technology* **83,** 1–11 (2002).

2. Demirbas, A. Progress and recent trends in biofuels. *Progress in Energy and Combustion Science* **33,** 1–18 (2007).

3. Rubin, E. M. Genomics of cellulosic biofuels. *Nature* **454,** 841–5 (2008).

4. Zinoviev, S. *et al.* Next-generation biofuels: Survey of emerging technologies and sustainability issues. *ChemSusChem* **3,** 1106–33 (2010).

5. Maki, M., Leung, K. T. & Qin, W. The prospects of cellulase-producing bacteria for the bioconversion of lignocellulosic biomass. *Int J Biol Sci* **5,** 500–16 (2009).

6. Buléon, A., Colonna, P., Planchot, V. & Ball, S. Starch granules: structure and biosynthesis. *Int J Biol Macromolec* **23,** 85–112 (1998).

7. Van der Maarel, M. J. E. C., van der Veen, B., Uitdehaag, J. C. M., Leemhuis, H. & Dijkhuizen, L. Properties and applications of starch-converting enzymes of the alpha-amylase family. *J Biotechnol* **94,** 137–55 (2002).

8. Kandra, L. α-Amylases of medical and industrial importance. *Journal of Molecular Structure: THEOCHEM* **666-667,** 487–498 (2003).

9. Fengel, D. & Wegener, G. *Wood: chemistry, ultrastructure, reactions*. (Walter de Gruyter, 1983).

10. Han, Y. W. & Srinivasan, V. R. Isolation and characterization of a cellulose-utilizing bacterium. *Applied Microbiology* **16,** 1140–5 (1968).

11. Geddes, C. C., Nieves, I. U. & Ingram, L. O. Advances in ethanol production. *Current Opinion in Biotechnology* **22,** 312–9 (2011).

12. Huttermann, A. & Majcherczyk, A. in *Wood Production, Wood Technology, and Biotechnological Impacts* (Kues, U.) 537–554 (Goettingen University Press, 2007).

13. Harada, H. & Cote Jr., W. A. in *Biosynthesis and Biodegradation of Wood Components* 1–42 (1985).

14. Fan, L., Lee, Y. & Gharpuray, M. The nature of lignocellulosics and their pretreatments for enzymatic hydrolysis. *Microbial Reactions* 157–187 (1982). at <http://link.springer.com/chapter/10.1007/3540116982_4>

15. Boerjan, W., Ralph, J. & Baucher, M. Lignin biosynthesis. *Annu Rev Plant Biol* **54,** 519–46 (2003).

16. Hendriks, A. T. W. M. & Zeeman, G. Pretreatments to enhance the digestibility of lignocellulosic biomass. *Bioresource Technology* **100,** 10–8 (2009).

17. Sjöström, E. *Wood chemistry: fundamentals and applications*. (Academic Press, 1993).

18. Ericsson, T., Petersson, G. & Samuelson, O. Galacturonic acid groups in birch xylan. *Wood Science and Technology* **223,** 219–223 (1977).

19. Laureano-Perez, L., Teymouri, F., Alizadeh, H. & Dale, B. E. Understanding factors that limit enzymatic hydrolysis of biomass. *Applied Biochemistry and Biotechnology* **124,** 1081–1099 (2005).

20. Sweet, M. S. & Winandy, J. E. Influence of Degree of Polymerization of Cellulose and Hemicellulose on Strength Loss in Fire-Retardant-Treated Southern Pine. *Holzforschung* **53,** 311–317 (1999).

21. Jarvis, M. Structure and properties of pectin gels in plant cell walls. *Plant, Cell & Environment* **7,** 153–164 (1984).

22. Brown Jr, R. The Biosynthesis of Cellulose. *Journal of Macromolecular Science, Part A: Pure and Applied Chemistry* **33,** 1345–1373 (1996).

23. Mosier, N. *et al.* Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresource Technology* **96,** 673–86 (2005).

24. Klemm, D., Heublein, B., Fink, H.-P. & Bohn, A. Cellulose: fascinating biopolymer and sustainable raw material. *Angewandte Chemie (International ed. in English)* **44,** 3358–93 (2005).

25. Hori, H. & Elbein, A. D. in *Biosynthesis and Biodegradation of Wood Components* 109–139 (1985).

26. Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research* **37,** D233–8 (2009).

27. Henrissat, B. & Davies, G. Structural and sequence-based classification of glycoside hydrolases. *Current Opinion in Structural Biology* **7,** 637–44 (1997).

28. Henrissat, B. & Bairoch, A. Updating the sequence-based classification of glycosyl hydrolases. *Biochem J* **316 (Pt 2,** 695–6 (1996).

29. Barras, F. *et al.* Stereochemistry of the hydrolysis reaction catalyzed by endoglucanase Z from *Erwinia chrysanthemi*. *FEBS Lett* **300,** 145–8 (1992).

30. Knowles, J. K. C., Lentovaara, P., Murray, M. & Sinnott, M. L. Stereochemical course of the action of the cellobioside hydrolases I and II of *Trichoderma reesei*. *Journal of the Chemical Society, Chemical Communications* 1401 (1988). doi:10.1039/c39880001401

31. Gebler, J. *et al.* Stereoselective hydrolysis catalyzed by related beta-1,4-glucanases and beta-1,4-xylanases. *J Biol Chem* **267,** 12559–61 (1992).

32. Dominguez, R. *et al.* A common protein fold and similar active site in two distinct families of beta-glycanases. *Nature Structural Biology* **2,** 569–76 (1995).

33. Irwin, D. *et al.* Roles of the catalytic domain and two cellulose binding domains of *Thermomonospora fusca* E4 in cellulose hydrolysis. *J Bacteriol* **180,** 1709–14 (1998).

34. Henrissat, B. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* **280 (Pt 2,** 309–16 (1991).

35. Vuong, T. V & Wilson, D. B. Glycoside hydrolases: catalytic base/nucleophile diversity. *Biotechnol Bioeng* **107,** 195–205 (2010).

36. McCarter, J. D. & Withers, S. G. Mechanisms of enzymatic glycoside hydrolysis. *Current Opinion in Structural Biology* **4,** 885–92 (1994).

37. Zechel, D. L. & Withers, S. G. Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Acc Chem Res* **33,** 11–8 (2000).

38. Davies, G. & Henrissat, B. Structures and mechanisms of glycosyl hydrolases. *Structure (London, England : 1993)* **3,** 853–9 (1995).

39. Breznak, J. & Brune, A. Role of microorganisms in the digestion of lignocellulose by termites. *Annual Review of Entomology* **39,** 453–487 (1994).

40. Bayer, E. A., Chanzy, H., Lamed, R. & Shoham, Y. Cellulose, cellulases and cellulosomes. *Current Opinion in Structural Biology* **8,** 548–57 (1998).

41. Meinke, A., Gilkes, N. R., Kilburn, D. G., Miller, R. C. & Warren, R. A. Multiple domains in endoglucanase B (CenB) from *Cellulomonas fimi*: functions and relatedness to domains in other polypeptides. *J Bacteriol* **173,** 7126–35 (1991).

42. Mejia-Castillo, T., Hidalgo-Lara, M. E., Brieba, L. G. & Ortega-Lopez, J. Purification, characterization and modular organization of a cellulose-binding protein, CBP105, a processive beta-1,4-endoglucanase from *Cellulomonas flavigena*. *Biotechnology Letters* **30,** 681–7 (2008).

43. Ilmberger, N. & Streit, W. R. Screening for cellulase-encoding clones in metagenomic libraries. *Methods in Molecular Biology (Clifton, N.J.)* **668,** 177–88 (2010).

44. Gilkes, N. R., Kilburn, D. G., Miller, R. C. & Warren, R. a. Structural and functional analysis of a bacterial cellulase by proteolysis. *J Biol Chem* **264,** 17802–8 (1989).

45. Gilkes, N. R., Warren, R. A., Miller, R. C. & Kilburn, D. G. Precise excision of the cellulose binding domains from two *Cellulomonas fimi* cellulases by a homologous protease and the effect on catalysis. *J Biol Chem* **263,** 10401–7 (1988).

46. Flint, H. J., Bayer, E. A., Rincon, M. T., Lamed, R. & White, B. A. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nature reviews. Microbiology* **6,** 121–31 (2008).

47. Hess, M. *et al.* Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* **331,** 463–7 (2011).

48. Torsvik, V. & Øvreås, L. Microbial diversity and function in soil: from genes to ecosystems. *Current Opinion in Microbiology* **5,** 240–5 (2002).

49. Goyal, A., Ghosh, B. & Eveleigh, D. Characteristics of fungal cellulases. *Bioresource Technology* **36,** 37–50 (1991).

50. Ljungdahl, L. G. in *Enzyme Systems for Lignocellulose Degradation* 5–16 (1989).

51. Sørensen, H. P. & Mortensen, K. K. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb Cell Fact* **4,** 1 (2005).

52. Gasser, B. *et al.* Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. *Microb Cell Fact* **7,** 11 (2008).

53. Wilson, D. B. Microbial diversity of cellulose hydrolysis. *Current Opinion in Microbiology* **14,** 259–63 (2011).

54. Dror, T., Rolider, A. & Bayer, E. Regulation of major cellulosomal endoglucanases of *Clostridium thermocellum* differs from that of a prominent cellulosomal xylanase. *J Bacteriol* **187,** 2261–2266 (2005).

55. Anderson, I. *et al.* Genomics of aerobic cellulose utilization systems in actinobacteria. *PloS one* **7,** e39331 (2012).

56. Abt, B. *et al.* Complete genome sequence of *Cellulomonas flavigena* type strain (134). *Stand Genomic Sci* **3,** 15–25 (2010).

57. Stackebrandt, E., Schumann, P. & Prauser, H. in *The Prokaryotes Volume 3: Archaea. Bacteria: Firmicutes, Actinomycetes* (Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H. & Stackebrandt, E.) 983–1001 (Springer New York, 2006). doi:10.1007/0-387-30743-5

58. Chaudhary, P., Kumar, N. N. & Deobagkar, D. N. The glucanases of *Cellulomonas*. *Biotechnology advances* **15,** 315–31 (1997).

59. Rajoka, M. I. & Malik, K. a. Cellulase production by *Cellulomonas biazotea* cultured in media containing different cellulosic substrates. *Bioresource Technology* **59,** 21–27 (1997).

60. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227,** 680–685 (1970).

61. Meinke, a, Gilkes, N. R., Kilburn, D. G., Miller, R. C. & Warren, R. a. Cellulose-binding polypeptides from *Cellulomonas fimi*: endoglucanase D (CenD), a family A beta-1,4-glucanase. *J Bacteriol* **175,** 1910–8 (1993).

62. Wong, W. K. *et al.* Characterization and structure of an endoglucanase gene cenA of *Cellulomonas fimi*. *Gene* **44,** 315–24 (1986).

63. Damude, H. G., Ferro, V., Withers, S. G. & Warren, R. A. Substrate specificity of endoglucanase A from *Cellulomonas fimi*: fundamental differences between endoglucanases and exoglucanases from family 6. *Biochem J* **315 (Pt 2,** 467–72 (1996).

64. Cockburn, D. W., Vandenende, C. & Clarke, A. J. Modulating the pH-activity profile of cellulase by substitution: replacing the general base catalyst aspartate with cysteinesulfinate in cellulase A from *Cellulomonas fimi*. *Biochemistry* **49,** 2042–50 (2010).

65. Tull, D. & Withers, S. G. Mechanisms of cellulases and xylanases: a detailed kinetic study of the exo-beta-1,4-glycanase from *Cellulomonas fimi*. *Biochemistry* **33,** 6363–70 (1994).

66. Lever, M. Colorimetric and Fluorometric Carbohydrate determination with p-Hydroxybenzoic Acid hydrazide. *Biochemical Medicine* **7,** 274–281 (1973).

67. Mewis, K. *et al.* Biomining active cellulases from a mining bioremediation system. *Journal of Biotechnology* 1–9 (2013). doi:10.1016/j.jbiotec.2013.07.015

68. Goacher, R. E., Jeremic, D. & Master, E. R. Expanding the library of secondary ions that distinguish lignin and polysaccharides in time-of-flight secondary ion mass spectrometry analysis of wood. *Analytical Chemistry* **83,** 804–12 (2011).

69. Wilkins, M. R. *et al.* Protein identification and analysis tools in the ExPASy server. *Methods in Molecular Biology (Clifton, N.J.)* **112,** 531–52 (1999).

70. Gilad, R. *et al.* CelI, a Noncellulosomal Family 9 Enzyme from *Clostridium thermocellum*, Is a Processive Endoglucanase That Degrades Crystalline Cellulose. *J Bacteriol* **185,** 391–398 (2003).

71. Li, Y., Irwin, D. C. & Wilson, D. B. Processivity, substrate binding, and mechanism of cellulose hydrolysis by *Thermobifida fusca* Cel9A. *Appl Environ Microbiol* **73,** 3165–72 (2007).

72. Gilkes, N. R., Langsford, M. L., Kilburn, D. G., Miller, R. C. & Warren, R. a. Mode of action and substrate specificities of cellulases from cloned bacterial genes. *J Biol Chem* **259,** 10455–9 (1984).

73. Wang, Y. *et al.* Chimeric cellulase matrix for investigating intramolecular synergism between non-hydrolytic disruptive functions of carbohydrate-binding modules and catalytic hydrolysis. *J Biol Chem* **287,** 29568–78 (2012).

74. Reyes-Ortiz, V. *et al.* Addition of a carbohydrate-binding module enhances cellulase penetration into cellulose substrates. *Biotechnology for Biofuels* **6,** 93 (2013).

75. Xue, G., Gobius, K. & Orpin, C. polysaccharide hydrolase cDNA (celD) from *Neocallimastix patriciarum* encoding three multi-functional catalytic domains with high endoglucanase, cellobiohydrolase. *Journal of General Microbiology* **138,** 2397–2403 (1992).

76. Flint, H. J., Martin, J., McPherson, C. A., Daniel, A. S. & Zhang, J. X. A bifunctional enzyme, with separate xylanase and beta(1,3-1,4)-glucanase domains, encoded by the *xynD* gene of *Ruminococcus flavefaciens*. *J Bacteriol* **175**, 2943–51 (1993).