MECHANISTIC STUDIES OF *THERMOBIFIDA FUSCA* EXOCELLULASE CEL6B

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Thu Van Vuong

February 2010

MECHANISTIC STUDIES OF *THERMOBIFIDA FUSCA* EXOCELLULASE CEL6B

Thu Van Vuong, Ph. D.

Cornell University 2010

Understanding the mechanism by which cellulases catalyze cellulose hydrolysis can greatly contribute to the development of biofuels. The thermophilic bacterium *Thermobifida fusca,* a major degrader of plant cell walls in certain environments, secretes seven different cellulases including exocellulase Cel6B. This cellulase acts by an inverting mechanism; however, its catalytic acid and base residues had not been identified. Biochemical approaches confirmed D274 to be the catalytic acid residue. A single catalytic base residue could not be determined, as sodium azide assays showed no activity rescue for any single mutations of candidate residues. However, a double mutation of D226A and S232A knocked out enzymatic activity and its activity was partially rescued by sodium azide. We therefore propose a novel hydrolysis mechanism for *T. fusca* Cel6B involving a proton-transferring network to carry out the catalytic base function.

T. fusca exocellulase Cel6B was also engineered to gain knowledge on the relationship between processivity and synergism as these properties are important for hydrolyzing crystalline cellulose. Mutations of several residues in the active site tunnel of Cel6B gave higher processivity. This improvement was confirmed by two assays: the ratio of soluble/insoluble reducing sugars as well as the ratio of oligosaccharide products. Surprisingly, the mutant enzyme, which has the highest processivity, showed the least synergism in mixtures with endocellulases, suggesting

that improving exocellulase processivity might not always be an effective strategy for producing improved cellulase mixtures for biomass conversion.

The highly processive Cel6B mutant enzymes were successfully fluorescently labeled, so these species can be used to visualize binding and track their movement on cellulose. The catalytic domains of Cel6B was found to bind non-productively to other polysaccharides; therefore, the balance between specific binding and non-specific adsorption should be always considered when engineering cellulases for hydrolyzing complex substrates. Using immuno-precipitation, Cel6B was demonstrated to contribute greatly to the hydrolysis of crystalline cellulose by *T. fusca*.

BIOGRAPHICAL SKETCH

Thu V. Vuong was born in a small village, approximately 40 miles from Hanoi, Vietnam. He received a Bachelor of Biotechnology from the College of Science, Vietnam National University, Hanoi, and then he worked for Department of Biochemistry and Plant Physiology, College of Science for a year before pursuing a Master's degree at Flinders University, Australia. After Thu received a Master of Biotechnology from Department of Medical Biotechnology, School of Medicine, Flinders University, he went to University of Illinois, Urbana-Champaign, Illinois for a summer ESL course before entering the graduate program of Environmental Toxicology at Cornell University to work with Dr. David Wilson on cellulases.

To my dear parents

11.20.2009

ACKNOWLEDGMENTS

It is my fortune to work with Dr. David Wilson. I would like to thank him for his encouragement and the wonderful opportunity to do research in his laboratory. I would also like to thank my committee members Dr. Anthony Hay and Dr. John Helmann for their valuable support. Many thanks go to Dr. Ann Lemley for agreeing to serve as a field-appointed member for my A-exam.

Sincere thanks go to all of the current and past members of the Wilson laboratory: Drs. Colleen McGrath, Yongchao Li, Ikram-ul-Haq, Jae Seon Park and Yuting Hu, graduate students John Dingee and Maxim Kostylev, visiting student José Carlos Filho, and undergraduates Felix Moser, Paul Wolski, Sheridan Reiger and Samuel Kim. Special thanks are due to Diana Irwin for sharing with me everything she knows about cellulases, for her never-lost enthusiasm and brilliant comments.

Many thanks are due to Dr. Larry Walker and his laboratory's members, particularly Drs. Jose Moran-Mirabal and Stephane Corgie, as well as Edward Evans and Sarah Munro for kind assistance when I worked at the Cornell Biofuels Research Laboratory. I would also like to thank Dr. Bik Tye for showing me how to be a great lecturer when I was a teaching assistant for her course in Principles of Biochemistry.

I would like to thank the Vietnam Education Foundation and the U.S. Department of Energy for research grants, as well as Cornell University Graduate School, the U.S. National Agricultural Biotechnology Council and Virginia Tech for conference grants.

I am very grateful to my parents, my young brother Thọ and my soon-to-be wife Vân for their love and inspiration. I would also like to thank those who have worked and communicated with me over my time at Cornell University, I am indebted to you for your friendship and support to my sense of being and contributing.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

2,4-DNPC	2,4-dinitrophenyl-β-D-cellobioside
BMCC	Bacterial microcrystalline cellulose
CBM	Carbohydrate binding module
cd	Catalytic domain
CD	Circular dichroism
CMC	Carboxymethyl cellulose
DNS	Dinitrosalicylic acid
FP	Filter paper
FPLC	Fast protein liquid chromatography
GH	Glycoside hydrolase
HPLC	High performance liquid chromatography
MALDI-TOF	Matrix assisted laser desorption/ionization time-of-flight
MUG2	4-methylumbelliferyl β-cellobioside
NaOAc	Sodium acetate
PC	Phosphoric acid-treated cotton
SC	Phosphoric acid-swollen cellulose
TIRFM	Total internal reflection fluorescence microscope
TLC	Thin layer chromatography

CHAPTER ONE
CELLULASES

Section 1.1. Cellulose

Cellulose is the most abundant organic polymer in the biosphere. The conversion of cellulose to ethanol or butanol is of great interest because it provides renewable energy, reducing dependence on fossil fuel and decreasing the volume of waste, thus beneficially improving our environment [1] . Cellulose is a linear unbranched polymer of D-glucose with β-1,4 glycosidic linkages, which are highly stable and resistant to chemical attack because of the high degree of hydrogen bonding with other chains of cellulose to form crystalline regions [1]. All glycosyl hydroxyl groups are oriented equatorially; therefore, the glucopyranose ring tends to be hydrophilic at the edge while being hydrophobic at the α and β faces [2]. Individual cellulose molecules possess two different ends: a non-reducing end with a free C_4 hydroxyl group and a reducing end with a free C_1 hydroxyl group (Figure 1.1).

The crystallinity of purified celluloses varies depending on the source and pretreatment; for instance, cellulose from *Valonia* is considered as 100% crystalline [3] while swollen cellulose is 0% crystalline [4].

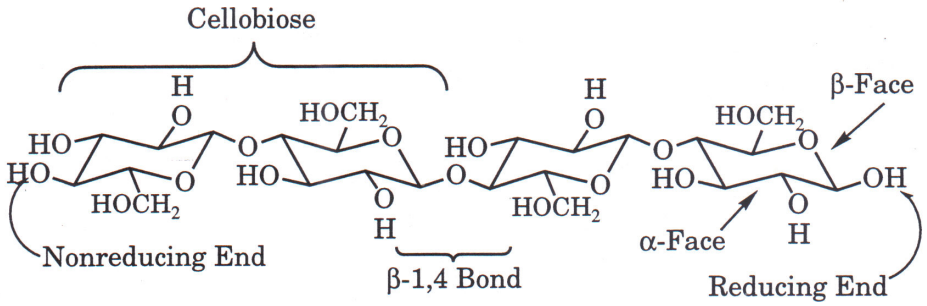


Figure 1.1: Representative structure of a cellulose strand.

Section 1.2. Cellulases and cellulase engineering

Cellulases are specialized enzymes catalyzing the hydrolysis of β-1,4-glycosidic linkages. These enzymes are produced not only by bacteria and fungi in the soil and in the ruminating chambers of herbivores, but also by insects (termites and wood roaches) and plants for leaf and flower abscission, ripening of fruits, differentiation of vascular tissue and cell wall growth [5].

Cellulases are mainly categorized into endocellulases (EC 3.2.1.4), which cleave cellulose chains internally, generating products of variable length with new chain ends, and exocellulases or cellobiohydrolases (EC 3.2.1.91), which act from the ends of cellulose chains, processively cleave off cellobiose as the main product [5-7]. Cellulases do not simply cleave glycosidic bonds but during the process of hydrolysis, they also extract and hold part of a single polysaccharide chain to separate it from the other molecules [8].

Cellulases are additionally categorized as processive or non-processive groups. Processive enzymes continue to bind and hydrolyze a single polysaccharide strand after hydrolyzing the first bond while non-processive cellulases hydrolyze a bond, disengage and rebind at a different site. Exocellulases are often highly processive while endocellulases are generally non-processive. However, there are a few cellulases with properties of both types; for instance, *Thermobifida fusca* Cel9A, a processive endocellulase, which cleaves a cellulose strand internally and then cleaves many bonds before disengaging [9,10].

There is a strong interest in engineering cellulases for higher activity as they show really low hydrolysis rate. *T. fusca* exocellulase Cel6B hydrolyzes cellulose at a rate of as many as 2 bonds per minute [11] while a protease can cleave up to one million peptide bonds per second [12]. A number of factors limit cellulases' rate of hydrolysis:

adsorption of cellulases onto the substrate, the separation of a single cellulose chain from the matrix, threading of the chain into the active site, conformational changes in the active site, processivity (if exocellulases) and desorption of cellulases from the substrate (if endocellulases) [13]. It should be noticed that cellulose is a very difficult substrate; the half life of cellulose at room temperature and pH 7.0 is millions of years [14]. Therefore, substrate-pretreatment is crucial for hydrolysis as it modifies cellulose crystallinity, degree of polymerization, accessible surface area [15]. The rate of hydrolysis can drop by two to three orders of magnitude at high degrees of conversion [16].

1.2.1. Modular structures of cellulases

Cellulases often consist of distinct domains including a catalytic domain and one or more carbohydrate-binding modules (CBMs). Based on sequence alignment and secondary structure prediction, the catalytic domains of cellulases were categorized into 11 glycoside hydrolases (GH) families (families 5 - 10, 12, 26, 44, 45, 48) (www.cazy.org) [17]. These domains are linked by linkers, which are rich in glycine, proline, serine and threonine residues, and are often O-glycosylated [18]. Figure 1.2 shows a modular structure of *Thermobifida fusca* Cel6B, which contains two functional domains, a C-terminal family-6 catalytic domain, linked to an N-terminal family-2 CBM through a Pro-Ser rich linker [11]. A number of cellulases have additional domains with unknown functions, for instance the FnIII modules in *T. fusca* Cel9A and Cel48A [10,19].

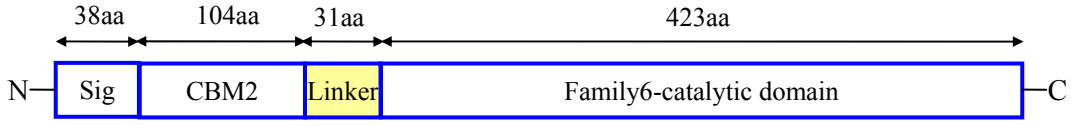


Figure 1.2: Gene structure of *Thermobifida fusca* exocellulase Cel6B (Sig – signal peptide, CBM2 – family-2 carbohydrate binding module).

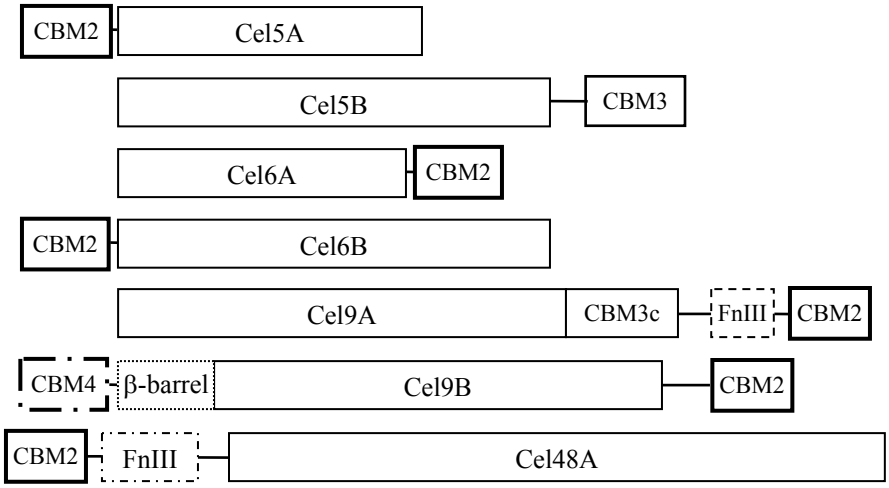


Figure 1.3: The arrangement of the domains and modules of *T. fusca* cellulases.

The position of CBMs may vary. A family 2 CBM is located at the N-terminus of Cel5A, Cel6B and Cel48A and at the C-terminus of Cel6A, Cel9A, and Cel9B in *T. fusca* (Figure 1.3). The family-2 CBMs in *T. fusca* cellulases were found to be important for crystalline cellulose degradation. Removal of this CBM had little effect on hydrolysis of soluble cellulose, but reduced enzymatic activity on crystalline substrates such as bacterial microcrystalline cellulose and filter paper [20].

The crystal structures of many cellulase catalytic domains have been published [21-25]. Structural analysis showed that the active sites of the exocellulases are enclosed by two long loops, forming a tunnel while the endocellulases have an open active site

groove (Figure 1.4). The catalytic domains of several *T. fusca* cellulases also showed non-productive adsorption to α-chitin [26].

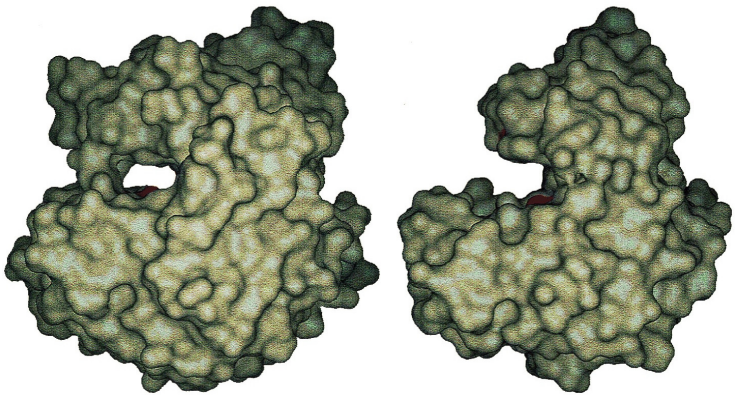


Figure 1.4: Van der Waals surface representations of the exocellulase *Humicola insolens* Cel6A (left) and the endocellulase *T. fusca* Cel6A (right) (modified from [23]).

1.2.2. Catalytic mechanisms

Enzymes from the same GH family are not necessarily all endo- or exocellulases; however, a property that is conserved in all members of most families is the stereochemistry of cleavage: inverting or retaining [27]. For instance, all family-6 GH members catalyze hydrolysis with inversion of the anomeric carbon configuration.

Both catalytic mechanisms are proposed to require two catalytic carboxylate residues. In the inverting mechanism, the catalytic base such as a deprotonated Asp or Glu removes a proton from a water molecule, making it a better nucleophile to directly attack at the anomeric C_1 carbon, breaking the covalent bond between C_1 and the glycosidic oxygen, thus inverting the linkage from β to α [5] while the catalytic acid residue is protonated, donating its proton to the glycosidic oxygen of the leaving group (Figure 1.5). Water molecules then change the protonation of the catalytic residues to

allow them to hydrolyze another bond.

In retaining GHs, a general acid/base catalyst works as an acid and base in two different steps: glycosylation and deglycosylation, respectively. In the first step, it facilitates departure of the leaving group by donating a proton to the glycosyl oxygen atom while the nucleophile forms an enzyme sequestered covalent intermediate. In the second step, the deprotonated acid/base acts as a general base to activate a water molecule that carries out a nucleophilic attack on the glycosyl-enzyme intermediate, retaining the stereochemistry at the anomeric center (Figure 1.5).

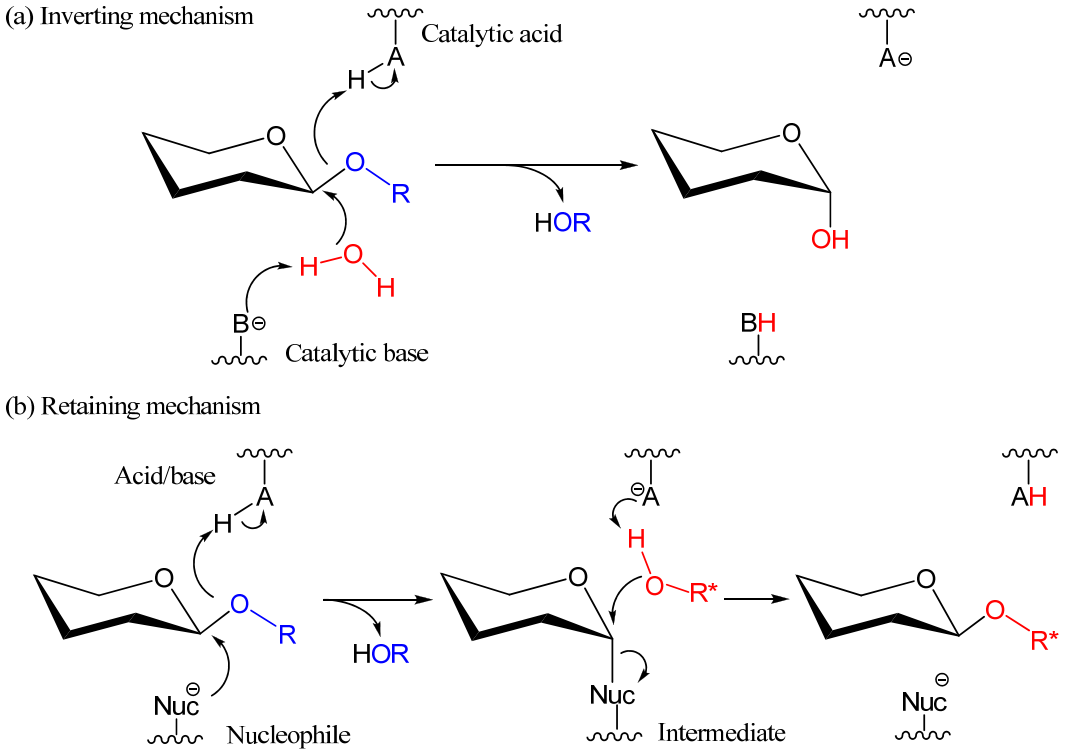


Figure 1.5: Proposed inverting (a) and retaining (b) catalytic mechanisms. AH: a catalytic acid residue, B^-: a catalytic base residue, Nuc: a nucleophile, and R: a carbohydrate derivative. HOR*: an exogenous nucleophile, often a water molecule.

As more glycosyl hydrolases are studied, exceptions to these catalytic mechanisms

have been discovered. In retaining GHs, the presence of a nucleophile is important as it directly attacks the anomeric center to form a glycosyl-enzyme intermediate. However, the carbonyl oxygen of the 2-acetamide group in the substrate of GH-18 chitinases [28], GH-20 hexosaminidases [29], GH-56 hyaluronidases [30], GH-84 O-GlcNAc-ases [31], GH-85 *endo*-β-*N*-acetylglucosaminidases [32,33], and GH-103 lytic transglycosylases [34] has been shown to act as the nucleophile to form an oxazoline intermediate.

Knowledge of the catalytic residues of glycoside hydrolases has been applied to engineer these enzymes for new functions. Removal of the catalytic base in inverting enzymes [35,36] or the nucleophile in retaining enzymes [37] forms a new enzyme class, glycosynthases, which catalyze the synthesis of glycosides from activated glycosyl donors such as glycosyl fluorides. Glycosynthases are getting more attention, particularly for the synthesis of glycosides of pharmaceutical interest [38].

1.2.3. Enzymology of cellulases

The variable structural complexity of cellulose and the mixed specificities of individual enzymes make the study of cellulase activity difficult. Therefore, enzyme assays are usually conducted on modified substrates, which could be categorized into four types: purified insoluble substrates approximated to a native substrate, modified insoluble substrates, soluble modified polysaccharides and soluble oligosaccharides [39]. Enzymatic activity is often quantified by the soluble products, particularly the reducing sugars. Two popular approaches for measuring reducing sugars are the 3,5-dinitrosalisylic acid (DNS) [40] and p-hydroxybenzoic acid hydrazide (PAHBAH) methods [41].

Soluble oligosaccharide substrates are well developed for kinetic measurements in mechanistic studies of enzyme action [39]. Oligosaccharides could be coupled with a

chromophoric aglycon such as methylumbelliferyl, which is useful for studying ligand binding into the active site using fluorescence titration [42].

1.2.4. Processivity

Exocellulases and processive endocellulases such as *T. fusca* Cel6B and Cel9A display the ability, known as processivity, to continue hydrolyzing a cellulose chain without disassociation. In this way, the enzyme remains close to the detached chain and prevents the chain from re-associating with the crystalline matrix. The ratio of soluble to insoluble reducing sugars is one criterion to evaluate the processivity of cellulases, as processive enzymes cleave many times along a cellulose chain to produce more soluble reducing sugars [43]. This measurement is helpful to distinguish between exocellulases and endocellulases [43]. However, soluble and insoluble reducing sugars can be separated easily only when crystalline cellulose such as filter paper or Avicel is used.

Improvement of processivity is a difficult task; for instance, processivity in *T. fusca* Cel9A requires coordination between the sliding of the substrate into the cleavage site and the release of the products. The R378K mutant enzyme showed the highest improvement in processivity among single mutations in the catalytic domain of Cel9A; however, a double mutant enzyme containing R378, which has two hydrogen bonds to Glc(+1) O2, and D261, which is located near Glc(-4), dramatically decreased processivity [44].

1.2.5. Synergism

Synergism can be defined as the ability of a mixture of cellulases to give higher activity than the sum of the individual activities. Synergism does not require a direct interaction between individual cellulases; however, two cellulases give synergism only

when they attack different sites on a cellulose chain and they produce new sites for each other [20]. Therefore, two exocellulases give synergism only when they are from different classes: non-reducing end or reducing end-directed exocellulases. Pre-treatment of crystalline cellulose with an endocellulase produces a better substrate for exocellulases; however, pre-treatment with an exocellulase does not create the same effect for endocellulases [20]. The activity of endocellulases also are increased in synergistic mixtures of endocellulases and exocellulases [43].

Section 1.3. Cellulase-producing organisms

There are two major cellulase systems, complexed and non-complexed [6]. Most anaerobic cellulase-producing organisms such as *Clostridium thermocellum* [45] and *C. cellulovorans* [46] produced exocellular, high molecular-weight complexes with full complement of hydrolytic enzymes, called cellulosomes. These cellulosomes are attached to the surface of the microorganism, allowing the microorganism to retain the hydrolyzed products efficiently. Aerobic organisms such as filamentous fungi and actinomycetes secrete individual hydrolytic enzymes, a non-complexed hydrolytic system [6]. Besides these systems, several other less studied mechanisms are also used by cellulolytic microorganisms such as the aerobic bacterium, *Cytophaga hutchinsonii* and the anaerobic bacterium, *Fibrobacter succinogenes* [47,48], these bacteria code only endocellulases, not processive endocellulases and exocellulases [49].

Our laboratory studies *Thermobifida fusca*, formerly known as *Thermomonospora fusca*, an actinomycete of the suborder streptosporangineae, as a model organism [50]. *T. fusca* is a Gram positive, spore-forming, filamentous soil bacterium. It is a moderately thermophilic, cellulolytic bacterium with an optimum growth temperature around 50^{o}C [51].

The genome of *T. fusca* (3.7Mb) was sequenced in 2000 by the Department of

Energy [20], which facilitated the discovery of a number of *T. fusca* enzymes participating in the hydrolysis of plant cell walls. An intracellular enzyme (a β-glucosidase) and several extracellular enzymes including seven different cellulases, two low-molecular weight cellulose binding proteins, one xylanase and one xyloglucanase have been cloned, purified and characterized [20,52]. The genes for four extracellular cellulases including endocellulases Cel5A, Cel6A and Cel9B, as well as a processive endocellulase Cel9B were cloned by screening *Escherichia coli* colonies containing *T. fusca* DNA inserts on carboxymethyl cellulose (CMC) overlay assays. The genes for two exocellulase Cel6B and Cel48A were detected by screening plasmid libraries with labeled oligonucleotides complementing the N-terminal sequence of each protein [20]. *T. fusca* Cel5B has recently been detected by zymogram analysis [52].

CHAPTER TWO

IDENTIFICATION AND BIOCHEMICAL CHARACTERIZATION OF CATALYTIC RESIDUES IN *T. FUSCA* CEL6B[*]

Section 2.1. Introduction

As presented in the preceding chapter, *T. fusca* produces a mixture of functionally distinct cellulases, which act synergistically. One of these cellulases, Cel6B is very important for achieving the maximum activity of synergistic mixtures, although its activity alone is relatively weak on all polysaccharide substrates [11]. *T. fusca* Cel6B is an exocellulase (EC.3.2.1.91) that processively hydrolyzes a number of β-1,4-glycosidic bonds from the non-reducing end of cellulose molecules before dissociation. The enzyme has higher thermostability and a broader pH optimum than the homologous fungal exocellulase *Trichoderma reesei* (also known as *Hypocrea jecorina*) Cel6A, which is present in most commercial cellulase preparations [11].

The three-dimensional structures of the catalytic domains of five family GH-6 cellulases have been determined. *Humicola insolens* Cel6A [23] and *T. reesei* Cel6A [21] are exocellulases while *T. fusca* Cel6A [22], *H. insolens* Cel6B [24] and *Mycobacterium tuberculosis* Cel6 [25] are endocellulases. The difference in the modes of action of these enzymes is clearly reflected in their structures. The active sites of the exocellulases are enclosed by two long loops forming a tunnel while the corresponding loops in the endocellulases are shorter, opening the active sites (Figure

---

[*] Reproduced in part with permission from "The absence of an identifiable single catalytic base residue in *Thermobifida fusca* exocellulase Cel6B" Thu V. Vuong and David B. Wilson, FEBS Journal, 276 (14), p.3837-3845 © 2009 Federation of European Biochemical Societies.

1.4). Increasing knowledge of cellulase structures and improvement of modeling software [53] have greatly facilitated rational protein design to study the catalytic mechanism of cellulases.

Members of glycosyl hydrolase family-6 were shown to utilize an inverting mechanism [27] (Figure 1.5). However, the detailed catalytic mechanism of family GH-6, particularly the existence of a catalytic base is still in doubt. Based on site-directed mutagenesis, residue D392 in *Cellulomonas fimi* Cel6A, corresponding to *T. reesei* Cel6A D401, *H. insolens* Cel6A D405, *H. insolens* Cel6B D316, *T. fusca* Cel6A D265 and *T. fusca* Cel6B D497 (Table 2.1), was concluded to be a classical Brønsted base [54]. However, crystallographic and kinetic studies in *T. reesei* Cel6A suggested that D175, not D401 was the catalytic base [55]. The *H. insolens* Cel6A D405A and D405N mutant enzymes still retained approximately 0.5-1% activity [56]. Although mutation of D316 in *H. insolens* Cel6B to alanine or asparagine led to an inactive enzyme [56], the three-dimensional structure determination showed that D316 is likely to be correctly positioned to act as a base only if a conformational re-arrangement of the -1 subsite sugar ring occurs [24]. In *T. fusca* Cel6A, D265 was not directly involved in hydrolysis, but participated in substrate binding [57]. Therefore, it is interesting to investigate the catalytic residues in *T. fusca* Cel6B and other family-6 GH members.

Recently, activity rescue of catalytic mutants by sodium azide has been demonstrated to be a useful tool for identification of the catalytic base in both retaining [58] and inverting GHs [59]. This approach distinguished the actual catalytic base from other catalytic residues in the inverting cellulase *T. fusca* Cel9A [44].

This chapter describes the work done to identify the catalytic residues, particularly the catalytic acid and catalytic base of *T. fusca* Cel6B and to determine whether the enzyme acts by the typical inverting mechanism.

Section 2.2. Experimental procedures

Strains and plasmids - *Escherichia coli* DH5α and BL21 RPIL DE3 (Agilent Technologies, CA, USA) were used as the host strains for plasmid extraction and protein expression, respectively. The entire Cel6B gene in plasmid pSZ143 [60], which was constructed from the pET26b+ vector (Novagen), was used as the template for mutagenesis. A plasmid (pTVcd), which contains only the catalytic domain of Cel6B, was constructed using *Not*I, then ligated and transformed into *E.coli* DH5α. This plasmid was used as the template to produce the D274A catalytic domain (D274Acd) mutant enzyme.

Site-directed mutagenesis - Complementary primers were designed using PrimerSelect, Lasergene v.8.0 (DNASTAR, WI, USA) to incorporate the desired mutations. High purity salt-free primers were then synthesized by Eurofins MWG Operon (AL, USA). PCR was performed for 18 cycles at 95^{o}C x 1min, 60^{o}C x 50s and 68^{o}C x 7min, using the QuikChange method (Agilent Technologies). The methylated DNA template was hydrolyzed by *Dpn*I. The PCR products were transformed into *E. coli* DH5α. Plasmids were then isolated and purified using the Qiagen Plasmid Miniprep Kit (Qiagen, CA, USA). Mutant plasmids were checked by both restriction enzyme digestion and DNA sequencing (Applied Biosystems Automated 3730 DNA Analyzer, Cornell University Life Sciences Core Laboratories Center, NY, USA). Mutant plasmids with the correct sequence were transformed and expressed in *E. coli* BL21 RPIL DE3.

Protein expression - The expression of proteins in *E. coli* BL21 RPIL DE3 was tested at different growth temperatures (30-37^{o}C), cell density before induction (OD_{600}nm 0.5-1.0) and length of induction (16-24hrs) with isopropyl-thio-β-D-galactosidase, IPTG. Proteins produced in the supernatant and shock fluid were compared by Western blotting.

Western blotting - Proteins were separated on SDS-polyacrylamide gels and electrophoretically transferred to an Immobilon-P membrane (Millipore, MA, USA). The primary antibody was rabbit polyclonal antiserum raised against Cel6B and the secondary antibody was goat anti-rabbit IgG alkaline phosphatase conjugate (Bio-Rad, CA, USA). Westerns blots were developed using nitro-blue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate (Bio-Rad protocol).

Enzyme purification - *E. coli* BL21 RPIL DE3 strains were grown at 37^{o}C overnight in 30mL of Luria broth with 60μg/mL kanamycin before being transferred into 1L of M9 medium with 0.5% glucose and 60μg/mL kanamycin. Cells were grown at 30^{o}C until OD_{600}nm is about 0.8, then IPTG was added to 0.8mM and the culture was grown at 30^{o}C for 20hrs. The supernatant was collected and Cel6B enzymes were purified using published chromatographic techniques [11], first on a CL-4B Phenyl-Sepharose column, and then on a Q-Sepharose column. Enzyme purity was assessed on SDS gels. Enzymes were buffer exchanged and concentrated using Vivaspin 30kDa MWCO centrifugal concentrators (Sartorius, NY, USA), and then filtered through Costar ® Spin-X centrifuge tube filters (0.45μm nylon membrane). The enzyme concentrations were calculated based on absorbance at 280nm and the corresponding molar extinction coefficients determined from the predicted amino acid compositions. All proteins were prepared at the same concentration with 5mM NaOAc pH5.5 plus 10% glycerol, and stored at -70^{o}C.

Polysaccharide assay - As recommended by Ghose (1987) [40], polysaccharide assays were conducted using a series of enzyme concentrations above and below the target for each substrate for a fixed time with saturating substrate. Activities of wild-type and mutant enzymes were determined on 2.5mg/mL bacterial microcrystalline cellulose (BMCC), 2.5mg/mL phosphoric acid-treated swollen cellulose (SC), 2.5mg/mL phosphoric acid-treated cotton (PC), and 10mg/mL carboxymethyl cellulose (CMC). PC was prepared using a previously described method for swollen cellulose [61]. All assays were run in triplicate for 16hrs at 50^{o}C in 50mM NaOAc pH5.5 at the final reaction volume of 400μL. Reducing sugars were measured using dinitro-salicylic acid (DNS) [40]. The DNS method fits the assay range well and does not require a protein blank. Nanomoles (nmol) of protein used were plotted versus the A_{600}nm, and KaleidaGraph (Synergy Software, PA, USA) was used to fit the curve to determine the amount of enzyme required for 6% substrate digestion of BMCC, SC and PC, and 1.5% digestion of CMC. If the activity was too weak to achieve the target digestion, activity was calculated at a high concentration of enzyme (1.5μM). Mutant enzyme activities were determined concurrently with wild-type.

2,4-DNPC assay - 2,4-dinitrophenyl-β-D-cellobioside (2,4-DNPC) was a gift from Dr. Stephen Withers (University of British Columbia, Vancouver). Reactions were carried out at 50°C in 50mM NaOAc pH 5.5, using 1.5μM enzyme and initial substrate concentrations of 20, 40, 80, 150 and 600μM. The change in absorbance at 400nm, measured for every 10min minus the blank, was used as the activity for the substrate concentration at the beginning of the next time point. The concentration of 2,4-dinitrophenol was determined at A_{400}nm, using an extinction coefficient of 10,900$M^{-1}.cm^{-1}$ [62].

Azide rescue assay - Different concentrations of sodium azide, up to 3M, were added to mixtures of 0.75-1.5μM enzyme and corresponding substrates (CMC or SC). The reaction tubes were incubated at 50°C in 50mM NaOAc, pH 5.5 for 16hrs and reducing sugars were measured with DNS.

Substrate binding assays - BMCC, SC and PC are insoluble cellulose, which can be used for substrate binding assays. Binding of 4μM enzyme to 0.1% BMCC, SC or PC was determined in 50mM NaOAc buffer pH5.5 and 10% glycerol in Eppendorf® Protein LoBind tubes (Eppendorf, NY, USA). Reactions were incubated for 1hr on a Nutator rocking table (Clay-Adams, MD, USA) at 4°C to limit hydrolysis. The insoluble substrate was separated from the supernatant by centrifugation at 16,000g for 5min, and the A_{280}nm of the supernatant was measured to determine the amount of unbound protein. CMC binding was evaluated by the relative migration of enzymes on native gels containing 0.5% CMC.

CMC viscosity - Viscometric activity was measured according to the method of Irwin et al. [43].

pH profile - Enzymes were assayed with SC as above, but in 12 different pH buffers mixed from 50mM citric acid, 50mM boric acid and 50mM NaH_2PO_4. All enzymes were normalized by their activity at pH5.5.

Thin layer chromatography - Thin layer chromatography was performed as previously described by Jung et al. [63]. Briefly, oligosaccharide products were separated by a solvent mixture of ethyl acetate: water: methanol (40:15:20, v/v) on Whatman® LK5D 150-A silica gel plates (Whatman, NJ, USA), and then visualized by dipping in a color-developing mix (100mL of acetic acid plus 1mL *p*-anisaldehyde plus 1mL of concentrated sulfuric acid) before being heated for 1hr at 95^{o}C.

Fluorescence quenching titration - Dissociation constants, K_d, for the binding of 4-methylumbelliferyl β-cellobioside (MUG2) to wild-type and mutant enzymes were determined by direct fluorescence titration at 5.5°C using an Aminco SLM8000C spectrofluorimeter (SLM-Aminco, IL, USA) as described [42]. The initial concentration of MUG2 was 1.7μM in 903μL of 50mM NaOAc pH5.5, and 53.7μM enzyme was added at 3.5μL/min. Excitation was at 316nm and emission was measured at 360nm.

Circular dichroism analysis - Spectra of 10μg/uL protein were recorded from 190 to 290nm on an Aviv CD400 Spectrometer (AVIV Biomedical Inc., NJ, USA) at a scanning rate of 1nm/s at 4°C.

Section 2.3. Results

Cel6B structural model - A structural model of the Cel6B catalytic domain was built based on the X-ray structures of *H. insolens* Cel6Acd (1OCB) and *T. reesei* Cel6Acd (1QK2) using the Swiss-Model Workspace to choose the residues for mutation. The reliability of the model was evaluated by the WhatCheck program to check a battery of physico-chemical constraints [64]. Energy minimization was computed by the Swiss-Model Workspace [53], and the final total energy of the model was -3641 KJ/mol.

Selection of amino acids for mutation - All highly conserved aspartic and glutamic residues including D226, D274, D497 and E495, which are approximately 6Å away from the -1 and +1 subsites, were mutated to alanine. Figure 2.1 shows the position of potential Cel6B catalytic residues and Table 2.1 shows the corresponding residues in four other family-6 GHs.

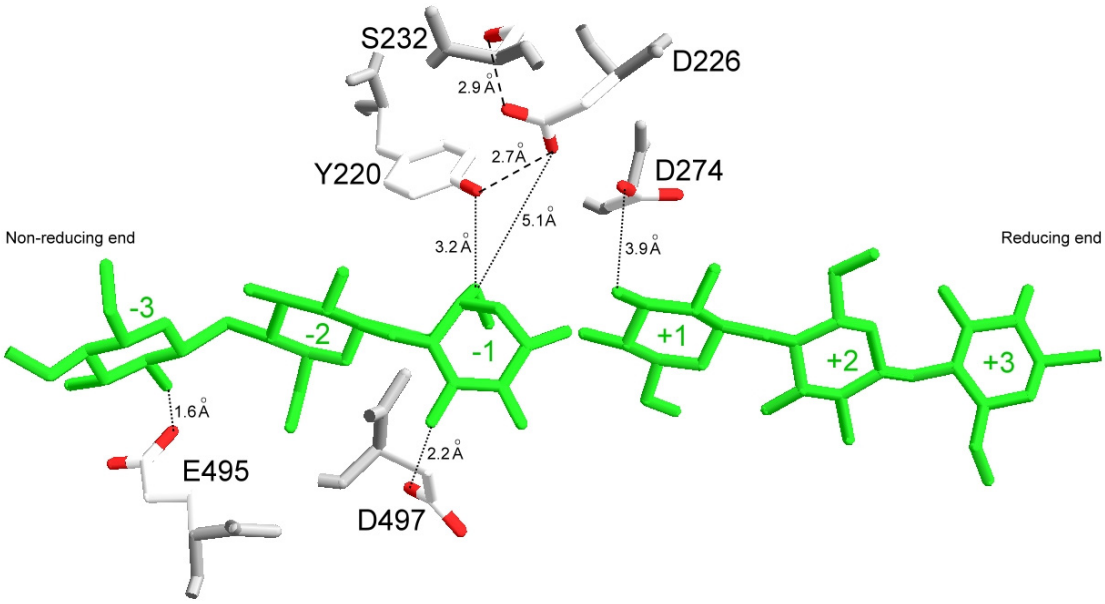


Figure 2.1: Location of *T. fusca* Cel6B putative catalytic residues, modeled to both *H. insolens* Cel6A (1OCB) and *T. reesei* Cel6A (1QK2). The model was built with Swiss-Workspace; dashed lines indicate hydrogen bonds, dotted lines indicate the distance in angstroms.

D274 was expected to be the catalytic acid, as shown in other family-6 cellulases [55,57,65]. D226 corresponds to a residue in *T. reesei* Cel6A that forms a carboxyl-carboxylate pair with the catalytic acid to increase its pK_a [66]. D497 is a candidate for a catalytic base, as it is located in the -1 subsite and almost opposite to the putative catalytic acid D274. The E495 side chain appears to be near the -3 subsite; however, as the structures of exocellulases are known to be somewhat flexible [67], this residue could also be a catalytic base.

S232 was chosen as it is positioned near the -1 subsite and the residue hydrogen-bonds to D226 OD2, thus might participate in a proton transferring network as has been postulated in *T. reesei* Cel6A [55]. The residue corresponding to Y220 in *T. fusca* Cel6A (Y73) was found to be essential for hydrolysis [68,69]. In retaining

enzymes of GH families 33, 34 and 83, a tyrosine was showed to act as a catalytic nucleophile [70].

Table 2.1: Amino acids chosen for mutation. The gene sequences were aligned and analyzed using Megalign (DNASTAR-Lasergene)

T. fusca Cel6B (exo)	Glycosyl subsite location	Proposed role	Corresponding residue in:			
			H. insolens Cel6A (exo)	*T. reesei* Cel6A (exo)	*C. fimi* Cel6A (endo)	*T. fusca* Cel6A (endo)
Y220	-1	Substrate distortion	Y174	Y169	D210	Y73
D226	-1	Increase of pK_a	D180	D175	D216	D79
S232	-1	Proton network	S186	S181	G222	S85
D274	+1	Catalytic acid	D226	D221	D252	D117
E495	-3	Catalytic base	E403	E399	E390	E263
D497	-1	Catalytic base/ Substrate binding	D405	D401	D392	D265

Enzyme expression and purification - Western blotting analysis suggested that the expression strains secreted the Cel6B enzymes into the supernatant most efficiently when grown at 30^{o}C, induced by 0.8mM IPTG at OD_{600}nm of 0.8 for 20hrs (data not shown). All mutant enzymes were expressed with a yield of 10-12mg/L and they all behaved similarly to wild-type Cel6B during purification. Circular dichroism spectra of all mutant enzymes were identical with that of wild-type (data not shown), indicating that the global secondary structure of the mutant proteins remained intact.

Enzyme activity - The activities of the mutant enzymes on different polysaccharides were determined and compared to those of wild-type Cel6B (Table 2.2). BMCC is a crystalline substrate with a degree of polymerization (DP) >1000 [2]. SC and PC are amorphous celluloses with DP ranging 1500-2600 (Jessica Hatch, personal communication) while CMC is soluble cellulose with random carboxylmethyl substitutions and a DP of 250-500 [71].

D274 mutation - The D274A mutant enzyme could not achieve the target digestion on any polysaccharide substrate, suggesting that it is essential for catalysis (Table 2.2). The substrate 2,4-DNPC has an excellent leaving group, which does not require a catalytic acid, thus an alanine mutation of the catalytic acid still has 2,4-DNPC activity [54]. The data from the 2,4-DNPC assays fit the Michaelis-Menten equation well. The k_{cat} of the D274A enzyme on 2,4-DNPC was more than 6-fold higher than that of wild-type Cel6B (Table 2.3).

The D274A mutant enzyme and its catalytic domain bound more to both BMCC and SC than wild-type and the wild-type catalytic domain, respectively (Figure 2.2). The percentage of D274Acd bound to BMCC was around 65% while only 20% of wild-type was bound. A previous study [60] showed that the fluorescence emission of 4-methylumbelliferyl ligands was strongly quenched upon binding to Cel6B and the enzyme did not hydrolyze these ligands. Therefore, fluorescence titration could be used to investigate ligand binding affinity of the active sites. Fluorescence titration showed that D274A bound 4-methylumbelliferyl β-cellobioside (MUG2) approximately the same as wild-type. The K_d of the D274A enzyme and wild-type for MUG2 were $1.5x10^{-8}$ and $3.6x10^{-8}$M, respectively (Table 2.2), showing that the D274A mutation did not impair ligand binding.

Table 2.2: Polysaccharide activity and ligand binding of the *T. fusca* Cel6B enzymes.

	Activity (μmole cellobiose min^{-1} μmole^{-1} enzyme)				K_d (μM) for MUG2 (x 10^{-2})
	BMCC	SC	PC	CMC	
Cel6B	0.93	2.25	3.37	0.57	3.6 ± 0.3
Y220A	--(0.02[a])	--(0.02[a])	nd	nd	57 ± 4
D226A	--(0.10[a])	--(0.17[a])	0.67	0.63	25 ± 1
S232A	0.57	1.78	4.14	--(0.25[a])	3.2 ± 0.3
D274A	--(0.01[a])	--(0.06[a])	--(0.05[a])	--(0.10[a])	1.5 ± 0.2
E495A	--(0.13[a])	0.87	2.46	0.37	~ 900[b]
D497A	--(0.12[a])	1.26	2.90	--(0.12[a])	~ 13,000[b]
D226A-S232A	--(0.03[a])	--(0.06[a])	--(0.05[a])	--(0.08[a])	73 ± 2

Activity was calculated at 6% digestion for BMCC, SC and PC and 1.5% digestion for CMC. The average coefficients of variation were 4, 5, 5.5 and 2.5 for BMCC, SC, PC and CMC, respectively. K_d was determined by fluorescence titration of 53.7μM enzyme to 1.7μM of 4-methylumbelliferyl β-cellobioside (MUG2). [a] Target digestion could not be achieved; activity was calculated at 1.5μM enzyme; nd- not detected. [b]Value is approximate as titration curve did not fit well due to poor binding.

Table 2.3: 2,4-DNPC kinetics of the *T. fusca* Cel6B wild-type and mutant enzymes. Initial 2,4-DNPC concentrations of 20-600μM were hydrolyzed by 1.5μM enzyme.

	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (min^{-1} μM^{-1}) ($x10^{-3}$)
Cel6B	0.34 ± 0.06	2.3 ± 1.9	146 ± 122
Y220A	0.09 ± 0.04	161 ± 37	0.56 ± 0.28
S232A	0.03 ± 0.01	44 ± 13	0.68 ± 0.30
D226A	0.11 ± 0.05	6.5 ± 1.3	16.9 ± 8.4
D274A	2.26 ± 0.14	1.5 ± 0.2	1,507 ± 186
D497A	0.006 ± 0.001	214 ± 65	0.03 ± 0.01

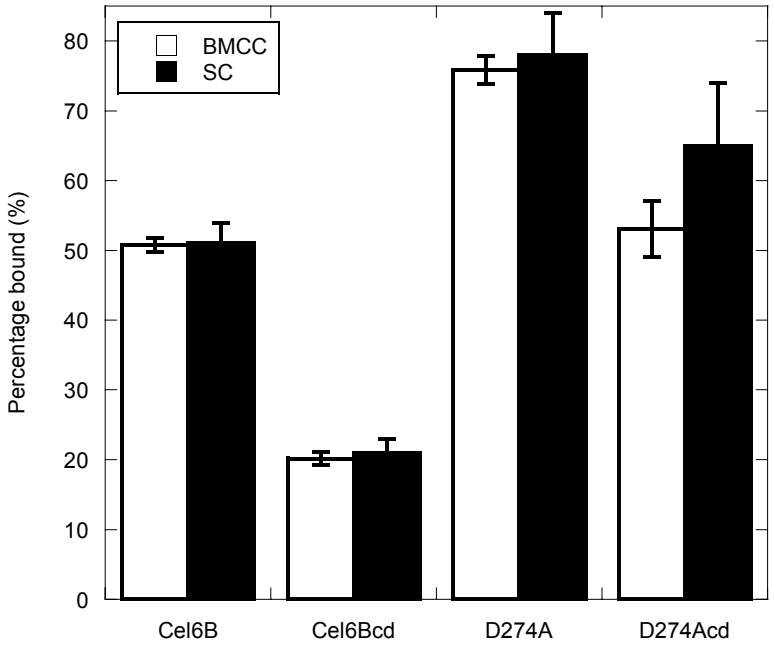


Figure 2.2: Binding of *T. fusca* Cel6B wild-type, the D274A enzyme and their catalytic domains (cd) to BMCC and SC. Substrate binding was conducted using 4μM of enzymes in 50mM NaOAc pH5.5 for 1hr at 4°C.

D226 mutation - The D226A enzyme had very low activity on insoluble cellulose (BMCC, SC and PC) (Table 2.2). TLC analysis indicated that Cel6B wild-type completely hydrolyzed cellotetraose, cellopentaose and cellohexaose within 20min while only a trace of products were produced by the D226A enzyme after 16hrs (Figure 2.3). The D226A protein bound more than wild-type to BMCC and SC (data not shown) and its K_d for MUG2 only slightly reduced, suggesting that activity loss on these substrates was not caused by loss of substrate binding.

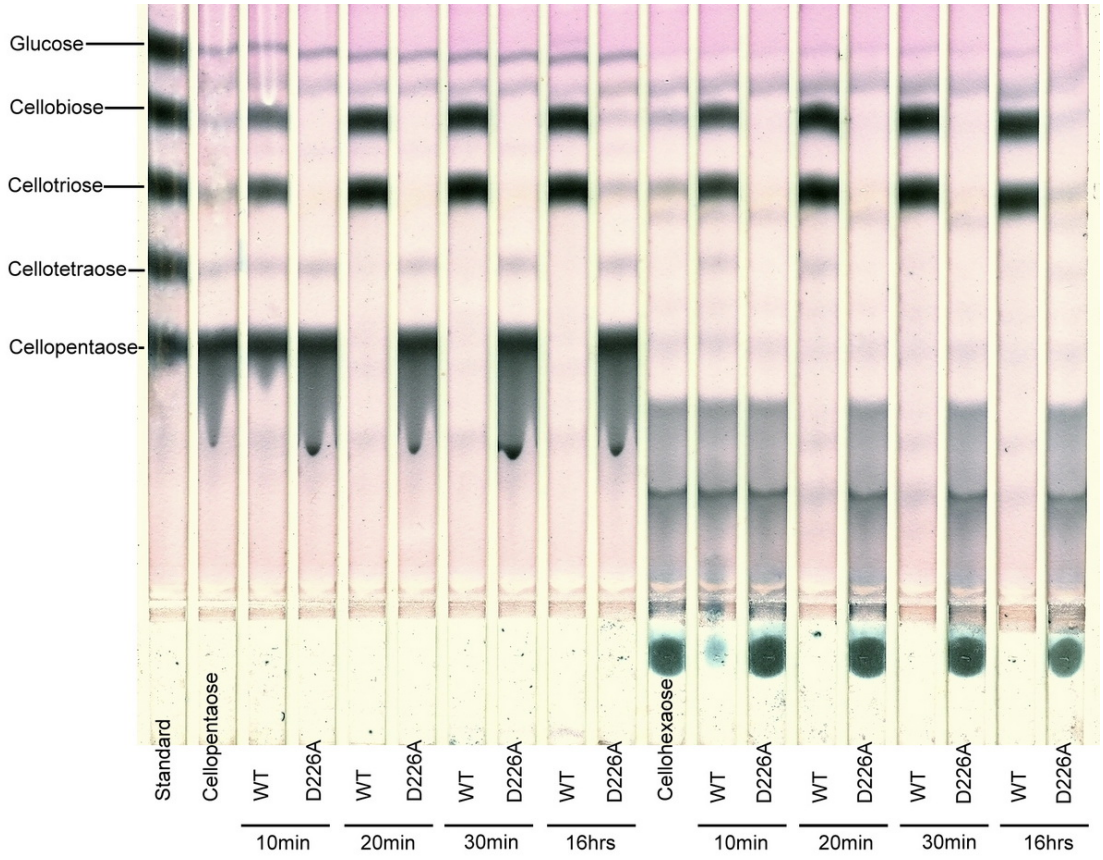


Figure 2.3: TLC analysis of cellopentaose and cellohexaose hydrolysis by Cel6B wild-type and D226A

In the absence of a catalytic base, an exogenous nucleophile such as sodium azide can partially rescue enzyme activity [58]. The activity of the D226A enzyme on SC was not improved by sodium azide even at 3M (Figure 2.4).

Surprisingly, the D226A enzyme had slightly higher activity on CMC than the wild-type. The mutant enzyme reduced the viscosity of a CMC solution faster than the wild-type although the decrease was much lower than that of a typical endocellulase (Figure 2.5). D226A required up to 10hrs while the endocellulase *T. fusca* Cel6A required only 20min to reduce the outflow time to 150s [43]. However, TLC analysis and matrix assisted laser desorption/ionization time-of-flight (MALDI-TOF) analysis of the CMC digestion products of D226A did not show cellobiose, the major product of wild-type Cel6B, or carboxymethyl cellobiose (Figure 2.6). The MALDI-TOF spectra showed cellotriose, cellotetraose, cellopentaose, cellohexaose, and their carboxymethyl derivatives.

To investigate the production of insoluble reducing sugars from CMC, TLC bands corresponding to the loading spot, cellotriose and cellobiose were eluted and reducing sugars from each fraction were measured. The majority of reducing sugars produced by the D226A enzyme were found at the loading spot while wild-type Cel6B produced primarily cellobiose (data not shown). To test whether the cleavage of CMC by D226A was dependent on the carboxymethyl groups of CMC, the enzymes were assayed on hydroxyethyl cellulose (HEC), which does not contain charged groups as does CMC. The D226A enzyme had several-fold higher HEC activity than wild-type (data not shown). CMC-native gels showed that D226A bound CMC as tightly as wild-type (data not shown).

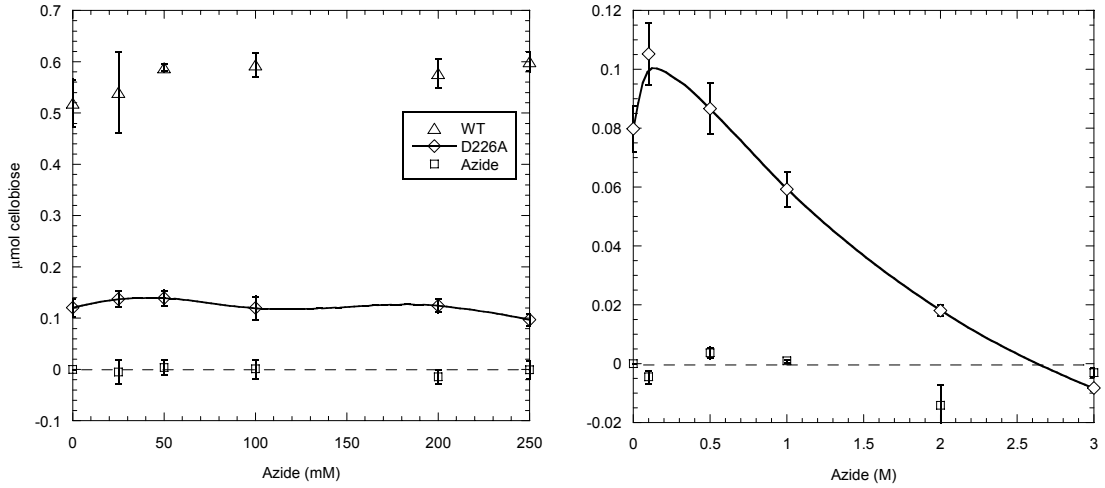


Figure 2.4: Sodium azide rescue test for D226A activity on SC; WT- wild-type. Sodium azide was added to 1µM enzyme and 0.25% SC.

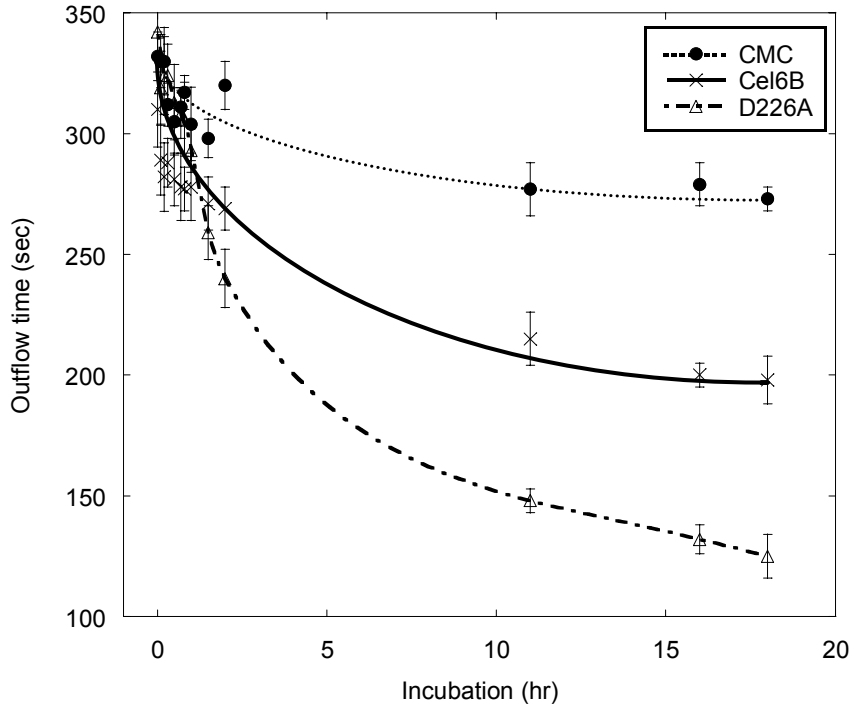


Figure 2.5: Ability of *T. fusca* Cel6B wild-type and the D226A enzyme to reduce the viscosity of CMC.

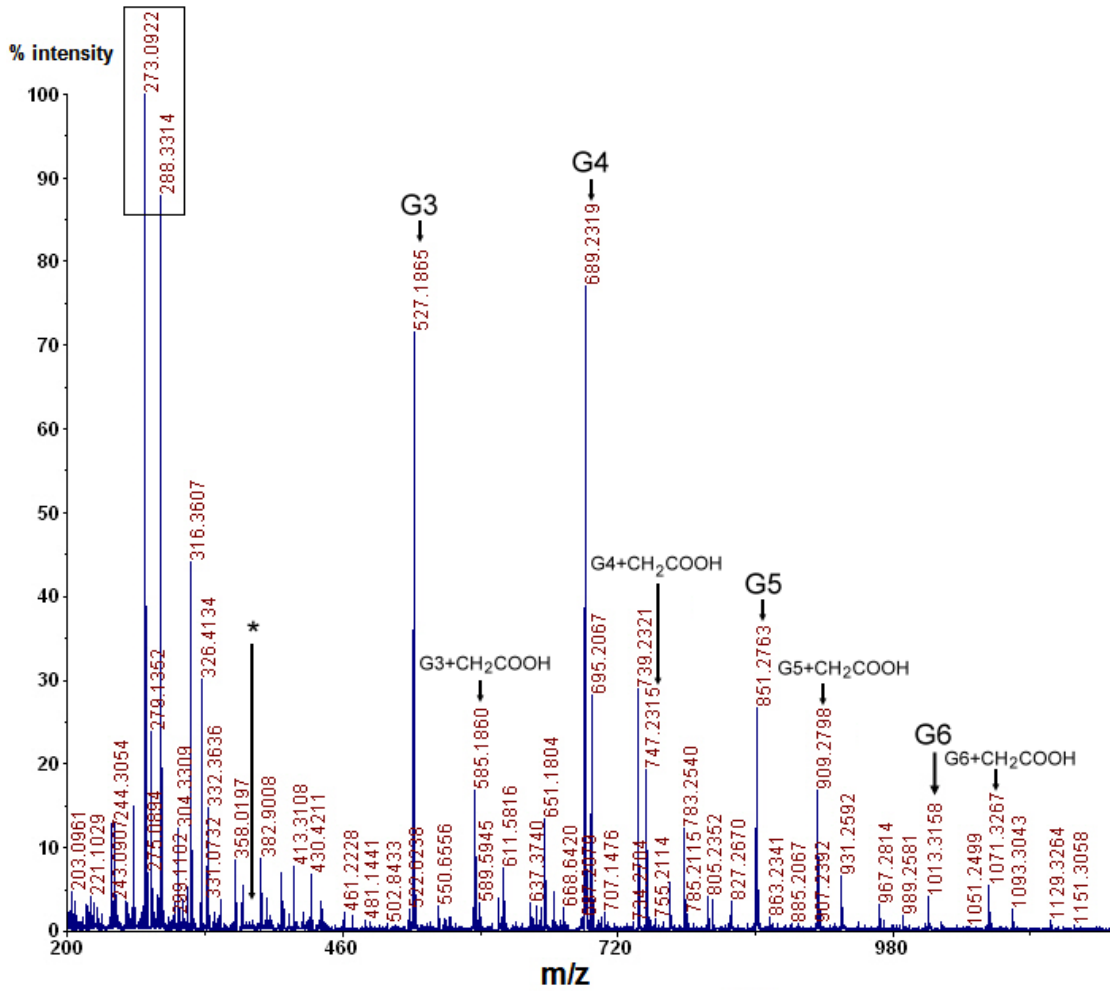


Figure 2.6: MALDI-TOF spectra for CMC products of the D226A enzyme. G3, G4, G5 and G6 are cellotriose, cellotetraose, cellopentaose and cellohexaose, respectively. *Supposed position of cellobiose (G2, experimental mass or EM= 365); neither G2-CH_2COOH (EM= 423) nor G2-CH_2COOH-CH_2COOH (EM= 504) were detected. Peaks 273.09 and 288.33 are matrix artifacts.

D497, E495, S232 and Y220 mutation - The D497A and E495A mutant enzymes had reduced activity on all substrates, particularly on crystalline BMCC (Table 2.2). The K_d for MUG2 of these enzymes decreased significantly (Table 2.3). The activity of the D497A enzyme on 2,4-DNPC was nearly 60-fold lower than that of the wild-type; whereas its K_m was over 90-fold higher.

The S232A mutant enzyme retained near wild-type activity on most substrates, but CMC activity was drastically reduced (Table 2.2). The HEC activity of the S232A enzyme was also lower than that of the wild-type (data not shown).

The Y220A mutant enzyme could not reach target digestion on either BMCC or SC, and no PC or CMC activity was detected (Table 2.2). The enzyme showed a slightly lower K_d for MUG2 than wild-type (Table 2.2) indicating good binding. However, the k_{cat} of Y220A on 2,4-DNPC was approximately 26% of wild-type and the K_m increased 70-fold (Table 2.3).

None of these four mutant enzymes was rescued by sodium azide (data not shown). To test whether any mutation caused a change in the pK_a of the catalytic acid eliminating activity rescue by sodium azide, the mutant enzymes, except for Y220A due to its extremely low activity, were normalized by activity at pH 5.5 and assayed for PC activity for 16hrs over the pH range from 2-12. None of the pH profiles showed a significant difference from wild-type (Figure 2.7).

D226A-S232A double mutation - The double mutation knocked out activity on all polysaccharides and slightly decreased ligand binding (Table 2.2). Binding to BMCC and SC by the mutant enzyme was similar to wild-type (Figure 2.8). Excitingly, CMC activity of the mutant enzyme was partially rescued at low concentrations of sodium azide (Figure 2.9).

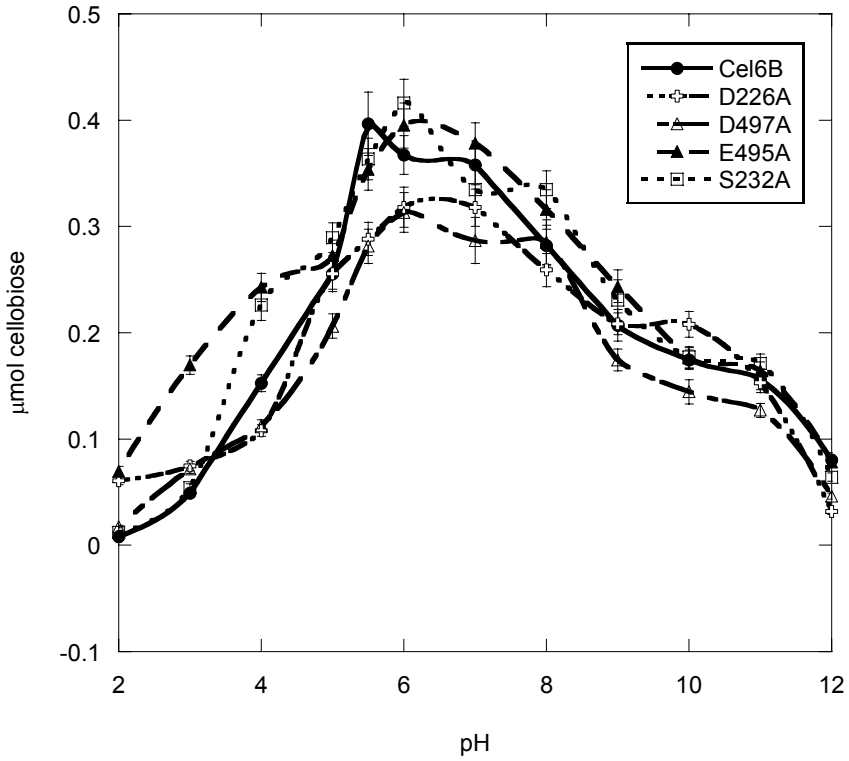


Figure 2.7: PC activity of Cel6B wild-type and mutant enzymes as a function of pH. All enzymes were normalized according to their activity at pH 5.5. Activity of Y220A and the double mutation D226A-S232A was too low to be assayed.

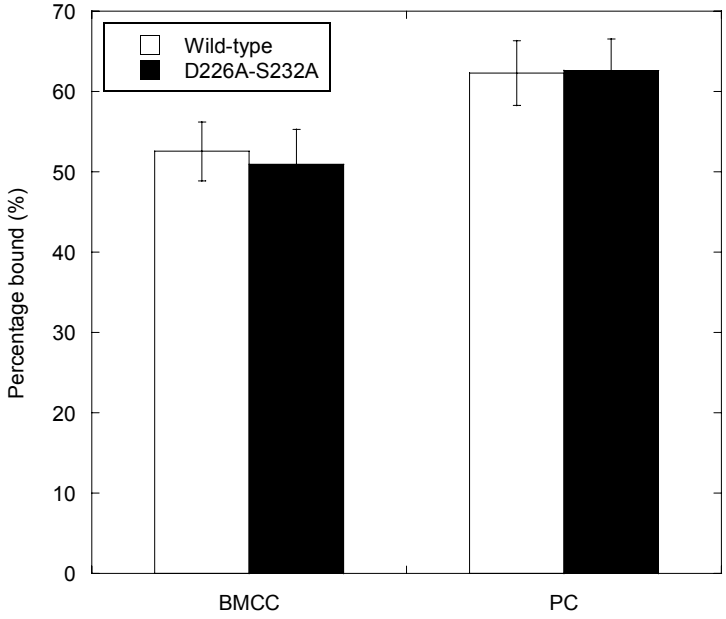


Figure 2.8: Binding of the wild-type and the double mutant enzyme D226A-S232A to BMCC and PC. Substrate binding was conducted using 4μM of enzymes in 50mM NaOAc pH5.5 for 1hr at 4^{o}C.

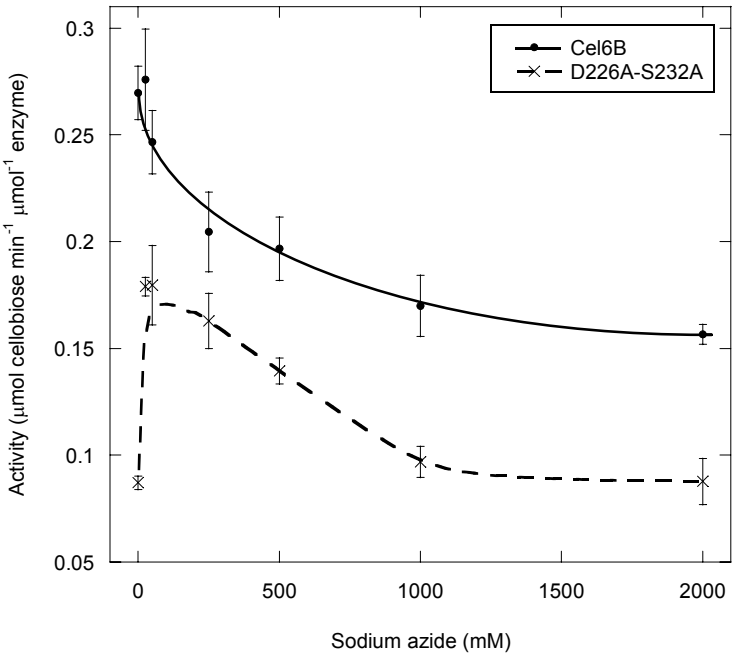


Figure 2.9: Sodium azide rescue for wild-type and D226A-S232A activity on CMC. Wild-type (0.75µM) and the D226A-S232A mutant enzyme (1.5µM) were added with different concentrations of sodium azide and assayed on 1% CMC in 50mM NaOAc pH5.5. Reducing sugars were measured after 16-hr incubation at 50^{o}C.

Section 2.4. Discussion

Key residues - The results of this study show the essential roles of D274 and Y220A as mutation of either residue resulted in nearly inactive *T. fusca* Cel6B. D274 functions as the catalytic acid since the D274A mutation increased activity on 2,4-DNPC, which does not require a catalytic acid. A drastic increase in the K_m for 2,4-DNPC together with a slightly lower K_d for MUG2 supports a role for Y220 in distortion of the glycosyl unit in subsite -1 rather than in simple binding. The Y220 equivalents, Y73 and Y169 in *T. fusca* Cel6A and *T. reesei* Cel6A, respectively have been shown to cause substrate distortion in the -1 subsite [42,69,72]. Structures of family-6 GHs [65,73] show no direct hydrogen bond between the residues corresponding to *T. fusca* Cel6B Y220 and bound substrates.

Absence of a single catalytic base - *T. fusca* Cel6B appears to lack a classic Brønsted base, as none of the single mutations in any aspartic and glutamic residues (D226A, D497A and E495A), which are within 6Å of the -1 and +1 subsites, abolished activity on all polysaccharide substrates, and none of the mutant enzymes showed activity rescue by sodium azide. Similarly, azide rescue assays on SC for *T. fusca* Cel6A mutations including all four highly-conserved aspartic residues (D79A, D117A, D156A, and D265A) did not show activity rescue (unpublished data). It should be noted that *T. fusca* Cel6A is an inverting endocellulase with short loops, providing a more open active site cleft for substrates and sodium azide; additionally *T. fusca* Cel6A showed nearly 300-fold higher SC activity than *T. fusca* Cel6B [74], thus even a subtle change in *T. fusca* Cel6A activity on SC by sodium azide could be detected easily.

None of the *T. fusca* Cel6B mutations significantly altered the pK_a of D274, even D226, corresponding to residue D175 in *T. reesei* Cel6A, which affected the

protonation of the catalytic acid [55]. Among aspartic mutations in *T. fusca* Cel6A, only the D156A mutation drastically raised the pK_a of the catalytic acid [57]. However, the corresponding residue in Cel6B, D323, is buried and least 7Å away from the substrate and the D274 side chain.

The retention of nearly 90% of wild-type activity on PC eliminates D497A as a Brønsted base, which had been suggested for the corresponding D392 in *C. fimi* Cel6A [54]. This result is also consistent with the elimination of D401 as a catalytic base in *T. reesei* Cel6A [55]. The drastic decrease in 2,4-DNPC and MUG2 binding to D497A supported a role for residue D497 in substrate binding, as seen for the *T. fusca* Cel6A D265A mutant [57]. The carboxylate group of *T. reesei* Cel6A D401 was seen to interact with the O3 hydroxyl of the glucosyl unit in the -1 subsite, and loss of this interaction may account for decreased binding [55].

Crystallographic analysis of the *T. reesei* Cel6A structure indicated that the residue corresponding to Cel6B E495 is a key sugar binding residue [66]. The E495A enzyme bound weakly to MUG2, showing the importance of the hydrogen bonds between residue E495 and the sugar hydroxyl group in the -3 subsite. When the residue was replaced with Asn or Asp, BMCC activity was partially retained [60].

The only published evidence for a catalytic base in GH family 6 is the loss of activity of the *C. fimi* Cel6A D392 mutant enzyme [54]. However, sodium azide rescue and substrate binding were not reported for D392A and there was no direct evidence for the correct folding of this mutant enzyme.

Proton transferring network - The activities of the D226A and S232A mutant enzymes were substrate-specific, in that they reached target digestion on only certain substrates. Although this finding eliminates these residues as single catalytic bases, it does not exclude the participation of these residues in the activation of the catalytic

water molecule via a proton transferring network, which acts as the catalytic base. Depending on the structure of substrates and the position of the network components, their significance in the network for hydrolysis might vary. When both residues were mutated, the network could not function, causing activity loss on all substrates. The success of sodium azide rescue for the double mutant enzyme, but not for the single mutant enzymes, further supports the model where a network of D226 and S232 acts as the catalytic base. In *T. reesei* Cel6A, based on structural analysis and simulations, D175 was suggested to act as the catalytic base via a water chain between D175 and S181 (D226 and S232 in *T. fusca* Cel6B) [55]. The nucleophilic water is fixed by the D497 backbone carbonyl (Figure 2.10); therefore, removal of either S232 or D226 side chains might not completely remove this water molecule from the proper position to interact with C1.

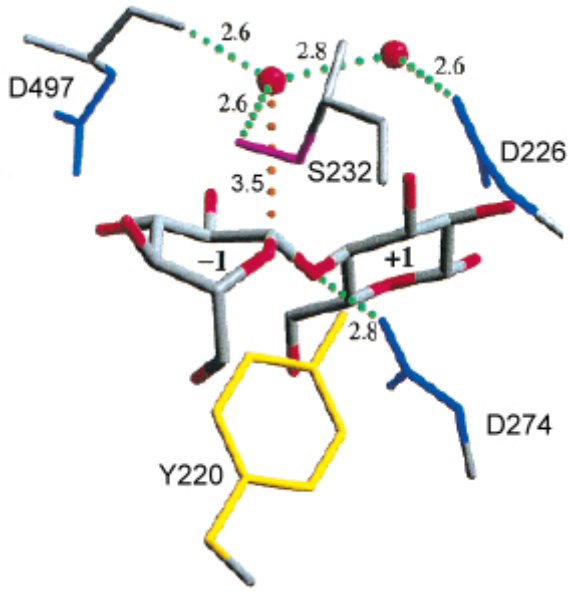


Figure 2.10: A snapshot of a molecular dynamics simulation in *T. reesei* Cel6A [55], where the residues were numbered using *T. fusca* Cel6B residues. Important hydrogen bonds are shown as green dots, and the interaction between water and C1 is shown in orange.

The participation of a non-acidic residue in a proton transferring network was reported in a *Bifidobacterium bifidum* GH-95 inverting α-fucosidase [75], where an asparagine can serve as an intermediate in the network, leading to activation of a catalytic water molecule. Analysis of the first structure of a GH-55 family member, *Phanerochaete chrysosporium* laminarinase did not identify a catalytic base residue [76], even though a candidate for the nucleophilic water was found. There are no acidic residues, but the side chains of Ser204 and Gln176 and the main chain carbonyl oxygen of Gln146 interact with this water molecule [76].

Reducing sugars would not be measured by DNS if azide adducts were formed, suggesting the azide ion act indirectly via a water molecule to perform hydrolysis. Furthermore, in another inverting *T. fusca* glycosyl hydrolase Lam81A, MALDI-TOF analysis showed no peak of azide sugar adduct in chemical rescue assays [77].

Cleavage of internal linkages - The unexpected exclusively internal cleavage of CMC by the D226A mutant enzyme is very interesting. This provides the first example of a mutant exocellulase, which could hydrolyze a soluble substrate with wild-type activity to produce mainly large soluble oligosaccharides and insoluble products, but could not hydrolyze crystalline substrate and oligosaccharides, nor produce cellobiose. Under the same experimental conditions, the CMC activity of the corresponding D79 mutation in *T. fusca* Cel6A was only 1% of wild-type [57]. As high activity was seen on both charged CMC and uncharged HEC, the activity with CMC is unlikely due to a substrate-assisted catalysis [78]. This mechanism currently has been shown only in GH retaining enzymes that cleave substrates having an acetamido group [70].

The increase in large oligosaccharide products caused by the D226A mutation may be explained by the smaller side chain, allowing modified glucose residues to bind in

the active site and thus the mutant enzyme may be able to move along a CMC molecule until it finds a group of unmodified glucose residues, where it can carry out internal cleavage. A study in a GH-18 enzyme [79] showed chitinase can processively move along the substrate without hydrolysis. Cleavage probably occurs at a lower rate than wild-type, but the great increase in potential cleavage sites due to the ability to move through modified residues compensates for this. This modification did not change the global conformation of the enzyme as circular dichroism did not reveal any global change; however, a local structural modification cannot be excluded.

Section 2.5. Conclusion

The data presented in this chapter as well as data obtained from other family-6 cellulases are consistent with the role of D274 as the catalytic acid of *T. fusca* Cel6B, and roles for E495 and D497 in substrate binding. Residue Y220 probably plays an important role in substrate distortion. Mutation of all putative catalytic base residues, within 6Å of the -1/+1 glucose binding subsites did not reveal a single catalytic base. Therefore, *T. fusca* Cel6B may function via a novel inverting mechanism without the aid of a single Brønsted base residue, but via a proton transferring network.

Section 2.6. Acknowledgements

I would like to thank Jessica Hatch and Dr. Arthur Stipanovic, Dept. of Chemistry, SUNY College of Environmental Science and Forestry for structural analysis of SC and PC.

CHAPTER THREE

T. FUSCA CEL6B LOOP RESIDUES AFFECTING SUBSTRATE SPECIFICITY, PROCESSIVITY AND SYNERGISM*

Section 3.1. Introduction

The previous chapter provides new knowledge about the catalytic mechanism of *T. fusca* Cel6B. This chapter investigates residues that might help to improve this enzyme for industrial applications.

Processivity and synergism are important properties of cellulases. Processivity indicates how far a cellulase molecule proceeds and hydrolyzes the substrate chain before dissociation. Processivity can be measured indirectly by the ratio of soluble products to insoluble products in filter paper assays [43].

Synergism between cellulases in the hydrolysis of cellulose was first demonstrated by Gilligan and Reese [80]. Four types of synergism have been demonstrated within cellulase systems: synergism between endocellulases and exocellulases, between reducing and non-reducing end-directed exocellulases, between processive endocellulases and endo- or exo-cellulases; and between β-glucosidases and other cellulases [81]. Synergism is dependent on the quality (physicochemical properties) of the substrate [82,83], the ratio of the individual enzymes [83], and the substrate saturation [84]. The degree of synergism increases with the crystallinity of cellulose; the synergism was high on highly crystalline cellulose, low in amorphous cellulose

---

* Reproduced in part with permission from "Processivity, synergism and substrate specificity of *Thermobifida fusca* Cel6B" Thu V. Vuong and David B. Wilson, Applied and Environmental Microbiology, 75 (21), p. 6655-6661 © 2009 American Society for Microbiology.

and absent on soluble cellulose [81]. Besides intra-organism synergism, synergism of mixtures from fungal and bacterial organisms or cross-synergism has been also observed [43,85].

Although cellulolytic mixtures have commercially applied in industrial processes, the rate of enzymatic hydrolysis of cellulose is relatively low [86]. Random mutagenesis approaches such as directed evolution [87] and rational protein design such as site-directed mutagenesis have been used to engineer cellulases to understand the hydrolysis mechanism [55,57,77] as well as to try to improve both the catalytic domain and carbohydrate-binding module [44,86].

There is some evidence for loop movement in exocellulases [23,24]. A comparison of native *H. insolens* Cel6A and its complexes with oligosaccharide ligands revealed the movement of two loops together so as to optimize the contacts between the enzyme and substrates [65]. The structures of *T. reesei* Cel6A in complex with different oligosaccharides have shown substantial mobility of a tunnel-forming loop, resulting in a breathing tunnel [73]. Loop movement can greatly influence enzymatic activity [24,56,60] and exocellulases might demonstrate transient endocellulolytic activity as a result of disrupting amino acids in tunnel-forming loops [88].

This chapter presents the results from engineering non-catalytic, loop residues of *T. fusca* Cel6B to obtain insight into the role of these residues in processivity and substrate specificity, as well as the relationship between processivity and synergism.

Section 3.2. Experimental procedures

Site-directed mutagenesis, enzyme purification and substrate binding assays - The protocols were the same as those presented in Chapter 2.

Polysaccharide hydrolysis assays - Activity measurement on CMC, BMCC, SW

and PC was presented in the previous chapter. Cel6B wild-type and mutant enzymes were also tested with Whatman filter-paper No.1 (FP) at 8mg/mL. All assays were run in triplicate for 16hrs at 50^{o}C in 50mM NaOAc pH5.5 at the final reaction volume of 400μL. Reducing sugars were measured using the DNS method [40].

Processivity assays - 1.5μM of enzyme was incubated with FP for 16hrs at 50^{o}C in 50mM NaOAc pH5.5. The filter paper circle was separated from the supernatant and washed with the buffer three times, and then the reducing sugar content of the supernatant (soluble) and the filter paper (insoluble) was determined by the DNS reagent. The ratio of soluble/ insoluble reducing sugars was used to calculate processivity [43].

HPLC analysis - HPLC was run on Shimadzu HPLC equipment consisting of a LC-20AD pump, a SIL-20A autosampler and a RID-10A refractive index detector. Separation was achieved on an Aminex® HPX-87P analytical column (300mm x 7.8mm, Bio-Rad), equipped with Micro-Guard® deashing cartridges (Bio-Rad). The column oven temperature was 84^{o}C. 50pmol of enzyme was incubated with 3.2mg of FP or 500nmol of cellotriose (Megazyme, Wicklow, Ireland) for 16hrs, or with 10nmol of cellohexaose (G6) (Megazyme) for 10min in 400μL of 50mM NaOAc pH5.5 at 50^{o}C. The samples were filtered through Millipore® 5K NMWL membrane filter devices before being injected at 0.6μL/min. Data were analyzed using OriginPro v.8.0 (Origin Lab, MA, USA).

Synergism assay - Synergism assays for wild-type and selected mutant enzymes were run in the presence of *T. fusca* Cel5A and/or Cel48A on FP and PC at different molar ratios. Synergism was measured based on the amount of enzyme needed to achieve 5% digestion on FP or 6% digestion on PC in 16hrs at 50^{o}C. The synergistic effects were calculated using the activity of the mixture divided by the sum of the

individual activities.

Thermostability assay - Cel6B wild-type and select mutant enzymes were pre-incubated in 50mM NaOAc pH 5.5, at temperatures from 45^{o}C to 70^{o}C for 16hrs, and then 1.5µM enzyme was assayed on SC at 50^{o}C for 16hrs to calculate T_{50}, the temperature where activity dropped by 50%.

Circular dichroism (CD) analysis - Spectra of 10µg/uL protein were recorded from 190 to 290nm on an Aviv CD400 Spectrometer (AVIV Biomedical, INC.) at a scanning rate of 1nm/s at 4^{o}C. The CD spectra were analyzed for percent secondary structure using CDNN CD spectra deconvolution software, which was developed by Böhm *et al*. [89].

Section 3.3. Results

Selection of mutations – Based on the structural models presented in the previous chapter, a number of loop residues were chosen for mutation (Table 3.1). N282 and R180 are located at the +2 and -4 glucose subsites, respectively whereas L230 is located in a turn on the top of the tunnel (Figure 3.1). The two potential sugar binding residues N282 and R180 were mutated to investigate the role of residues near the tunnel entrance and exit on processivity. W464, which corresponds to W371 in *H. insolens* Cel6A was suggested to participate in substrate binding (Figure 3.2) [56] while residues D512 and M514 might affect loop flexibility.

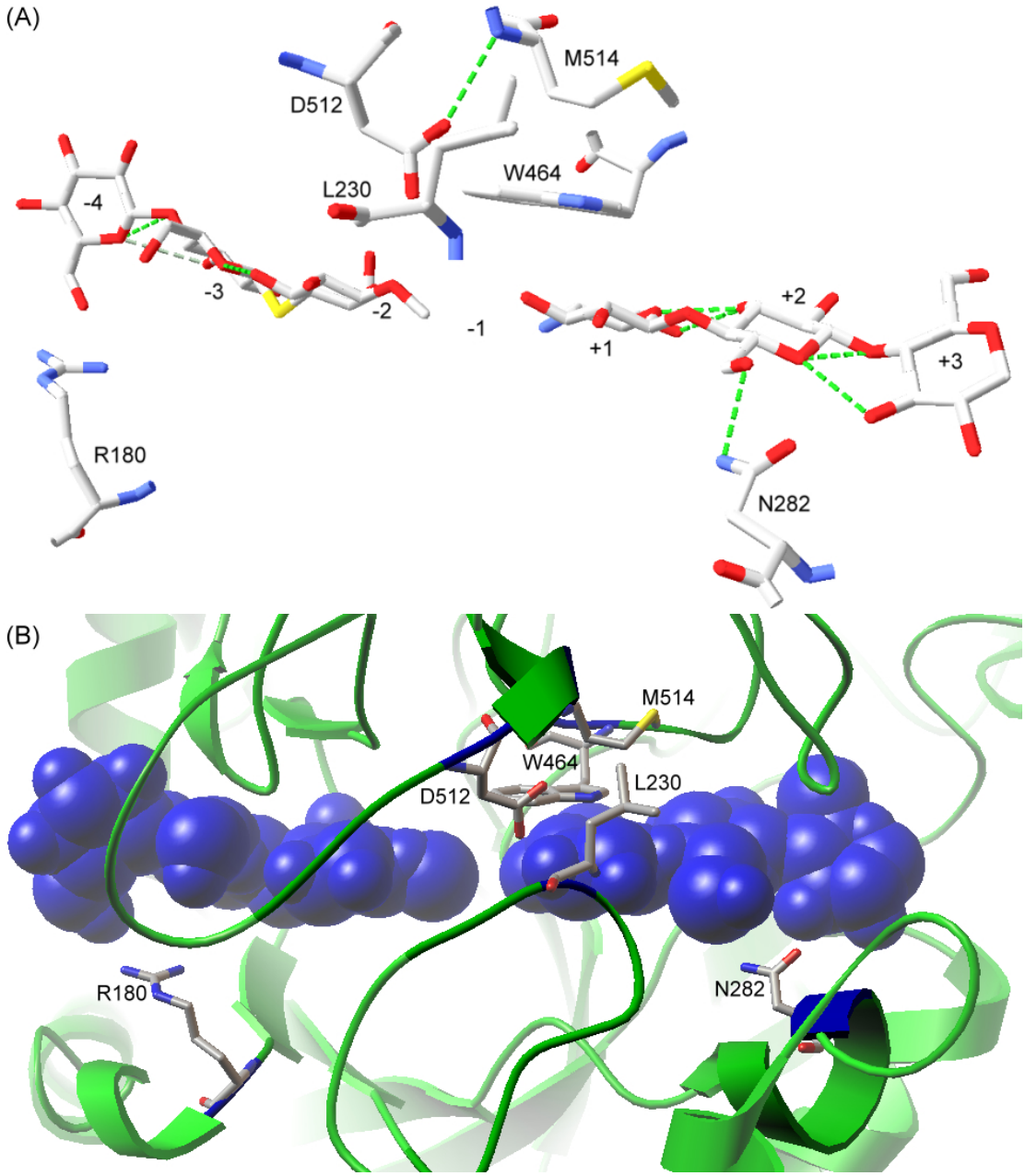


Figure 3.1: Location of *T. fusca* Cel6B residues for mutation, modeled to *H. insolens* Cel6A (1OCB, 1.75Å resolution) by the Swiss-Model Workspace. (A): 1D-view, dashed lines show hydrogen bonds; (B): 3D-view, showing the active site tunnel with two molecules of fluoresceinylthioureido-derivatized tetrasaccharide (for neatness, the +4 glycosyl residue was removed).

Table 3.1: Amino acid residues chosen for site-directed mutagenesis.

Residue	Location	Subsite	Corresponding residue in:	
			T. reesei	*H. insolens*
R180	On the top of the tunnel exit	-4	K129	E131
L230	On a turn, top of the tunnel	+1	L179	A184
N282	In an α-helix, near the top of tunnel entrance	+2	N229	N234
W464	Side wall of the tunnel	+1	W367	W371
D512	On the top of the tunnel	+1/-1	D412	D416
M514	Side wall of the tunnel	+1	H414	H418

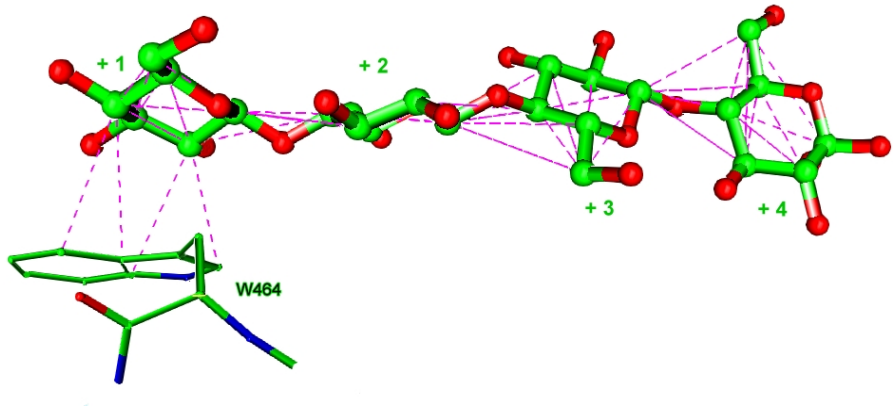
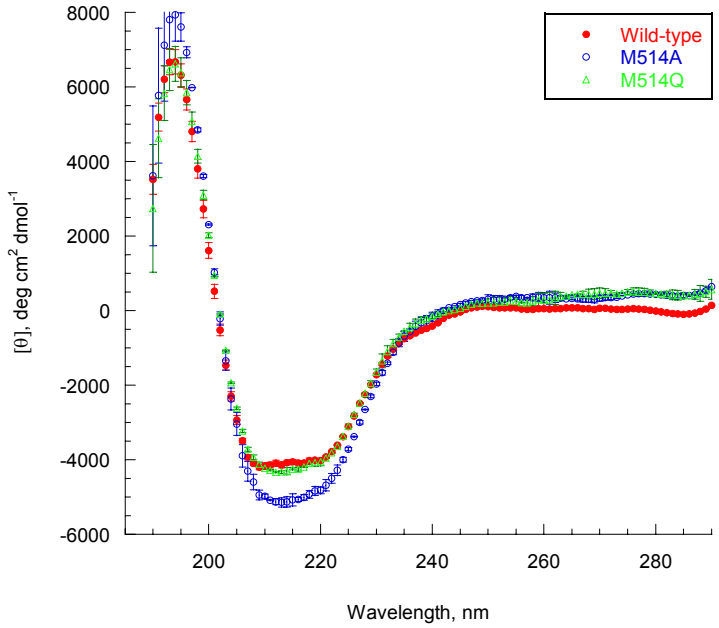


Figure 3.2: Interaction of residue W464 with a substrate as modeled by Ligand Explorer (dash - hydrophobic interaction) using the structure of *H. insolens* Cel6A (2BVW).

Enzyme activity and processivity - All mutant enzymes behaved like wild-type Cel6B during purification. The circular dichroism spectra of all mutant enzymes, except M514A and M514Q (Figure 3.3) were identical with that of wild-type, indicating that the global secondary structure of the mutant proteins remained intact.

(A)



(B)

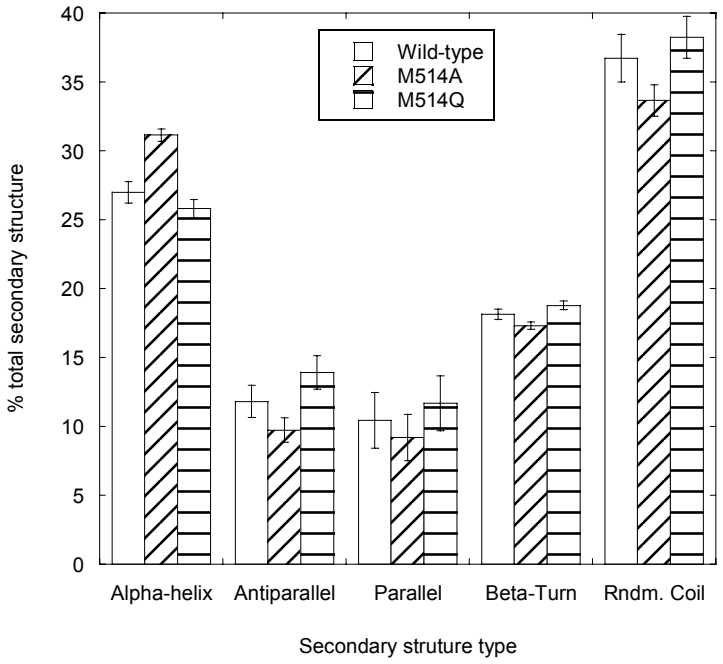


Figure 3.3: Circular dichroism analysis of the wild-type and mutant Cel6B. (A) Circular dichroism spectra of the wild-type, M514A and M514Q mutant enzymes. (B) Relative amount of each type of secondary structure of each enzyme. The error bars represent the standard deviation for three independent trials.

The purified enzymes were assayed on five polysaccharides, and their activities were expressed as percentage of wild-type activity to facilitate comparison (Table 3.2). Besides cellulose substrates presented in the previous chapter, the enzymes were also assayed on filter paper (FP), which is a crystalline substrate, made from long-fiber cotton pulp, with a degree of polymerization (DP) >1000 [2].

Mutant enzymes in residues near the tunnel exit (R180K and R180A) and the tunnel entrance (N282A and N282D) had on average a 2-fold increase in processivity (Table 3.2). The L230A mutation slightly increased processivity, and increased PC activity over 250% (Table 3.2). HPLC was used to investigate the production of oligosaccharides by the N282A and L230A enzymes on FP. While cellobiose (G2) is the main product, small amounts of cellotriose (G3) and glucose (G1) were also produced (Table 3.3). As cellobiose is the repeating unit of cellulose (5), it is thought that the first hydrolytic step can produce either G3 or G2, but the subsequent steps yield only G2. TLC and HPLC of the products of G3, G4, G5 and G6 hydrolysis showed G4, G5 and G6 were completely hydrolyzed within minutes while a small amount of G1 was detected from G3 hydrolysis after 16hr-incubation (data not shown). Therefore, the (G2-G1)/(G3+G1) ratio can provide an assay of processivity. Both N282A and L230A produced approximately 2.5-fold more oligosaccharides than wild-type, and their (G2-G1)/(G3+G1) ratios were 1.8-fold higher (Table 3.3). However, when hydrolyzing G6, the N282A enzyme was less active and produced a lower ratio of G2/G3 than the L230A enzyme (Table 3.3).

Table 3.2: Activities and processivity of the Cel6B mutant enzymes on polysaccharide substrates.

	Activity (μmole cellobiose min^{-1} $\mu mole^{-1}$ enzyme)[a]					Processivity
	BMCC	SC	PC	CMC	FP[b]	
Wild-type	0.93	2.25	3.37	0.57	0.22	7.2
	Percentage of wild-type activity					
Mutant enzymes for processivity						
R180A	85	125	130	96	91	16.4
R180K	59	123	180	104	100	13.6
L230A	108	137	252	126	159	9.9
N282A	105	86	313	196	145	20.9
N282D	116	110	323	158	145	13.1
Mutant enzymes for substrate specificity						
W464A	26	86	368	718	132	7.9
W464Y	54	79	159	195	114	8.9
D512A	34	174	240	568	95	5.7
M514A[c]	131	131	--	151	91	5.9
M514Q[c]	125	128	--	174	118	9.1

[a]Activity was calculated at 6% digestion for BMCC, SC and PC and 1.5% digestion for CMC. The average coefficients of variation were 4, 5, 5.5, 2.5, 3 and 4% for BMCC, SC, PC, CMC, FP, and processivity (soluble/insoluble reducing sugars), respectively. [b]Activity was calculated at 1.5μM of enzyme. [c]Activity was measured right after purification when PC had not been prepared.

Table 3.3: Oligosaccharide production on filter paper (FP) and cellohexaose (G6) by the Cel6B wild-type (WT) and mutant enzymes[a].

Enzyme	FP hydrolysis (16hr-incubation)				G6 hydrolysis (10min-incubation)	
	G1	G2	G3	$\frac{(G2-G1)}{(G3+G1)}$	Unhydrolyzed G6	$\frac{G2}{G3}$
WT	0.22 ± 0.07	3.35 ± 0.07	0.16 ± 0.02	8.2	0.65 ± 0.13	0.97 ± 0.01
L230A	0.33 ± 0.09	9.22 ± 0.02	0.25 ± 0.02	15.3	0.51 ± 0.07	1.14 ± 0.04
N282A	0.22 ± 0.04	8.29 ± 0.04	0.34 ± 0.06	14.4	0.90 ± 0.04	0.86 ± 0.01

[a]Oligosaccharides (nmol) were determined by HPLC. G1, G2 and G3 are glucose, cellobiose and cellotriose, respectively.

Synergism with other T. fusca enzymes - Selected mutant enzymes including L230A and N282A were assayed in the presence of *T. fusca* endocellulase Cel5A to test for synergism in FP hydrolysis at a molar ratio of 4:1, which was previously found to be optimal [43,90]. Although the FP activity of those individual mutant enzymes were up to 150% of the wild-type activity, their mixtures with Cel5A did not give higher synergism than the wild-type mixture (Table 3.4). A similar pattern was observed for a range of incubation time (4-16hrs) (Figure 3.4), at two different molar ratios (9:1 and 19:1) and in mixtures with *T. fusca* Cel9A-68, a processive endocellulase (data not shown). The same result was observed for mixtures of Cel6B enzymes with Cel5A for PC hydrolysis (Table 3.5). Although the N282A mutant enzyme alone had over 310% of wild-type PC activity, the activity of the mixture was only 63% of that of the wild-type mixture (Table 3.5). Processivity on PC was not reported due to difficulty in measuring insoluble reducing sugars from this substrate.

Table 3.4: Synergism of Cel6B enzymes with *T. fusca* endocellulase Cel5A and exocellulase Cel48A in FP hydrolysis.

Mixture	Mol ratio in mixture	Specific activity[a]	% wild-type/wild-type mixture activity	Synergism factor[b]	Processivity
Cel6B		0.22	100		7.2
L230A		0.35	159		9.9
N282A		0.32	145		20.9
Cel5A		0.93			
Cel6B+Cel5A	4:1	2.39	100	6.2	8.4
L230A+Cel5A	4:1	2.41	101	5.5	7.6
N282A+Cel5A	4:1	2.05	86	3.8	5.6
Cel6B+Cel48A	1:1	1.24	100	2.0	7.3
L230A+Cel48A	1:1	1.69	136	3.5	9.6
N282A+Cel48A	1:1	1.45	117	2.8	12.9
Cel6B+Cel48A+Cel5A	4:8:1	6.69	100	6.4	
L230A+ Cel48A+Cel5A	4:8:1	5.70	85	4.5	
N282A+ Cel48A+Cel5A	4:8:1	2.85	43	3.7	

[a]μmole cellobiose min^{-1} μmole^{-1} enzyme; the average coefficient of variation was 4%. [b]Activity of the mixture divided by the sum of the activities of the mixture components.

Table 3.5: Synergism of Cel6B enzymes with *T. fusca* endocellulase Cel5A in PC hydrolysis.

Mixture	Mol ratio in mixture	Specific activity[a]	% wild-type/wild-type mixture activity	Synergism factor[b]
Cel6B		3.4	100	
L230A		8.5	252	
N282A		10.6	313	
Cel5A		49.4		
Cel6B+Cel5A	4:1	460	100	3.3
L230A+Cel5A	4:1	362	79	2.5
N282A+Cel5A	4:1	289	63	2.2

[a]μmole cellobiose min^{-1} $\mu mole^{-1}$ enzyme; the average coefficient of variation was 4%. [b]Activity of the mixture divided by the sum of the activities of the mixture components.

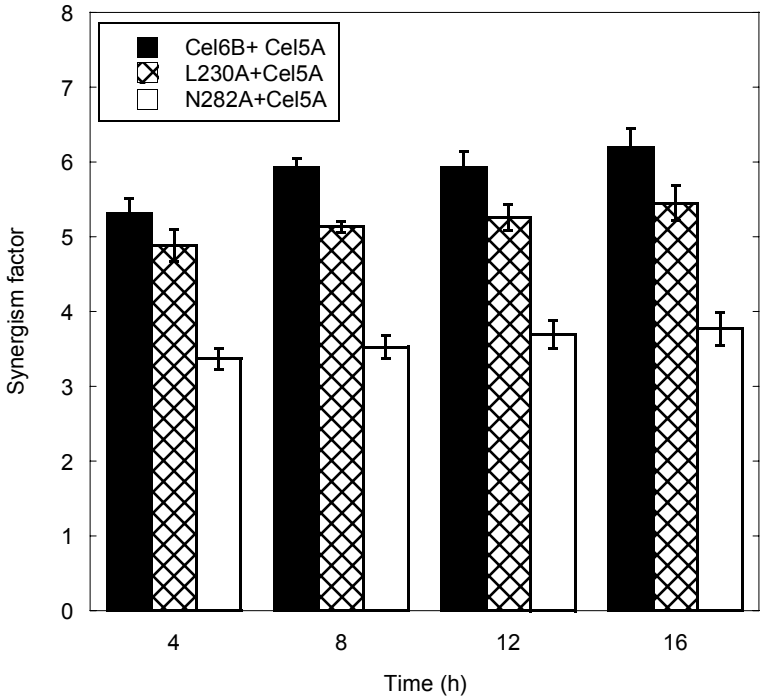


Figure 3.4: Time course of FP synergism of exocellulase Cel6B and endocellulase Cel5A mixtures.

To test the correlation between processivity and synergism, the processivity of the mixtures was measured. Mixtures of mutant enzymes with Cel5A showed lower processivity than wild-type mixtures (Table 3.4). Cel6B mutant enzymes were also mixed with *T. fusca* Cel48A, an exocellulase that attacks the reducing end of cellulose. The Cel6B mutant enzymes, which individually showed higher processivity, gave higher synergism and processivity with Cel48A than the wild-type mixture (Table 3.4).

Substrate specificity - The W464A and W464Y mutant enzymes had reduced activity with BMCC, but had higher activity with PC and CMC (Table 3.2). The binding of the W464A enzyme to BMCC was lower than that of the wild-type enzyme (Figure 3.5). The activity of the W464A enzyme with CMC increased sevenfold, and that of the W464Y enzyme nearly doubled. The high CMC activity of the W464A

mutant enzyme was retained at lower concentrations of CMC (0.25 and 0.5%) (Figure 3.6). TLC analysis confirmed the high CMC activity of this enzyme, and showed that the products were not changed by the mutation (data not shown). Residue D512 is structurally close to W464; its mutation to Ala also gave higher activity on PC and CMC (240 and 568%, respectively), but only 34% wild-type activity on BMCC (Table 3.2).

To test whether the high CMC activity of the W464A enzyme was due to higher binding, CMC binding assays were carried out. When the wild-type and W464A enzymes were run on native gels, they both migrated to the same position (Figure 3.7A). However, only the W464A enzyme migrated in a CMC-containing native gel at 4^{o}C or at room temperature. *T. fusca* Cel5A, an endocellulase with five orders of magnitude higher CMC activity than Cel6B (data not shown) migrated as fast as the W464A enzyme (Figure 3.7B). Congo Red staining showed that both Cel5A and W464A produced yellow traces when running on CMC native gels at room temperature (Figure 3.7B), indicating hydrolysis of CMC. Because the CMC activity of the W464A enzyme was much lower than that of Cel5A, a yellow trace was observed only when a very high concentration of the W464A enzyme (0.9nmoles) was used.

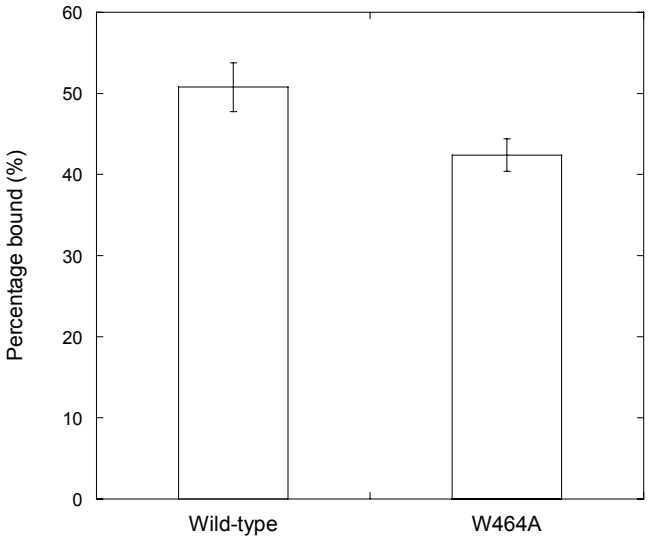


Figure 3.5: Binding of the wild-type and W464A enzymes to BMCC. Substrate binding was conducted using 4µM of enzymes in 50mM NaOAc pH5.5 for 1hr at 4°C.

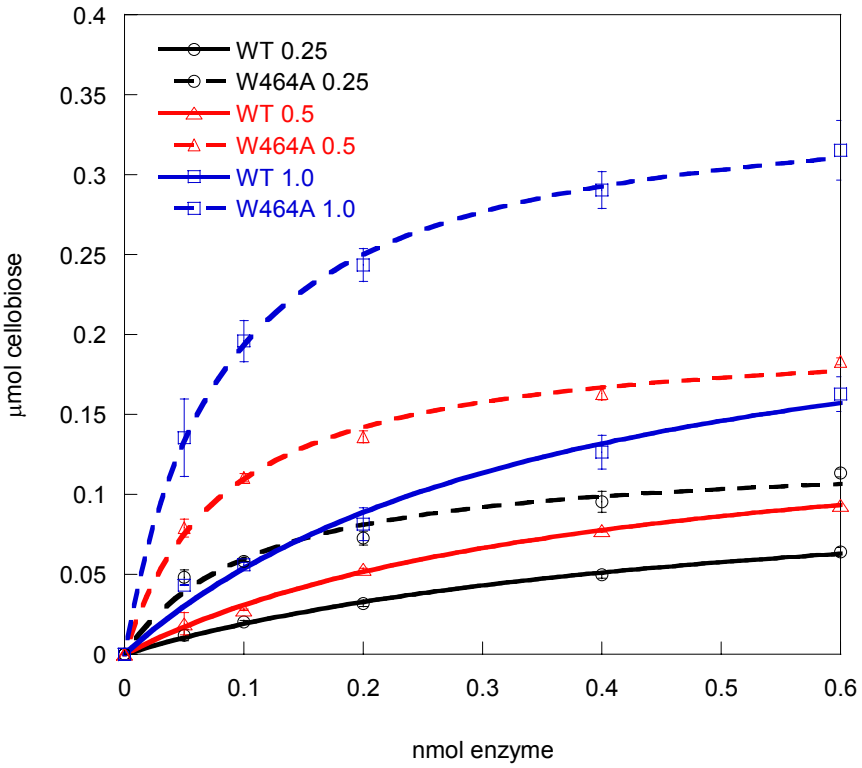


Figure 3.6: Activities of the W464A enzyme and the wild-type enzyme (WT) on different concentrations of CMC (0.25, 0.5 and 1%).

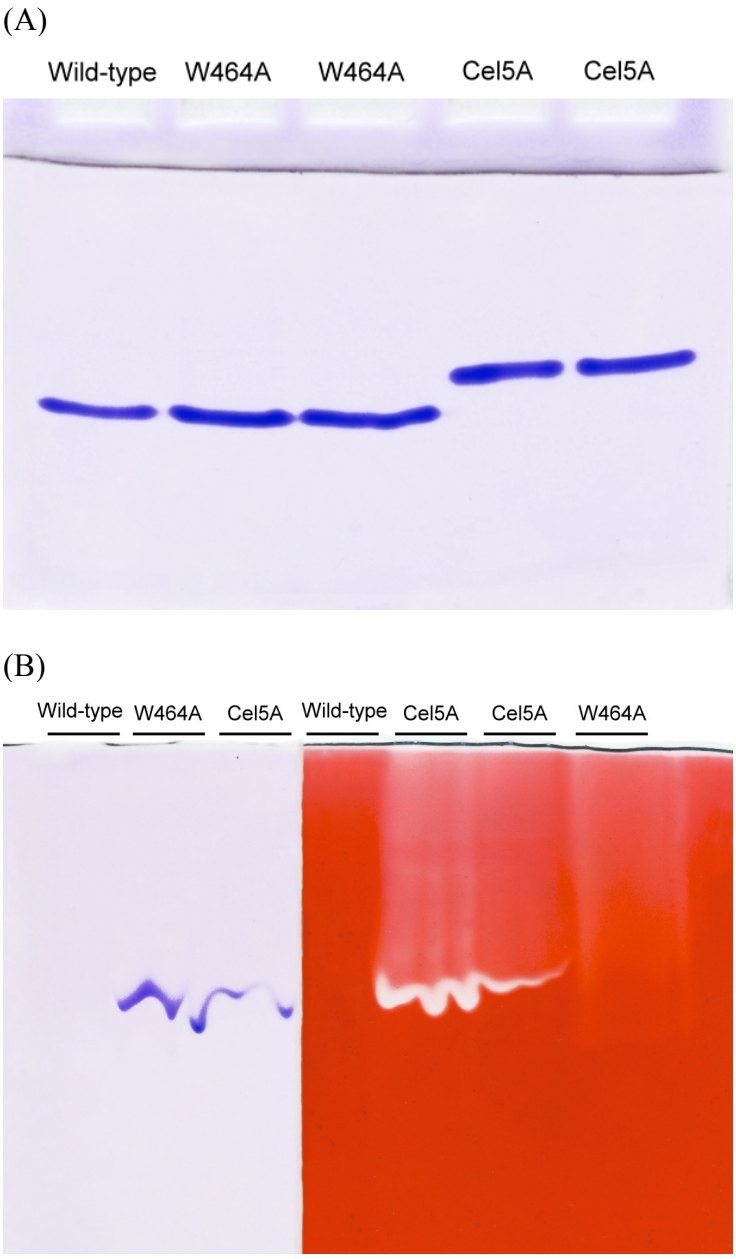


Figure 3.7: *T. fusca* Cel6B and Cel5A enzymes on a native gel (A) and a CMC-containing native gel (B) at room temperature. (A): 0.09nmol of each enzyme loaded. (B): Left- Coomassie staining, 0.09nmol of each enzyme loaded; Right- Congo Red staining, from left to right, Cel6B (0.9nmol), Cel5A (0.009 and 0.0018nmol) and W464A (0.9nmol).

The BMCC activity of the two M514 mutant enzymes unexpectedly decreased with increasing storage time (in 5mM NaOAc pH 5.5 and 10% glycerol at -70^{o}C). BMCC assays conducted right after purification showed that the mutant enzymes had slightly higher activities than wild-type (Table 3.2). Five months after the first enzymatic assay, the M514A specific activity decreased to approximately 2% of wild-type activity (data not shown). The loss of BMCC activity in the M514A enzyme correlated with increased CMC activity; and the loss of enzymatic activity of the M514A enzyme was always higher than that of the M514Q enzyme (data not shown). SDS-PAGE gels showed no difference in mobility or band pattern between boiled and unboiled samples of the wild-type and the mutant enzymes (data not shown), eliminating enzymatic degradation during storage. Both mutant enzymes were unstable at 55°C and above. Their T_{50} was 58°C while that of wild-type was 64°C (Figure 3.8). Circular dichroism analysis after two years of storage showed that the M514A enzyme spectrum, particularly from 207-228nm and the M514Q spectrum from 260-290nm are different from wild-type, indicating a structural modification of each enzyme (Figure 3.3A). CD spectra analysis indicated the α-helix content of the M514A mutant enzyme was higher while the content of random coil was slightly lower than wild type (Figure 3.3B).

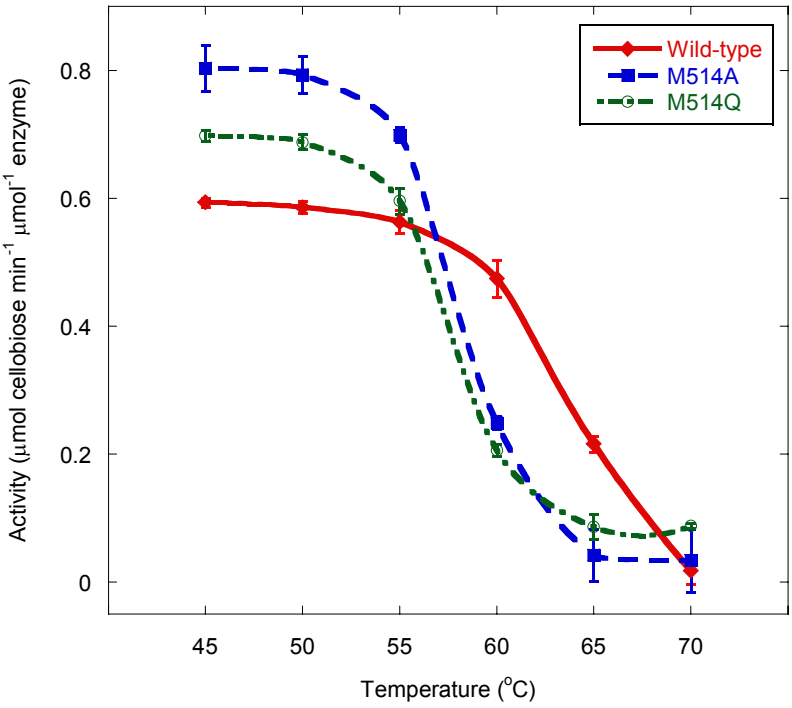


Figure 3.8: Thermostability of wild-type Cel6B and M514 mutant enzymes (right after purification). The enzymes were pre-incubated in 50mM NaOAc pH5.5 at 45-70°C for 16hrs, then assayed on SC at 50°C for 16hrs.

Section 3.4. Discussion

Effects of tunnel entrance and exit residues on processivity - Mutation of either N282 or R180, which are at opposite ends of the active site tunnel, to amino acids with shorter side chains increased the processivity of Cel6B, and mutation to the smallest side chain (Ala) gave the largest increase. Each subsite of a cellulase can accommodate both faces of the pyranoside ring and tolerate the C_6 hydroxyl group when the substrate moves along the catalytic site [56]. The decrease in size of the side chains at these positions might allow the cellulose molecule more freedom to advance through the tunnel in case of N282A and facilitate the release of cellobiose for R180A. A study on *Aspergillus niger* endopolygalacturonases also showed that a region far away from the scissile bond (subsite -5) strongly influences processivity [91]. High processivity does not always indicate higher activity as is found for the FP activity of the R180 mutant enzymes. Processivity may be more about disassociation rather than the rate of hydrolysis.

Increased processivity of the N282A mutant enzyme also was shown by its high (G2-G1)/(G3+G1) ratio. Although the processivity of N282A as measured by the ratio of soluble/insoluble reducing sugars was higher than that of L230A, their oligosaccharides ratios were close to each other. This might be due to a difference in their initial substrate binding preference. As very small amounts of G3 and G1 were produced, a small change in substrate binding, which might be detected by G6 hydrolysis, could significantly affect the ratio of (G2-G1)/(G3+G1). The lower G2/G3 ratio from G6 hydrolysis by N282A indicates that its true ratio of (G2-G1)/(G3+G1) in FP hydrolysis is higher than measured, consistent with its higher processivity from the ratio of soluble/insoluble reducing sugars. The similarity of the two approaches for measuring processivity is further supported by the L230A mutant enzyme. The higher

(G2-G1)/(G3+G1) ratio of L230A is due to its lower initial binding preference leading to G3 as it produced less G3 from G6.

The ratio of oligosaccharides can not be used to assess processivity in mixtures with endocellulases, as G3 is produced by internal cleavage as well as in the initial hydrolysis step. A different ratio, G2/(G1+G3) was used to measure processivity for *T. reesei* exocellulase Cel7A [92] as G1 was assumed to be released only from the initial attack (G3 hydrolysis by Cel7A was not addressed in this study). High processivity means that the enzyme has been optimized for the movement of a cellulose chain in the active site; however, this change can reduce hydrolysis activity on easily diffusible soluble substrates [93], which is in agreement with the slow hydrolysis of G6 by N282A.

A link between processivity and synergism - Although the individual mutant enzymes had higher FP activity than wild-type, their mixtures with *T. fusca* Cel5A did not show increased synergism on FP. Walker et al. [90] found that there was no binding competition between Cel6B and Cel5A. Jeoh et al. [94] found that substrate binding of Cel6B and Cel5A in mixtures was higher than that of the individual enzymes. A very low enzyme to substrate ratio was used here so that competition for adsorption is unlikely. The fact that the rate of hydrolysis of an exocellulase increased on endocellulase-pretreated cellulose [5] as well as the fact that synergism occurs between cellulases from unrelated organisms [43] shows that synergism does not require a direct interaction between the cellulases. A simple synergism model is that endocellulases act on accessible sites, producing new ends for the attack by exocellulases, which in turn open up new sites for endocellulases. However, our data suggested a more complicated synergism, in which a more active exocellulase does not give higher synergism even at low enzyme to substrate ratios. As endocellulase

Cel5A produces shorter cellulose chains, exocellulase mutants with increase in processivity are not needed for maximizing synergism. In contrast, these mutant enzymes did give increased synergism with the reducing-end attacking exocellulase Cel48A.

Different effects of mutations on various substrates - CMC activity is not always a good indicator of higher activity on crystalline substrates. Soluble CMC is expected to bind readily in the active site; however, its high proportion of modified residues may cause CMC to bind in a distorted manner. The tunnel structure of *H. insolens* Cel6A restricts the polysaccharide strand in the tunnel [65] so the increased CMC activity of the W464A enzyme might be due to easier movement of modified sugars through the tunnel after the removal of the bulky side chain. Increased CMC activity and decreased BMCC activity also have been seen in several *T. fusca* Cel9A mutations [44] when substrate binding Trp residues were mutated to smaller side chains.

The decrease in both activity and binding to BMCC caused by the W464A mutation indicates that W464 helps to bind a cellulose chain into the active site and this function may not be required for binding of easy accessible substrates like CMC and PC. Structural analysis showed that the corresponding residue in *T. reesei* Cel6A, W367 interacts with the α-face of a glucosyl ring during productive binding of a cellulose chain [21] and the position of *H. insolens* Cel6A W371 was shifted upon ligand binding [56].

The specific loss of BMCC activity in the M514 mutant enzymes upon storage suggests that a change in structure occurs on storage that inactivates the rate limiting step for crystalline cellulose activity. Residue M514 is located in a tunnel-forming loop at subsite +1, next to residue C515, which forms a disulfide bond with residue C465 in this loop. The increased flexibility of this disulfide bond in the mutant

enzyme might cause a conformational change as shown by circular dichroism and reduced thermostability.

Section 3.5. Conclusion

Mutation of several residues located in the active site tunnel of *Thermobifida fusca* exocellulase Cel6B increased processivity on filter paper. Surprisingly, mixtures of these Cel6B mutant enzymes and a *T. fusca* endocellulase Cel5A did not show increased synergism or processivity, and the mutant enzyme which had the highest processivity gave the poorest synergism. This study suggests that improving exocellulase processivity might not always be an effective strategy for producing improved cellulase mixtures for biomass conversion. The inverse relationship between bacterial microcrystalline cellulose and carboxymethyl cellulose activities of many of the mutant enzymes indicated differences in the mechanisms of hydrolysis for these substrates, supporting the possibility of engineering Cel6B to target selected substrates.

CHAPTER FOUR

ADDITIONAL EXPERIMENTS*

Section 4.1. Introduction

In the course of completing the experimental work presented in the previous chapters, several additional research projects were conducted to investigate other aspects of *T. fusca* exocellulase Cel6B. Each individual project is not large enough to deserve a separate chapter. However, together they provide experimental knowledge on fluorescence-labeling of *T. fusca* exocellulase Cel6B and non-productive binding of this enzyme to other polysaccharides. These experiments also provide more details about the role of Cel6B and other proteins in *T. fusca* supernatants. Therefore, they are included in this dissertation.

Section 4.2. Fluorescence labeling of *T. fusca* Cel6B enzymes

Chapter 3 has described several Cel6B mutant enzymes with higher processivity; however, it is unclear what causes dissociation of the processive enzymes. Fluorescence labeling of these enzymes to track their movement may answer this question while offering an optical approach for measuring binding and processivity. Tracking a quantum dot-labeled CBM2 from *Acidothermus cellulolyticus* using single-molecule spectroscopy indicated a linear motion of this CBM along the cellulose fiber [95]. Therefore, the Cel6B wild-type and mutant enzymes with higher processivity including N282A and L230A were fluorescently labeled with amine-reactive Alexa

---

* Reproduced in part from "Engineering *Thermobifida fusca* cellulases: catalytic mechanisms and improved activity" Thu V. Vuong and David B. Wilson, accepted in: Protein Engineering. Nova Science Publishers, Inc.

Fluor 647® (AF647) succinimidyl ester (Invitrogen, CA, USA).

4.2.1. Fluorescence-labeling protocol

This labeling method was based on the method of Moran-Mirabal et al. [96]. The binding matrix is an 8:1 (w/w) mixture of CF11:BMCC at 32mg/ml in 50mM NaOAc buffer pH 5.5. 1mL of the matrix was pipetted into each Costar® 0.45µm nylon centrifugation tube, and centrifuged for 1min at 10,000g to remove the buffer. The Cel6B enzymes were bound to the matrix for 1hr with end-over-end agitation at 4°C in 300µL of the labeling buffer (35mM boric acid buffer and 50mM NaCl pH 8.3). Unbound enzymes were removed by two consecutive centrifugations (1min at 5,000g). The fluorophore was added at a 100:1 molar ratio of fluorophore and enzyme. The mixture was incubated for 24hrs at 4^{o}C without agitation. Unreacted dye was removed by six consecutive centrifugations (2min at 5,000g), followed by addition of 500µL of the labeling buffer. The enzymes were recovered by 3 elutions with chilled ethylene glycol (EG). Two washes were done with the addition of 400µL EG, followed by incubation in ice for 10min, and centrifugation for 5min at 6,000g and 8,000g, respectively. The third wash was conducted with 200µL EG and centrifugation for 10min at 10,000g. The flow-through of each wash was immediately diluted 7.5x with 20mM chilled MES buffer pH6.0, and loaded onto a Vivaspin 4 (10kDa MWCO, Sartorius, NY, USA) columns before being centrifuged at 7,000g for 10min at 4°C.

Separation of labeled species was done by FPLC at 21°C with the MES buffer as mobile phase, using an ÄKTA Explorer 10S FPLC system and a Resource Q column (1ml, GE Healthcare, NJ, USA). The column was equilibrated with 250mM NaCl, and 1ml labeling mixture was injected. Enzymes were eluted by applying a 250mM linear salt gradient (3ml/10mM NaCl) at 1.5ml/min flow rate. Absorbance was recorded at

280nm for proteins, and at 650nm for AF647. Eluted proteins were collected in 1ml fractions, and partitioned into labeled enzyme species according to the chromatograms. Fractions were collected, concentrated, and then buffer was exchanged with 50mM NaOAc pH 5.5 using Vivaspin 4 (10kDa MWCO) columns. Labeled species were divided into aliquots at a concentration of 200nM and stored at -20^{o}C.

4.2.2. Labeled enzyme concentration

Protein concentrations and degrees of labeling (DoL, or moles dye per mole protein) were calculated as followed:

$$\text{Enz. conc.(M)} = \frac{[\text{Abs.280} - \text{Abs.Max}_{\text{dye}} \times \text{C.F}] \times \text{Dilution}}{\varepsilon_{\text{protein}}}$$

$$\text{Moles dye per mole protein} = \frac{\text{Abs.Max}_{\text{dye}} \times \text{Dilution}}{\varepsilon_{\text{dye}} \times \text{Enz. conc.}}$$

Where Abs.280 is the absorbance at 280nm, $\text{Abs.Max}_{\text{dye}}$ is the absorbance at the fluorophore's excitation wavelength, C.F is the fluorophore correction factor, and $\varepsilon_{\text{protein}}$ and ε_{dye} are the extinction coefficients of the protein and the fluorophore, respectively.

4.2.3. Enzymatic activity of labeled enzymes

The activities of selected labeled species were assayed with BMCC and PC. Costar® Spin-X 0.45μm nylon tubes were washed with MQ water by centrifuging 5,000g for 5min. 300μL of 5mg/mL BMCC or PC along with a glass bead were added to the Spin-X tubes. The final concentration of each labeled species was 25nM and the final reaction volume was 600μL. The reaction was incubated at 50^{o}C for 24hrs with 360^{o} rotation. The tubes were then centrifuged 5,000g for 5min. Cellulases were

removed by binding to washed BMCC for 1hr at 4^{o}C, and the oligosaccharides were separated and quantified by HPLC, as described in Chapter 3.

4.2.4. Labeled Cel6B characterization

The Cel6B enzymes were successfully fluorescence-labeled with AF647 to different degrees of labeling (Figure 4.1). As the enzymes were bound to substrate before being labeled, fluorophores were unlikely to interact with key residues for catalysis. The AF647 fluorophore reacts only with accessible lysine residues to form stable dye-protein conjugates. There are four accessible lysine residues in Cel6B as calculated with a minimum of 30% surface exposure; therefore, the maximum expected DoL is 4. The highest DoL of labeled species was 3.4 (Table 4.1), indicating that the dye was able to interact with most of accessible lysine residues of each cellulase molecule.

Labeled species were assayed on both BMCC and PC. Low degrees of labeling did not affect catalytic activity, but higher DoL did in some cases. The peak 11 of Cel6B wild-type, which has a DoL of 2.8 (Table 4. 1), showed lower activity on PC compared with wild-type enzyme (Figure 4. 2). A similar trend of labeled wild-type activity was observed on BMCC (data not shown). Surprisingly, the total amount of (G2 + G3) produced on PC by the peaks 1 and 6 of labeled L230A was about 30% higher than the unlabeled L230A enzyme. However, this increase in activity was not seen with BMCC. These labeled species also produced a high ratio of G2/G3 with PC (data not shown). Neither L230A nor N282A produced more oligosaccharides on PC than the wild-type enzyme (Figure 4.2). This variance might be due to the fact that we could not detect and measure glucose because of HPLC artifacts, thus glucose was not included in the calculation.

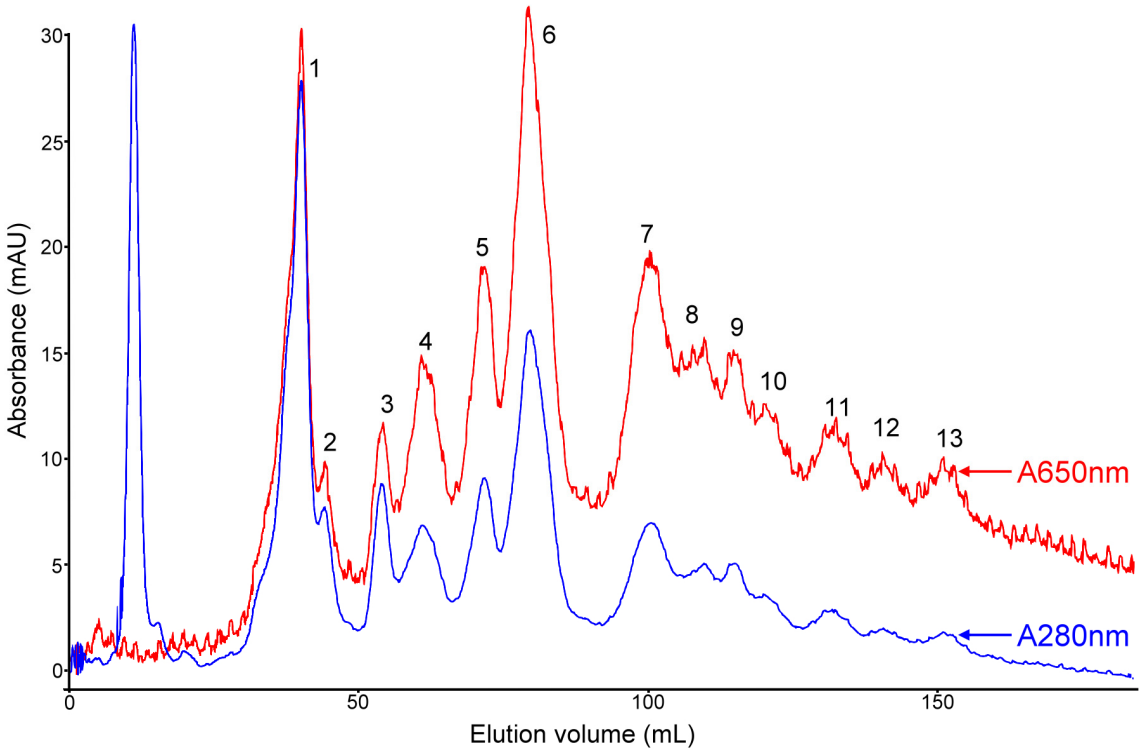


Figure 4.1: FPLC chromatogram showing thirteen fractions (1-13) of an AF647-labeled Cel6B mutant enzyme (L230A) with different degrees of labeling, ranging from 1.0 to 3.3. Unlabeled protein (absorbance at 280nm) elutes earlier than labeled products (absorbance at 650nm).

Table 4.1: Fractions of the Cel6B wild-type and mutant enzymes with different degrees of labeling (DoL). The samples are denominated by the enzyme (wild-type, L230A or N282A), fluorophore (647), and peak number (PXX).

Peak Name	DoL	Peak Name	DoL	Peak Name	DoL
Wild-type 647 P1	1.1	L230A 647 P1	1.0	N282A 647 P1	1.1
Wild-type 647 P2	0.9	L230A 647 P2	0.9	N282A 647 P2	0.9
Wild-type 647 P3	1.3	L230A 647 P3	1.0	N282A 647 P3	1.1
Wild-type 647 P4	1.8	L230A 647 P4	1.7	N282A 647 P4	1.8
Wild-type 647 P5	2.0	L230A 647 P5	1.8	N282A 647 P5	1.9
Wild-type 647 P6	2.1	L230A 647 P6	2.0	N282A 647 P6	2.0
Wild-type 647 P7	2.9	L230A 647 P7	2.5	N282A 647 P7	2.7
Wild-type 647 P8	2.6	L230A 647 P8	2.6	N282A 647 P8	2.5
Wild-type 647 P9	2.5	L230A 647 P9	2.3	N282A 647 P9	2.4
Wild-type 647 P10	2.6	L230A 647 P10	2.6	N282A 647 P10	2.7
Wild-type 647 P11	2.8	L230A 647 P11	2.5	N282A 647 P11	3.3
Wild-type 647 P12	3.1	L230A 647 P12	-	N282A 647 P12	3.1
Wild-type 647 P13	3.4	L230A 647 P13	2.7	N282A 647 P13	2.9

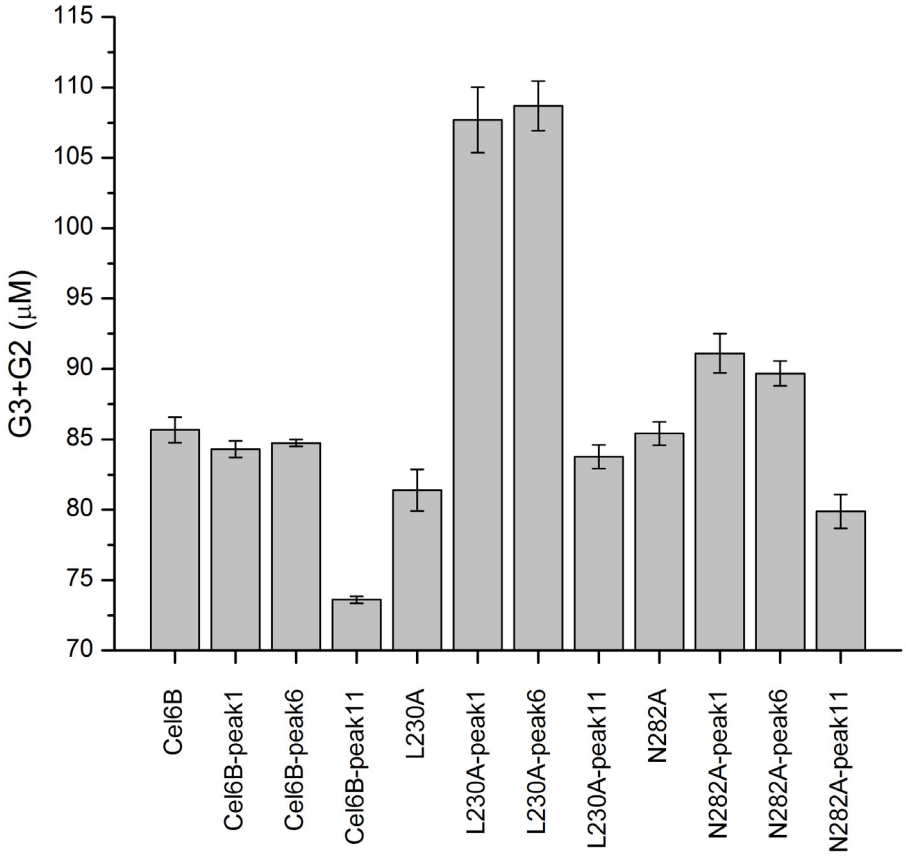


Figure 4.2: Total oligosaccharides produced by labeled Cel6B species (25nM) on PC.

In future work, the labeled enzymes with low DoL will be assayed to compare their processivity, and their displacement and movement on cellulose can be tracked by a total internal reflection fluorescence microscope (TIRFM). Additionally, the labeled L230A peak 1 will be denatured, digested with trypsin, and analyzed by mass spectrometry (MALDI-TOF/TOF) to determine which lysine residue was labeled. Labeled mutant enzymes with higher activity will be mixed with an endocellulase (*T. fusca* Cel6A or Cel5A) to test for synergism.

Section 4.3. Non-productive binding of *T. fusca* catalytic domains

Plant cell walls consist of different polysaccharides integrated with each other; therefore, it is possible that cellulases bind non-specifically to other components of plant cell walls before they can locate and bind cellulose. Most substrate binding is due to carbohydrate-binding modules, but the catalytic domains of cellulases are found to play an important part in cellulose binding as well [44]. Besides productive binding on cellulose, the catalytic domains of several *T. fusca* cellulases also bind to α-chitin without hydrolysis [26].

In this section, computational docking of oligosaccharides to the catalytic domain of *T. fusca* Cel6B and others was conducted to calculate their free energies of binding and dissociation constants, and then binding of the catalytic domains to polysaccharides was measured.

4.3.1. Computational docking

The structures of oligosaccharide ligands were obtained from different sources to prepare for docking. Laminaribiose (the PDB entry code: 2BN0), laminaritriose (2CL8) and laminarihexaose (1W9W) were found in the protein database of the Research Collaboratory for Structural Bioinformatics (RCSB, http://www.pdb.org). Laminaritetraose and laminaripentaose were created by MarvinSketch 5.1.3_2 (ChemAxon) as their x-ray structures were not available. Xylooligosaccharides including xylobiose (1B3W), xylotriose (1UX7), xylotetraose (1UYZ) and xylopentaose (1UXX) as well as cellooligosaccharides: cellobiose (3ENG), cellotriose (1UYY), cellotetraose (1F9D), cellopentaose (2EEX) and cellohexaose (2EJ1) were from the Hetero-compound Information Centre - Uppsala (http://alpha2.bmc.uu.se/hicup/).

AutodockTools 1.5.2 revision 2 (http://autodock.scripps.edu/) running on Python 2.5 was used to prepare ligands and enzymes for docking. Different pdbqt files of each oligosaccharide were set up ranging from zero rotatable bonds to several potential torsions in order to find the most suitable conformations. The x-ray structure of Cel6Acd [22] and the structural model of Cel6Bcd, which was presented in the previous chapters, were used for docking. The map files were centered at the active-site clefts with grid boxes covering the entire clefts. The map files were created by Autogrid 4. The number of genetic algorithm runs was 10, with a population size of 150. The maximum number of energy evaluations was 250,000, with the maximum number of generations of 27,000 and a rate of crossover of 0.8. A Lamarckian genetic algorithm was used for docking stimulation [97]. Docking files were generated by Autodock 4 and graphically viewed using AutodockTools.

4.3.2. Enzyme production and binding assays

The catalytic domains of two *T. fusca* endocellulases Cel5A and Cel6A as well as two exocellulases Cel6B and Cel48A were purified from the supernatants using a Phenyl-Sepharose column, followed by a Q-Sepharose column, as described in previous work [11,19,43,63]. Binding assays of these catalytic domains to pachyman (2mg/mL, Megazyme®), lichenan (1mg/mL, from *Cetraria islandica*, Sigma®) and xylan (2mg/mL, from Birchwood, Sigma®) were conducted like insoluble cellulose binding assays presented in Chapter 2.

4.3.3. Oligosaccharide and polysaccharide binding

Thirteen oligosaccharides were computationally docked to the active sites of *T. fusca* endocellulase Cel6A and exocellulase Cel6B (Table 4.2). The calculated K_d values of Cel6A with cellobiose and cellotriose, which are 167.8 and 3.4, respectively,

were much lower than the experimental ones, which are approximately 400 and 100, respectively [42]. This difference might be due to the fact that the torsion of ligands had been optimized for the best docking. Computational docking showed that both cellulases bound the most tightly to cellotetraose, suggesting that these enzymes have at least four glucose subsites. Structural analysis of *T. fusca* Cel6A identified four binding subsites, from -2 to +2 [69] while *H. insolens* Cel6A, a homolog of *T. fusca* Cel6B, showed up to 6 binding subsites, from -2 to +4 [24]. These family-6 cellulases appear to have fewer binding subsites for non-cellulosic substrates, as they tend to bind more tightly to the trisaccharides from these substrates. Generally, exocellulase Cel6B with an active site tunnel bound oligosaccharides better than endocellulase Cel6A with an open active site cleft (Figure 4.3).

Docking results suggested that the active sites can bind other oligosaccharides, besides cellooligosaccharides. Therefore, binding of catalytic domains to other polysaccharides including xylan (polymer of β-1,4-D xylose), lichenan (polymer of β-1,3:1,4-D glucose) and pachyman (polymer of β-1,3-D glucose) was measured (Table 4.3). Binding to these substrates by the catalytic domains is not similar. The endocellulases appeared to bind more to linear polysaccharides (xylan and pachyman) than did the exocellulases. Although computational docking suggested that the Cel6B active site bound more to xylooligosaccharides than Cel6Acd, experiments showed Cel6Bcd bound less xylan than the endocellulase catalytic domain (Table 4.3). It might be that small oligosaccharides could enter and bind the tunnel-shaped active site cleft of Cel6B more efficiently than their polymers. Binding of Cel6B to these polysaccharides is mostly non-productive, as the enzyme could not hydrolyze them (Figure 4.4).

Therefore, the active sites of *T. fusca* Cel5A, Cel6A, Cel6B and Cel48A can bind

to other polysaccharides, besides cellulose. Competition binding with these polysaccharides and cellulose are required to know whether they bind to the same subsites in the active site clefts.

Table 4.2: Calculated free energies of binding ΔG_b and dissociation constants K_d of Cel6A and Cel6B catalytic domains.

Docked ligand	Cel6Acd		Cel6Bcd	
	Docked energy ΔG_b (kcal/mol)	K_d (μM)	Docked energy ΔG_b (kcal/mol)	K_d (μM)
Cellobiose	-5.2	167.8	-6.4	21.2
Cellotriose	-23.8	3.4	-9.1	0.2
Cellotetraose	-13.6	0.1	-10.7	0.01
Cellopentaose	-5.7	70.6	-6.2	27.8
Cellohexaose	-4.1	989.2	-	-
Xylobiose	-5.3	136.9	-6.0	37.1
Xylotriose	-5.7	71.5	-7.8	1.8
Xylotetraose	-4.2	881.4	-9.5	0.1
Laminaribiose	-5.3	123.6	-6.6	14.0
Laminaritriose	-6.0	37.3	-8.9	0.3
Laminaritetraose	-4.8	315.6	-7.3	4.2*
Laminaripentaose	-4.0	1110	-6.0	42.6*
Laminarihexaose	-3.9	1470	-3.9	1420*

(-): docking was unsuccessful; (*): unusual docking position.

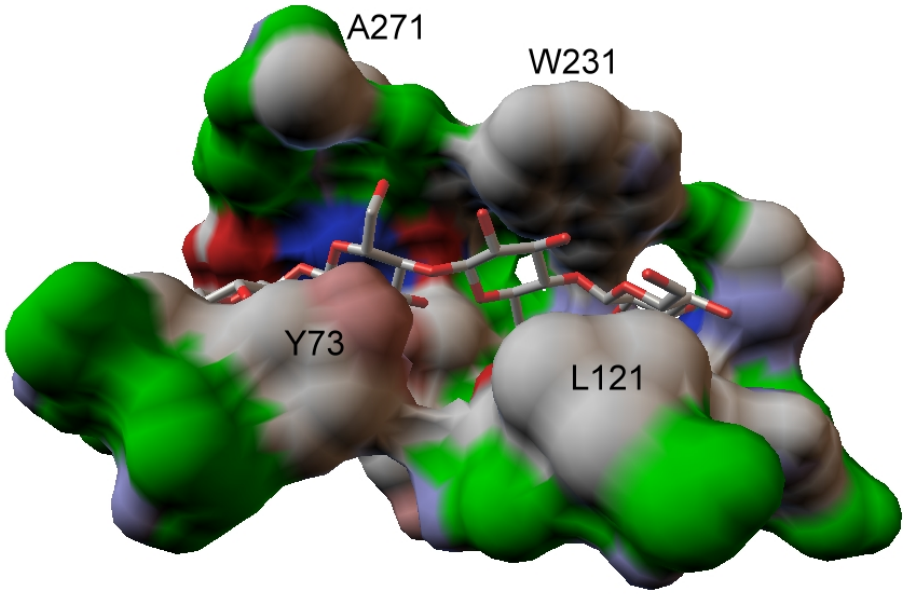


Figure 4.3: Docking of cellotetraose (K_d of 0.1μM) to the open active site of *T. fusca* Cel6Acd (green sphere- backbone, side chain colored in CPK).

Table 4.3: Bound percentage of *T. fusca* catalytic domains to different polysaccharides. Binding was conducted at 4°C for 1hr.

	Xylan (1,4)	Pachyman (1,3)	Lichenan (1,3:1,4)
Cel6Acd	52.5 ± 5.6	43.6 ± 3.0	40.6 ± 3.4
Cel5Acd	48.0 ± 5.0	43.3 ± 1.6	53.1 ± 3.6
Cel6Bcd	12.5 ± 2.8	14.3 ± 4.0	39.6 ± 5.6
Cel48Acd	24.9 ± 7.6	7.3 ± 4.8	44.8 ± 5.8

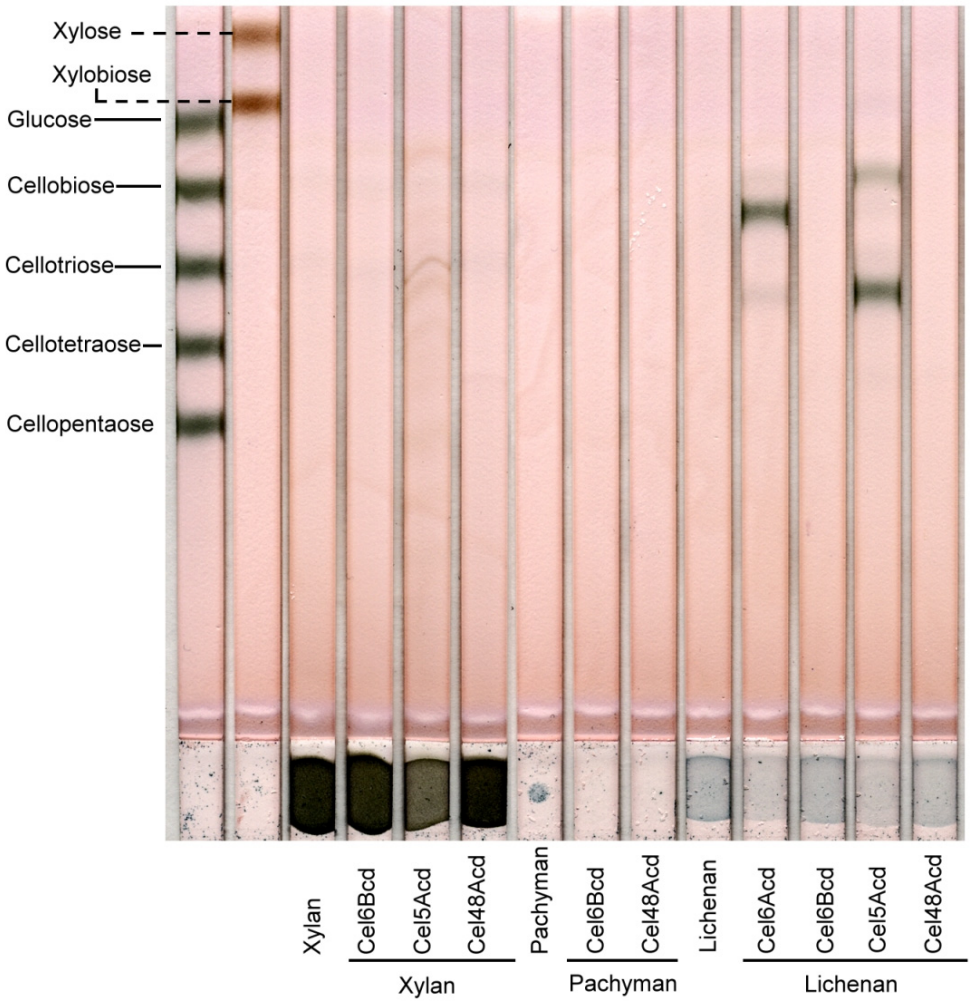


Figure 4.4: TLC analysis of xylan, pachyman and lichenan hydrolysis by the catalytic domains of *T. fusca* cellulases.

Section 4.4. Significance of individual cellulases in *T. fusca* supernatant

T. fusca secretes a number of cellulases [20,52] as well as other cellulose binding proteins including E7 and E8 [98] to hydrolyze cellulose effectively. When cultures grown on Solka Floc, the level of the two *T. fusca* exocellulases Cel6B and Cel48A was 4 to 7 times higher than that of endocellulases [99], which is also consistent with the corresponding levels of transcription [100]. These secreted components were purified and characterized; however, their roles in *T. fusca* supernatants were not well understood.

4.4.1. Immunoprecipitation assays

T. fusca was grown on Solka Floc-switchgrass for 96hrs, and then 80uL of the supernatant (1.6mg/mL) was added to 200μL of protein antisera against *T. fusca* Cel5Acd, Cel6Acd, Cel6B, Cel9Acd, Cel9B, Cel48A, E7 and E8, plus 720μL of the TBS buffer (20mM Tris pH7.5 and 0.5M NaCl) in Eppendorf® Protein LoBind tubes (Eppendorf, NY, USA). The antibody blanks and substrate blank were always included. Reactions were incubated for 1hr at 37^{o}C, and then 16hrs at 4^{o}C. Pellets were removed by centrifuged at 16,000g for 10min, and different amounts of the supernatant were assayed with BMCC (5mg/mL) in 50mM NaOAc pH5.5 at 50^{o}C for 16hrs. Reducing sugars were measured by the DNS method.

4.4.2. Specificity of the Cel6B and Cel9B antibodies

Eight *T. fusca* proteins including Cel5A, Cel6A, Cel6B, Cel9A, Cel9B, Cel48A, E7 and E8 were purified as described in previous work [11,19,43,63]. The mixture of all proteins was prepared in a molar ratio of Cel5A: Cel6A: Cel6B: Cel9A: Cel9B: Cel48A: E7: E8 of 5: 5: 20: 5: 3.5: 20: 3.5: 3.5. This mixture, along with mixtures without Cel6B or Cel9B was used for immunoprecipitation assays with the full-length Cel6B and Cel9B antibodies as presented above.

4.4.3. Effects of antibodies to supernatant activity

Although activity of Cel6B on soluble and insoluble substrates is lower than *T. fusca* endocellulases [20], removal of this exocellulase from the *T. fusca* supernatant grown on Solka Floc + switchgrass drastically decreased activity of the supernatant (Figure 4.5). This effect was also seen when the amount of Cel6B antibody was reduced in half (data not shown). Cel6B and Cel48A have low activity on all substrates and both are abundant in the supernatant [99]; however, Cel6B seems to be more important than Cel48A for cellulose hydrolysis of *T. fusca*.

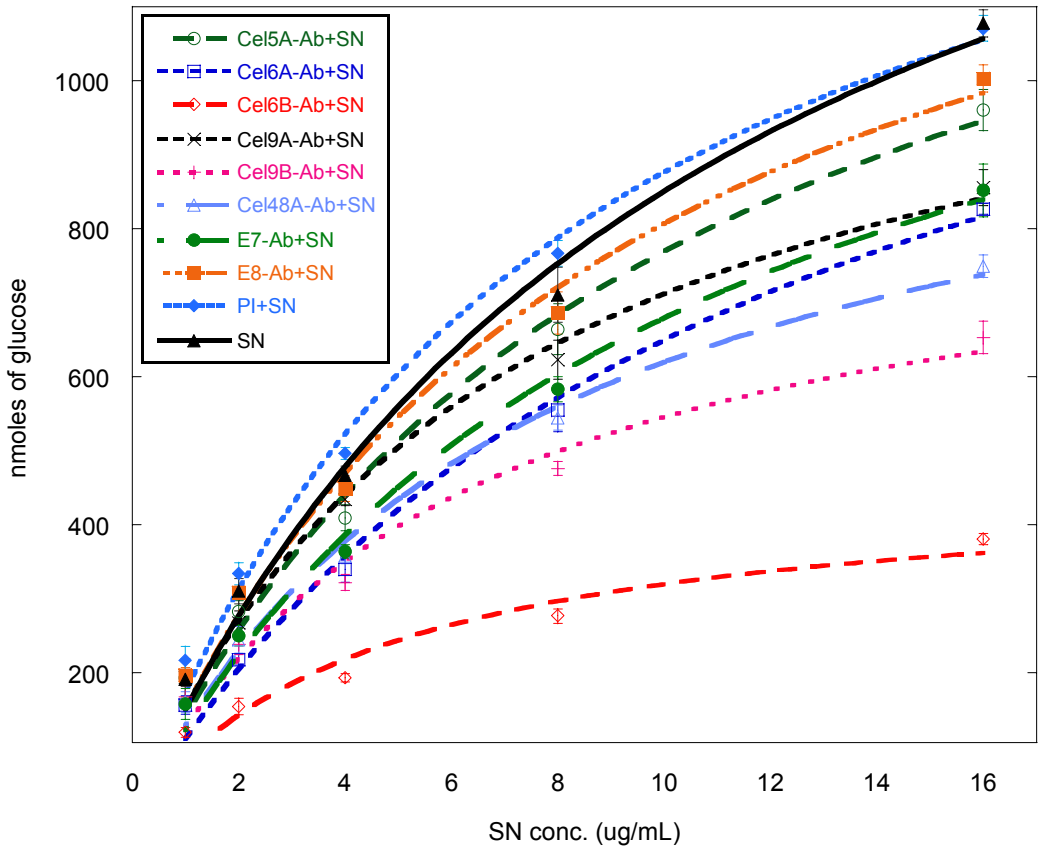


Figure 4.5: Effects of immunoprecipitation of *T. fusca* proteins in the supernatant on BMCC activity. The supernatant (SN) was incubated with the corresponding antibodies (Ab) or a pre-immune (PI) serum for 1hr at 37°C and then 16hrs at 4°C.

All *T. fusca* cellulases used in these experiments have a family-2 CBM; therefore, the specificity of the Cel6B and Cel9B antibodies was tested. The antibody of Cel6B showed a high binding specificity and it did not affect the enzymatic activity of the mixture without Cel6B (Figure 4.6). In contrast, the decrease in activity of the supernatant in the addition of Cel9B antibody unfortunately was due to the low specificity of this antibody (data not shown).

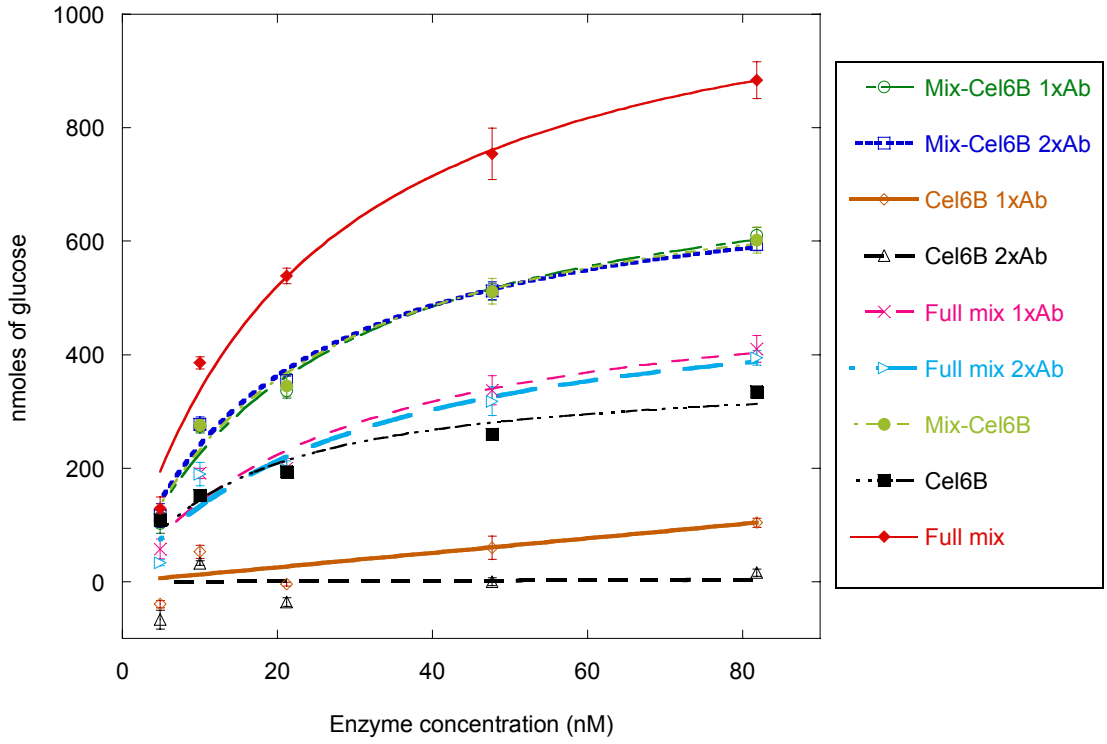


Figure 4.6: Specificity of the *T. fusca* full-length Cel6B antibody (Ab). Different amount (1xAb or 2xAb) of the antibody was incubated with a mixture of all eight *T. fusca* proteins (Full mix), a mixture of seven proteins excluding Cel6B (Mix-Cel6B), and only the Cel6B enzyme.

Future experiments will focus on investigating simultaneous removal of several components of the *T. fusca* supernatant, including removal of a pair of E7 and E8 and a pair of endocellulase Cel5A and endocellulase Cel6A.

Section 4.5. Conclusion

The Cel6B wild-type and mutant enzymes were fluorescently labeled. These labeled cellulases will be a helpful tool for tracking their binding and processivity. The catalytic domains of Cel6B bound non-productively to other polysaccharides; therefore, the balance between different types of binding should always be considered when design the enzyme for higher activity on complex substrates. Exocellulase Cel6B was found to play a very important role for *T. fusca* in hydrolyzing crystalline cellulose, although its specific activity on crystalline cellulose is low.

REFERENCES

1. Clarke AJ: *Biodegradation of cellulose, enzymology and biotechnology*. Lancaster, Pennsylvania: Technomic Publishing Company; 1997.

2. Tarchevsky IA, Marchenko GN: *Cellulose: Biosynthesis and structure*. New York: Springer-Verlag; 1991.

3. Boisset C, Fraschini C, Schulein M, Henrissat B, Chanzy H: Imaging the enzymatic digestion of bacterial cellulose ribbons reveals the endo character of the cellobiohydrolase Cel6A from *Humicola insolens* and its mode of synergy with cellobiohydrolase Cel7A. *Applied and Environmental Microbiology* 2000, 66:1444-1452.

4. Kulshreshtha AK, Dweltz NE: Paracrystalline lattice disorder in cellulose. I. Reappraisal of the application of the two-phase hypothesis to the analysis of powder X-ray diffractograms of native and hydrolyzed cellulosic materials. *Journal of Polymer Science: Polymer Physics Edition* 1973, 11:487-497.

5. Wilson DB, Irwin DC: Genetics and properties of cellulases. In *Advances in Biochemical Engineering/Biotechnology*. Edited by Scheper T: Springer-Verlag; 1999:1-21.

6. Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS: Microbial cellulose utilization: Fundamentals and biotechnology. *Microbiology and Molecular Biology Reviews* 2002, 66:506-577.

7. Bayer EA, Chanzy H, Lamed R, Shoham Y: Cellulose, cellulases and cellulosomes. *Current Opinion in Structural Biology* 1998, 8:548-557.

8. Skopec CE, Himmel ME, Matthews JF, Brady JW: Energetics for displacing a single chain from the surface of microcrystalline cellulose into the active site of *Acidothermus cellulolyticus* Cel5A. *Protein Engineering* 2003, 16:1005-1015.

9. Zhou W, Irwin DC, Escovar-Kousen J, Wilson DB: Kinetic studies of *Thermobifida fusca* Cel9A active site mutant enzymes. *Biochemistry* 2004, 43:9655-9663.

10. Sakon J, Irwin D, Wilson DB, Karplus PA: Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. *Nature Structural Biology* 1997, 4:810-818.

11. Zhang S, Lao G, Wilson DB: Characterization of a *Thermomonospora fusca* exocellulase. *Biochemistry* 1995, 34:3386-3395.

12. Sajid M, McKerrow JH: Cysteine proteases of parasitic organisms. *Molecular and Biochemical Parasitology* 2002, 120:1-21.

13. Bansal P, Hall M, Realff MJ, Lee JH, Bommarius AS: Modeling cellulase kinetics on lignocellulosic substrates. *Biotechnology Advances* 2009, 27:833-848.

14. Wilson BD: Aerobic microbial cellulase systems. In *Biomass recalcitrance: deconstructing the plant cell wall for bioenergy*. Edited by Himmel ME: Blackwell Publishers; 2008: 374-392.

15. Chandra RP, Bura R, Mabee WE, Berlin A, Pan X, Saddler JN: Substrate pretreatment: the key to effective enzymatic hydrolysis of lignocellulosics? *Advances in Biochemical Engineering/Biotechnology* 2007, 108:67-93.

16. Bommarius AS, Katona A, Cheben SE, Patel AS, Ragauskas AJ, Knudson K, Pu Y: Cellulase kinetics as a function of cellulose pretreatment. *Metabolic Engineering* 2008, 10:370-381.

17. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B: The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Research* 2009, 37:D233-238.

18. Tomme P, Warren RA, Gilkes NR: Cellulose hydrolysis by bacteria and fungi. *Advances in Microbial Physiology* 1995, 37:1-81.

19. Irwin DC, Zhang S, Wilson DB: Cloning, expression and characterization of a family 48 exocellulase, Cel48A, from *Thermobifida fusca*. *European Journal of Biochemistry* 2000, 267:4988-4997.

20. Wilson BD: Studies of *Thermobifida fusca* plant cell wall degrading enzymes. *The Chemical Record* 2004, 4:72-82.

21. Rouvinen J, Bergfors T, Teeri T, Knowles JK, Jones TA: Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science* 1990, 249:380-386.

22. Spezio M, Wilson DB, Karplus PA: Crystal structure of the catalytic domain of a thermophilic endocellulase. *Biochemistry* 1993, 32:9906-9916.

23. Varrot A, Hastrup S, Schulein M, Davies GJ: Crystal structure of the catalytic core

domain of the family 6 cellobiohydrolase II, Cel6A, from *Humicola insolens*, at 1.92Å resolution. *Biochemical Journal* 1999, 337 (Pt 2):297-204.

24. Davies GJ, Brzozowski AM, Dauter M, Varrot A, Schulein M: Structure and function of *Humicola insolens* family 6 cellulases: Structure of the endoglucanase, Cel6B, at 1.6Å resolution. *Biochemical Journal* 2000, 348:201-207.

25. Varrot A, Leydier S, Pell G, Macdonald JM, Stick RV, Henrissat B, Gilbert HJ, Davies GJ: *Mycobacterium tuberculosis* strains possess functional cellulases. *Journal of Biological Chemistry* 2005, 280:20181-20184.

26. Li Y, Wilson DB: Chitin binding by *Thermobifida fusca* cellulase catalytic domains. *Biotechnology and Bioengineering* 2008, 100:644-652.

27. Gebler J, Gilkes NR, Claeyssens M, Wilson DB, Beguin P, Wakarchuk WW, Kilburn DG, Miller RC, Jr., Warren RA, Withers SG: Stereoselective hydrolysis catalyzed by related β-1,4-glucanases and β-1,4-xylanases. *Journal of Biological Chemistry* 1992, 267:12559-12561.

28. Honda Y, Kitaoka M, Hayashi K: Kinetic evidence related to substrate-assisted catalysis of family 18 chitinases. *FEBS Letters* 2004, 567:307-310.

29. Langley DB, Harty DWS, Jacques NA, Hunter N, Guss JM, Collyer CA: Structure of N-acetyl-β-D-glucosaminidase (GcnA) from the endocarditis pathogen *Streptococcus gordonii* and its complex with the mechanism-based inhibitor NAG-thiazoline. *Journal of Molecular Biology* 2008, 377:104-116.

30. Markovic-Housley Z, Miglierini G, Soldatova L, Rizkallah PJ, Müller U, Schirmer T: Crystal structure of hyaluronidase, a major allergen of bee venom. *Structure* 2000, 8:1025-1035.

31. Dennis RJ, Taylor EJ, Macauley MS, Stubbs KA, Turkenburg JP, Hart SJ, Black GN, Vocadlo DJ, Davies GJ: Structure and mechanism of a bacterial β-glucosaminidase having *O*-GlcNAcase activity. *Nature Structural & Molecular Biology* 2006, 13:365-371.

32. Ling Z, Suits MDL, Bingham RJ, Bruce NC, Davies GJ, Fairbanks AJ, Moir JWB, Taylor EJ: The X-ray crystal structure of an *Arthrobacter protophormiae* endo-β-N-acetylglucosaminidase reveals a $(\beta/\alpha)_8$ catalytic domain, two ancillary domains and active site residues key for transglycosylation activity. *Journal of*

Molecular Biology 2009, 389:1-9.

33. Rich JR, Withers SG: Emerging methods for the production of homogeneous human glycoproteins. *Nature Chemical Biology* 2009, 5:206-215.

34. Reid CW, Legaree BA, Clarke AJ: Role of Ser216 in the mechanism of action of membrane-bound lytic transglycosylase B: Further evidence for substrate-assisted catalysis. *FEBS Letters* 2007, 581:4988-4992.

35. Honda Y, Kitaoka M: The first glycosynthase derived from an inverting glycoside hydrolase. *Journal of Biological Chemistry* 2006, 281:1426-1431.

36. Wada J, Honda Y, Nagae M, Kato R, Wakatsuki S, Katayama T, Taniguchi H, Kumagai H, Kitaoka M, Yamamoto K: 1,2-α-L-fucosynthase: A glycosynthase derived from an inverting α-glycosidase with an unusual reaction mechanism. *FEBS Letters* 2008, 582:3739-3743.

37. Mackenzie LF, Wang Q, Warren RAJ, Withers SG: Glycosynthases: Mutant glycosidases for oligosaccharide synthesis. *Journal of the American Chemical Society* 1998, 120:5583-5584.

38. Rakić B, Withers SG: Recent developments in glycoside synthesis with glycosynthases and thioglycoligases. *Australian Journal of Chemistry* 2009, 62:510-520.

39. Koivula A: Structure-function studies of two polysaccharide-degrading enzymes: *Bacillus stearothermophilus* α-amylase and *Trichoderma reesei* cellobiohydrolase II. Helsinki: University of Helsinki: 1996.

40. Ghose TK: Measurement of cellulase activities. *Pure and Applied Chemistry* 1987, 59:257-268.

41. Lever M: A new reaction for colorimetric determination of carbohydrates. *Analytical Biochemistry* 1972, 47:273-279.

42. Barr BK, Wolfgang DE, Piens K, Claeyssens M, Wilson DB: Active-site binding of glycosides by *Thermomonospora fusca* endocellulase E2. *Biochemistry* 1998, 37:9220-9229.

43. Irwin DC, Spezio M, Walker LP, Wilson DB: Activity studies of eight purified cellulases: Specificity, synergism, and binding domain effects. *Biotechnology and Bioengineering* 1993, 42:1002-1013.

44. Li Y, Irwin DC, Wilson DB: Processivity, substrate binding, and mechanism of cellulose hydrolysis by *Thermobifida fusca* Cel9A. *Applied and Environmental Microbiology* 2007, 73:3165-3172.

45. Bayer EA, Lamed R, White BA, Flint HJ: From cellulosomes to cellulosomics. *The Chemical Record* 2008, 8:364-377.

46. Arai T, Matsuoka S, Cho H-Y, Yukawa H, Inui M, Wong S-L, Doi RH: Synthesis of *Clostridium cellulovorans* minicellulosomes by intercellular complementation. *Proceedings of the National Academy of Sciences* 2007, 104:1456-1460.

47. Xie G, Bruce DC, Challacombe JF, Chertkov O, Detter JC, Gilna P, Han CS, Lucas S, Misra M, Myers GL, et al.: Genome sequence of the cellulolytic gliding bacterium *Cytophaga hutchinsonii*. *Applied and Environmental Microbiology* 2007, 73:3536-3546.

48. Qi M, Jun H-S, Forsberg CW: Characterization and synergistic interactions of *Fibrobacter succinogenes* glycoside hydrolases. *Applied and Environmental Microbiology* 2007, 73:6098-6105.

49. Wilson D: Evidence for a novel mechanism of microbial cellulose degradation. *Cellulose* 2009, 16:723-727.

50. Zhang Z, Wang Y, Ruan J: Reclassification of *Thermomonospora* and *Microtetraspora*. *International Journal of Systematic Bacteriology* 1998, 48:411-422.

51. Wilson DB: Biochemistry and genetics of actinomycete cellulases. *Critical Reviews in Biotechnology* 1992, 12:45-63.

52. Posta K, Béki E, Wilson DB, Kukolya J, Hornok L: Cloning, characterization and phylogenetic relationships of Cel5B, a new endoglucanase encoding gene from *Thermobifida fusca*. *Journal of Basic Microbiology* 2004, 44:383-399.

53. Arnold K, Bordoli L, Kopp J, Schwede T: The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 2006, 22:195-201.

54. Damude HG, Withers SG, Kilburn DG, Miller RC, Jr., Warren RA: Site-directed mutation of the putative catalytic residues of endoglucanase CenA from *Cellulomonas fimi*. *Biochemistry* 1995, 34:2220-2224.

55. Koivula A, Ruohonen L, Wohlfahrt G, Reinikainen T, Teeri TT, Piens K, Claeyssens M, Weber M, Vasella A, Becker D, et al.: The active site of cellobiohydrolase Cel6A from *Trichoderma reesei*: the roles of aspartic acids D221 and D175. *Journal of the American Chemical Society* 2002, 124:10015-10024.

56. Varrot A, Frandsen TP, von Ossowski I, Boyer V, Cottaz S, Driguez H, Schulein M, Davies GJ: Structural basis for ligand binding and processivity in cellobiohydrolase Cel6A from *Humicola insolens*. *Structure* 2003, 11:855-864.

57. Wolfgang DE, Wilson DB: Mechanistic studies of active site mutants of *Thermomonospora fusca* endocellulase E2. *Biochemistry* 1999, 38:9746-9751.

58. Zechel DL, Reid SP, Stoll D, Nashiru O, Warren RA, Withers SG: Mechanism, mutagenesis, and chemical rescue of a β-mannosidase from *Cellulomonas fimi*. *Biochemistry* 2003, 42:7195-7204.

59. Shallom D, Leon M, Bravman T, Ben-David A, Zaide G, Belakhov V, Shoham G, Schomburg D, Baasov T, Shoham Y: Biochemical characterization and identification of the catalytic residues of a family 43 β-D-xylosidase from *Geobacillus stearothermophilus* T-6. *Biochemistry* 2005, 44:387-397.

60. Zhang S, Irwin DC, Wilson DB: Site-directed mutation of noncatalytic residues of *Thermobifida fusca* exocellulase Cel6B. *European Journal of Biochemistry* 2000, 267:3101-3115.

61. Zhang YHP, Cui J, Lynd LR, Kuang LR: A transition from cellulose swelling to cellulose dissolution by *o*-phosphoric acid: Evidence from enzymatic hydrolysis and supramolecular structure. *Biomacromolecules* 2006, 7:644-648.

62. Kempton JB, Withers SG: Mechanism of *Agrobacterium* β-glucosidase: kinetic studies. *Biochemistry* 1992, 31:9961-9969.

63. Jung ED, Lao G, Irwin D, Barr BK, Benjamin A, Wilson DB: DNA sequences and expression in *Streptomyces lividans* of an exoglucanase gene and an endoglucanase gene from *Thermomonospora fusca*. *Applied and Environmental Microbiology* 1993, 59:3032-3043.

64. Hooft RWW, Vriend G, Sander C, Abola EE: Errors in protein structures. *Nature* 1996, 381:272.

65. Varrot A, Schulein M, Davies GJ: Structural changes of the active site tunnel of

Humicola insolens cellobiohydrolase, Cel6A, upon oligosaccharide binding. *Biochemistry* 1999, 38:8884-8891.

66. Wohlfahrt G, Pellikka T, Boer H, Teeri TT, Koivula A: Probing pH-dependent functional elements in proteins: modification of carboxylic acid pairs in *Trichoderma reesei* cellobiohydrolase Cel6A. *Biochemistry* 2003, 42:10095-10103.

67. Varrot A, Frandsen TP, Driguez H, Davies GJ: Structure of the *Humicola insolens* cellobiohydrolase Cel6A D416A mutant in complex with a non-hydrolysable substrate analogue, methyl cellobiosyl-4-thio-β-cellobioside, at 1.9Å. *Acta Crystallographica Section D* 2002, 58:2201-2204.

68. Andre G, Kanchanawong P, Palma R, Cho H, Deng X, Irwin D, Himmel ME, Wilson DB, Brady JW: Computational and experimental studies of the catalytic mechanism of *Thermobifida fusca* cellulase Cel6A (E2). *Protein Engineering Design and Selection* 2003, 16:125-134.

69. Larsson AM, Bergfors T, Dultz E, Irwin DC, Roos A, Driguez H, Wilson DB, Jones TA: Crystal structure of *Thermobifida fusca* endoglucanase Cel6A in complex with substrate and inhibitor: The role of tyrosine Y73 in substrate ring distortion. *Biochemistry* 2005, 44:12915-12922.

70. Vocadlo DJ, Davies GJ: Mechanistic insights into glycosidase chemistry. *Current Opinion in Chemical Biology* 2008, 12:539-555.

71. Shimada M, Takahashi M: Biodegradation of cellulosic materials. In *Wood and cellulosic chemistry*. Edited by Hon DNS, Shiraishi N: Marcel Dekker; 1991:621-663.

72. Koivula A, Reinikainen T, Ruohonen L, Valkeajarvi A, Claeyssens M, Teleman O, Kleywegt GJ, Szardenings M, Rouvinen J, Jones TA, et al.: The active site of *Trichoderma reesei* cellobiohydrolase II: the role of tyrosine 169. *Protein Engineering Design and Selection* 1996, 9:691-699.

73. Zou J-y, Kleywegt GJ, Stahlberg J, Driguez H, Nerinckx W, Claeyssens M, Koivula A, Teeri TT, Jones TA: Crystallographic evidence for substrate ring distortion and protein conformational changes during catalysis in cellobiohydrolase Cel6A from *Trichoderma reesei*. *Structure* 1999, 7:1035-1045.

74. Zhang S, Barr BK, Wilson DB: Effects of noncatalytic residue mutations on substrate specificity and ligand binding of *Thermobifida fusca* endocellulase Cel6A. *European Journal of Biochemistry* 2000, 267:244-252.

75. Nagae M, Tsuchiya A, Katayama T, Yamamoto K, Wakatsuki S, Kato R: Structural basis of the catalytic reaction mechanism of novel 1,2-α-L-fucosidase from *Bifidobacterium bifidum*. *Journal of Biological Chemistry* 2007, 282:18497-18509.

76. Ishida T, Fushinobu S, Kawai R, Kitaoka M, Igarashi K, Samejima M: Crystal structure of glycoside hydrolase family 55 β-1,3-glucanase from the Basidiomycete *Phanerochaete chrysosporium*. *Journal of Biological Chemistry* 2009, 284:10100-10109.

77. McGrath CE, Vuong TV, Wilson DB: Site-directed mutagenesis to probe catalysis by a *Thermobifida fusca* β-1,3-glucanase (Lam81A). *Protein Engineering, Design and Selection* 2009, 22:375-382.

78. Abbott DW, Macauley MS, Vocadlo DJ, Boraston AB: *Streptococcus pneumoniae* endohexosaminidase D: Structural and mechanistic insight into substrate-assisted catalysis in family 85 glycoside hydrolases. *Journal of Biological Chemistry* 2009.

79. Sørbotten A, Horn SJ, Eijsink VGH, Varum KM: Degradation of chitosans with chitinase B from *Serratia marcescens*. *FEBS Journal* 2005, 272:538-549.

80. Gilligan W, Reese ET: Evidence for multiple components in microbial cellulases. *Canadian Journal of Microbiology* 1954, 1:90-107.

81. Bhat MK, Bhat S: Cellulose degrading enzymes and their potential industrial applications. *Biotechnology Advances* 1997, 15:583-620.

82. Medve J, Karlsson J, Lee D, Tjerneld F: Hydrolysis of microcrystalline cellulose by cellobiohydrolase I and endoglucanase II from *Trichoderma reesei*: Adsorption, sugar production pattern, and synergism of the enzymes. *Biotechnology and Bioengineering* 1998, 59:621-634.

83. Henrissat B, Driguez H, Viet C, Schulein M: Synergism of cellulases from *Trichoderma reesei* in the degradation of cellulose. *Nature Biotechnology* 1985, 3:722-726.

84. Woodward J, Lima M, Lee NE: The role of cellulase concentration in determining

the degree of synergism in the hydrolysis of microcrystalline cellulose. *Biochemical Journal* 1988, 255:895-899.

85. Gow LA, Wood TM: Breakdown of crystalline cellulose by synergistic action between cellulase components from *Clostridium thermocellum* and *Trichoderma koningii*. *FEMS Microbiology Letters* 1988, 50:247-252.

86. Escovar-Kousen JM, Wilson D, Irwin D: Integration of computer modeling and initial studies of site-directed mutagenesis to improve cellulase activity on Cel9A from *Thermobifida fusca*. *Applied Biochemistry and Biotechnology* 2004, 113-116:287-297.

87. Qin Y, Wei X, Song X, Qu Y: Engineering endoglucanase II from *Trichoderma reesei* to improve the catalytic efficiency at a higher pH optimum. *Journal of Biotechnology* 2008, 135:190-195.

88. Zhang YH, Lynd LR: Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems. *Biotechnology and Bioengineering* 2004, 88:797-824.

89. Bohm G, Muhr R, Jaenicke R: Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Engineering Design and Selection* 1992, 5:191-195.

90. Walker LP, Belair CD, Wilson DB, Irwin DC: Engineering cellulase mixtures by varying the mole fraction of *Thermomonospora fusca* E5 and E3, *Trichoderma reesei* CBHI, and *Caldocellum saccharolyticum* glucosidase. *Biotechnology and Bioengineering* 1993, 42:1019-1028.

91. Pages S, Kester HC, Visser J, Benen JA: Changing a single amino acid residue switches processive and non-processive behavior of *Aspergillus niger* endopolygalacturonase I and II. *Journal of Biological Chemistry* 2001, 276:33652-33656.

92. von Ossowski I, Ståhlberg J, Koivula A, Piens K, Becker D, Boer H, Harle R, Harris M, Divne C, Mahdi S, et al.: Engineering the exo-loop of *Trichoderma reesei* cellobiohydrolase, Cel7A. A comparison with *Phanerochaete chrysosporium* Cel7D. *Journal of Molecular Biology* 2003, 333:817-829.

93. Harjunpaa V, Teleman A, Koivula A, Ruohonen L, Teeri TT, Teleman O, Drakenberg T: Cello-oligosaccharide hydrolysis by cellobiohydrolase II from

Trichoderma reesei. Association and rate constants derived from an analysis of progress curves. *European Journal of Biochemistry* 1996, 240:584-591.

94. Jeoh T, Wilson DB, Walker LP: Cooperative and competitive binding in synergistic mixtures of *Thermobifida fusca* cellulases Cel5A, Cel6B, and Cel9A. *Biotechnology Progress* 2002, 18:760-769.

95. Liu Y-S, Zeng Y, Luo Y, Xu Q, Himmel M, Smith S, Ding S-Y: Does the cellulose-binding module move on the cellulose surface? *Cellulose* 2009, 16:587-597.

96. Moran-Mirabal JM, Corgie SC, Bolewski JC, Smith HM, Cipriany BR, Craighead HG, Walker LP: Labeling and purification of cellulose-binding proteins for high resolution fluorescence applications. *Analytical Chemistry* 2009, 81:7981-7987.

97. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ: Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry* 1998, 19:1639-1662.

98. Moser F, Irwin D, Chen S, Wilson BD: Regulation and characterization of *Thermobifida fusca* carbohydrate-binding module proteins E7 and E8. *Biotechnology and Bioengineering* 2008, 100:1066-1077.

99. Spiridonov NA, Wilson DB: Regulation of biosynthesis of individual cellulases in *Thermomonospora fusca*. *Journal of Bacteriology* 1998, 180:3529-3532.

100. Chen S, Wilson DB: Proteomic and transcriptomic analysis of extracellular proteins and mRNA levels in *Thermobifida fusca* grown on cellobiose and glucose. *Journal of Bacteriology* 2007, 189:6260–6265.