Engineering cellulases for enhanced degradation of insoluble substrates

By

Vimalier Reyes-Ortiz

A dissertation submitted in partial satisfaction of the requirements for the degree of

Joint Doctor of Philosophy

with University of California, San Francisco

in

Bioengineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in Charge:

Professor Danielle Tullman-Ercek, Chair
Professor Patricia Babbitt
Professor Clayton J. Radke

Fall 2013
Abstract

**Engineering cellulases for enhanced degradation of insoluble substrates**

Vimalier Reyes-Ortiz

Joint Doctor of Philosophy in Bioengineering with UCSF

University of California, Berkeley

Professor Danielle Tullman-Ercek, Chair

There is tremendous social and political interest in the production of sustainable and carbon neutral liquid fuels that can meet our transportation energy demands. Second-generation biofuels, derived from cellulosic non-edible plant matter (biomass), represent a possible solution to this issue. The US Department of Energy and the Department of Agriculture estimate that the US can produce the energy equivalent 42 percent of the total US annual transportation consumption from cellulosic biomass, without having a grossly negative impact on food supply.

Biomass can be deconstructed to glucose, a type of sugar that can be fermented to make the biofuels. Glucose monomers are located within a strong crystalline polymer called cellulose, which is surrounded by other complex polymeric structures called hemicellulose and lignin. In order to obtain this sugar the biomass therefore has to pass through mechanical, physicochemical, and enzymatic treatments. The mechanical and physicochemical stages of the process increase the accessibility of cellulose to cellulolytic enzymes responsible of converting the cellulose to glucose. Optimization of these enzymes can have a significant impact on the economics and feasibility of biofuels. A detailed understanding of how carbohydrate-binding domains (CBMs) increase the activity of cellulases remains elusive. We have completed a series of studies that shed light on how certain CBMs affect the activity of cellulases from three different perspectives:

1) Specific enzyme-substrate interactions: Neutron reflectivity is used to analyze the effect of endocellulases on cellulose film density and thickness. Two endocellulases, each tethered or untethered to a single CBM, are compared in this study.

2) General synergy between CBMs from three different families and a specific cellulase: We set out to determine if there is any correlation between CBM family and activity level or substrate preference, as measured by enhanced activity of a single cellulase. From a library of chimeric enzymes tested for activity on crystalline and non-crystalline substrates, we chose to examine in more detail three enzymes that exhibited approximately equal activity enhancements on amorphous cellulose. These chimeras consisted of a model endocellulase tethered to a CBM belonging to family one, two or three. Neutron reflectivity was used to study the interaction of the three chimeric endocellulases with amorphous cellulose model films.

3) The effect of enzymatic ratios in simple cellulosytic cocktails and the impact of CBM-containing enzymes on those ratios: We focused on the optimization of simple cellulosytic cocktails for real-world operational conditions. With basic conditions established, optimization is achieved by changing the ratios of the different components within the cocktail. The resulting optimal ratios and efficiency of glucose release was examined for cocktails comprising either commercial enzymes or enzymes identified to be active under extreme, but industrially relevant conditions.
**Table of Contents**

Table of Contents .................................................................................. i
Table of Figures .................................................................................. iv
Table of Tables .................................................................................. vi
Acknowledgements ............................................................................ vii

Chapter 1 – Introduction ..................................................................... 1
  1.1 – A need for alternative fuels: Biofuels Abstract .................................. 1
    1.1.1 – Energy needs ........................................................................... 1
    1.1.2 – Biofuels and biomass recalcitrance ........................................... 1
  1.2 – Biomass pretreatments .................................................................. 3
    1.2.1 – Chemical pretreatment methods ............................................. 3
      1.2.1.1 – Acid and alkaline pretreatment .......................................... 3
      1.2.1.2 – Organic bases pretreatments .......................................... 4
      1.2.1.3 – Ionic liquids pretreatments ........................................... 4
    1.2.2 – Physicochemical pretreatments methods .................................. 4
      1.2.2.1 – Steam based pretreatments ........................................... 4
    1.2.3 – Current status of biomass pretreatments ................................... 5
  1.3 – Enzymes as part of degradation process ....................................... 5
    1.3.1 – Introduction to carbohydrate-active enzymes ........................... 5
    1.3.2 – Enzymes for biomass degradation (lignocellulolytic cocktails) .... 6
    1.3.3 – An introduction to cellulases for cellulose degradation .......... 7
  1.4 – Need for optimization of cellulolytic enzymes ............................... 8
    1.4.1 – Examples of enzyme optimization ........................................ 8
      1.4.1.1 – Brief on enzyme optimization ....................................... 8
        1.4.1.1.1 – Introduction to carbohydrate binding modules .......... 9
    1.6 – Introduction to CBM studies in this dissertation .......................... 11

Chapter 2 – Addition of a carbohydrate-binding module enhances cellulase penetration into cellulose substrates ................................................. 12
  2.1 – Abstract .................................................................................... 12
  2.2 – Background .............................................................................. 12
  2.3 – Results and discussion ................................................................ 13
    2.3.1 – Construction and characterization of chimeric cellulases .......... 13
    2.3.2 – Enzymatic hydrolysis of insoluble substrates .......................... 15
    2.3.3 – Amorphous Cellulose Film Characterization ........................... 16
    2.3.4 – NR analysis of films incubated with cellulases ....................... 17
    2.3.5 – Proposed mechanism of CBM enhancement ........................... 26
  2.4 – Conclusions .............................................................................. 27
  2.5 – Methods ................................................................................... 28
    2.5.1 – Strains and growth conditions ............................................ 28
    2.5.2 – Cellulase genes and vector construction ............................... 28
    2.5.3 – Enzyme expression and purification ...................................... 29
    2.5.4 – Ionic liquid pretreatment .................................................... 30
    2.5.5 – Enzymatic hydrolysis assays .............................................. 30
4.5.2 – Enzyme expression and purification..........................................................55
4.5.3 – [C$_2$ mim][OAc] pretreated switchgrass...............................................55
4.5.4 – Dispensing of biomass...........................................................................55
4.5.5 – Saccharification assay: Cel5A vs Cel5A-CBM2a cocktails......................56
4.5.6 – Saccharification assay: four-enzyme cocktails.......................................57
4.5.7 – Saccharification: Cocktail optimization experiments..........................57
4.5.8 – Saccharification measurements..............................................................57
    4.5.8.1 – Quantification of soluble sugars by DNS assay..............................57
    4.5.8.2 Glucose detection by YSI.................................................................57
    4.5.8.3 Glucose detection via HPLC.............................................................58
4.5.9 – Design of experiments for mixtures (containing non-chimeric
    endocellulases)............................................................................................58
4.5.10 – Analysis of experiments for mixtures....................................................59
4.6 – Acknowledgements....................................................................................59

**Chapter 5 – References** ..................................................................................60

**Chapter 6 – Appendix** ...................................................................................68

6.1 Abbreviations................................................................................................68
# Table of Figures

Figure 2.1. Temperatures and pH profiles for Cel9A, Cel5A and chimeras..........................14
Figure 2.2. SDS-Gels and chimeric cellulase stability.........................................................14
Figure 2.3. Saccharification assays of cellulases on MCC and IL-MCC...............................15
Figure 2.4. Saccharification of chimeric cellulases: comparison of linker length..............16
Figure 2.5. Saccharification of chimeric cellulases: different enzymatic loadings.........17
Figure 2.6. Cellulose fraction profiles of cellulases at 20°C.............................................18
Figure 2.7. Reflectivity profiles cellulose of cellulases at 20°C..........................................19
Figure 2.8. Scattered light density profiles of Cel9A-CBM and Cel5A-CBM at 20°C........20
Figure 2.9. Cellulose fraction profiles of knock-out cellulases at 20°C.............................20
Figure 2.10. Reflectivity and scattered light density profiles of Cel9A_{KO}-CBM and
             Cel5A_{KO}-CBM at 20°C.......................................................................................22
Figure 2.11. Reflectivity and cellulose fraction profiles of CBM at 20°C and 65°C.........23
Figure 2.12. Reflectivity and cellulose fraction profiles of untethered Cel9A and
             CBM at 20°C...........................................................................................................24
Figure 2.13. Cellulose fraction profiles of cellulases at optimal temperatures...............24
Figure 2.14. Reflectivity profiles of all cellulases at their optimal temperatures...........25
Figure 2.15. Proposed model for chimeric interaction with cellulose model films.......27
Figure 2.16 Uncertainty limits of cellulose fraction profiles for Cel5A and
             Cel5A-CBM............................................................................................................32
Figure 3.1. Neutron reflectivity data and normalized cellulose volume fraction
 profiles of Cel5A with the cellulose film at 75°C..........................................................27
Figure 3.2. Neutron reflectivity data and normalized cellulose volume fraction
 profiles of Cel5A-CBM3 with the cellulose film at 75°C...............................................38
Figure 3.3. Neutron reflectivity data and normalized cellulose volume fraction
 profiles of Cel5A-CBM1 with the cellulose film at 75°C.............................................39
Figure 3.4. Neutron reflectivity data and normalized cellulose volume fraction profiles of Cel5A-CBM2 with the cellulose film at 75°C. ........................................40

Figure 4.1. Glucose release by enzymatic cocktails containing Cel5A or Cel5a-CBM2a tested on Avicel, IL-Avicel and ILSG. ..................................................49

Figure 4.2. Percentage of glucose release by four-enzyme cocktails containing a BG, CBH and Cel5A/Cel9A or Cel5a-/Ce9A-CBM2a. These cocktails were tested against ILSG.................................................................51

Figure 4.3. Percentage of cellulose degradation by enzymatic cocktails comprising three commercial cellulases at two different temperatures and two different pHs........................................................................52

Figure 4.4. Percentage of cellulose degradation by enzymatic cocktails comprising the in-house CBH, BG and Cel5A or Cel9A........................................53

Figure 4.5. Heat maps predicting optimal activity based on enzyme concentration for the treatment of ILSG in the presence of 20% IL. These enzymatic cocktails consist of our in-house a CBH, a BG and Cel5A or Cel9A..........................................................54

Figure 4.6. Sample points for augmented simplex lattice design. .................................58
Table of Tables

Table 2.1. Comparison of changes in cellulose model films between cellulases.............18

Table 3.1. Comparison of activity of Cel5A and Cel5A chimeric cellulases on Avicel and ionic pretreated switchgrass (ILSG) ..................................................35

Table 3.2. Simple description of the three CBMs used in chapter.................................43

Table 4.1. Output of the DOE software for cocktails; values indicate the relative concentration of the enzyme in each mixture (Endo:CBH:BG)...................... 56
Acknowledgements

I am deeply grateful to the dedicated individuals who have somehow contributed to my development as a scientist. I want to thank my advisor, Danielle Tullman-Ercek, for accepting me into her laboratory when I was looking for a new “home lab”. Her patience and direction have been an invaluable and formative part of my experience. To the Tullman-Ercek lab, my academic siblings, thank you for being there and reaching out to me, even when my time spent in Lewis was never long. I also want to make special mention of Michael S. Kent, who was key on this achievement and whose patience and advice I greatly appreciate.

Thanks to the Graduate Program in BioEngineering at UC Berkeley and UC San Francisco, and to the Joint BioEnergy Institute for providing me with resources that helped me navigate toward this goal. To Terry Johnson, Steven Conolly, Dominique Loque, John Christopher Anderson, Michelle C. Y. Chang, Patricia Babbitt, Henrik V. Scheller, John Dueber, Clayton J. Radke and several members of their labs, who extended a helpful and friendly hand, thank you. I cannot forget to mention Rebecca Pauling, Beatriz Lopez-Flores and Carlos Ríos Velázquez, who also helped me stay confident during these years.

I most thank my wife and pillar, Diana Herrera, for her love, patience and support during these years. She has been subjected to more than anyone around me, and she has done nothing but being supportive and patient throughout. I need to thank those friends and family members who were there in one way or the other. I have been very fortunate to be surrounded by encouraging people like you.

Finally, I would also like to acknowledge the Biotechnology for Biofuels journal for granting me permission to use published material as part of this thesis. This work conducted by the Joint BioEnergy Institute was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under contract No. DE-AC02-05CH11231.
Chapter 1. Introduction

1.1 A need for alternative fuels: biofuels

1.1.1 Energy needs

The ability to control energy sources for the powering of heavy machinery was a key driving force of the industrial revolution. The impact of changing from the combination of human and horse power to horsepower enabled humanity to travel distances and perform unprecedented work. Research and applied technology have given rise to more efficient, more powerful and more affordable machines. However, the energy that powers many of these machines has not changed in more than a century [1]; most of today’s machinery are linked directly or indirectly to the use of petroleum, a very versatile fuel that has led to us to an oil-dependent economy [2] in the United States and around the world.

In 2009, The International Energy Agency (IEA) calculated the world’s energy demand to be close to 12 billion metric ton oil equivalents (t.o.e.) and estimated it to be up to 18 billion t.o.e by 2035 under the current policy scenarios [3]. An increase in oil demand will also mean an increase in emissions of carbon dioxide from 29 to 43 billion metric tons per year [2]. From all the carbon dioxide emitted around the globe, 27 percent comes from the use of fuels for transportation in the US. [4], making our nation also responsible for the impact on climate change.

The US is highly dependent on petroleum. For instance, petroleum-derived liquid fuels are the main source of energy for the US transportation system. Moreover, most of the world’s oil reserves are located outside US territory. This obligates the US to depend on oil import (about 60 percent) for the operation of its economy and gives rise to balance-of-trade and national-security challenges. Two main disadvantages of petroleum-derived fuels underline the interest in moving toward alternative sources of energy: 1) the security of supply and 2) the climate change caused by the emissions from oil use. Biofuels – fuels derived from biomass - are a promising alternative energy source and their application as liquid fuels would address both of these issues [3–5].

Biofuels could be converted to a host of other petroleum-derived products (e.g. plastics and industrial chemicals) in addition to liquid fuels. Some low-cost bio-products that can replace their petroleum counterparts have already been commercialized. Examples of these include bio-ethylene derived from sugar cane (produced by Chemtex) and bio-1,3-propanediol (produced by DuPont). Other companies are starting projects; Solvay Indupa and Archer Daniel Midland (ADM) have commissioned polyvinyl chloride (PVC) and propylene glycol plants in Brazil, respectively [6]. As with any early technology, continuing research is needed to improve biorefinery technologies for better profitability and increased adoption [7].

1.1.2 Biofuels and biomass recalcitrance

Biofuels are fuels derived from biomass. For second-generation biofuels, this biomass consists of the lignocellulosic (non-edible) part of woody crops and agricultural residues. These fuels can be made wherever the plants used as the biomass source are grown. A study by the Department of Energy (DOE) and the Department of Agriculture (USDA) estimates that the US can produce approximately 1.3 billion tons of cellulosic biomass per year, without having a grossly negative impact on food supply [5].
amount of cellulosic biomass can produce the energy equivalent of three billion barrels of oil, or 42 percent of the total US annual consumption. While biofuels cannot fulfill the entire energy demand of the US, their contribution can be significant.

Biofuels can be carbon neutral, which means that the carbon produced during combustion is consumed by the plants for organic synthesis. The overall process can be summarized in a few steps. First, green plants harness energy from the sun, converting it from photon energy to chemical energy by a process called photosynthesis. This energy is used, in part, to synthesize all the carbohydrates that constitute the plant matter. The carbon for these compounds is derived from the carbon dioxide taken from the air. The grown plants are then collected and mechanically ground to smaller pieces (1.0 - 0.1 mm). Next, microorganisms use enzymes to convert the biomass into sugars. These sugars are in turn fermented and converted into fuels by these or other microorganisms. Finally, when the fuel is burned, growing plants will reabsorb the emitted carbon dioxide.

Although the conversion of biomass to biofuels is simple in principle, the complex nature of the biomass structure makes it difficult to achieve an economically feasible process. A key challenge is to cost-effectively disrupt the interaction of the components of the plant cell wall: cellulose, hemicellulose and lignin. Hemicellulose accounts for 10-25% of the total carbohydrates in a plant. Lignin can count for 17-25% of the lignocellulosic biomass [8]. Finally, cellulose, which is the main source for fermentable sugars, is very resistant to sugar solubilization and separation from the rest of the biomass.

Cellulose consists of a linear chain of several hundred to thousands of β-(1,4)-linked –D glucose units. These fermentable units are organized side-to-side in chains that are stacked in a parallel manner by hydrogen bonds. When glucose is arranged in such tight cellulose bundles, the resulting structure is referred to as crystalline cellulose. The crystalline arrangement can exist in multiple allomorphs based on the differing hydrogen bond networks and unit cell dimensions, and include cellulose I-alpha, I-β II, III, and IV. Only cellulose I-alpha and I-β are produced in a natural way [9]. The other three are created as a result of cellulose treatment or regeneration, during processes that assist in the conversion of cellulose to glucose [10]. Cellulose characterization and deconstruction is therefore vital to the biofuels industry [11, 12].

Hemicellulose is similar to cellulose, as its prefix indicates. However, it has a branched non-crystalline structure. This difference in structure is mainly due to the presence of other six-carbon (C6) sugars such as mannose, galactose and rhamnose together with five-carbon (C5) sugars such as xylose and arabinose [8]. As such the “hemicellulose” term has become relatively generic. Within the hemicellulose category are more specific polymer arrangements of sugars such as xylan, glucuronoxylan, arabinoxylan, xyloglucan and glucomannan [8]. All hemicelluloses are randomly branched in a structure that is weaker than cellulose. It is easier to break hemicellulose into its simpler sugar units, but this results in a heterogeneous mixture of C6 and C5 sugars. The C5 sugars can be metabolized via the pentose phosphate pathway but the technology is not yet as efficient as the fermentation of glucose into alcohols, and typically occurs sequentially rather than simultaneously with glucose utilization.

Lignin is a highly recalcitrant network of cross-linked aromatic alcohols [13]. Its complex network makes it difficult to characterize [13]. Although lignin can be an energy
rich molecule, it cannot be used as a source for fuel production. Therefore, it is typically burned to generate energy for the rest of the processes involved in biofuel production.  

As described above, cellulose is a polymer of glucose. Glucose is a substrate for the production of fuel substitutes using engineered metabolic pathways [14, 15]. Sugars such as glucose are also used to produce molecules such as succinic acid, which can be subsequently converted to high-value chemicals such as 2-pyrrolidone, 1,4-butanediol, and tetrahydrofuran [7].

1.2 Biomass pretreatments

It is well known that cellulose can be converted into glucose monomers through a hydrolytic process known as saccharification. In nature, this process is carried out by enzymes called glucanases. These enzymes can be efficient and specific, and often have optimal activity under moderate temperature and pH conditions; they are therefore the preferred method for industrial cellulose degradation as well. However, the crystallinity of the cellulose, together with the network of lignin and hemicellulose within biomass, are physical barriers for the action of the glucanases. Thus the biomass is first subjected to pretreatments, including mechanical and thermochemical processes, prior to saccharification with glucanases.

The first stage of biomass pretreatment involves its mechanical reduction. The purpose of this stage is to increase the surface area of the biomass to enable increased interaction with the chemicals used during the next stages of pretreatment [12, 16, 17]. For instance, during this stage woody biomass is first chopped to reduce the size of the wooden pieces of 10-30 mm. These chips are then further reduced in size by hammer and knife mills, to 0.2 -2 mm [18]. It is important that the energy used during mechanical reduction of biomass is significantly less than the energy in the biomass.

Chemical pretreatment of biomass is the step after mechanical reduction. Many of these processes are well characterized because of their common use in paper manufacturing. However, due to the different needs of biofuel production, biofuels-targeted pretreatments are focused on increasing the accessibility of lignocellulolytic enzymes into the cellulosic material while having the highest possible sugar yields at the possible lowest cost. This can be achieved by exposing the biomass to chemicals such as concentrated acid, diluted acid, bases, organic solvents and ionic liquids and/or to physicochemical treatments including steam explosion, ammonia fiber explosion (AFEX) and hydrothermolysis.

1.2.1 Chemical Pretreatment Methods

1.2.1.1 Acid and Alkaline pretreatments

Concentrated acid pretreatment consists of exposing the biomass to high concentrations of strong acids such as 72% sulfuric acid (H₂SO₄) and 42% hydrochloric acid (HCl). Concentrated HCl penetrates the biomass and solubilizes the crystalline cellulose during several phases [12, 19]. Normally, the hemicellulose is first removed at 1% HCl at 130°C. The wood can then be dried and treated with 40% HCl. The HCl and the lignin are recovered by vacuum distillation and wash, respectively. The dilute acid pretreatment is a variation of the concentrated acid pretreatment, developed as a result of the low efficiency of acid recovery during concentrated acid pretreatments. This process includes a wood grinding stage (1 mm) prior to acid treatment. This facilitates permeation
of the acid into the biomass. Hemicellulose is then removed with 0.05 N H$_2$SO$_4$ at 140°C for approximately an hour. Alkaline pretreatments are based on the ability of alkaline agents (e.g. NaOH, hydrazine and anhydrous NH$_3$) to reduce cellulose crystallinity by disrupting the intracrystalline regions. Cellulose structure can also change from its allomorph I to II and/or III [10]. The advantage of these processes is that they can be conducted at low temperatures, but a disadvantage is that they can take from hours to days.

**1.2.1.2 Organic-based pretreatments**

The most common organic solvent for biomass production is *organosolv*, a mixture of water and organic solvents such as acetone, methanol, ethanol, ethylene glycol, formic acid and acetic acid. The ratio of each solvent can vary depending on the pretreatment or manufacturer. The water ratio can also vary from 40 - 80%. During this pretreatment the biomass is exposed to the mixture at 100 – 250°C. Temperatures above 185°C cause the release of organic acids necessary for the pretreatment. An inorganic acid catalyst (H$_2$SO$_4$) is added to the mixture if the treatment is run below 185°C [20]. There is a variation of the organosolv treatment called *Lignol* [21]. In this process the cellulosic fraction is directly converted to sugars for fermentation without the need of enzymatic hydrolysis. The biomass is cooked with approximately 50 wt% ethanol at low pH (2.0 - 4.0) and high temperatures (180 – 195°C) for 0.5 -15 hours. The ratio of liquid to biomass can vary from 4:1 to 10:1 (w/w). The lignin is cleaved and dissolved during the process and the cellulosic fraction can be hydrolyzed to sugars in short times (12 – 24 hrs). A lignin-rich black liquor is produced from this method. The lignin can be precipitated, filtered, washed, dried and sold at a price significant enough to improve the economics of this process. This is different from previous methods in which the lignin is burned for generation of heat due to its poor purity and quality.

**1.2.1.3 Ionic Liquid pretreatment**

Ionic liquids (ILs) can also be used for the pretreatment of biomass due to their ability to dissolve lignocellulose [22, 23]. These liquids consist of organic salts with melting points typically below 100°C. They are not flammable and have low vapor pressure. Both the anion and the cation of the IL are believed to play a role in the dissolution of the biomass. In a normal treatment, the biomass will be dissolved in ILs (3 – 10 wt%) at 120 – 160°C for 3 – 6 hours. Addition of an antisolvent such as water causes precipitation of the cellulosic fraction. This cellulosic fraction can be recovered by filtration or centrifugation. The recovered cellulosic fraction can be hydrolyzed into sugars by enzymatic treatments [22, 24, 25]. Recovery of the ionic liquids present the major drawback for this treatment.

**1.2.2 Physicochemical pretreatment methods:**

**1.2.2.1 Steam based treatments:**

Steam explosion pretreatments have been shown effective in multiple types of biomass such as hardwoods, softwoods and agricultural residues. During steam pretreatment, the biomass is exposed to steam at 160 – 260°C for a few seconds and then discharged [26, 27]. The decompression during the process significantly increases the biomass surface area. The hemicellulose is quickly dissolved in this process by the acids
released from the biomass. Lignin is melted and re-polymerized in new locations within the biomass. A pre-saccharification wash is necessary to remove degradation products and acids that are released during the steam pretreatment and that can inhibit the lignocellulosic enzymes.

Ammonia fiber explosion is another version of the steam explosion. In this case, ammonia is added as a reaction catalyst. Prior to steam explosion, 0.5 - 1.0 kg of anhydrous ammonia per kg of dry biomass is added to pre-wetted biomass (60% moisture content) at 650 psi and 130°C for 15 mins. The pressure is then released and the ammonia is removed from the biomass by air drying [28]. A less-common but effective method is the hydrothermolysis [29]. This process is based on the ability of water to dissolve organic materials at its critical point. Separation of biomass polymers is possible by exposure to pure water at 340°C [30]. The main advantage of this process is that it does not involve addition of any chemicals.

1.2.3 Current status of biomass pretreatments:

All the pretreatments mentioned above are efforts toward the search of the ideal pretreatment, which should have the following characteristics: “1) be effective on a wide range of feedstocks, 2) accept minimally prepared biomass, 3) obtain each of the major biomass components with high quality and purity, 4) avoid degradation of sugars or production of inhibitors during the fermentation process 5) result in a cellulosic stream that can be hydrolyzed effectively and with minimum enzyme loading, 6) require low energy inputs and 7) have minimal operating and capital costs” [12]. No current pretreatment satisfies these requirements, and researchers are still looking for answers that can help us achieve these high standards.

1.3 Enzymes as part of degradation process

1.3.1 Introduction to carbohydrate-active enzymes

Many carbohydrate-active enzymes are responsible for the deconstruction (or construction) of the plant cell wall. Description and classification of these enzymes can be found in the carbohydrate-active enzymes database (CAZy; www.cazy.org) [31]. CAZy has classified many of these enzymes into four main groups: glycoside hydrolases (GH), glycosyltransferases (GT), polysaccharide lyases (PL), carbohydrate esterases (CE), and the newest classification, auxiliary activity (AA) enzymes. This last classification includes different types of enzymes such as lytic polysaccharide monooxygenases (LPMO) and redox enzymes involved in lignin breakdown [32]. Other classified domains in CAZy include non-catalytic carbohydrate-binding modules (CBM).

Glycoside hydrolases (GH), also known as glycosidases and glycosyl hydrolases, are the enzymes responsible for the degradation of glycosyl polymers. They catalyze the hydrolysis of the glycosidic linkage of O-, N- and S-linked glycosides. As the reaction suggests, a water molecule cleavages a bond that results in glycoside polymers with reduced-/reducing-ends and/or free sugars, depending on the complexity of the sugar polymers. Exo- and endo- are very general classifications among all GHs. The terms endo- and exo- refers to the ability of GHs to cleave a substrate within the middle or at the end of a glycan chain. Examples of GHs can be found in almost every organism able to obtain sugars from glycosides. Some of these enzymes are further discussed below.
In contrast to GHs, the glycosyltransferases (GTs) are the enzymes responsible for the synthesis of the polysaccharides comprising the cell wall structure. This means that they catalyze the formation of the glycosidic linkage between mono-saccharides (activated nucleotide sugars) to the rest of the saccharide chain (a glycosyl acceptor molecule) to form the glycoside. Resulting linkages can be part of an O-, N-, S-, or C-glycoside. These enzymes show very conserved three-dimensional structures grouped as GT-A, GT-B, and a newer GT-A-similar fold [33, 34].

Polysaccharide lyases can cleave several activated glycosidic linkages that are present in acidic polysaccharides. Their mechanism of action is not hydrolytic but rather a β-elimination in which an unsaturated hexenuronic acid residue and a new reducing end are generated. Interestingly, these enzymes exist in variable folds types, which suggest their evolution originated from different scaffolds [35, 36].

Carbohydrate esterases (CE) are those that catalyze the de-O or de-N-acylation of substituted saccharides. The fact that an ester can be viewed as the addition of an acid and an alcohol, divides the substrates of these enzymes into two groups: 1) sugars that play the role of an acid, and 2) sugars that play the role of an alcohol. Examples of CEs include the chitin deacetylase from *Mucor rouxii*, the acetyl xylan esterase from *Streptomyces lividans*, and the acetyl xylan esterase from *Bacillus pumilus*. The *B. pumilus* esterase is classified as a CE from family 7 and is very specific towards xylan [37]. The other two, classified under CE family 4, are active on xylan as well as other chitinous substrates. Interestingly, these enzymes seem to be affected by the presence of metals (e.g. Co²⁺), which is not uncommon among carbohydrate-active enzymes.

The auxiliary activity (AA) enzymes are the latest addition to the CAZy repertoire. These AA enzymes are specifically classified based on biochemical characterization [32] and their classification is not strictly dependent on a particular catalytic reaction or a specific substrate. There are currently 10 AA families consisting of non-hydrolytic but oxidative enzymes such as laccases, ferroxidases, peroxidases, and oxidoreductases among others. The newest families of the group are family nine and ten (AA9 and AA10), each including one characterized example of a lytic polysaccharide monoxygenase, which are able to oxidatively cleave polysaccharide chains. Interestingly, the proteins previously categorized within the GH61 and CBM33 families are now have been recategorized in the AA9 and AA10 families, respectively.

All enzymes can also be classified by an enzyme commission (EC) number. The EC is based only on the enzymatic reaction catalyzed by the enzyme. Enzymes capable of catalyzing more than one reaction bear more than one EC number. For the purpose of this dissertation we concentrate on those GHs with EC numbers related to the breaking of β-1,4-glycoside bonds found in cellulose. These enzymes are called cellulases and fall under EC 3.2.1.4. The order of the numbers “3, 2, 1 and 4” go from more general to more specific classification and indicates that these enzymes have activity of hydrolases, glycosylases, glycosydases and cellulases, respectively.

1.3.2 Enzymes for biomass degradation (lignocellulolytic cocktails):

Complete enzymatic degradation of untreated biomass can be a very slow process that may require the action of many different GHs, CEs and AAs. Pretreating the biomass (see sections above) can increase saccharification yields by up to 10 fold [38], and also reduce the amount of saccharification enzymes to a fraction of what would be otherwise.
Goutami et al. [39] reported optimal cellulose degradation of IL-pretreated substrates with a total of six core enzymes listed as follows: two cellobiohydrolases (types I and II) one endoglucanase (EG1), one β-glucosidase (BG), one endo-β-1,4-xylanase (EX) and a β-xylosidase (BX). Different ratios of these enzymes were needed for optimal saccharification of cellulosic substrates that were pretreated differently. Some of the pretreated substrates required the additional auxiliary enzymes in for optimal saccharification. Current work performed at the Joint BioEnergy Institute has demonstrated the possibility of reaching optimal saccharification of IL-pretreated substrates with only three enzymes (See chapter 4). These three enzymes are the same three cellulases (among the glycosyl hydrolases) suggested by Henrissat et al. [40] to be responsible for the degradation of cellulose. We believe that the reason why only three enzymes are needed for optimal saccharification of IL-pretreated biomass is that the main component of this specific substrate is cellulose.

1.3.3 An introduction to cellulases for cellulose degradation

Cellulases are glycosyl hydrolases that specifically hydrolyze β-(1,4)-glucosidic bonds in cellulose polymers [31]. The hydrolysis products are simpler glucose oligomers called cellodextrins. Three major types of cellulases are thought to synergistically depolymerize cellulose into its glucose sugar monomers: endocellulases, exocellulases or cellulbiohydrolases and β-glucosidases [40]. Based on the best-known model, hydrolysis of cellulose is initiated by random action of an endocellulases (EG, E.C. 3.2.1.4) in the intermediate (and probably amorphous) regions of the cellulose fibers. The purpose of this action is not to liberate free cellodextrins (glucose oligomers) into solution, but rather to make the ends of the cellulose fibers available to cellulbiohydrolase (CBH, E.C. 3.2.1.91) action. Cellulbiohydrolases then release cellobiose from non-reducing ends off the cellulose chain [41, 42]. This theory has been debated because of the findings of cellulbiohydrolases that can degrade highly organized cellulose without assistance of an EC [43, 44]. The β-glucosidases are then responsible for the conversion of cellobiose, a glucose dimer, into two glucose monomers. As such, β-glucosidases do not play a role in solubilizing the highly organized crystal structure of these cellulose substrates. However, they are important to the cellulose degradation process, as β-glucosidases directly produce the desired sugar product (glucose) and in some cases also alleviate cellulbiohydrolase product inhibition (β-glucosidases lower cellobiose concentrations).

Although necessary, the synergistic role of each cellulase in the degradation of cellulose (as well as pretreated cellulose) is highly debated. A seminal work by Henrissat and coworkers [40] demonstrated that the activity of cellulases (or combinations of cellulases) on cellulosic substrates depends on the specificity and the specific synergy of the cellulases. Specifically, these researchers observed synergistic interactions between endocellulases, cellulbiohydrolases from T. reesei while depolymerizing cellulosic substrates, but found that no enzyme mixture was optimal against all types of substrates. [40]. This work reported little production of soluble cellodextrins from insoluble cellulose substrates when exposed to individual cellulases. However, the same substrates showed higher reactivity when exposed to mixtures of EC and cellulbiohydrolases. Synergistic degradation was also observed with mixtures of two different cellulbiohydrolases (type I and type II).
Non-catalytic modules may also affect glucanase synergy, particularly those modules that affect the binding of their enzymatic domains to a specific substrate [45–47]. For example, some cellulases contain carbohydrate-binding modules (CBMs), which can increase cellulase activity by enhancing the binding of the enzyme to the substrate [48]. These modular CBMs are not only present in cellulases, but also in many different types of glycosyl hydrolases and other carbohydrate-active enzymes. Classification and the modes of action of CBMs will be discussed in the next section 1.4.1.1.1.

1.4 Need for optimization of cellulolytic enzymes

As the previous paragraphs have suggested, complete depolymerization of cellulosic biomass to its minimal units (e.g. glucose) may be achieved by mixing the right combination of enzymes (and modular domains) with an appropriately pretreated feedstock. As straightforward as this appears, techno-economic analyses conclude that the use of material resources, energy and time for saccharification significantly increases the production cost of biofuels. The estimated cost for enzymes required to produce one gallon of ligno-cellulosic ethanol is $1.47 [49]. This elevates the selling price of ethanol-based fuels (E85) to over $5.00 per gallon, a price that cannot compete against petroleum-based fuels in today’s market [20]. Reduction of time (and cost) for biomass saccharification remains a big challenge due to the many parameters involved in the process. These variables are not only associated with biomass structural features but also with enzyme-related factors such as the enzyme source, enzyme production titers, product inhibition, thermal inactivation, activity balance for synergism, specific activity, non-specific binding, enzyme processibility and enzyme compatibility.

Research is being done to overcome the variables that limit the feasibility of enzymatic saccharification for biofuel production. Of primary interest is the optimization of enzyme cocktails to curb product inhibition [50], the engineering of enzymes that are less susceptible to inhibition and deactivation [51], the optimization of enzyme production [52], and the discovery and optimization of new enzymes for cellulose degradation [45, 53].

1.4.1. Examples of enzyme optimization

1.4.1.1 Brief on enzyme optimization

Banerjee et al. [39] have taken a significant step toward the optimization of enzymatic cocktails for dedicated substrates. Using a high-throughput analysis platform called GENPLANT, they were able to combine purified enzymes with statistical experimental design to determine the best enzymatic combinations for the saccharification of different cellulosic pretreated substrates.

Jeoh et al. [51] have demonstrated that the host used for enzymatic production can affect an enzyme’s activity and susceptibility to inhibition. This work is a strong example of the relevance of producing fungal enzymes in native organisms. Heinzelman and co-workers [53] used SCHEMA, a structure-guided recombination method, to create thermostable fungal cellulases able to operate under industrial conditions (e.g. high temperature). The SCHEMA approach uses protein structure data to define boundaries of contiguous amino acid “blocks” to design chimeric cellulases with significant sequence diversity. Finally, researchers such as Mingardon et al, Kim et al, and Reyes-Ortiz et al,
have used the approach of modular CBM addition to improve the activity of certain enzymes [45, 54, 55].

1.4.1.1 Introduction to carbohydrate binding modules (CBMs)

Lignocellulolytic enzymes are commonly composed of various modular domains. Some of those domains are called carbohydrate-binding module (CBMs) for their capacity of binding to carbohydrate molecules with no catalytic action involved. It is well established that the presence of these CBMs can increase the efficiency of the enzymes complexes that have them. The most common hypothesis is that CBMs can do this by increasing the local concentration of the enzyme on the surface of insoluble substrates. Adsorption of CBMs to cellulose depends on three critical aromatic residues that are aligned with every second glucose ring in the glucan chain. The positioning of these three aromatic residues allows for van der Waals’ interactions with the pyranose rings exposed on the surface of the cellulose [56, 57]. Different types of CBMs, classified as A, B and C, are able to interact with different types/parts of cellulose.

Type A CBMs are surface-binding CBMs. This category includes members from families 1, 2a, 3, 5 and 10, which are known to bind crystalline cellulose and/or chitin. The binding site of these CBMs, a set of aromatic residues, form a planar surface that is predicted to be complementary to the crystalline surface of cellulose or chitin [31, 48]. Studies done by Tormo et al and Lethio et al suggest that the binding of this CBM is specific to the 110 face of the cellulose crystal [57, 58]. Type B CBMs can be found with cellulases, mannanases and xylases [48]. These CBMs are also known as glycan-chain-binding CBMs. In contrast to type-A CBMs, type-B CBMs show a concave binding site. This binding site is described as a cleft and is able to accommodate the individual sugar units of a cellulose fiber. Although the depth of the cleft can vary, these types of CBMs are always observed to bind glycan chains (longer than three saccharides) and not crystalline cellulose. In some cases they have been found to bind reducing ends of polysaccharides, suggesting that they might be targeting damaged regions of the plant cell wall [59, 60]. Hydrogen bonds are primarily responsible for the binding of type-B CBMs to glycan fibers. Type-C CBMs are also called lectin-like CBMs. These CBMs have small concave binding sites that enable them to recognize very small saccharides chains such as mono-, di- or tri-saccharides. This concave shape is similar to that of type-B CBMs, but the length of the cleft is what restrains this CBM from binding longer polymer saccharides. Due to their appearance, these CBMs were first described as lectins capable of binding short sugar polymers. However, their presence as cellulase domains resulted in their classification among the CBMs. Deeper knowledge of these CBMs is lagging behind their type A and B counterparts due to the relatively few known type-C CBMs. Those that have been characterized (such as those from families 13 and 31) appear to be more common in bacterial toxic enzymes that appear to attach cell surface matrix glycans [48].

CBMs have also been classified within over 60 CBM families that share similar amino acid sequences [31]. Family classification is not completely related to substrate binding and so there is not a clear correlation between CBM type and family. However, some trends have been observed. Members of the family 4, 17 and 28 tend to exhibit higher affinity toward non-crystalline cellulose [61–64], while CBMs from families 1, 2a and 3a tend to bind crystalline cellulose [65]. Specificity is one of the most intriguing
aspects of these modules. One of the first studies on CBM specificity was performed by Carrard et al. [66], who appended CBMs belonging to families 1 and 3 to the same catalytic domain of an endoglucanase from *C. thermocellum*. Both CBMs seemed to increase the degradation of cellulose by the cellulose, but in different ways, implying that both CBMs can recognize different parts of the same cellulose substrates. Another interesting study was performed by Boraston et al. [67]. In this work, CBMs from families 17 and 28 were shown to bind non-crystalline cellulose in a non-competitive way.

The presence of a CBM is not typically required for activity of the CD. However, removal of the CBM from the CD can cause a decrease in activity against insoluble cellulose [68–70]. Activity of the same CDs remained unaffected on soluble substrates when the CBM is removed [55, 69, 70]. Bolam et al. [71] set out to investigate the role of a cellulose-specific CBM (cellulose binding domain, CBD) in enhancing the activity of an endoglucanase on bacterial microcrystalline cellulose and acid-swollen cellulose (an amorphous cellulose). In order to do this, a CBD, originally found on a xylanase A from *Pseudomonas fluorescens* was fused to the catalytic domain of an endoglucanase from *Clostridium thermocellum*. The findings indicate that the additional CBD enhanced the activity of the endoglucanase CD against insoluble cellulose, but not soluble cellulose (Carboxymethyl cellulose, CMC). This suggests that the CBMs increase cellulose degradation by increasing the local concentration of the enzymes on the surface of insoluble cellulose. However, CBMs isolated from bacteria and fungi have also been shown to facilitate cellulose degradation by a non-catalytic physical disruption of the cellulose fibers [72, 73].

Din et al. [72] documented the first disruptive function of a CBM type A from family 2 (CBM2a) located at the N-terminus of an endoglucanase (Cel6A) from *Cellulomonas fimi*. In this study, both the catalytic domain and CBM of the same cellulase were exposed to cotton fibers. Intriguingly, there was a release of small cotton particles as well as surface roughening after the fibers were exposed to the CBM. Exposure of the same kind of fibers to the CD alone did not give rise to these same observations. Instead, the CD seemed to polish the surfaces of the cotton fibers. The authors of this work suggested that the CBM appeared to have a non-catalytic but disruptive effect on the crystalline structure of cellulose. The disruptive capabilities of the CBM also were required to enhance the activity of the catalytic domain of the cellulase Cel6A from *C. fimi* on insoluble substrates. In unrelated work, Teeri et al. [73] proposed that CBMs are able to bind to cellulose and penetrate in those areas that showed network discontinuity. This could liberate cellulose fibrils that are not covalently bound to the cellulose, but interact with the underlying cellulose layers. Teeri et al also suggested that further penetration of the CBM into the cellulose substrates could cause release of the ends of the cellulose fibers. These fibers remain covalently bound to the rest of the fiber, but the release of the ends results in roughening of the cellulose. Another interesting finding of the CBM disruptive capabilities is described by Lee et al. [74]. With the help of atomic force microscopy, Lee et al. observed slightly elongated holes that were left through the surface of cotton fibers after being exposed to an inactive cellubiohydrolase from *T reseei*. These holes were not observed when the fibers were incubated with an inactive enzyme with no CBM. Therefore, the elongated holes were interpreted to be a result of CBM penetration in the cotton fibers.
Another mechanism for CBM-enhanced catalytic activity is proposed by Gilad et al. [75]. In their model the CBMs align cellulose fibers for better docking of the CDs, hence, enhancing its activity. In this work, a processive cellobiohydrolase is capable of solubilizing different insoluble cellulose substrates. However, this activity seems to be lost when the C-terminus of CBM3b is truncated from the wild type cellulase. The cellulase then fails to bind the crystalline substrates, resulting in a decreased activity. When the CBM3c was also truncated, the cellulase activity decreased even more. This data supports the relevance of accessory CBMs on the processivity of these enzymes by facilitating the docking of the CD with the cellulose fibrils. This case was further studied and supported by Beckham et al. [76]. In his work, Beckam uses computational models to study the role of the CBM into the degradation of cellulose. His results also suggest that the CBM enhances the activity of cellulases by aligning the cellulose fibril and facilitating the docking of the CD.

1.6 Introduction to CBM studies in this dissertation

A detailed understanding of how CBMs increase the activity of cellulases remains elusive. We have completed a series of studies that shed light on how certain CBMs affect the activity of cellulases from three different perspectives:

1) Specific enzyme-substrate interactions: Neutron reflectivity is used to analyze the effect of endo cellulases on cellulose film density and thickness. Two endo cellulases, each tethered or untethered to a single CBM, are compared in this study. We then expanded our method to investigate a broader set of CBMs with the same endo cellulases.

2) General synergy between CBMs from three different families and a specific cellulase: We set out to determine if there is any correlation between CBM family and activity level or substrate preference, as measured by enhanced activity of a single cellulase. We chose three different CBMS belonging to families one, two and three, and tested them on crystalline and non-crystalline substrates. These chimeric cellulases were then exposed to amorphous cellulose model films where their interactions were studied by the use of neutron reflectivity.

3) The effect of enzymatic ratios in simple cellulolytic cocktails and how the addition of the CBMs affect those ratios: This study is focused on the optimization of simple cellulolytic cocktails for specific operational conditions. With basic conditions established, optimization is achieved by changing the ratios of the different components within the cocktail.
Chapter 2 – Addition of a carbohydrate-binding module enhances cellulase penetration into cellulose substrates

This chapter is a postprint of a paper submitted to and accepted for publication as


© 2013 Reyes-Ortiz et al.; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The copy of record is available at http://www.biotechnologyforbiofuels.com/content/6/1/93

2.1 – Abstract

Cellulases are of great interest for application in biomass degradation, yet the molecular details of the mode of action of glycoside hydrolases during degradation of insoluble cellulose remains elusive. To further improve these enzymes for application at industrial conditions, it is critical to gain a better understanding of not only the details of the degradation process, but also the function of accessory modules. Toward that end, we fused a cellulose-binding module (CBM) from family 2a to two thermophilic endoglucanases. Catalytic activity of the chimeric enzymes was enhanced up to three fold on insoluble cellulose substrates as compared to wild type. We then applied neutron reflectometry to determine the mechanism of these enhancements. Importantly, we demonstrate that the wild type enzymes affect primarily the surface properties of an amorphous cellulose film, while the chimeras containing a CBM alter the bulk properties of the amorphous film. These findings suggest that the CBM improves the efficiency of these cellulases by enabling digestion within the bulk of the film.

2.2 – Background

The utilization of enzymes for the conversion of biomass into fermentable products has been demonstrated to be a viable and promising approach toward the development of cost-effective biofuels [16, 49]. Cellulases hydrolyze the β-1,4-glycosidic bonds in cellulose, the most abundant biomass polymer. Despite advances in protein expression and enzyme optimization, there is still a need for significant improvement with respect to cellulase enzymatic activity at industrial process conditions, such as high temperature [11, 12, 53, 77, 78]. To that end, there have been some reported successes in engineering enzymes to be more active and stable at high temperatures [45, 53, 79], but still more are needed to further lower the costs of biofuel production [4, 11, 49]. Thermophilic cellulases, in particular, are promising due to their temperature stability but do not natively have the high activities required for efficient biomass hydrolysis.
A cellulase consists of a catalytic domain (CD) that often is linked to other modular accessory domains, including carbohydrate-binding modules (CBMs) [80], in many possible orientations and combinations [31, 48, 81]. CBMs are thought to 1) enhance the adsorption of CDs to their substrate [71]; 2) enable the alignment of cellulose fibrils for better docking of the CD [75, 76]; and 3) modify substrate surfaces to facilitate enzymatic hydrolysis [72, 82, 83]. However, some molecular details are still unclear, particularly in relation to how the CBM affects enzyme-substrate interactions and activity on solid substrates.

Several studies have been carried out for the purpose of understanding cellulase-substrate interactions [84–91]. Ellipsometry and quartz crystal microbalance with dissipation (QCM-D) were used to correlate changes in mass and energy dissipation to structural changes of crystalline cellulose model films incubated with cellulases [86]. Josefsson et al. performed similar studies with QCM-D and atomic force microscopy (AFM) which suggested that endoglucanases cause swelling of crystalline cellulose films [87]. Recently, analytical techniques have been applied to understand the interactions with amorphous films as well [90, 91]. For example, Suchy et al. combined QCM-D, AFM, and X-ray photoelectron spectroscopy (XPS) to demonstrate that an endoglucanase and a cellobiohydrolase work uniformly within the entire volume of swollen amorphous cellulose films [90]. Neutron reflectometry (NR) has also been applied to differentiate between cellulases that are more active at the at the surface of an amorphous cellulose film and those that are active in the bulk of the same substrate [83, 91]. We reasoned that these analytical techniques might be similarly applied to study the interactions of the CBM, CD, and cellulose substrate. This study directly addresses the effect of addition of a CBM to a CD in terms of changes in enzyme activity on insoluble substrate and on alterations to both the surface and bulk properties of non-crystalline cellulose films. To examine these effects, we genetically fused a CBM from family 2a to two thermophilic endoglucanases, Cel9A from Alicyclobacillus acidocaldarius and Cel5A from Thermotoga maritima, which do not naturally have a CBM. We observed that the addition of the CBM increases cellulase activity by up to three fold on insoluble cellulosic substrates. Data acquired with NR demonstrates that the wild type enzymes are active mainly on the surface of amorphous cellulose films while their respective CBM-containing chimeras substantially alter the bulk properties of cellulose films. These findings suggest that addition of the CBM enhances the efficiency of the two CDs not only by enhancing adsorption to the surface of the film but also by enabling increased penetration into and digestion within the bulk of the film. We therefore propose a new model for CBM-enhanced insoluble cellulose degradation.

2.3 – Results and discussion
2.3.1 – Construction and characterization of chimeric cellulases

In this work, we set out to systematically investigate the effect of addition of a CBM to the CD of two thermophilic cellulases, Cel9A from A. acidocaldarius and Cel5A from T. maritima. Cel9A and Cel5A are endoglucanases that release cellobiose and cellotriose, respectively, as their primary hydrolysis product [92–94]. We fused these CDs to a family 2a CBM from the thermophilic exoglucanase E3 (UniProt Q60029) from Thermomonospora fusca. This CBM is natively at the N-terminus of a CD from glycosyl hydrolase family 6 [95, 96] and is expected to bind crystalline cellulose [97]. The
chimeras were constructed in the format such that the CBM is at the C-terminus, with a 31 amino acid linker, denoted as “CD-CBM”. The CBM2a chimera was also made with two additional linker lengths (12 and 47 amino acids) in order to examine the role of linker length on activity. Finally, mutants for which the catalytic activity is eliminated (knockouts) were made for both the wild type and chimeras with the 31 amino acid linkers (“CD_ko-CBM” and “CD_ko”). The enzymes, chimeras, and knockouts were overexpressed in Escherichia coli and purified to >80% purity.

Catalytic activity of the chimeric cellulases on carboxymethylcellulose closely matched that of the corresponding wild type enzymes (Figure 2.1). Optimal temperatures (T_{opt}) and pHs were determined to be 75°C and pH 4.8 for Cel5A and its corresponding chimeras, and 65°C and pH 5.5 for Cel9A and its chimeras (Figure 2.1) These values are similar to those reported previously for the wild type enzymes [98, 99]. Polyacrylamide gel electrophoresis followed by Coomassie staining indicates no degradation of the chimera under standard assay conditions (Figure 2.2).
Enzymatic activities were measured via DNS assay on microcrystalline cellulose (MCC, Avicel pH-101, Sigma) and ionic-liquid pretreated microcrystalline cellulose (IL-MCC), to represent model crystalline and non-crystalline cellulosic substrates. The chimeric cellulases gave rise to up to approximately three-fold increases in sugar release from IL-MCC compared to their corresponding wild type CDs alone (Figure 2.3). Ionic liquid pretreatment is a promising method for biomass preparation that causes the structure of the cellulose to change from mostly crystalline to a mixture of cellulose II and amorphous cellulose [22], thereby enhancing downstream enzymatic hydrolysis [22, 100]. Since CBMs have different binding preferences [31, 48, 62, 63], we conclude that this CBM2a appears to be more specific for cellulose sites accessible within this less crystalline substrate [22].

**Enzymatic hydrolysis of insoluble substrates**

The enzymatic activities were measured via DNS assay on microcrystalline cellulose (MCC, Avicel pH-101, Sigma) and ionic-liquid pretreated microcrystalline cellulose (IL-MCC), to represent model crystalline and non-crystalline cellulosic substrates. The chimeric cellulases gave rise to up to approximately three-fold increases in sugar release from IL-MCC compared to their corresponding wild type CDs alone (Figure 2.3). Ionic liquid pretreatment is a promising method for biomass preparation that causes the structure of the cellulose to change from mostly crystalline to a mixture of cellulose II and amorphous cellulose [22], thereby enhancing downstream enzymatic hydrolysis [22, 100]. Since CBMs have different binding preferences [31, 48, 62, 63], we conclude that this CBM2a appears to be more specific for cellulose sites accessible within this less crystalline substrate [22]. It should be noted that while the Cel9A CD liberated soluble cellodextrins from both MCC and IL-MCC, the Cel5A CD alone did not exhibit any significant activity above background on either substrate (Figures 2.3 and 2.4). These results served as motivation to test for enzymatic hydrolysis of IL-MCC at higher enzyme loadings. At these higher enzyme loadings, the Cel5A CD did give rise to measurable cellodextrin release. Furthermore, the chimeric cellulases were again approximately three-fold more active than their respective CDs at every enzyme concentration (Figure 2.5, results reported in total production of cellodextrin). These experiments suggest that the CBM is the factor responsible for enhancing the cellulase activity of both chimeric constructs. This result is in agreement with the results of Kim et al., who also found up to three-fold increases in activity upon fusion of a CBM to certain CDs [45].

As a control, we also measured the enzymatic activities on these substrates using the same concentration of CBM with either Cel9A or Cel5A added as separate (non-
fusion) proteins. For these controls we observed the same level of activity as for each respective CD alone (Figure 2.3); therefore, physical linkage to the CBM is required for enhanced hydrolysis. Hydrolysis assays were also performed with chimeric cellulases that had linker lengths of 12 and 47 amino acids. As there were no significant differences in activity between the 31 amino acid linker and the other two linkers for either chimera (p>0.05 by two-tailed student’s t-test with equal variance for each; Figure 2.4), the remaining experiments were carried out using the original 31 amino acid linker.

2.3.3 – Amorphous Cellulose Film Characterization

To further explore the mechanism for the enhanced hydrolysis activity of the chimeras, we used neutron reflectivity to analyze changes in cellulose thin films upon incubation with the enzymes. The cellulose films were regenerated from trimethylsilyl cellulose (TMSC), and were largely amorphous as reported previously [91, 101], permitting comparison to the model substrate IL-MCC used in the hydrolysis assays. The films swelled in aqueous buffer (with no enzyme added) by a factor of ~ 1.9 - 2.2. Two concentrations of TMSC were used in order to determine if the roughness of the film varied appreciably with film thickness, but no consistent trend was observed. The film thickness was 250 Å - 270 Å for 10 mg/ml TMSC, and 310 Å– 340 Å for 12 mg/ml. The density of the dried films varied from 1.25 to 1.35 g/cm³. Other characteristics of these films have been reported previously [83]. NR data revealed that the films were highly smooth, and that swelling of the films was uniform except for some variation at the film-substrate interface.
2.3.4 – NR analysis of films incubated with cellulases

To aid discussion of the NR results, we summarize here several effects expected upon interaction of endoglucanases with amorphous cellulose films. i) Activity of endoglucanases at the surface of a film will release mass and result in a decrease in film thickness. ii) Activity of endoglucanases in the bulk of a film will result in an increase in water content. As endoglucanases digest they cleave $\beta$-1,4-glycosidic bonds randomly along cellulose chains. Enzymatic cleavage of bonds interior to cellulose chains will create free chain ends, whereas cleavage near the ends of cellulose chains will result in the release of soluble fragments. Since chain ends resulting from hydrolysis are hydrophilic by virtue of the hydroxyl group, both cases will result in increased water content. iii) Activity of endoglucanases in the bulk of a film may lead to an increase in film thickness. We assume that amorphous cellulose in an aqueous buffer will behave as a glassy polymer. For glassy polymers, the occupied volume will increase with an increase in the number of chain ends. This, and the increased hydrophilicity of chain ends, will both contribute to an increase in thickness upon the action of endoglucanases within the film. We note that hydrolysis occurring at chain ends, while increasing the water content, will not result in an increase in film thickness. iv) Adsorption of CDs and CBMs to cellulose will generally be weaker at higher temperatures due to entropic considerations.

We note that film expansion from activity within the bulk can be compensated by enzyme activity at the surface. Therefore, for enzymes that penetrate into the bulk of the
Figure 2.6 Chimeric cellulases (CD-CBM) show degradation and swelling of amorphous cellulose films at room temperature. Cellulose volume profiles are shown for each cellulase. Amorphous cellulose films were incubated at room temperature with a) Cel9A or Cel9A-CBM, and b) Cel5A or Cel5A-CBM. Plots show the normalized cellulose volume fraction versus normalized distance from the silicon oxide surface. Due to variance in film preparation, the baseline (film without enzyme addition) for every cellulose film assay is shown. Arrows indicate the cellulose-buffer interface. Brown dotted line shows the extrapolation of the cellulose profile for Cel9A-CBM after the enzyme layer at the surface of the film was removed.

<table>
<thead>
<tr>
<th>Cellulase</th>
<th>Percentage of mass area removed</th>
<th>Film swelling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RT</td>
<td>T opt</td>
</tr>
<tr>
<td>Cel9A</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>Cel9A-CBM</td>
<td>12</td>
<td>32</td>
</tr>
<tr>
<td>Cel5A</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Cel5A-CBM</td>
<td>16</td>
<td>35</td>
</tr>
</tbody>
</table>

Table 2.1 Observations of cellulose films exposed to different cellulases. Chimeric cellulases show higher cellulose removal and swelling of the amorphous cellulose film. The table above summarizes the neutron reflectivity (NR) results of the amorphous cellulose films exposed to several wild type and chimeric cellulases. The mass/area removed determined by integrating the cellulose volume profiles. Film swelling is the increase of film thickness above baseline observed after the film is exposed to the cellulase. RT and T opt stand for “room temperature” and “optimal temperature”, respectively. All standard deviations are within 10%.
film, the magnitude of the change (increase or decrease) in film thickness will be determined by i) the amount of bond scission interior to chains that creates chain ends relative to the amount of bond scission at chain ends that results in soluble fragments, and ii) the relative levels of activity within the bulk and at the surface of the film.

Figure 2.7 Neutron reflectivity data from cellulose films exposed to Cel9A (a), Cel5A (b) and the chimeric cellulases Cel9A-CBM (c) and Cel5A-CBM (d) at 5 mM and 20°C. The fringe pattern, related to the thicknesses of the various layers, is distinctly different for the two chimeras: the fringes decay uniformly with qz for Cel5A-CBM, whereas for Cel9A-CBM the first three fringes are damped relative to the fringes at higher qz. The latter pattern indicates a distinct layer of enzyme adsorbed at the surface of the film.
Figure 2.8 Scattering length density (SLD) profiles with respect to distance for Cel9A-CBM (a) and Cel5A-CBM (b) at 20°C. The SLD profile for Cel9A-CBM shows an additional layer at the film-solution interface that we attribute to adsorbed protein.

Figure 2.9 Cellulose volume profiles are shown for chimeric cellulases with inactive catalytic domains (CD<sup>KO</sup>-CBM). Amorphous cellulose films were incubated at room temperature with a) Cel9A<sub>KO</sub>-CBM, and b) Cel5A<sub>KO</sub>-CBM. Plots show the normalized cellulose volume fraction versus normalized distance from the silicon oxide surface. Due to variance in film preparation the baseline (film without enzyme addition) for every cellulose film assay is shown. Arrows indicate the cellulose-buffer interface. Brown dotted line shows the extrapolation of the cellulose profile for Cel9A-CBM after the enzyme layer at the surface of the film was removed.
Finally, NR is highly sensitive to changes in the breadth of the film/solution interface. Broadening of that interface could occur due to enzymes digesting as they enetrate into the film, or could also result from film swelling. Lateral surface roughness and a composition gradient normal to the interface have identical effects on the reflectivity data [57, 58]. With these considerations in mind we now discuss the NR results for the CDs and chimeras. In each case, initial (prior to adding enzymes) and final volume fraction profiles are compared. The final profile is that measured after the activity had slowed such that little or no change in the NR data was observed on the time scale of several hours. For Cel9A and Cel5A incubated against cellulose films at RT, the cellulose volume fraction profiles (Figures 2.6, 2.7a and 2.7b) show relatively small changes. Cel9A and Cel5A removed only 5% and 1% of the cellulose, respectively (Table 1). The volume fraction profiles indicate that cellulose was removed primarily from the surface of the films. In contrast, far more substantial changes were observed in films incubated with the chimeras. First, much greater mass loss resulted (12% and 16% for Cel9A-CBM and Cel5A-CBM, respectively). Second, much greater digestion occurred within the bulk of the films, as reflected by large increases in water content. Third, for Cel5A-CBM the cellulose-buffer interface moved to greater depths, indicating film expansion at RT (Figure 2.6). This expansion, despite the substantial loss in cellulose mass, is qualitatively different from the changes observed for the Cel5A CD alone. Taken together, these results demonstrate that the chimeric enzymes penetrate into, and are active within, the cellulose films to a far greater extent than the CDs alone. We emphasize that the final profiles were measured after activity had slowed dramatically, and so the lack of activity of the CDs within the bulk of the film is a true limitation of the enzyme and is not due to a slower rate or insufficient time. Moreover, we repeated the experiment with five-fold less Cel9A-CBM (1 mM) and the final profile was nearly the same as at 5 mM (data not shown).

For Cel9A-CBM a fringe pattern that is distinctly different from that for Cel5A-CBM resulted, that indicates a layer of protein adsorbed to the film surface (Figures 2.7c, 2.7d and 2.8). The additional layer at the film surface is clearly identified as a layer of protein because the scattering length density (SLD) value for that layer is much lower than that of the swollen cellulose film. This additional layer is not observed for the Cel9A CD alone nor for the Cel5A-CBM chimera. Therefore this layer is likely due to strong CBM binding at the surface along with an inherent property of the Cel9A CD that inhibits its penetration (e.g. larger size that slows its entry into the bulk of the film). Thus for Cel9A-CBM the results indicate substantial penetration into the bulk as well as a distinct layer of enzyme remaining at the surface. Film expansion was not observed for Cel9A-CBM as for Cel5A-CBM, but it is possible that for Cel9A-CBM film expansion was compensated by activity at the surface of the film due to the adsorbed layer of enzymes. The cellulose profile shown in Figure 2.6 for Cel9A-CBM was obtained by truncating the full SLD profile (Figure 2.8) at the interface between the cellulose film and the layer of adsorbed enzyme. As controls, we also examined the effect of incubating the film with the knockout chimeras, which are inactive due to mutations in the active site residues (Figures 2.9 and 2.10). With each knockout chimera, the film expands somewhat (~30 Å for Cel5A\text{ko}-CBM and ~15 Å for Cel9A\text{ko}-CBM), supporting the conclusion that the protein is penetrating into the bulk of the film. In addition we observe again that a layer of protein adsorbs to the film surface for Cel9A\text{ko}-CBM, but not for Cel5A\text{ko}-CBM.

21
(Figures 2.10c and 2.10d). There is negligible change in the cellulose volume fraction and water content of the films, as expected for inactive enzymes. While some water displacement is expected from protein entering the film, this will be offset by the increased hydrophilicity of the protein, such that little change in the SLD results. We also incubated a film with the same concentration of CBM alone at RT and at 65 °C (Figure 2.11), and incubated another film with a mixture of Cel9A and CBM (Figure 2.12). For CBM alone a small but definite film expansion (~ 10 Å) was observed at each temperature. Again, this is consistent with CBM localization within the bulk of the film, and slight swelling due to the hydrophilicity of the protein. For the mixture of Cel9A and CBM the effects were far weaker than for the chimera. The water content in the bulk of the film was only slightly greater than for CBM alone, and the increase in film thickness
was comparable to that for the CBM alone. These results suggest that localization of the CBM within the bulk film may slightly increase the penetration of Cel9A, but not nearly to the level as for the chimera.

At $T_{\text{opt}}$ (65°C for Cel9A and 75°C for Cel5A) the results again show much greater activity within the bulk of the film for the chimeras relative to the wild-type enzymes as indicated by greatly increased water content (Figures 2.13 and 2.14). In addition, a large increase in the interfacial gradient occurred with the chimeras, whereas little change in the interfacial gradient resulted with the CDs. The increase in breadth of the film-solution interface is very obvious in the NR data as the fringes are severely damped at higher $q_z$ values (Figure 2.14). The large interfacial gradient in each case is consistent with the hydrolysis of bonds as the chimeras penetrate into the film as well as film swelling. At $T_{\text{opt}}$ the change in film thickness between Cel5A-CBM and Cel5A is distinctly different.

**Figure 2.11:** Neutron reflectivity data (a,c) and normalized cellulose volume fraction profiles (c,d) for a cellulose film before and after incubation with 5μM CBM at 20°C (a,c) and 65°C (b,c).
Figure 2.12 Neutron reflectivity data (a) and normalized cellulose volume fraction profiles (b) for a cellulose film before and after incubation with a solution containing 5 mM Cel9A and 5 mM CBM at 20°C.

Figure 2.13 Chimeric cellulases (CD-CBM) show higher degradation and swelling of amorphous cellulose films at T_{opt}. Cellulose volume profiles at T_{opt} are shown for each cellulase. Amorphous cellulose films were incubated with a) Cel9A or Cel9A-CBM, and b) Cel5A or Cel5A-CBM. Due to variance in film preparation the baseline (film without enzyme addition) for every cellulose film assay is shown. Arrows indicate the cellulose-buffer interface.
than at RT. We suggest that this may be due to the competing effects of bulk and surface activity. We note that there is a process and rate constant associated with each domain of the chimeras, and the relative rates of CD activity (R\textsubscript{CD}) and CBM-driven penetration of the chimera into the film (R\textsubscript{pen}) will affect the resulting film structure. When R\textsubscript{CD} / R\textsubscript{pen} \ll 1 the chimeras penetrate and equilibrate throughout the film before hydrolytic bond cleavage occurs. In that case bond cleavage occurs uniformly throughout the film, and swelling resulting from bond scission within the bulk will be maximal. The data for Cel5A-CBM at RT appears to correspond to this case. When R\textsubscript{CD} / R\textsubscript{pen} > 1 the chimeras hydrolyze bonds as they penetrate into the bulk, beginning at the surface. The result for

---

**Figure 2.14** Neutron reflectivity data from cellulose films before and after incubation with 5 mM Cel9A at 65°C (a), 5 mM Cel9A-CBM at 65°C (b), 5 mM Cel5A at 75°C (c) and 5 mM Cel5A-CBM at 75°C (d).
Cel5A-CBM at $T_{\text{opt}}$ seems to correspond to this case in that at $T_{\text{opt}}$ the film thickness for Cel5A-CBM is substantially reduced relative to the thickness of the as-prepared film. A similar trend with temperature was reported previously in NR data of Cel45A from *Humicola insolens* digesting amorphous cellulose films [91]. Native Cel45A contains a CBM. For Cel9A-CBM at $T_{\text{opt}}$ a distinct layer of enzyme is not detectable at the surface as is the case at RT. Greater penetration by Cel9A-CBM at the higher temperature could be explained by greater mobility of the cellulose chains. While there is greatly increased water content in the bulk of the film indicating penetration and digestion by Cel9A-CBM, the film thickness only decreased slightly from the as-prepared film and there is a very large gradient at the film-solution interface. In this case digestion in the bulk of the film by Cel9A-CBM appears to have led to sufficient film expansion to compensate for surface activity. A particularly large film expansion is also consistent with the large gradient at the film-solution interface.

Finally, since the binding domain is the same for the two chimeras, comparing the results reveals differences due to the two CDs. The first striking difference is the presence of a dense adsorbed protein layer at the surface for films incubated with Cel9A-CBM or the Cel9A-CBM$_{\text{KO}}$ at RT, but not for those incubated with Cel5A-CBM or Cel5A$_{\text{KO}}$-CBM. We suggest that this is due to the greater size of Cel9A (MW = 62 kDa) compared with that of Cel5A (MW = 40 kDa), which impedes Cel9A from entering the film. The second difference is the disparity in the thickness values at $T_{\text{opt}}$ along with the greater interfacial roughness for Cel9A-CBM. One possible explanation consistent with these data is that the activity of Cel9A results in greater swelling in the bulk of the film than does Cel5A. The third difference is that the proportional increase in mass released from the amorphous films for chimera compared with CD was greater for Cel5A than for Cel9A. At $T_{\text{opt}}$ for Cel9A and Cel5A the cellulose mass decreased by 11% and 4%, respectively, while for Cel9A-CBM and Cel5A-CBM the cellulose mass decreased by 32% and 35%, respectively (Table 1). So while three times as much mass was released by Cel9A-CBM as by Cel9A, roughly nine times as much mass was released by Cel5A-CBM as by Cel5A. This difference is due to the very low amount of mass released by Cel5A, while the chimeras released about the same amount of mass. The lower mass loss with Cel5A is likely due to a lower adsorption affinity for Cel5A than for Cel9A.

While the fold increase in cellulose mass released for Cel9A-CBM / Cel9A from NR is in good agreement with that obtained from the bulk saccharification assay on the IL-MCC substrate, the fold increase in cellulose mass released for Cel5A-CBM / Cel5A from NR is much greater than that released in the bulk saccharification assay on IL-MCC. This could be due to differences in the two assays. We note that IL-MCC is likely a mixture of amorphous and crystalline (cellulose II) cellulose, whereas the NR study involved films of amorphous cellulose with no detectable cellulose II. Another important difference in the assays is the fact that the bulk saccharification assays involved shaking/mixing whereas the NR study was performed in the absence of mixing.

### 2.3.5 – Proposed mechanism of CBM enhancement

Reese *et al.* first put forth a mechanism for cellulose degradation that involved a domain for cleavage or disruption of nonglycosidic linkages, in addition to the hydrolytic domain that cleaves the $\beta$-1,4-glucosidic bond [102]. The non-hydrolytic domain could
be CBMs, expansins, swollenins, or some other factor (for an extensive review, see [103]). For crystalline cellulose, the work of many research groups has led to a proposed mechanism in which the CBM is a swelling factor that helps to separate the glucan chains from the crystalline surface via the disruption of hydrogen bonds, resulting in layer-by-layer degradation [74, 82]. Nonetheless, CBMs that are specific for non-crystalline cellulose also have been shown to increase cellulase activity on amorphous cellulose (Figure 2.3, and references [45, 67]). The mechanism for this enhancement is largely unexplored, but is primarily attributed to the higher binding affinities that the CBMs confer to the enzymes [47, 67, 75, 104].

In the present study the wild type enzymes are primarily active at the amorphous cellulose film surface. However, for the chimeric fusions containing a CD and the CBM from T. fusca E3 we observe a swelling of the cellulose film and substantial changes to its bulk properties upon incubation. These results point to a mechanism of amorphous cellulose disruption by the CBM akin to the cellulose “swelling factor” proposed in the model by Reese et al. [102]. Taken together, the cellulose hydrolysis and NR data indicate that the addition of the CBM increases binding to cellulose and facilitates penetration into the bulk of the solid cellulose substrate (Figure 2.15). This enzyme penetration is driven by the interaction of the CBM with the cellulose chains. The CBM may disrupt hydrogen bonding between cellulose chains, or the increased affinity for cellulose provided by the CBM may overcome the energetic barrier to deform the cellulose chains as needed to allow the CD to penetrate into the film. As the enhanced activity requires physical tethering of the two domains, we suggest that the binding affinity of the CBM also still plays a role, providing the fused CD substantial time to access the newly available sites in the bulk of the cellulose substrate.

**Figure 2.15** Proposed model for chimeric cellulase degradation of amorphous cellulose substrates. Wild-type cellulases (without CBM) are limited to surface degradation, while physical linkage to CBM drives penetration into the bulk of the substrate. Hydrolysis within the bulk creates sites for water coupling and enhanced enzymatic access to additional substrate, resulting in swelling.

### 2.4 – Conclusions

The fusion of a CBM to the wild type cellulases Cel9A and Cel5A enhanced their activity as much as three fold on the insoluble lignocellulosic substrates MCC and IL-MCC. This activity enhancement can be explained as a change in mechanism in which the CBM increases binding to cellulose and dramatically enhances the penetration of the
cellulases into the bulk of the cellulose. We and others [45] also observe that CBM addition has varying synergistic effects depending on the CD chosen. Additional work is needed to determine the parameters required for a CBM to enhance access to the bulk cellulose, and to explore the effect of CBM choice and orientation on our proposed mechanism.

2.5 – Methods
2.5.1 – Strains and growth conditions
All cloning was carried out in DH10B and expression in BL21 (DE3) Star cells (Invitrogen). Cells were grown at 37°C in 2XYT media with 100 µg/ml kanamycin or 100 µg/ml carbenicillin as appropriate, unless otherwise noted.

2.5.2 – Cellulase genes and vector construction
Cel9A from Alicyclobacillus acidocaldarius (Cel9A) was initially obtained as a gift from K. Eckert (Humboldt-Universität zu Berlin, Institut für Biologie/Bakterienphysiologie, Germany) [92, 94]. Its open reading frame was amplified and cloned into the pENTR221 vector (Invitrogen) modified to contain a thrombin cleavage site and ten consecutive histidine residues on its N terminus (N-terminal 10X His-tag) for affinity purification with nickel columns. The creation of the cel5A gene from Thermotoga maritima (encoding Cel5A) was described previously [93]. Amplification and the cloning process of the CD was analogous to that of the Cel9A CD.

The Cel9A- and Cel5A-CBM chimeras were created by combining three distinct sequences: genes encoding the CD, a synthetically designed linker region, and the CBM. The CBM (UniProt Q60029), found in an exocellulase from Thermotoga fusca, was synthesized by GenScript Corporation. The linker was designed to be rich in proline and threonine, and is based on the sequence of a linker in endoglucanase N in Erwinia carotovora. Only the ends of this sequence were modified for overlapping PCR and efficient cloning. The DNA encoding the linker was created using PCR assembly of multiple oligonucleotides (IDT, integrated Technologies). (Oligonucleotide sequences: forward (5’-GCTAACCTGGCCGGTGGTGGATAGTCTCCGACTACCCCTACCACCACCCCGACCGAACCAGACT AAC-3’) and reverse (5’-ACCACCGGCTGGCCGGCCGTGGTGCCTGGCCCGCTTAGTCCGTTCCGTCGGG-3’). All three parts (Cel9A/Cel5A, CBM and linker) were amplified and then assembled by single overlap extension PCR. The resulting product was ligated into the modified pENTR221 vector as described above.

Chimeras also were constructed with 12 amino acids and 47 amino acids linkers, based on the sequence of the linker in CelE of Caldicellulosiruptor sp. Tok7B.1 and the endoglucanase/exoglucanase B from Caldocellum saccharolyticum, respectively. The linkers were altered slightly to include identical sequences at either end (5’-GCTAACCTGGCCGGTGGTGGTGGATAGTCTCCGACTACCCCTACCACCACCCCGACCGAACCAGACT AAC-3’) and reverse (5’-ACCACCGGCTGGCCGGCCGTGGTGCCTGGCCCGCTTAGTCCGTTCCGTCGGG-3’).

All three parts (Cel9A/Cel5A, CBM and linker) were amplified and then assembled by single overlap extension PCR. The resulting product was ligated into the modified pENTR221 vector as described above.

Chimeras also were constructed with 12 amino acids and 47 amino acids linkers, based on the sequence of the linker in CelE of Caldicellulosiruptor sp. Tok7B.1 and the endoglucanase/exoglucanase B from Caldocellum saccharolyticum, respectively. The linkers were altered slightly to include identical sequences at either end (5’-GCTAACCTGGCCGGTGGTGGTGGATAGTCTCCGACTACCCCTACCACCACCCCGACCGAACCAGACT AAC-3’) and reverse (5’-ACCACCGGCTGGCCGGCCGTGGTGCCTGGCCCGCTTAGTCCGTTCCGTCGGG-3’). All three parts (Cel9A/Cel5A, CBM and linker) were amplified and then assembled by single overlap extension PCR. The resulting product was ligated into the modified pENTR221 vector as described above.

All three parts (Cel9A/Cel5A, CBM and linker) were amplified and then assembled by single overlap extension PCR. The resulting product was ligated into the modified pENTR221 vector as described above.
added to the mix. The large linker (47 amino acids) was assembled via overlap PCR using the following four primers:

forward-1
(5'-GCTAACCTGGCCGTGGGTTACCGACTCCCAACCCCAACTCC-3')

reverse-1
(5'-GTCGGCGGTGGAAGTGATACCGGTTGAGTTGGGGGTGGTT-3')

forward-2
(5'-CGACTCCGACGCCGCGCATTACCAGCAACCCGACTCCCAACCCCTACT-3')

reverse-2
(5'-ACCACCGGTGCGCCTAGACACCCGGAGTTGGGTTG-3')

The fully assembled CD-linker-CBM chimeras were then constructed as described above.

Catalytic knockout versions of each enzyme were also constructed using site-directed mutagenesis to introduce the following changes: E515Q in Cel9A and E136Q in Cel5A; the corresponding amino acids were also changed in Cel9A-CBM and Cel5A-CBM chimeras. Site-directed mutagenesis was achieved by carrying out PCR of the entire plasmid using the following primers: Cel9A forward primer (5'-GGACAGCTACTCGACCAACCAAGTCGCCGTCTACTGGAATTC-3'); Cel9A reverse primer (5'-GAATTCCAGTAGACCGCGACTTGGTTGGTCGAGTAGCTGTCC-3'); Cel5A forward primer (5'-GTGGTTCTGAAATTCTGAACCAGCCGATGGGAACCTGAC-3'); and Cel5A reverse primer (5'-GTCAAGGTTCCATGCGGCTGGTTCAGAATTTCGAAAA-3'). All constructs were later passed into pDEST42 following standard Gateway cloning protocols (Invitrogen). The final constructs were verified by sequencing.

2.5.3 – Enzyme expression and purification

Cellulases were expressed in BL21 (DE3) Star cells in 2X YT media supplemented with 100 µg/ml of carbenicillin. Cells were grown at 37°C until reaching an OD600 of 0.7 at which time 0.5 mM IPTG was added to induce expression, and they were incubated with shaking at 18°C for 24 h. Cells were then pelleted down at 12,000 X g for 10 min at 4°C. The recombinant proteins were extracted from cell pellets using BugBuster (Novagen) according to the protocol with the following additions: 25 U benzonase/mL BugBuster (Novagen), 1 U r-lysozyme/mL BugBuster (Sigma), and 1X protein inhibitor cocktail V (EDTA-free) (Calbiochem). Cell debris was pelleted by centrifugation at 16,000 X g for 15 min at 4°C. The lysate was then incubated in a 50°C water bath for 15 min. Any denatured proteins were pelleted by centrifugation at 16,000 X g for 15 min at 4°C. The cellulases were affinity purified from the soluble fraction using gravity nickel columns (His GraviTrap, GE Healthcare) according to the manufacturer recommendations. Briefly, the columns were pre-equilibrated with 10 ml of washing buffer (20 mM sodium phosphate, 1 M NaCl, 20 mM imidazole, pH 7.4). The supernatant of the lysate described above was run twice through the equilibrated columns. Columns were washed with 10 ml of washing buffer. Enzymes were eluted by increasing the imidazole concentration to 500 mM. The eluted cellulases were buffer-exchanged using gravity desalting columns (PD-10 Desalting Columns, GE Healthcare) into a buffer consisting of 20 mM Tris, 150 mM NaCl, pH 7.4. Purity of the cellulases was determined by SDS-PAGE. Concentration of the enzymes was determined by bicinehonic acid (BCA) colorimetric assay (Pierce Thermo Scientific). The final protein concentrations were at least 1.3 mg/mL for each protein. The polydispersity of the
purified cellulases was measured by dynamic light scattering (DLS) on a DynaPro Plate Reader (Wyatt Technologies).

### 2.5.4 – Ionic liquid pretreatment

Ionic liquid (IL) pretreatment was carried out as previously described [105]. Briefly, the microcrystalline cellulose samples were treated with 1-ethyl-3-methylimidazolium acetate (Sigma-Aldrich, St Louis, MO, USA) at a loading of 3% (w/v) cellulose and heated at 160°C for three hours. The pretreated material was washed with deionized hot water (80°C). Samples were centrifuged at 10,000 X g for 20–25 min and were washed in this way five times. This has been shown to be sufficient for removal of ionic liquids [105]. Samples were freeze-dried in a lyophilizer (Labconco 12L Freeze Dryer).

### 2.5.5 – Enzymatic hydrolysis assays

Enzymatic hydrolysis assays were carried out as described before [98] with slight changes to the protocol. To determine the optimal temperatures and pHs of the cellulases, enzyme was added at 25 nM to a solution of 50 mM citric buffer with 1% carboxymethyl cellulose (CMC) in 96-well PCR plates to a total volume of 200 µl. The solution pH was varied incrementally from 4 to 8 (Figure 2.1) and the samples were incubated in a thermocycler for 30 minutes at temperatures varying from 10 to 95°C. After incubation, the soluble sugar content of all samples was quantified by colorimetric dinitrosalicylic acid (DNS) assay (see below) [106].

For enzymatic hydrolysis assays on insoluble substrates, enzyme was added at 200 nM to citric buffer containing 20 mg/ml substrate to a total volume of 500 µl in micro-tubes with caps. Cellulases were tested at the optimal pH: 4.8 for Cel5A and 5.5 for Cel9A. All samples were incubated for 24 hrs at 50°C and 1400 rpm in a bench-top thermomixer (Eppendorf AG 22331 Hamburg). There are some reports of shear stress inactivation, particularly of exoglucanases [107–111]. We did not observe any significant inactivation at this speed compared to lower speeds for the endoglucanases in this study. After incubation, all samples were spun down and the supernatant collected to determine the concentration of soluble sugars using the DNS assay described below.

### 2.5.6 – Quantification of soluble sugar concentrations

Sugar concentration of the samples was determined by DNS assay as previously described [106]. Briefly, 60 µl of 2X DNS reagent (1 g DNS, 30 g KNa tartrate, 20 ml 2 N NaOH in 100 ml total volume) were mixed with 60 µl of the samples supernatant. Samples were well mixed and incubated at 95°C for five minutes. Samples were allowed to cool down to room temperature and their absorbance read at a wavelength of 540 nm (Molecular Devices, SpectraMax M2). Concentrations were calculated by comparison of absorbance to a cellobiose standard curve.

### 2.5.7 – Preparation of cellulose films

The preparation of uniform regenerated cellulose films sufficiently smooth for NR measurements has been previously reported [83]. In summary, regenerated cellulose films for NR were prepared by spincoating and conversion of precursor films of trimethylsilylcellulose (TMSC) on polished silicon wafers (diameter = 75 mm, thickness
Preparation of TMSC has been described previously \cite{112, 113}. The wafers were cleaned in a solution of sulfuric acid/30% by volume hydrogen peroxide, 7:3 by volume (Piranha solution), followed by UV/ozone treatment for 20 min.

TMSC was spin-coated onto the cleaned silicon substrates with a spinning speed of 4000 rpm from solutions of 10 mg/ml or 12 mg/ml in toluene. The TMSC films were converted into cellulose by exposing them to vapors of 0.5 N HCl solution for 15 minutes in an enclosed container. This practice resulted in complete conversion and ultrasmooth films, as reported previously \cite{83}. Film thicknesses ranged from 240 to 310 Å for the different solution concentrations as determined from X-ray reflectivity and NR.

\subsection*{2.5.8 – Neutron Reflectivity}

NR studies were performed on the SPEAR (Lujan Center/LANSCE), Liquids (SNS/ORNL), and NG1 (NIST) reflectometers. The SPEAR and Liquids reflectometers operate in the time-of-flight mode where a band of wavelengths impinge onto the film-buffer interface by passing through the silicon wafer and are resolved at the detector based on their time-of-flight. Data collected from several incident angles were merged together. The measurements on NG1 were performed using a wavelength of 0.475 nm and varying angles of incidence. The data are plotted as the ratio of reflected to incident intensity as a function of momentum transfer \( q_z = \frac{4\pi}{\lambda} \sin \theta \), where \( \theta \) is the angle of incidence with respect to the plane of the film and \( \lambda \) is the wavelength \cite{114}. The precise form of this curve is determined by the profile of the in-plane averaged scattering length density (SLD) normal to the surface. The SLD is determined by the atomic composition and the density \cite{115}. The spacing between minima or maxima (fringes) on the \( q_z \) scale is related to the film thickness. The magnitude of the fringes is related to the volume fraction of water and cellulose within the film. Progressive dampening of fringes with increasing \( q_z \) indicates roughening or broadening of the solution film interface. The NR data were analyzed using the Ga_refl program based on the optical matrix method. Ga_refl is available at www.ncnr.nist.gov. Analyses were performed with free-form models involving a small number of slabs. Fitting reflectivity data results in defining a family of SLD curves that are consistent with the data. The uncertainty in the fitted profiles was determined by a Monte Carlo resampling procedure in which a large number (1000) of statistically independent sets of reflectivity data were created from the original data set and the error bars from the counting statistics. The result is a range of values for each fit parameter that is consistent with the statistics of the original data. This method has been reported in detail elsewhere \cite{116}. The SLD profile bands for Cel5C-CBM at RT are given in the supporting information as a representative example (Figure 2.16).

The cellulose volume fractions (Figures 2.6, 2.9, 2.13, 2.11, and 2.12) were determined from the SLDs of the buffer solution, the swollen cellulose film, and pure cellulose using the following relation:

\[
(SLD)_{\text{meas}} = \phi_{\text{cellulose}} (SLD)_{\text{cellulose}} + (1-\phi_{\text{cellulose}})(SLD)_{\text{buffer}}
\]

where \((SLD)_{\text{meas}}\) is the measured SLD for the swollen film, \( \phi_{\text{cellulose}} \) is the volume fraction of cellulose, \((SLD)_{\text{cellulose}} = 1.67 \times 10^{-6} \ \text{Å}^{-2}\) is the SLD of pure cellulose (C<sub>6</sub>H<sub>10</sub>O<sub>5</sub>)[83] and \((SLD)_{\text{buffer}} = -0.54 \times 10^{-6} \ \text{Å}^{-2}\) is the SLD for aqueous buffer. In using this equation,
additivity of volumes is assumed. The amount of cellulose per unit area within each film was obtained by integrating the cellulose volume fraction profiles.

Before each measurement, the regenerated cellulose film was allowed to equilibrate with sodium acetate buffer for 20 min, after which several scans were collected. After equilibration of the film in buffer, a 5 µM protein solution was injected into the measurement cell and incubated with the films in absence of flow until little change in NR was observed on a timescale of several hours. The incubation time therefore varied somewhat among the samples. The NR studies were performed at room temperature and also at the optimal temperature for each CD (65°C and 75°C for Cel9A and Cel5A, respectively). The sample cell was heated by circulating a heating fluid through copper blocks placed underneath and on top of the sample cell. The sample cell and copper blocks were enclosed in a Styrofoam box containing thin aluminum foil windows. The conceived of the project, assisted in the design and interpretation of experiments, and helped to draft and edit the manuscript. All authors read and approved the final manuscript.

2.6 – Acknowledgements

This work conducted by the Joint BioEnergy Institute was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231

Figure 2.16 Plot showing volume fraction profiles with the uncertainty limits indicated as multicolored bands for Cel5A-CBM (yellow/red/black) and Cel5A (cyan/blue) at 20°C. The uncertainty bands shown in this plot are representative of the uncertainty in the SLD plots for each NR experiment performed in this study.
Chapter 3 – A comparison of the effect of different carbohydrate-binding modules on the interaction between an endoglucanase and insoluble cellulose

3.1 – Abstract:
Carbohydrate binding modules (CBMs) are known to increase the activity of ligno-cellulolytic enzymes through different mechanisms including: 1) enhancing enzyme adsorption to the substrate; 2) enabling alignment of cellulose fibrils for better enzyme docking; and 3) modifying substrates surfaces to facilitate enzymatic hydrolysis. However, much is still unknown, particularly in relation to how the CBM affects enzyme-substrate interactions and activity on solid substrates. To address this gap, we have applied our previously established neutron reflectometry (NR) assay to examine the interactions of a cellulase containing or lacking CBMs with cellulose model films. We chose to study the thermophilic Cel5A cellulose from \textit{T. maritima}, which has a catalytic domain (CD) but not a native CBM. The wild type enzyme was compared to three different chimeric versions containing peptide linkages to individual CBMs from families 1, 2, and 3. The addition of each CBM increases the activity of the CD by approximately the same magnitude, but results indicate that each CBM has a unique effect on the interaction of the enzyme with a cellulose film.

3.2 – Background
Saccharification is the process of generating glucose from cellulosic substrates, and can be orchestrated by the synergistic action of a group of enzymes called cellulases (See chapter 1 for a deeper description of saccharification). This process is of extreme interest to the commodity and specialty chemical industries, because glucose can be microbially and chemically converted into profitable bioproducts. Enhancement of cellulases can be reflected in the reduced cost of bioproducts. As discussed in previous chapters, cellulases have been engineered to optimize enzymatic efficiency through different approaches such as increased thermostability and addition of carbohydrate binding modules (CBMs) (See chapter 2 and \cite{45, 53}).

CBMs are classified according to sequence similarity by the CAZy (www.cazy.org) database into 67 families (see section 1.4.1.1.1 for more detail). Some CBMs have been experimentally characterized for substrate binding preferences including type of carbohydrate substrate, chain length, and dominant binding residues \cite{31, 62, 63, 67}. Such information about the binding specificity and substrate preference as it relates to sequence results in continuous updating of the CAZy database. Families of CBMs can exhibit some commonalities such as the type of organism in which they are found, size, tertiary structure and predicted carbohydrate affinity. Family one CBMs (CBM1) are almost exclusively found in fungi. Comprising approximately 40 amino acids, the CBMs of this family are the smallest of all CBMs. They have a cysteine-knot tertiary structure \cite{48}. In contrast, CBMs from families 2, 3, 17 and 28 usually share a β-sandwich fold and have approximately 200 amino acids. Members within all the mentioned families have exhibited affinity towards cellulose. Furthermore, cellulose-binding members from families 4, 17 and 28 have so far been shown to have affinity exclusively towards amorphous cellulose \cite{62, 63, 67, 117}. 
In the early 50’s Reese et al. first proposed a mechanism for cellulose degradation that involved a domain for cleavage or disruption of nonglycosidic linkages by an enzyme he called C1, in addition to the hydrolytic domain that cleaves the β-1,4-glucosidic bond [102]. Din et al. first called that C1 disruptive agent a CBM. Their work reported a disruptive function by a CBMs type A from family 2 (CBM2a) located at the N-terminal of an endoglucanase A from Cellulomonas fimi (See chapter 1 for further discussion) [72]. These interesting results demonstrated that the CBM2a was capable of roughening and releasing small particles from cotton fibers. Exposure of the same kind of fibers to the catalytic domain (CD) alone seemed to polish the surfaces of the cotton fibers. In subsequent work, Teeri et al. [73] proposed that CBMs were able to bind to cellulose and penetrate into those areas that showed network discontinuity. They also suggested that further penetration of the CBM into the cellulose substrates could release the ends of the cellulose fibers, which remained covalently bound to the rest of the fiber, but resulted in roughening of the cellulose. Another interesting finding of the CBM disruptive capabilities is described by Lee et al. [74]. With the help of atomic force microscopy, Lee et al. observed slightly elongated holes that were left through the surface of cotton fibers after being exposed to an inactive CBH from T. reseei. These holes were not observed when incubating the fibers with an inactive enzyme with no CBM. Therefore, the elongated holes were interpreted to be a result of CBM penetration in the cotton fibers. In Chapter 2, we used neutron reflectivity (NR) to examine the impact of the CBM on the interaction of its corresponding enzyme and the cellulosic substrate, and found that a CBM from family 2 substantially enhanced the ability of the enzyme to penetrate into the bulk of an amorphous cellulose film.

This study extends that work to compare the catalytic activity as well as substrate surface behavior between the fusions of the same CD to CBMs from families 1, 2, and 3. The addition of these three CBMs to the CD enhanced cellulase activity two and a half times on ILSG as measured by reducing ends released. Cellulose profiles acquired with neutron reflectivity (NR) suggest that the addition of each CBM to the CD not only increases its activity but also modifies its interaction with the substrate in different manners.

3.3 – Results and discussion

3.3.1 – Construction and characterization of chimeric cellulases

We have identified an endocellulase, Cel5A from Thermotoga maritima, which can operate at high temperatures and in the presence of ionic liquids (1-ethyl-3-methylimidazolium acetate) (chapter 4). Cel5A consists of a single CD and releases cellobiose as its primary hydrolysis product [93]. We therefore proposed to enhance the activity of Cel5A by adding a CBM to the CD (which originally has no CBM). Primary selection of these CBMs was based on two main parameters: 1) the CBMs had to have at least a predicted binding affinity towards cellulose and 2) had to be thermotolerant. CBMs meeting these requirements were expected to increase activity of the Cel5A CD if tethered to it. Our collaborator at the Joint BioEnergy Institute generated a library of 68 chimeric constructs comprising combinations of the Cel5A CD with CBMs belonging to families 1, 2, 3, 4, 17, and 28. Very few of these CD-CBMx constructs exhibited an increase in activity over the CD alone (data not shown) against switchgrass pretreated with 1-ethyl-3-methyl-imidazolium acetate (ILSG). However certain members of CBM
families one (CBM1), two (CBM2) and three (CBM3) were among the few chimeric cellulases that approximately doubled the activity of the CD on this substrate (Table 3.1).

To confirm the results of the library screen, the Cel5A CD was fused to three CBMs that conferred enhanced activity; we examined one CBM in detail from each family identified in the screen (CBM1, CBM2, and CBM3). The CBM1 (Genbank BAA74956.1) is originally from the fungal endoglucanase Egl3 from *Humicola grisea var. thermoidea* [118]. The CBM2 (Genbank ABH87623.1) is a subunit of the bacterial endoglucanase GuxA from *Acidothermus cellulolyticus* [119]. The CBM3 (Genbank ABN51281.1) is originally from the bacterial endoglucanase Cel9I from *Clostridium thermocellum*. All three genes were synthesized by GenScript Corporation [120]. The chimeras were constructed in the format such that the CBM is at the C-terminus, with a 31 amino acid linker, denoted as “CD-CBM”. The enzymes, chimeras, and knockouts were overexpressed in *Escherichia coli* and purified to >80% purity.

Catalytic activity of the chimeric cellulases on carboxymethylcellulose closely matched that of the corresponding wild type enzyme (data not shown). Optimal temperatures ($T_{opt}$) and pHs were determined to be 75°C and pH 4.8 for each chimera [93].

3.3.2 – Enzymatic hydrolysis of insoluble substrates

Enzymatic activities were measured via DNS assay on microcrystalline cellulose (MCC, Avicel pH-101, Sigma) and ionic-liquid pretreated switchgrass (ILSG). The MCC was used to represent a model of highly organized crystalline cellulose. The ILSG was used to test the enzymes on a “real world” cellulosic substrate. Ionic liquid pretreatment is a promising method for biomass preparation that causes the structure of the cellulose to change from mostly crystalline to a mixture of cellulose II and amorphous cellulose [22], thereby enhancing downstream enzymatic hydrolysis [22, 100]. Pretreatment of switchgrass with ionic liquid can produce biomass with approximately 50% cellulose (a mixture of amorphous and type II allomorph). The other 50% is formed by variable residues of hemicellulose (~ 6%), lignin (10 - 15 %), and ash (remaining fraction). Addition of CBMs 1, 2 and 3 to the Cel5A CD increased cellulase activity by two to three-fold from ILSG, but on MCC there was little difference between the Cel5A CD and the rest of the chimeras (Table 3.1). The results of these experiments suggest that the CBMs are the factor responsible for enhancing the cellulase activity of Cel5A on ILSG.

<table>
<thead>
<tr>
<th>Cellulase Construct</th>
<th>umol of celllobiose equivalent</th>
<th>nmol of enzyme</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MCC</td>
<td>ILSG</td>
</tr>
<tr>
<td>Cel5A</td>
<td>0.182 ± 0.04</td>
<td>2.224 ± 0.13</td>
</tr>
<tr>
<td>Cel5A-CBM1</td>
<td>0.183 ± 0.09</td>
<td>4.311 ± 0.07</td>
</tr>
<tr>
<td>Cel5A-CBM2</td>
<td>0.231 ± 0.11</td>
<td>5.270 ± 0.26</td>
</tr>
<tr>
<td>Cel5A-CBM3</td>
<td>0.216 ± 0.13</td>
<td>4.225 ± 0.19</td>
</tr>
</tbody>
</table>

*Table 3.1* Comparison of activity of Cel5A and Cel5A chimeric cellulases on two insoluble substrates: microcrystalline cellulose (MCC, Avicel) and ionic-pretreated switchgrass (ILSG)
This result is in agreement with the work of the previous chapter, as well as the results of Kim et al., who also found an increase in activity upon fusion of a CBM to certain CDs [45].

3.3.3 – Amorphous Cellulose Film Characterization

We hypothesized that the mechanism by which the CBM was enhancing the activity of the CD on insoluble substrates was different among all three CBMs. To explore this hypothesis, we then applied NR to examine whether CBMs are capable of dictating how a single CD interacts with same type of substrate. Specifically, we used NR to analyze changes in cellulose thin films upon incubation with Cel5A and the Cel5A chimeras. The cellulose model films were regenerated from trimethylsilyl cellulose (TMSC) (See methods). This method resulted in thin cellulose films that were largely amorphous [91, 101]. This allows for comparison between our model film and ILSG used for the hydrolysis assays. The films swelled in aqueous buffer (with no enzyme added) by a factor of ~ 1.9 - 2.2. Thickness of the cellulose films was measured to be 250 Å - 270 Å, and the density of the dried films varied from 1.25 to 1.35 g/cm³. NR data revealed that the films were highly smooth, and that swelling of the films was uniform except for some variation at the film-substrate interface.

3.3.4 – NR analysis of films incubated with cellulases
3.3.4.1 Cel5A and Cel5A-CBM Chimeras

Each plot reported in this paper comprises initial and final cellulose volume fraction profiles, in which the final profile was determined when little or no change in the NR data was observed on the time scale of several hours. All the measurements shown here were performed at 75°C at a concentration of 200 nM. Experiments at T_{opt} result in greater changes in the cellulose films compared to experiments performed at room temperature even at much higher concentrations (see chapter two). The three chimeric cellulases incubated against cellulose films at 75°C showed distinctly different reflectivity and cellulose volume fraction profiles. This demonstrates that the type of CBM affects the spatial distribution of enzyme throughout the film during the digestion process.

All cellulose volume fraction profiles for the chimeric cellulases at 0.2 µM and 75°C indicate a greater loss of cellulose compared to that for Cel5A CD alone at 5 µM and 75°C (see chapter 2 and Figures 3.1 – 3.4). The cellulose volume fraction profile for Cel5A-CBM3 (figure 3.2) indicates a decrease in cellulose of approximately 13%, as compared to 5% for the CD alone at 25-fold higher concentration (see chapter 2). This is further confirmation that the linking of CBM3 to Cel5A CD increased its activity against this amorphous cellulose substrate. The cellulose mass losses for the other two chimeric cellulases in this study are greater than that for Cel5A-CBM3. The cellulose volume fraction profile for Cel5A-CBM2 corresponds to a loss of approximately 17% of the film (Figure 3.3), and for Cel5A-CBM1 approximately 29% of the cellulose was released from the film (Figure 3.4). This demonstrates that Cel5A-CBM1 has the highest level of activity on the amorphous model film.

In contrast to the Cel5A CD reflectivity curve (Figure 3.1), for all the chimeric cellulases (Figures 3.2 – 3.4) the reflectivity curves show substantially dampened fringes. Dampening of the fringes can result from either increased roughening of the film-solution
Figure 3.1 Neutron reflectivity data (a) and normalized cellulose volume fraction profiles (b) with an interpretation of the interaction of Cel5A with the cellulose film at 75°C. This figure was modified from Reyes-Ortiz et al. 2013 (Chapter 2).
Figure 3.2 Neutron reflectivity data (a) and normalized cellulose volume fraction profiles (b) with an interpretation of the interaction of Cel5A-CBM3 with the cellulose film at 75°C.
Figure 3.3 Neutron reflectivity data (a) and normalized cellulose volume fraction profiles (b) with an interpretation of the interaction of Cel5A-CBM2 with the cellulose film at 75°C.
Figure 3.4 Neutron reflectivity data (a) and normalized cellulose volume fraction profiles (b) with an interpretation of the interaction of Cel5A-CBM1 with the cellulose film at 75°C.
interface or from loss of contrast between the film and the solution. The latter effect results from an increase in the water content within the bulk of the film. However, the NR data distinguish between these two cases. Roughening of the film-solution interface results in stronger dampening of the fringes at higher \( q_z \) values, whereas loss of contrast from increased water content in the bulk of the film results in dampening of the fringes over the entire range of \( q_z \). This will be discussed further below.

Another major qualitative difference between the Cel5A reflectivity profile and the reflectivity profiles belonging to the chimeric cellulases is the shifting of the fringes to lower or higher \( q_z \) for the chimeras (Figures 3.2 - 3.4). These shifts indicate increased or decreased film thickness, respectively. While the figures show only the final states, for each chimera scans were collected as a function of time prior to achieving the final state. This gives a more complete picture of the differences in digestion resulting from the CBMs. For Cel5A-CBM1 a continuous increase in thickness of the cellulose film was observed over the entire time course. The film thickness prior to enzyme addition was approximately 550 Å, as compared to approximately 600 Å after enzymatic exposure (Figure 3.4). Correspondingly, the water content within the film also increased.

Perhaps the most interesting case is that of Cel5A-CBM3 (Figure 3.2). In this case the fringes initially shifted to higher \( q_z \), indicating digestion at the surface. However, in subsequent scans the fringes shifted substantially to lower \( q_z \) indicating that at this point the enzyme penetrated and digested within the bulk of the film, causing the film to swell. The extent of swelling was sufficient to more than compensate for the initial decrease in film thickness, such that in the final state the film thickness is slightly greater than that of the starting film. Another distinctive feature for Cel5A-CBM3 is that breadth of the film-solution interface did not increase. Therefore these results show that the CBMs result in distinctly different spatial distribution of enzyme throughout the films.

In contrast, for Cel5A-CBM2 (Figure 3.3) the fringes shifted continuously to higher \( q_z \) indicating a reduction of the film thickness. This indicates that the enzyme digested the film from the surface. In this case, the fringes are strongly damped at higher \( q_z \) values indicating a large roughening of the film-solution interface. The NR data are analyzed using the Ga_refl program based on the optical matrix method (www.ncnr.nist.gov). These analyses were performed with free-form models involving the smallest possible number of slabs. New slabs are added as needed defined in order to fit the data, but the number of slabs can be kept as low as one, if the data fit is consistent. The uncertainty in the fitted profiles is determined by a Monte Carlo resampling procedure in which a large number (1000) of statistically independent sets of reflectivity data are created from the original data set and the error bars from the counting statistics. The result is a range of values for each fit parameter that is consistent with the statistics of the original data. This method has been reported in detail elsewhere [116]. Unlike all other data analyzed for this study, in order to fit the data for Cel5A-CBM2, the cellulose film has to be divided into two slabs. This suggests that the cellulase penetrates only a short distance into the cellulose film.

3.3.5 – Proposed mechanism of CBM enhancement

To understand the disruptive capabilities of CBMs we tethered three different CBMs to the CD of Cel5A. Two of these chimeric cellulases cause swelling of
the cellulose film upon incubation. Like our previous results with the CBM2a from Thermotoga fuscum, these two CBMs appear to be the cellulose “swelling factor” proposed in the model by Reese et al. [102]. Our results indicate that the addition of the CBMs can increase cellulosytic activity by facilitating access of the cellulase to inner parts of the cellulose film, in addition to simply increasing the binding affinity.

In the previous case of Cel5A-CBM2a (see chapter 2) this facilitated access was interpreted as a disruptive and non-catalytic CBM action that allowed for penetration of the chimeric cellulase into the film. The term “penetration” was then used to describe the increased access to cellulose sites at all points within the film. Results from the Cel5A-CBM1 and Cel5A-CBM3 (Figures 3.2, 3.4) pointed toward the same direction due to the increase in activity and almost immediate swelling of the film. Figure 3.1 shows that Cel5A CD is able to reach into the film as well. However, no swelling and less degradation is observed with the wild type enzyme. This suggests that the chimeric cellulases must be accessing alternative sites throughout the film. We conclude that the differences observed between the cellulose profiles of the CD and those two CD-CBM chimeras are due to the disruptive activity of the CBM1 and CBM3.

Cel5A-CBM2 (Figure 3.3) seems also to increase activity of the CD, but the film incubated with Cel5A-CBM2 exhibited unique reflectivity profiles. As mentioned before, the reflectivity data collected for this CBM had to be divided into several slabs for its best fit. The best fit of the Cel5A-CBM2 illustrates two abrupt changes in cellulose fraction values along the thickness of the film (Figure 3.3b). This profile exhibits a decrease in cellulose fraction through the whole film. However, there is another big decrease of cellulose at approximately 300 Å away from the silicon oxide wafer. Two different actions may provide an explanation for this profile: The first explanation is that this CBM allows for additional penetration of the chimeric cellulase into the upper layers of the film. This specific depth might depend on the characteristic of the cellulose film, which cannot be clarified by NR techniques. The second explanation is that a higher affinity of the CBM for amorphous cellulose prevents the cellulase from reaching deeper points of the film. Results with increased enzyme concentration suggest that this is a plausible explanation. When the enzyme concentration was increased from 0.2 to 0.4 µM, the Cel5A-CBM2 also showed digestion within the bulk of the film. This suggests that once the binding sites in the upper layers of the film are saturated, the chimeric cellulase may enter further into the film and digest the cellulose. It should be noted that this is a different behavior than that observed with the profiles for the Cel5A-CBM2a studied in chapter 2, which exhibited degradation throughout the film. With this in mind, we can only state that the CBMs are able to disrupt the film in qualitatively different ways.

3.4 – Conclusions
Addition of CBMs from families one, two and three to the catalytic domain of Cel5A enhanced its activity as much as 2.5 fold on ionic liquid pretreated switch grass (Table 3.1, Figure 3.2 - 3.4). The activity enhancement is suggestive of a change in the interaction of the chimeric cellulases and the cellulose substrates compared to the interactions of wild type Cel5A alone with the same substrate. The difference in interaction can be attributed to the disruptive capabilities of the CBMs, which is associated with penetration of the cellulase into the cellulose film. However, each CBM
seems to do this in a different qualitative manner. These cases have to be further explored in order to gain a complete understanding of how the CBM facilitates digestion of insoluble cellulose substrates.

3.5 – Methods
3.5.1 – Strains and growth conditions
All cloning was carried out in DH10B and expression in BL21 (DE3) Star cells (Invitrogen). Cells were grown at 37°C in 2XYT media with 100 µg/ml kanamycin or 100 µg/ml carbenicillin as appropriate, unless otherwise noted.

<table>
<thead>
<tr>
<th>CBM Family</th>
<th>GenBank Number</th>
<th>Organism</th>
<th>Origin</th>
<th>Cellulase</th>
<th>Amino Acid Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BAA74956.1</td>
<td>Humicola grisea var. thermoidea</td>
<td>Fungal</td>
<td>endo-1,4-glucanase (Egl3)</td>
<td>CTAERWAQCG GNQWSGCTTC VAGSTCTKIND WYHQCL</td>
</tr>
<tr>
<td>2</td>
<td>ABH87623.1</td>
<td>Acidothermus cellulolyticus</td>
<td>Bacterial</td>
<td>endoglucanase (GuxA)</td>
<td>PSSSGVACRAT YVVNSDWGSGF TATVTVNTGSR ATNGWTVAVSF GGNQTVTNYW NTALTQSGASV TATNLSYNNVQI PQQSTTFGFGN SYSGTNAAPTLS CTAS</td>
</tr>
<tr>
<td>3</td>
<td>ABN51281.1</td>
<td>Clostridium thermocellum</td>
<td>Bacterial</td>
<td>endo-β-1,4-glucanase I (Cel9I)</td>
<td>VPEDEIFVEAGV NASGNNFIEIKAI VNNKSGWPARV CENLSFRFYFINIE EIVNAGKSASDL QVSSSYQNGAK LSDKHYKDNIVY YVEVDSGTKIY PGGQSAYKKEV QFRISAPEGTVF NPENDYSYQGL SAGTVKSEYIP VYDAQVLVFGREPG</td>
</tr>
</tbody>
</table>

Table 3.2 Simple description of the three CBMs used in this study.
3.5.2 – Cellulase genes and vector construction

The creation of the cel5A gene from Thermotoga maritima (encoding Cel5A) was described previously [93]. Its open reading frame was amplified and cloned into the pENTR221 vector (Invitrogen) modified to contain a thrombin cleavage site and ten consecutive histidine residues on its N terminus (N-terminal 10X His-tag) for affinity purification with nickel columns. The Cel5A-CBM chimeras were created by combining genes encoding the CD, a synthetically designed linker region, and the different CBMs. The CBM1 (Genbank BAA74956.1) is original from a fungal endoglucanase (Egl3) Humicola grisea var. thermoidea [118]. CBM2 (Genbank ABH87623.1) is original from a bacterial endoglucanase (GuxA) from Acidothermus cellulolyticus [119]. CBM3 (Genbank ABN51281.1) is original from a bacterial endoglucanase (Cel9I) from Clostridium thermocellum [120]. All three genes were synthesized by GenScript Corporation (see table 3.2 for more details). The linker was designed to be rich in proline and threonine, and is based on the sequence of a linker in endoglucanase N in Erwinia carotovora. Only the ends of this sequence were modified for overlapping PCR and efficient cloning. The DNA encoding the linker was created using PCR assembly of multiple oligonucleotides (IDT, integrated Technologies). (Oligonucleotide sequences: forward (5'- GCTAACCTGGGCGGTGGTGATACTCCGACCTACCCCTACCACCCCCGACCACCGAACC-3') and reverse (5’ACCACCGGTGGCCGCCGTGGTCGGTCCGGGTAGTCGGTGTCGGTCA-3’). All three parts (Cel5A, CBM and linker) were amplified and then assembled by single overlap extension PCR. The resulting product was ligated into the modified pENTR221 vector as described above. The fully assembled CD-linker-CBM chimeras were then constructed as described above.

3.5.3 – Enzyme expression and purification

Cellulases were expressed in BL21 (DE3) Star cells in 2X YT media supplemented with 100 µg/ml of carbenicillin. Cells were grown at 37°C until reaching an OD600 of 0.7 at which time 0.5 mM IPTG was added to induce expression, and they were incubated with shaking at 18°C for 24 h. Cells were then pelleted down at 12,000 X g for 10 min at 4°C. The recombinant proteins were extracted from cell pellets using BugBuster (Novagen) according to the protocol with the following additions: 25 U benzonase/mL BugBuster (Novagen), 1 U r-lysozyme/mL BugBuster (Sigma), and 1X protein inhibitor cocktail V (EDTA-free) (Calbiochem). Cell debris was pelleted by centrifugation at 16,000 X g for 15 min at 4°C. The lysate was then incubated in a 50°C water bath for 15 min. Any denatured proteins were pelleted by centrifugation at 16,000 X g for 15 min at 4°C. The cellulases were affinity purified from the soluble fraction using gravity nickel columns (His GraviTrap, GE Healthcare) according to the manufacturer recommendations. Briefly, the columns were pre-equilibrated with 10 ml of washing buffer (20 mM sodium phosphate, 1 M NaCl, 20 mM imidazole, pH 7.4). The supernatant of the lysate described above was run twice through the equilibrated columns. Columns were washed with 10 ml of washing buffer. Enzymes were eluted by increasing the imidazole concentration to 500 mM. The eluted cellulases were buffer-exchanged using gravity desalting columns (PD-10 Desalting Columns, GE Healthcare) into a buffer consisting of 20 mM Tris, 150 mM NaCl, pH 7.4. Purity of the cellulases
was determined by SDS-PAGE. Concentration of the enzymes was determined by bicinchonic acid (BCA) colorimetric assay (Pierce Thermo Scientific). The final protein concentrations were at least 1.3 mg/mL for each protein. The polydispersity of the purified cellulases was measured by dynamic light scattering (DLS) on a DynaPro Plate Reader (Wyatt Technologies).

### 3.5.4 – Preparation of cellulose films

The preparation of uniform regenerated cellulose films sufficiently smooth for NR measurements has been previously reported [83]. In summary, regenerated cellulose films for NR were prepared by spincoating and conversion of precursor films of trimethylsilylcellulose (TMSC) on polished silicon wafers (diameter = 75 mm, thickness = 5 mm). Preparation of TMSC has been described previously [112, 113]. The wafers were cleaned in a solution of sulfuric acid/30% by volume hydrogen peroxide, 7:3 by volume (Piranha solution), followed by UV/ozone treatment for 20 min.

TMSC was spin-coated onto the cleaned silicon substrates with a spinning speed of 4000 rpm from solutions of 10 mg/ml or 12 mg/ml in toluene. The TMSC films were converted into cellulose by exposing them to vapors of 0.5 N HCl solution for 15 minutes in an enclosed container. This practice resulted in complete conversion and ultrasmooth films, as reported previously [83]. Film thicknesses ranged from 240 to 310 Å for the different solution concentrations as determined from X-ray reflectivity and NR.

### 3.5.5 – Neutron Reflectivity

NR studies were performed on the SPEAR (Lujan Center/LANSCE), Liquids (SNS/ORNL), and NG1 (NIST) reflectometers. The SPEAR and Liquids reflectometers operate in the time-of-flight mode where a band of wavelengths impinge onto the film-buffer interface by passing through the silicon wafer and are resolved at the detector based on their time-of-flight. Data collected from several incident angles were merged together. The measurements on NG1 were performed using a wavelength of 0.475 nm and varying angles of incidence. The data are plotted as the ratio of reflected to incident intensity as a function of momentum transfer $q_z=(4\pi/\lambda)\sin\theta$, where $\theta$ is the angle of incidence with respect to the plane of the film and $\lambda$ is the wavelength [114]. The precise form of this curve is determined by the profile of the in-plane averaged scattering length density (SLD) normal to the surface. The SLD is determined by the atomic composition and the density [115]. The spacing between minima or maxima (fringes) on the $q_z$ scale is related to the film thickness. The magnitude of the fringes is related to the volume fraction of water and cellulose within the film. Progressive dampening of fringes with increasing $q_z$ indicates roughening or broadening of the solution film interface. The NR data were analyzed using the Ga_refl program based on the optical matrix method. Ga_refl is available at www.ncnr.nist.gov. Analyses were performed with free-form models involving a small number of slabs. Fitting reflectivity data results in defining a family of SLD curves that are consistent with the data. The uncertainty in the fitted profiles was determined by a Monte Carlo resampling procedure in which a large number (1000) of statistically independent sets of reflectivity data were created from the original data set and the error bars from the counting statistics. The result is a range of values for each fit parameter that is consistent with the statistics of the original data. This method has been reported in detail elsewhere [116].
The cellulose volume fractions (Figures 3.1 - 3.4) were determined from the SLDs of the buffer solution, the swollen cellulose film, and pure cellulose using the following relation:

\[
(SLD)_{\text{meas}} = \phi_{\text{cellulose}} (SLD)_{\text{cellulose}} + (1 - \phi_{\text{cellulose}}) (SLD)_{\text{buffer}}
\]  

(1)

where \((SLD)_{\text{meas}}\) is the measured SLD for the swollen film, \(\phi_{\text{cellulose}}\) is the volume fraction of cellulose, \((SLD)_{\text{cellulose}} = 1.67 \times 10^{-6} \text{ Å}^{-2}\) is the SLD of pure cellulose \((C_6H_{10}O_5)\), \([83]\) and \((SLD)_{\text{buffer}} = -0.54 \times 10^{-6} \text{ Å}^{-2}\) is the SLD for aqueous buffer. In using this equation, additivity of volumes is assumed. The amount of cellulose per unit area within each film was obtained by integrating the cellulose volume fraction profiles.

Before each measurement, the regenerated cellulose film was allowed to equilibrate with sodium acetate buffer for 20 min, after which several scans were collected. After equilibration of the film in buffer, a 0.2 µM protein solution was injected into the measurement cell and incubated with the films in absence of flow until little change in NR was observed on a timescale of several hours. The incubation time therefore varied somewhat among the samples. The NR studies were performed at the optimal temperature for Cel5A (75°C). The sample cell was heated by circulating a heating fluid through copper blocks placed underneath and on top of the sample cell. The sample cell and copper blocks were enclosed in a Styrofoam box containing thin aluminum foil windows.

3.6 – Acknowledgements

This work conducted by the Joint BioEnergy Institute was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231
Chapter 4 — Optimization of a recombinant cellulase cocktail for saccharification of ionic liquid pretreated biomass

4.1 Abstract
Lignocellulolytic cocktails consist primarily of a mixture of enzymes capable of solubilizing the plant cell walls. It is understood that different enzymes can have different roles in this synergistic phenomena. However, the role of each enzyme is not the same in the hydrolysis of every substrate. Thus it is possible to optimize cellulolytic cocktails for different feedstock/pretreatment. Among the benefits of creating an optimal cocktail is the resulting lowered cost of biomass saccharification. We have optimized a minimal cellulolytic cocktail for the hydrolysis of biomass pretreated with 1-ethyl-3-methylimidazolium acetate and in the presence of the same ionic liquid (IL). Minimal lignocellulosic cocktails comprising three individual pure enzymes were engineered with a design-of-experiments (DOE) software. Commercial enzymes (used as a benchmark) exhibited 90% glucose release at pH 5 and 55°C, but did not reach the same level of activity at the industrially desired higher temperatures or pHs. In addition, no glucose release could be measured from these enzymes in the presence of 1-ethyl-3-methylimidazolium. Enzymes that are more robust to such conditions have been previously characterized individually by our lab. Minimal cocktails made with these enzymes are able to degrade up to 50% of the ILSG at a temperature of 70°C and a total IL concentration of 20%. This DOE approach can be used for the formulation of more complex cellulolytic cocktails for the deconstruction of pretreated lignocellulosic substrates, particularly at the higher temperatures, pHs, and IL concentrations of relevance to industrial applications.

4.2 Background
The development of alternative, renewable transportation fuels is critical to energy, environmental, and economic security. Advanced biofuels created from lignocellulosic biomass will reduce our reliance on fossil fuels with minimal adverse impact on the food supply. However, significant roadblocks hamper the development of cost- and energy-efficient processes to convert lignocellulose into advanced fuels [4, 16, 49, 121], notably including the expense of enzymes used to deconstruct the biomass into fermentable sugars.

Pretreating the biomass can increase saccharification yields by up to ten fold [Ref Sousa 2009], thereby reducing the concentration of the cellulolytic enzymes to a fraction of what would be required otherwise. Ionic liquids (ILs) are a promising pretreatment, capable of completely dissolving lignocellulosic materials. The liberated cellulose can then be precipitated and saccharified [25, 122]. For example, pretreatment of microcrystalline cellulose (Avicel) with 1-ethyl-3-methylimidazolium acetate ([C2mim][OAc]) enhances the subsequent enzymatic hydrolysis to fermentable sugars by up to ten fold [25].

While saccharification yields are greatly improved by IL pretreatment, other variables still negatively affect yields, including enzyme inhibition, enzyme deactivation and enzyme specificity. Economic feasibility in terms of enzyme production costs must also be considered. To this end, there is considerable interest in optimizing enzymatic efficiency, from the engineering of enzymes for decreased susceptibility to inhibition and
deactivation [16, 51], to the discovery and optimization of new enzymes for cellulose degradation [45, 53]. Since a minimum of three enzymes are required to convert insoluble cellulose to glucose (see chapter 1.3.3 for more detail), one promising strategy is to optimize the enzyme mixtures, known as cocktails, for maximal glucose production [50]. Enzymatic treatment of pretreated biomass is usually performed with commercially available cellulolytic cocktails of partially defined mixtures. Some of these mixtures can contain over 80 proteins [123]. Not all the proteins present in the cocktail are necessary for the saccharification of every pretreated biomass. Different feedstocks can respond differently to each of the pretreatments available [39]. Therefore, a better understanding of the optimal cocktail components and concentrations for saccharification of different types of pretreated biomass is needed.

As discussed previously, biomass pretreatment with ionic liquids enables efficient saccharification of lignocellulosic substrates. However, traces of IL in the pretreated biomass can inhibit the enzymes responsible for the saccharification process, necessitating extensive washing of the IL-pretreated biomass before enzymatic treatment. The development of a cellulolytic enzymatic cocktail tolerant to low concentrations of ionic liquids would enable more cost-efficient bioprocessing. To that end, with this work we designed a cellulolytic cocktail tolerant to at least 20% IL. The cocktail is also functional at temperatures of approximately 70°C, which offers the added benefit of potentially increasing the rate of hydrolysis while maintaining a sterile environment.

4.3 Results and Discussion
4.3.1 Chimeric endocellulases have varying impact on cellulase cocktail performance
4.3.1.1. Chimeric endocellulases confer enhanced glucose release on all substrates at low enzyme loadings

Three enzymes were used to construct a cellulolytic cocktail, selected from a set of IL- and thermotolerant enzymes assembled by our collaborators at the Joint BioEnergy Institute. The three enzymes selected are the cellobiohydrolase CelB from *Caldicellulosiruptor saccharolyticus*, a β-glucosidase *Thermotoga petrophila* (BG) and an endoglucanase, Cel5A from *Thermotoga maritima*. Two versions of the endoglucanase were tested, the wild type Cel5A, which consisted of the CD alone, and its chimeric version in which a CBM2a is attached to create Cel5A-CBM2a (see Chapter 2 for more information on this chimera and its construction). Cocktails of the three enzymes (including either Cel5A or Cel5A-CBM2a) were incubated with Avicel, IL-Avicel or ILSG at different enzymatic component concentrations (Figure 4.1). All substrates treated with Cel5A-CBM2a-containing cocktail exhibited higher glucose yields over the same substrates treated with the Cel5A-containing cocktail. Furthermore, an increase in activity was observed with an increase in enzyme loading. These results suggest that the increase in cellulolytic activity could be attributed to the addition of the CBM to the Cel5A CD. Therefore, addition of a CBM2a to the CD of Cel5A can increase effectiveness of a cellulolytic cocktail across many different substrates.
4.3.1.2 Chimeric endocellulases do not significantly enhance glucose release at high enzyme loadings

Given the success of the cocktail containing the chimeric Cel5A-CBM2a, we next created a four-enzyme cocktail, containing two endocellulase fusions to this same CBM. For the second endocellulase CD, we chose Cel9A from *Alicylobacillus acidocaldarius*. The other components were identical to the original cocktails: Cel5A (or Cel5A-CBM2a), CelB, and BG. Thus four basic cocktails were created: CelB and BG with Cel5A or Cel5A-CBM2a and Cel9A or Cel9A-CBM2a. For this experiment, we investigated only ILSG as a substrate, and also used enzyme loadings relevant for industrial applications, approximately 20-fold higher than the enzyme loading in the first-generation cocktail. Importantly, we employed design-of-experiments (DOE) software to examine the impact of each enzyme’s concentration in the cocktail. A total of 16 different mixtures (but 20 samples) with different enzyme proportions were run in this assay. Total enzyme loading was kept constant (20 mg of total enzyme per gram of glucan) for all cocktails.

Control samples with only BG gave rise to no detectable glucose while cocktails without BG exhibited close to zero levels of released glucose as well (Figure 4.2). It is expected that the samples with pure BG would not have detectable glucose, since this enzyme is only responsible for the conversion of soluble cellobiose into glucose. Since ILSG is an insoluble substrate makes it impossible for the BG to produce glucose from it. Likewise, low levels of glucose in samples with no BG are due to the absence of an enzyme to convert cellobiose to glucose. To ensure that this is indeed the correct explanation, the measurements were confirmed using a standard assay for reducing ends with a cellobiose standard. High levels of soluble celldextrins were detected in samples that had no BG, confirming that the other cellulases were active.

The highest percentage of glucose release (64%) was observed from the samples that had the highest concentration of the cellobiohydrolase and minimal, but equal, concentration of the other three enzymes (Figure 4.2). In fact, all cocktails that gave rise to over 50% glucose release had a high concentration of cellobiohydrolase present. This suggests that this specific cellobiohydrolase was the most relevant enzyme for the liberation of the glucan contained in ILSG. In contrast, the addition of the CBM to the Cel5A and Cel9A endocellulases had little to no effect in the glucose yield of these...
simple cellulolytic cocktails. This is contradictory to what was observed in the previously described experiments of the Cel5A and Cel5A-CBM2a cocktails. However, there were significant differences between the two experiments that could give rise to such discrepancies. For instance, in these experiments, wild type and chimeric enzymes were added in the same concentration by mass basis, as opposed to by molar basis in the previous experimental design. Since the chimeras have a higher molecular weight, less chimeric enzyme is added for a given concentration. Most notably, however, there was approximately 20 times less total enzyme per g glucan in the previous experiments. This suggests that while the chimeric cellulases have enhanced activity at low ratios of enzyme to substrate, the effect is diminished at higher ratios, which are currently more relevant for industrial use. With this in mind, we decided to perform the rest of the cocktail-optimization experiments with out the implementation of chimeric endocellulases.

4.3.1.3 Cellulase cocktails comprising commercial enzymes do not have activity in the presence of 20% IL.

To further investigate the optimization of cocktails for the saccharification of ILSG, we set out to determine how much the presence of residual IL impacts enzyme activity. As a benchmark, three commercial enzymes were chosen and obtained from Megazymes: endocellulase E-CELTE from Talaromyces emersonii, cellobiohydrolase E-CBHI from Trichoderma sp. MA 3642 and a β-glucosidase from Thermotoga maritima.

As in the previous section, DOE software was used to efficiently pinpoint the optimum enzyme concentrations in the cocktail. Initially we examined the cocktail composition on ILSG in the absence of IL, requiring 13 variations of the three-enzyme cocktail. These 13 variations were spread through 17 samples as suggested by the DOE software. All samples were tested under different pH and temperature conditions (See methods and Figure 4.3). The glucose released from all the commercial cocktail samples ranged between 0 – 90 %. Maximum glucose production was detected in cocktails with equal proportion of all enzymes at 55°C and pH 5 (Figure 4.3a). Other conditions did not result in as much glucose release, but the proportion of enzymes that generated the maximum glucose release for a given pH remained constant across both tested pHs at 55°C (Figure 4.3b). This suggests that the presence of all three enzymes in equal mass proportions is necessary for optimal degradation of ILSG at this temperature. This optimal cellulase ratio, as measured by glucose production, changed when the temperature was increased to 70°C (Figure 4.3 c, d). At this temperature, the optimum cocktail had the highest concentration of endocellulase and a small amount of β-glucosidase, and resulted in 27% glucose release. This is likely due to the lower thermostability of the cellobiohydrolase E-CBHI. The reported optimal temperature for the cellobiohydrolase is 70°C, but it is unstable above 65°C. Optimal temperatures for the endocellulase and the β-glucosidase were 70°C and 90°C, respectively, and both enzymes are stable at 70°C.

Importantly, none of the commercial enzymes presented activity of glucose release in the presence of the IL 1-ethyl-3-methyl-imidazolium acetate at the lowest concentration tested (10%). This suggests that IL concentrations of 10% are able to completely inhibit the activity of these enzymes.
4.3.1.4 Cellulase cocktails comprising Cel5A/Cel9A exhibit activity at 70°C and in the presence 20% IL.

In an analogous experiment, three of our thermophilic, IL-tolerant enzymes were used to create a cocktail to determine if the IL tolerance extends to the cocktail environment. This cocktail contains the same enzymes as that tested in Section 4.3.1.1 (endoglucanase Cel5A from T. maritima; cellobiohydrolase CelB from C. saccharolyticus, and the β-glucosidase from T. petrophila), and will be referred to as the Cel5A cocktail for the remainder of the text. In parallel, we tested a cocktail with the Cel9A endoglucanase (Section 4.3.1.2) in place of the Cel5A endoglucanase, and this cocktail will be referred to as the Cel9A cocktail for the remainder of the text. Biomass pretreated with 1-ethyl-methyl-imidazolium acetate was incubated with 17 mixtures of each of the two cocktails comprising 13 different proportions of the constituent enzymes, as devised by the DOE software. These samples were all incubated at 70°C, pH 5.0, and either 0 or 20% IL. In the absence of IL, the cocktails reached 70% and 65% glucose release for the Cel5A and Cel9A cocktails, respectively. This is less than that exhibited by the cocktails designed with commercial enzymes. (data not shown).

In the presence of IL, however, the Cel5A and Cel9A cocktails performed extremely well (Figure 4.4, 4.5). A maximum of 52% and 43% glucose release was
measured from the samples treated with Cel5A and Cel9A cocktails, respectively under the conditions of 20% IL and 70°C. This demonstrates that all the enzymes used maintain their significant tolerance to the IL, corroborating an earlier report by [50]. This tolerance is considerably higher than that of the commercial enzymes, which exhibited no detectable glucose release in the presence of IL.

Increased glucose release from the Cel5A cocktails as compared to the Cel9A cocktails is most likely related to the lower thermostability of Cel9A. This set of experiments was carried out at 70°C, and the reported optimal temperature for the Cel9A
is 65°C, while the optimal temperature for Cel5A is between 75 and 80°C. Notably, upon analysis of the results from the 17 test cocktails, the DOE software predicted optimal deconstruction of ILSG for both Cel5A and Cel9A cocktails to be 3:10:7 for EG:CBH:BG (Figures 4.4, 4.5). Although higher glucose yields were generated by the Cel5A cocktails, the optimal enzyme proportions were very similar to those of Cel9A.

4.4 Conclusions

The idea that every enzyme present in a lignocellulolytic cocktail contributes differently to the full breakdown of the biomass gives rise to the hypothesis that unique optimal enzymatic proportions exist for each enzyme cocktail/substrate combination. We have demonstrated that a DOE approach can be used for the formulation of optimal cellulolytic cocktails for the hydrolysis of ionic-liquid pretreated switchgrass in the presence of up to 20% IL (Figures 4.5, 4.6). Cocktails were designed using three individual pure enzymes, and compared to a cocktail comprising commercial enzymes. The commercial cocktail did not exhibit major activity at temperatures over 60°C or at any concentrations of IL, whereas cocktails made with enzymes isolated for their enhanced tolerance were able to degrade up to 50% of the ILSG at a temperature of 70°C and a total IL concentration of 20%. Given the success of this designer cocktail, there is a clear value to further optimizing the cocktail, including the addition of other enzymes and accessory proteins, to create more complex and capable cocktails for saccharification of ILSG. We will continue to use this DOE approach to develop optimal enzymatic cocktails for the deconstruction of specific types pretreated lignocellulosic substrates.
4.5 Methods
4.5.1 – Cellulase genes and vector construction

Three commercial and four additional enzymes were used in this study. The commercial enzymes included endoglucanase E-CELTE from *Talaromyces emersonii*, cellobiohydrolase E-CBHI from *Trichoderma sp* and the β-glucosidase from *Thermotoga maritima*. All these enzymes were obtained from Megazyme. Cel9A from *Alicyclobacillus acidocaldarius* (Cel9A) was initially obtained as a gift from K. Eckert (Humboldt-Universität zu Berlin, Institut für Biologie/Bakterienphysiologie, Germany) [92, 94]. Creation of the cel5A gene from *Thermotoga maritima* (encoding Cel5A), the truncated (CBM3-GH5) form of the cellobiohydrolase from *Caldicellulosiruptor saccharolyticus* and the β-glucosidase *Thermotoga petrophila* from were described previously [124]. All open reading frames were amplified and cloned into the pENTR221 vector (Invitrogen) modified to contain ten consecutive histidine residues on its N terminus (N-terminal 10X His-tag) for affinity purification with nickel columns.

Creation of the Cel9A- and Cel5A-CBM chimeras was described in chapter 1. Briefly, the three distinct sequences: genes encoding the CD, a synthetically designed linker region, and the CBM. The CBM (UniProt Q60029), found in an exocellulase from *Thermotoga fusca*, was synthesized by GenScript Corporation. The linker was designed to be rich in proline and threonine, and is based on the sequence of a linker in endoglucanase N in *Erwinia carotovora*. Only the ends of this sequence were modified for overlapping PCR and efficient cloning. The DNA encoding the linker was created using PCR assembly of multiple oligonucleotides (IDT, integrated Technologies). All three parts (Cel9A/Cel5A, CBM and linker) were amplified and then assembled by single
overlap extension PCR. The resulting product was ligated into the modified pENTR221 vector as described above. All constructs were later passed into pDEST42 following standard Gateway cloning protocols (Invitrogen). The final constructs were verified by sequencing.

4.5.2 Enzyme expression and purification

The endocellulases Cel5A and Cel9A cellulosases were expressed in BL21 (DE3) Star cells in 2XYT media supplemented with 100 µg/ml of carbenicillin. The recombinant cellobiohydrolase (CBM3-GH) and the β-glucosidase were also expressed in BL21 (DE3) Star cells, but grown in LB media supplemented with 50 µg/mL of carbenicillin. Cells were grown at 37°C and a constant shaking of 200rpm until reaching an OD600 of 0.7 – 1.0 at which time 0.5 mM IPTG was added to induce expression. The endocellulases were incubated overnight with shaking (200 rpm) at 18°C for 24 h. The cellobiohydrolase and the β-glucosidase were incubated overnight at 30°C with constant shaking of 200rpm. Cells were then pelleted down at 12,000 X g for 10 min at 4°C. All cellulases were extracted from cell pellets using BugBuster (Novagen) according to the protocol with the following additions: 25 U benzonase/mL BugBuster (Novagen), 1 U r-lysozyme/mL BugBuster (Sigma), and 1X protein inhibitor cocktail V (EDTA-free) (Calbiochem). Cell debris was pelleted by centrifugation at 20,000 X g for 20 min at 4°C. The lysate was then incubated in a 50°C water bath for up to 30 min to precipitate most of the heat-labile E. coli native proteins. The cellulases were affinity purified from the soluble fraction using gravity nickel columns (His GraviTrap, GE Healthcare) according to the manufacturer recommendations. Briefly, the columns were pre-equilibrated with 10 ml of washing buffer (20 mM sodium phosphate, 1 M NaCl, 20 mM imidazole, pH 7.4). The supernatant of the lysate described above was run twice through the equilibrated columns. Columns were washed with 10 ml of washing buffer. Enzymes were eluted by increasing the imidazole concentration to 500 mM. The eluted cellulases were buffer-exchanged using gravity desalting columns (PD-10 Desalting Columns, GE Healthcare) into a buffer consisting of 20 mM Tris, 150 mM NaCl, pH 7.4. Purity of the cellulases was determined by SDS-PAGE. Concentration of the enzymes was determined by bicinchonic acid (BCA) colorimetric assay (Pierce Thermo Scientific).

4.5.3 [C2mim][OAc] pretreated switchgrass

Switchgrass (Putnam) was first milled with a Thomas–Wiley Mini Mill fitted with a 40-mesh screen (Model 3383-L10 Arthur H. Thomas Co., Philadelphia, PA, USA). The ionic liquid used for the pretreatment was 1-ethyl-3-methylimidazolium acetate ([C2mim][OAc]) (Sigma–Aldrich, St. Louis, MO, USA) as the IL solvent. One liter of 3% (m/v) Putnam switchgrass was mixed with [C2mim][OAc] in a globe reactor (Syrris). The samples were incubated with stirring at a temperature of 160°C for 3h. After incubation, the solubilized biomass in IL was transferred to a glass container and stored at 4°C until used.

4.5.4 Dispensing of biomass

IL-pretreated switchgrass was aliquoted into 96-well deep well blocks (Life Technologies, CS15196) using Biomek FX (Beckman Coulter) robot and a custom script.
In summary: The 1-ethyl-3-methylimidazolium acetate solution containing 3% putnam switchgrass was heated at 60°C and 75 µL of this solution and 400 µL of water were dispensed in each well of the 96-well deep well block. A precipitate biomass (50% cellulose) forms as the water is added to the samples. The plates were sealed with pliable sealing using a Plateloc thermal microplate sealer (Agilent Technologies). The plates were then shaken on the side for 0.5 h at 150 rpm (Multitron II) and 30°C. After shaking, the content in the plates is washed by replacing the water volume in each well. Plates were autoclaved for 30 minutes at 121°C and 16 psi (Tuttnauer 387EA). Water exchange and autoclave steps were performed twice. The water is finally removed and replaced with more water to a final volume of 275 µL of fresh DI water per well containing approximately 1 mg of regenerated biomass.

4.5.5 Saccharification assay: Cel5A vs Cel5A-CBM2a cocktails

Comparison of cellulolytic cocktails containing Cel5A from CD and Cel5A-CBM2a (CBM from an exocellulase from *Thermotoga fusca*, see chapter 1) was performed by mixing molar fractions of 40% endocellulase Cel5A with 40% cellobiohydrolase CelB from *Caldicellulosiruptor saccharolyticus* and 20% of the β-glucosidase *Thermotoga petrophila*. Three different total enzyme concentrations (100, 200 and 400 nM) were used to compare different enzyme loadings. For saccharification

<table>
<thead>
<tr>
<th>Std</th>
<th>Run</th>
<th>Block</th>
<th>Type</th>
<th>Component 1</th>
<th>Component 2</th>
<th>Component 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A:Endo</td>
<td>B:CBH</td>
<td>C:BG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>mg/g glucan</td>
<td>mg/g glucan</td>
<td>mg/g glucan</td>
</tr>
<tr>
<td>1</td>
<td>14</td>
<td>Block 1</td>
<td>Vertex</td>
<td>20.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>Block 1</td>
<td>Vertex</td>
<td>0.00</td>
<td>20.00</td>
<td>0.00</td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>Block 1</td>
<td>Vertex</td>
<td>0.00</td>
<td>0.00</td>
<td>20.00</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>Block 1</td>
<td>ThirdEdge</td>
<td>13.33</td>
<td>6.67</td>
<td>0.00</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>Block 1</td>
<td>ThirdEdge</td>
<td>13.33</td>
<td>0.00</td>
<td>6.67</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>Block 1</td>
<td>ThirdEdge</td>
<td>6.67</td>
<td>13.33</td>
<td>0.00</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>Block 1</td>
<td>Center</td>
<td>6.67</td>
<td>6.67</td>
<td>6.67</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>Block 1</td>
<td>ThirdEdge</td>
<td>6.67</td>
<td>0.00</td>
<td>13.33</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>Block 1</td>
<td>ThirdEdge</td>
<td>0.00</td>
<td>13.33</td>
<td>6.67</td>
</tr>
<tr>
<td>10</td>
<td>15</td>
<td>Block 1</td>
<td>ThirdEdge</td>
<td>0.00</td>
<td>6.67</td>
<td>13.33</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>Block 1</td>
<td>AxialCB</td>
<td>13.33</td>
<td>3.33</td>
<td>3.33</td>
</tr>
<tr>
<td>12</td>
<td>10</td>
<td>Block 1</td>
<td>AxialCB</td>
<td>3.33</td>
<td>13.33</td>
<td>3.33</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>Block 1</td>
<td>AxialCB</td>
<td>3.33</td>
<td>3.33</td>
<td>13.33</td>
</tr>
<tr>
<td>14</td>
<td>17</td>
<td>Block 1</td>
<td>Vertex</td>
<td>20.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>15</td>
<td>11</td>
<td>Block 1</td>
<td>Vertex</td>
<td>0.00</td>
<td>20.00</td>
<td>0.00</td>
</tr>
<tr>
<td>16</td>
<td>16</td>
<td>Block 1</td>
<td>Vertex</td>
<td>0.00</td>
<td>0.00</td>
<td>20.00</td>
</tr>
<tr>
<td>17</td>
<td>4</td>
<td>Block 1</td>
<td>Center</td>
<td>6.67</td>
<td>6.67</td>
<td>6.67</td>
</tr>
</tbody>
</table>

Table 4.1 Output of the DOE software for cocktails; values indicate the relative concentration of the enzyme in each mixture in units of mg of enzyme per gram of glucan.
assays on insoluble substrates, cocktails were added to 50 mM citric buffer containing 25 mg/ml cellulose substrate to a total volume of 500 µl in micro-tubes with caps. Cellulases were tested at the optimal pH, 4.8. All samples were incubated for 24 hrs at 55°C and 1400 rpm in a bench-top thermomixer (Eppendorf AG 22331 Hamburg). Three types of insoluble substrates were used for this experiment: Avicel (pH 101, Sigma), IL-Avicel and ILSG. Sugar release was measured by DNS assay and glucose release by YSI (Sections 4.5.8.1 and 4.5.8.2, respectively).

4.5.6 Saccharification assay: four-enzyme cocktails

DOE software was used to design 20 sample cocktails containing 16 different proportions of four cellulases. The resulting ratios are based in a simplex-lattice design, which is a space-filling, rectangular grid of runs. The vertices of the rectangle represent the highest pure concentrations of a single enzyme. Upper and lower proportions for each enzyme were 20 and 0 mg of enzyme per g of glucan, respectively. For saccharification assays on insoluble substrates, cocktails were added to citric buffer at pH 5.0 containing 25 mg/ml ILSG to a total volume of 500 µl in micro-tubes with caps. All samples were incubated for 24 hrs at 55°C and 1400 rpm in a bench-top thermomixer (Eppendorf AG 22331 Hamburg). Saccharification results (expressed in glucose concentrations (mM)) as measured by HPLC (Section 4.5.8.3) from all the enzyme mixtures were then analyzed by the Design-Expert software (Section 4.5.10).

4.5.7 Saccharification: Cocktail optimization experiments

Cocktails with of varying enzymatic ratios (Endo:CBH:BG) were prepared as guided by the output of the DOE software using a cubic approach (See table 4.1 for enzyme ratios, and section 4.5.9 for information on the DOE method). Enzymatic hydrolysis of ILSG assays were carried out as described in chapter 1 and 2 with slight changes to the protocol. Reactions were run in 96-deep-well blocks (Life Technologies, CS15196) in 50 mM citrate buffer (pH 5 or 7, depending on desired pH) and a total volume 500uL. Enzyme concentration and amount of biomass per well were 20 mg of total enzyme / g of glucan and 1.0 mg of ILSG (50% glucan), respectively. The reactions were topped with 200 µL of mineral oil (Sigman, M8662) to avoid evaporation at high temperatures. All samples were incubated for 24 hrs at either 55°C or 70°C (as indicated in figures) and 700 rpm (to avoid water mixing with oil) in an incubator (Multitron II). Subsequently, all samples were mixed in 50 mM citrate buffer and 50% glycerol and kept at -20°C for storage. Released sugars were measured by HPLC (Section 4.5.8.3), and the results were again analyzed by the Design-Expert software (Section 4.5.10).

4.5.8 Saccharification measurements

4.5.8.1 Quantification of soluble sugars by DNS assay

Sugar concentration of the samples was determined by DNS assay as previously described [106]. Briefly, 60 µl of 2X DNS reagent (1 g DNS, 30 g KNa tartrate, 20 ml 2 N NaOH in 100 ml total volume) were mixed with 60 µl of the samples supernatant. Samples were well mixed and incubated at 95°C for five minutes. Samples were allowed to cool down to room temperature and their absorbance read at a wavelength of 540 nm (Molecular Devices, SpectraMax M2). Concentrations were calculated by comparison of absorbance to a cellobiose standard curve.
4.5.8.2 Glucose detection by YSI

Glucose release from saccharification samples was measured via YSI 2900 (Biochemistry Analyzer). Samples were first filtered at 1000 X G (Eppendorf, 5810) using 0.45 µm polypropylene filers (Unifilter, 7700-1305). The filtered samples were then transferred to a 96-well PCR plates where their glucose content was determined by the use of a commercial glucose standard (YSI 2700).

4.5.8.3 Glucose detection via HPLC

Glucose release from saccharification samples was measured via high performance liquid chromatography (HPLC). HPLC analysis was performed using an Agilent 1260 Infinity system using a Biorad 300mm Aminex 87 H column with a Biorad cation H guard column. The detector is an Agilent 1260 refractive index detector heated to 35 C. The samples were run using an isocratic 4mM sulfuric acid eluent at 0.6 mL/min at 60 C for 16 minutes. Standards were prepared in a volumetric flask using Supelco monosaccharides in 4 mM sulfuric acid. Chemstation software was used to calculate the concentration of the samples from a five point standard calibration curve. Eluent was prepared using JT Baker HPLC grade water and 95% sulfuric acid from JT Baker.

4.5.9 Design of experiments for mixtures (containing non-chimeric endocellulases)

Statistical experiment design approaches for mixture proportioning were used to develop a multi-cellulase cocktail for the saccharification of [C2mim][OAc] pretreated switchgrass, with the goal being to maximize glucose yields using a minimum of total enzyme. A standard simplex lattice design of experiments augmented with axial check blends and the overall centroid was used for optimizing enzyme mixtures (Figure 4.6). Using this DOE approach makes it possible to estimate the properties of an entire system from a limited number of observations and has proven useful for any multicomponent system where intensive properties are to be related to composition (probably need a couple of refs, including Walton’s group, John Cornell and Greg Piepel 2008 might also be good to include). In a simplex design, the space of possible enzyme combinations to sample is greatly reduced, because sampling space is constrained to only those mixtures for which the sum of the proportions of all enzymes in the mix sum to one, and hence mixture component space may be represented by a regular simplex. Points of composition in this simplex are explored in accordance with a lattice arrangement, and responses are represented by simplified general polynomials. Since endoglucanases, cellobiohydrolases
and β-glucosidases act synergistically to catalyze the conversion of cellulose to glucose, we chose to design our experiments under the assumption that the underlying response surface is a cubic polynomial that includes second and third order terms representing nonlinear blending and synergism. Augmenting the standard simplex lattice design with axial check blends and the overall centroid provides additional data points for better fitting of the polynomial response surface.

4.5.10 Analysis of experiments for mixtures

Data were collected as described and input into the Design-Expert software program. An analysis of variance (ANOVA) was run to determine the degree to which the cubic model fit the data, and a best-fit (in the least squares sense) surface was fit to the sample data points. The resulting cubic response function relating glucose yield to cellulase proportions was then numerically maximized under the constraint that the sum of the proportions of the mixture components sums to one. The function was maximized from 200 random starting solutions (random assignment of starting proportion of each cellulase), and the solution (set of enzyme proportions) that maximized glucose yields was chosen as the optimum mixture of cellulases.

4.6 Acknowledgements:

This work conducted by the Joint BioEnergy Institute was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231
Chapter 5 — References

75. Gilad R, Rabinovich L, Yaron S, Bayer EA, Lamed R, Gilbert HJ, Shoham Y: **Cell, a noncellulosomal family 9 enzyme from Clostridium thermocellum, is a processive endoglucanase that degrades crystalline cellulose.** *Journal of Bacteriology* 2003, 185:391–398.
81. Carbohydrate Active Enzymes Database [http://www.cazy.org/].


Chapter 6 — Appendix

6.1 Abbreviations

AA – amino acid
AA – auxiliary activity (enzyme)
BG – β-glucosidase
CAZy – carbohydrate active enzymes (http://www.cazy.org)
CBM – carbohydrate binding module
CBH – cellobiohydrolase
CD – catalytic domain
CE – carbohydrate esterases
DNS – 3,5-dinitrosalicylic acid
EG – endoglucanase
EC – endocellulase
GH – glycoside hydrolase
GT – glycosyltransferases
IL – ionic liquid
IL-MCC – ionic-liquid pretreated microcrystalline cellulose
LPMO – lytic polysaccharide monooxygenases
MCC – microcrystalline cellulose (Avicel)
NR – neutron reflectivity
PCR – polymerase chain reaction
PL – polysaccharide lyases
SLD – scattered light density