Title
Molecular Level Investigations and Mathematical Modeling of Cellulase-catalyzed Hydrolysis of Cellulose

Author
Fox, Jerome M.

Publication Date
2012

Peer reviewed|Thesis/dissertation

Molecular Level Investigations and Mathematical Modeling of Cellulase-catalyzed Hydrolysis of Cellulose

by

Jerome Michael Fox

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

of

Doctor of Philosophy

in

Chemical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in Charge

Professors Douglas S. Clark and Harvey W. Blanch, Co-Chairs

Professor Carolyn R. Bertozzi

Professor Alexis T. Bell

Fall 2012

Abstract

Molecular Level Investigations and Mathematical Modeling of Cellulase-catalyzed Hydrolysis of Cellulose

by

Jerome Michael Fox

Doctor of Philosophy in Chemical Engineering

University of California, Berkeley

Professors Douglas S. Clark and Harvey W. Blanch, Co-Chairs

Cellulose-to-glucose conversion costs comprise one of the largest expenses in the production of lignocellulosic biofuels, a renewable alternative to traditional fossil fuels. Efforts to reduce these costs by improving the activity of cellulase enzymes, which act within multi-enzyme mixtures to catalyze the hydrolysis of cellulose, have been hindered by uncertainty surrounding the mechanistic origins of rate-limiting phenomena and by an incomplete understanding of complementary enzyme function. In this work, we employed mechanistic models of enzymatic action alongside experimental studies of enzyme-enzyme synergy to investigate kinetic impediments encountered by cellulase enzymes as they catalyze the hydrolysis of cellulose. Using several mechanistic models of enzymatic hydrolysis, we show how hydrolysis kinetics and optimal cellulolytic mixture compositions are dependent on the nature of the cellulosic substrate (i.e., particle shape, surface area, degree of polymerization) and the conditions under which it is depolymerized (i.e., hydrolysis and fermentation process conditions). By developing a method to estimate catalysis-specific products within multi-enzyme reactions, we show that cellobiohydrolase enzymes, which catalyze processive hydrolysis from cellulose chain ends, encounter rate limitations that result, not from intrinsic kinetics, but from slow rates of chain complexation and from morphological obstacles to processivity. And using photoactivated localization microscopy (PALM) to produce pointillistic maps of carbohydrate-binding modules (CBMs) bound to cotton, we develop an order parameter to quantify the different spatial arrangements of adsorbed CBMs and use that order parameter to explain synergy between cellulase enzymes designed to target different surface structures. The results of this work reveal strategies for using morphological targeting to enhance enzyme-enzyme cooperativity within cellulolytic mixtures, thereby improving their overall activity and lowering the cost of cellulose hydrolysis.

Dedication

This dissertation is dedicated to my family.

Chapter 1: Introduction to Cellulosic Biofuels

1.1 Making the Case for Renewable Energy

Global dependence on oil, coal, and natural gas has led to some of the most significant environmental, socioeconomic, and geopolitical problems of the 20th century. Today, 86% of the world's energy needs are met with these exhaustible resources (*1*). Alternatives to traditional fossil fuels could halt global climate change, reduce indoor air pollution, generate local job growth, facilitate energy independence and strengthen the trade balance for petroleum importing countries, and dampen energy market volatility (*2-8*). In this section, we explore some of the main policy arguments for renewable energy and, by examining their relationship to carbon emissions and energy prices, we motivate cellulosic biofuels research that is focused on lowering production costs.

Anthropogenic climate change is the most compelling argument for the pursuit of renewable energy technology. Since the beginning of the industrial era (~1750), atmospheric concentrations of carbon dioxide (CO_2), methane (CH_4), halocarbon gases, and nitrous oxide (N_2O) have grown as a result of fossil fuel combustion, agricultural practices, and other human activities (*9*). By enhancing radiative forcing at the tropopause, these greenhouse gases have increased the average temperature of the Earth's surface and brought about abrupt changes in global weather patterns. Strategies focused on reducing the incidence of the most catastrophic climate change events (i.e., major ecological disruptions, glacial melt, drought, forest fires) are focused on stabilizing atmospheric CO_2 levels at around 500 ppm (*2*). In 2004, the atmospheric CO_2 concentration was 375 ppm, and global emissions were about 7 billion tons of carbon-equivalents per year (GtC/yr). At that time, if CO_2 was to be stabilized at 500 ppm, global emissions had to be held constant for the next fifty years (*2*). In October of 2012, CO_2 levels reached 391 ppm, and annual emissions increased to over 9 billion tons of carbon per year (*10*). With greenhouse gas emissions accelerating, climate change abatement requires many forms of renewable energy to be developed and implemented in the next 10-30 years.

Air pollution motivations for clean fuel are most strongly rooted in health effects, not carbon dioxide. Over 2 billion people in the developing world are exposed to high levels of indoor air pollution that could be reduced or eliminated with cleaner cooking technologies (*4*). Incomplete combustion of wood, agricultural residues, and other solid cooking fuels results in the release of carbon monoxide, hydrocarbons, particulate matter, and nitrogen oxides that cause a variety of respiratory diseases (*3*). Cleaner fuels such as kerosene, methane, or ethanol can enhance air quality within these households. Renewable fuels such as methane derived from biogas digesters or ethanol derived from cassava or other plant sources are only advantageous if they can be sold at more competitive prices than fossil-fuel-derived equivalents; interestingly, several analyses suggest that renewables can be priced competitively in energy- and infrastructure-poor regions of China and Africa (*11, 12*) . Indoor air pollution may not be a good policy argument for renewable energy in its own right, but it is an argument for cleaner burning monocomponent fuels, and many of these can be produced from renewable sources in a low cost manner.

Job growth and national security arguments for renewable energy are also tied to price . Both motivations arise from a desire for expanded domestic energy production, which can create jobs and prevent supply disruptions in petroleum-importing countries (in addition to strengthening their trade balance) (*7*). Renewable energy technologies, however, are not necessarily advantageous for accomplishing these goals. In the United States, coal and shale gas reserves are adequate to last several generations (*13*). These domestically produced fossil fuels can create local jobs and promote national security by reducing reliance on foreign supplies. Renewable energy technologies will not be advantageous in accomplishing these two goals until global fossil fuel supplies begin to dwindle, extraction prices begin to increase, and/or renewable alternatives become cost competitive. Arguments can still be made that early development of renewable energy is required in light of diminishing fossil fuel reserves; however, with no consensus on the imminence of reserve depletion, these arguments are difficult to make on the political stage.

Dampened volatility in energy prices is a technology-dependent argument. The price of gasoline at the pump is set by supply and demand in international oil market. If U.S. domestic oil production levels are increased to meet national demand, global supply will change only slightly, and the shift in price paid at U.S. pumps will likely be slight (*14*). Biofuel prices will be similarly tethered to the existing transportation fuel market. Alternative forms of transportation energy, by contrast, may not be. For electric cars, for example, "fuel" prices will be set by the local electricity price. The influence of domestic renewable energy production on domestic energy prices clearly depends on the markets involved, which are, to an extent, technology-specific.

Although climate change is the strongest policy motivation for development of renewable energy technologies when they are not cost competitive with fossil fuels, non-climate change arguments are likely to play an important role in creating support for clean energy policies. Accordingly, their validity is important for the promotion of renewable energy development and deployment in the near future. There are two main mechanisms by which non-climate-change arguments for renewable energy can become stronger: (1) a carbon tax can be used to incorporate climate change externalities into the price of energy, thereby creating cost parity between renewable energy technologies and fossil fuels, or (2) the price of renewable energy technologies can be reduced through technological advancements, thereby allowing them to compete with fossil fuels. As there is little evidence to suggest that the first strategy is politically feasible in an international market, the second is the more realistic avenue to a clean energy future. By making renewable technologies cost competitive with conventional fossil fuels, non-climate-change-related arguments for renewable energy development become stronger, clean energy policies become more politically feasible, and renewable energy is likely to undergo more rapid market penetration.

1.2 Lignocellulosic Biofuels

Lignocellulosic biofuels are short chain hydrocarbons produced from the carbohydrate building blocks of plant cell walls. For automobile fuels, the most common examples are ethanol, butanol, and methyl esters; for jet fuels, cyclic isoprenoids and alkanes (*15*). In principle, any combustible compound produced from plant cell wall sugars is a lignocellulosic

biofuel. As more catalytic strategies are developed to interconvert and upgrade short chain hydrocarbons, the number of lignocellulosic fuel possibilities will continue to expand. The immense advantage of such fuels over other renewables is their ability to closely approximate the behavior of existing liquid fuel components and thereby minimize the need to dramatically modify the existing transportation infrastructure.

The renewable nature of lignocellulosic fuels is based on the carbon cycle with which they are associated. Carbon dioxide from the Earth's atmosphere is converted into carbohydrate polymers in plants via photosynthesis. Biomass thus accumulated is harvested, depolymerized, and converted into fuel products via chemical or biochemical catalysis. Fuels thus generated are combusted in automobile or airplane engines and released as carbon dioxide and water back into the atmosphere. The net greenhouse gas emissions resulting from biofuel production and use are plant-, process-, and fuel-specific (*16*), but most studies suggest that they can be negative or neutral if regionally-specific production strategies are employed (*17*).

Lignocellulosic biomass consists primarily of lignin (10-25%), hemicellulose (20-35%) and cellulose (35-50%) (*18*). Fuels can be produced from the sugars in hemicellulose, a branched heteropolymer of pentose and hexose sugars, and cellulose, an unbranched polymer of β-[1,4]-linked glucose. Lignin, a cross-linked polymer of aromatic alcohols, is burned or converted to phenolic resins or polymers in most proposed fuel processes (*19*); however, it can also be converted to fuels via thermochemical processes (*20*). Sources of lignocellulosic biomass include miscanthus, switchgrass, sweet sorghum, sugarcane bagasse, corn stover, wheat straw, pine, poplar, or eucalyptus (*18*). These feedstocks differ in their water and nutrient requirements; their use as energy crops will likely be dictated by the climatic conditions of specific biofuel production facilities, which must be located near feedstock farms to minimize greenhouse gas emissions associated with biomass transport (*17, 21-23*).

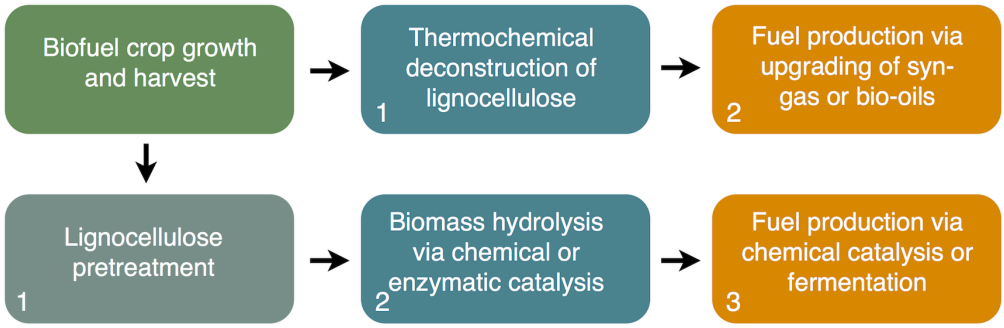


Figure 1.1 Process steps for the production of lignocellulosic biofuels.

Most proposed lignocellulosic biofuels production processes rely on either thermochemical or hydrolytic processes to convert biomass to fuel. Thermochemical strategies involve deconstruction of lignocellulosic substrates at high temperatures and pressures to yield upgradable intermediates such as bio-oils or syngas (CO/H_2 mixtures) (*20*); in a second step, these products undergo chemical/catalytic upgrading via processes such as Fischer-Tropsch synthesis or hydrodeoxygenation (*20*) (Fig. 1.1). Hydrolytic strategies, by contrast, rely on dissection of the biomass and generally consist of three main steps (Fig. 1.1) (*24*): (1) biomass pretreatment, (2) biomass depolymerization, and (3) fuels synthesis. In the first step, biomass is broken apart to increase surface area, reduce lignin and hemicellulose content, and increase

cellulose accessibility. In the second step, cellulose and residual hemicellulose are solubilized via chemical or enzymatic catalysis. In the third step, soluble sugars are converted into fuel products via chemical catalysis or microbial fermentation. Modern biofuels research is focused on minimizing the carbon emissions, maximizing the product yields, and reducing the costs associated with each of these steps.

1.3 The Plant Cell Wall: Structure and Composition

Before we detail the processes required to deconstruct the plant cell wall, we briefly describe its architecture. There are about 35 types of plant cells (*25*); all contain primary cell walls, but only a subset contains the secondary cell wall. Primary cell walls, which surround growing plant cells, are made up primarily of cellulose, hemicellulose, and pectin (*25, 26*). Secondary cell walls, which are mainly deposited after plant cell enlargement terminates, are made up primarily of cellulose, hemicellulose, and lignin. The composition of lignocellulosic substrates is dictated by the composition of secondary cell walls.

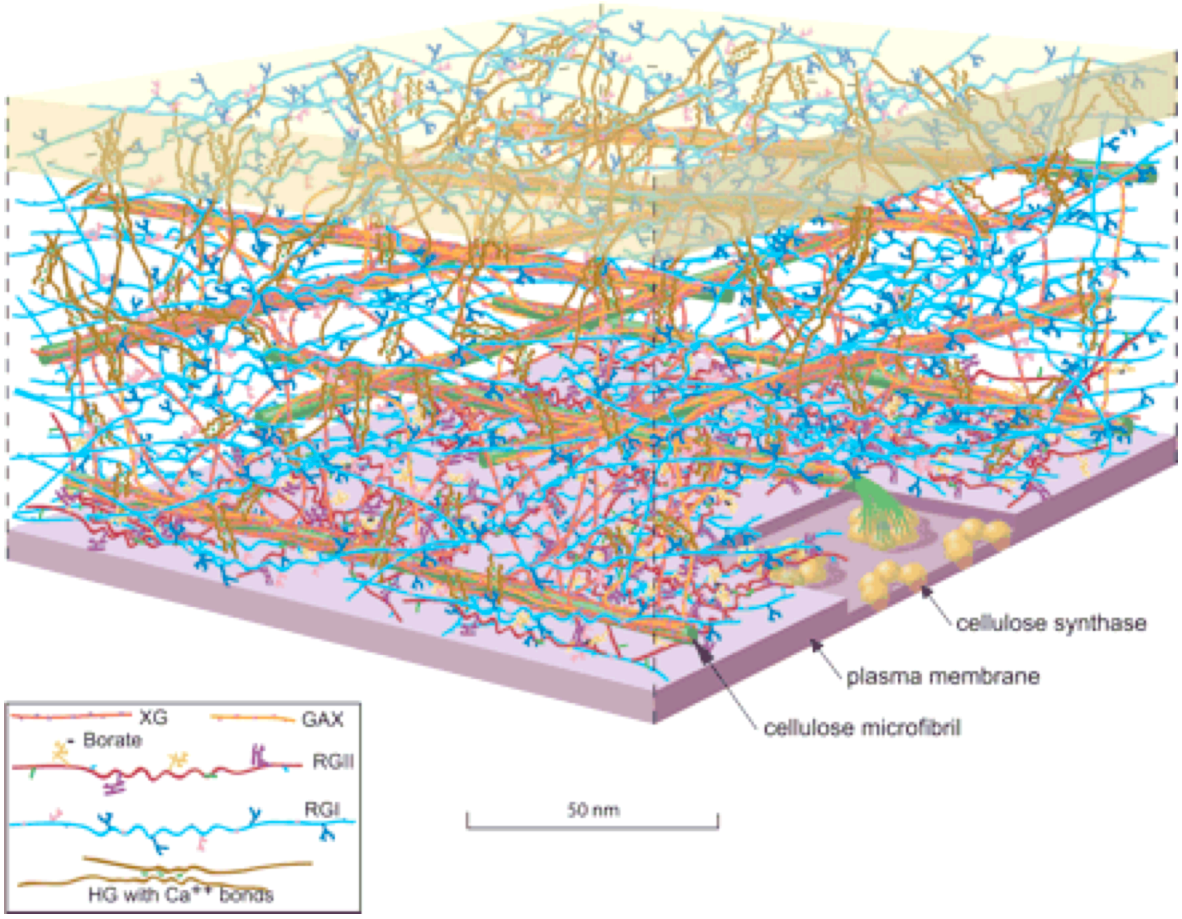


Figure 1.2 Scale model of the primary wall polysaccharides in an *Arabidopsis* leaf cell (reproduced from reference (*26*)). Polymers correspond to xyloglucan (XG), glucoronoariabinoxylan (GAX), rhamnogalacturonan I (RGI), rhamnogalacturonan II (RGII), and homogalacturonan with calcium bridges between carboxyl groups (HG with Ca++ bonds). The abundance of particular polymers is based on their relative abundance in the plant cell wall. Secondary cell walls are similarly organized, but they contain large quantities of lignin interspersed throughout the cellulose microfibrils. Pectin is also a minority component in secondary cell walls.

Cellulose, the main structural component of the plant cell wall, is a condensation polymer of D-anhydroglucopyranose monomers linked via β-[1,4]-glycosidic bonds (*27*). Individual cellulose chains associate with each other via inter- and intrachain hydrogen bonds (O-H-O) and intersheet hydrogen bonds (C-H-O) to form microfibrils that are 2-4 nm wide and 0.1-100 μm long (Fig. 1.2) (*28*). Van der Waal forces from pyranose ring stacking also play a significant role in interchain associations (*24, 28*). As plant cells grow, these microfibrils are deposited into a matrix of hemicellulose and pectin polymers (Fig. 1.2). Naturally occurring crystal structures of cellulose take on two phases: Iα, which is triclinic unit cell, or Iβ, which is monoclinic (*28-30*); the Iβ allomorph is dominant in higher plants. Hemicelluloses that become trapped in microfibrils as they form can distort these crystal structures, thereby brining about a spectrum of cellulose morphologies within the developing wall (*25*).

Hemicellulose chains have branches that prevent them from forming microfibrils through self-association. Xyloglucan, a β-[1,4]-linked glucose backbone substituted with xylose sidechains, and arabinoxylan, a [1,4]-linked β-D-xylan backbone substituted with arabinose sidechains, are the two most abundant hemicelluloses (*25*). These hemicellulose are believed to crosslink cellulose microfibrils via non-covalent interactions (Fig. 1.2) (*31*); their reorganization during plant cell growth allows for cellulose fibrils to move apart, thus facilitating the integration of new cell wall material.

Pectins play a major role in primary plant cell walls. These polysaccharides have uronic acids as major components (*26*). The most common pectins are (1) homogalcaturonan (HG), an unbranched polymer of α-[1,4]-linked galacturonic acid; (2) rhamnogalactruonan I (RGI), a backbone of α-[1,4]-L-rhamnose-β-[1,4]-D-galacturonic acid with arabinan and galactan sidechains, and (3) rhamnogalacturonan II (RGII), a sidechain similar in structure to RGI. Pectin domains serve as additional cross-links to strengthen the plant cell wall (Fig. 1.2); together with hemicellulose, they control the plant cell wall porosity (*25*). Pectins make up a small percentage of secondary cell walls, and thus comprise only 1-4% of lignocellulosic biomass (*18*).

Lignin is a complex polymer of p-hydroxyphenyls, guaiacyls, and syringyls (*32*). The precise ratio of these components varies by source. Lignin has a high level of heterogeneity that is the result of the random radical-radical coupling reactions that drive its formation (*32*). The primary role of lignin is to stabilize the secondary plant cell wall, which requires immense mechanical strength in growing plants.

Hemicellulose, cellulose, pectin, and lignin form complex cross-linked polymeric networks that can support the growth of individual cells while guiding the development of large, multi-cellular structures. Efficient deconstruction of those networks presents a major challenge for biofuels production.

1.4 Biomass Pretreatment

During pretreatment, physical, chemical, or biological means are used to increase the surface area of lignocellulosic material or to otherwise modify its structure, making it more susceptible to depolymerizing agents (catalysts or enzymes). Cellulose is the most recalcitrant component of lignocellulosic biomass, and is therefore the primary target of most pretreatment

processes. This polymer forms tight aggregates that protect most of its β-glycosidic linkages from hydrolysis. Pretreatment processes serve to increase cellulose accessibility by (1) removing other cell wall polymer obstructions (the "lignin-hemicellulose barrier") and/or by (2) breaking apart its native crystalline structure, thereby exposing β-glycosidic linkages that might otherwise be protected from hydrolytic cleavage (*18, 24*).

Commonly proposed pretreatments include dilute acid, ammonia fiber expansion, ionic liquid, hot water, and steam explosion. In dilute acid pretreatment, lignocellulosic materials are subjected to 1-2% w/w H_2SO_4/biomass at 160°-220°C for 1-30 minutes (*33*). This process removes most of the hemicellulose by converting it to monomeric sugars. During ammonia fiber expansion (AFEX), lignocellulosic substrates are subjected to 50-100% w/w ammonia/biomass at 40-180°C and 200-1000 psi for 5-45 minutes followed by rapid decompression (*24, 34*). This process only partially hydrolyzes hemicellulose, but it depolymerizes lignin and reduces the cellulose crystallinity. During ionic liquid pretreatment, biomass is dissolved in ionic liquid solutions containing little or no water. With the addition of water, cellulose can be pseudo-selectively precipitated; the regenerated product is less crystalline and contains significantly less lignin and hemicellulose (*35, 36*). Hot water pretreatment is an alternative to caustic chemical strategies that involves exposing lignocellulose to 160-230°C water at high pressures for 10-30 minutes (*24, 33*). Under these conditions, hemicellulose and amorphous cellulose are partially hydrolyzed. Steam explosion is a variation on hot water pretreatment where the pressure is explosively released, thus weakening the cellulose fibers and increasing particle surface area.

All pretreatments reduce the lignin and/or hemicellulose content of biomass and increase the susceptibility of cellulose to hydrolysis. Though they differ in their effectiveness, their most important differences relate to cost and a tendency to generate toxic byproducts, which lower the efficiency of fuel synthesis. Pretreatment choice therefore depends on overall process considerations. Several recently published technoeconomic analyses explore these considerations in greater detail.

1.5 Biomass Depolymerization

The depolymerization of biomass into soluble sugars can be accomplished with chemical or enzyme catalysts. Inorganic catalysts such as mineral acids (e.g., H_2SO_4), sulfonated resins (e.g., Amberlyst), or inorganic oxides (e.g., ZrO_2) will rapidly hydrolyze cellulose and hemicellulose at high temperatures (>100°C) (*20, 37*). Unfortunately, these methods require significant amounts of reagent (mineral acids in the first case, and ionic liquid solvents in the latter two cases), rely on high temperatures, and/or generate degradation products (hydroxymethylfurfural, levulinic acid, and insoluble humins in the case of H_2SO_4) (*20, 38*). Enzymatic depolymerization strategies, by contrast, use aqueous reaction environments and are highly selective (no byproducts). Hemicellulase and cellulase enzymes, which are naturally secreted by biomass-utilizing organisms, can hydrolyze hemicellulose and cellulose under mild conditions (30-65°, pH of 4-6) .

The lack of degradation products associated with enzymatic depolymerization make this biocatalytic strategy favorable for microbial fuel production processes. These are the focus of the present work.

Enzymatic depolymerization is dominated by cellulose saccharification. Hemicellulose remaining after pretreatment is structurally diffuse and rapidly broken down by hemicellulases, which hydrolyze glycosidic bonds and ester linkages of acetate or ferulic acid side groups (*39*). Cellulose, however, remains in closely associated aggregates, and is slowly depolymerized by multi-enzyme mixtures of hydrolytic and oxidative enzymes. As the majority of hemicellulose is removed during pretreatment, and as any remaining hemicellulose can be rapidly broken down by low concentrations of hemicellulases, the enzymatic mixtures employed in the depolymerization step are dominated by cellulolytic enzymes (*24*).

Cellulolytic enzymes account for a significant fraction of the cost associated with cellulosic biofuel production and are thus the focus of many ongoing research efforts. Recent estimates of the cost contribution of these enzymes to ethanol produced from corn stover range from $0.68 to $1.47/gallon (*40*). To reduce enzyme costs, many industrial and academic efforts have focused on improving the activity of cellulolytic mixtures through either mixture optimization or improvement of enzyme activity and/or thermostability (*41-44*). Before we discuss these improvement strategies, we will describe the cellulolytic enzyme systems in greater detail.

Cellulases are the most active component of the enzyme systems employed by biomass-degrading organisms. These enzymes, which catalyze the hydrolysis of the β-1,4-glycosidic linkages between adjacent glucose units in cellulose, are deployed within multi-enzyme complexes called cellulosomes or within secreted multi-component mixtures. Cellulosomes, which are found in cell wall protuberances of anaerobic bacteria, consist of numerous hemicellulase and/or cellulase enzymes linked to a scaffold protein called scaffoldin (*45, 46*). The architecture of the *Clostridium thermocellum* cellulosomes is a representative example. Its scaffoldin contains a number of domains that are linked by short peptide linkers: a single carbohydrate-binding module, which allows it to adsorb to the cellulose surface, and 9 cohesin domains, which allow it to template the hemicellulase and cellulase substituents via cohesion-dockerin interactions (*46*). Non-complexed cellulases, by comparison, can have similar modular architectures, but never contain more than one catalytic domain. These enzymes are secreted within multi-enzyme mixtures by biomass-degrading fungi and bacteria. Their canonical architecture consists of a catalytic domain connected to a CBM via a 6-109 peptide linker (*47, 48*); however, variants containing no CBMs, multiple CBMs, and/or no linker also exist. The specific activities (biomass hydrolyzed/mg enzyme/hr) of cellulosomes and multi-component cellulolytic mixtures are comparable (*49*), but the large size of cellulosomes (2-16 MDa) makes them far more difficult to produce at the >100 g/L titers which have been achieved for secreted non-complexed enzymes (*50*). Thus, most work on improving cellulolytic enzyme performance for use in industrial biofuel production has focused on non-complexed enzyme systems.

The enzymes of non-complexed cellulolytic mixtures employ one of two catalytic strategies to depolymerize cellulose: hydrolysis and oxidative cleavage. Below, we introduce the enzymes relevant to each strategy.

1.5.1 Hydrolytic Processes

The non-complexed cellulolytic system contains three main classes of hydrolytic enzymes: (1) cellobiohydrolases (or exoglucanases, EC 3.2.1.91), which hydrolyze cellulose chains processively from either the reducing- or non-reducing ends (*51, 52*); (2) endoglucanases (EC 3.2.1.4), which hydrolyze cellulose chains non-processively at any exposed glycosidic linkage (*53*); and (3) β-glucosidases (3.2.1.21), which hydrolyze soluble cello-oligosaccharide chains (degree of polymerization shorter than 7) at their terminal glycosidic bonds (*54*). All of these glycosyl hydrolases have active sites consisting of two glutamate and/or aspartate residues positioned to promote general acid-catalyzed hydrolysis of the β-1,4-bonds that link the repeating glucose subunits of cellulose. Two distinct mechanisms are employed (Fig. 1.3B): (1) retaining, which results in a net retention of the anomeric carbon's configuration, and (2) non-retaining, which results in an inversion of configuration. In the retaining mechanism (S_N1), the catalytic residues are separated by about 5.5 Å, and both interact directly with the carbohydrate substrate (Fig. 1.3): (1) a glycyosidic oxygen is protonated by an aspartate/glutamate; (2) a basic glutamate attacks the anomeric carbon, thus providing nucleophilic assistance to the aglycon leaving group and thereby generating a glucosyl-enzyme intermediate; and (3) the glucosyl-enzyme intermediate is hydrolyzed by a water (*55*). In the non-retaining mechanism (S_N2), the catalytic residues are separated by a larger distance of 10 Å, which allows for direct carbohydrate-interaction by only one catalytic residue; simultaneous protonation of the glycosidic oxygen by an aspartate/glutamate and nucleophilic attack by a glutamate-activated water result in cleavage of the glycosidic bond (*55*). All glycosyl hydrolases have been classified into over 130 families based on their amino acid sequence similarities in the Carbohydrate Active Enzyme Database (cazy.org); enzymes within each family often share substrate specificities and catalytic mechanisms. Cellobiohydrolases, endoglucanases, and β-glucosidases each have representatives in multiple families.

Cellobiohydrolases and endoglucanases catalyze hydrolysis via several discrete physical steps (Fig. 1.3A). A cellobiohydrolase (1) adsorbs to the cellulose surface via its CBM domain (*56-58*); (2) it complexes with a cellulose chain end via its catalytic domain, possibly after diffusing along the surface (*59*); (3) it catalyzes an initial hydrolysis event, generating glucose, cellobiose, or cellotriose (*60-62*); (4) it processively catalyzes 10-100 subsequent hydrolysis events, generating only cellobiose (*63, 64*); (5) it decomplexes from the cellulose chain (*65*); (6) it recomplexes with a cellulose chain or desorbs from the surface. Endoglucanases undergo similar steps, but they are generally thought to engage in non-processive random attack.

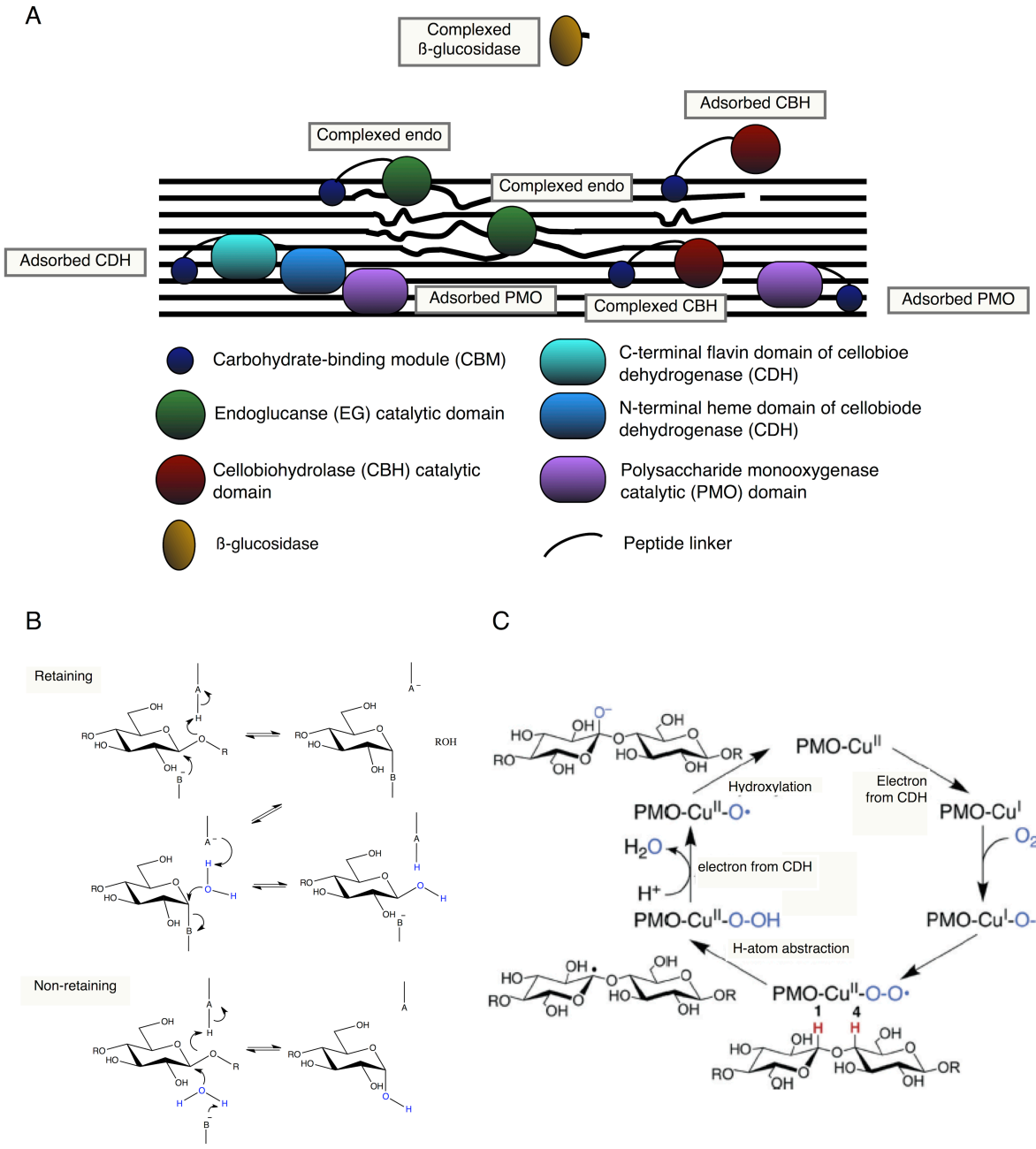


Figure 1.3 Mechanisms of key components of secreted non-complexed cellulolytic enzyme mixtures. (A) Endoglucanase, cellobiohydrolase, β-glucosidase, cellobiose dehydrogenase, and polysaccharide monooxygenase enzymes engaged in hydrolysis. Various cellulose morphologies and enzyme architectures are represented. On the left side of the diagram, the N-terminal heme domain is depicted as being in close proximity to a PMO as it is might be when it transfers an electron to the PMO. (B) The retaining and non-retaining mechanisms of glycosyl hydrolase enzymes (A=acid catalyst; B=base): for retaining, (1) a glycyosidic oxygen is protonated; (2) a basic glutamate attacks the anomeric carbon; (3) the resulting glucosyl-enzyme intermediate is hydrolyzed by a water (*55*). In the non-retaining mechanism, simultaneous protonation of the glycosidic oxygen by an aspartate/glutamate and nucleophilic attack by a glutamate-activated water result in cleavage of the glycosidic bond. (C) The propose mechanism of the PMO: the PMO Cu(II) is reduced to Cu(I); O_2 binds Cu(I); this superoxo species abstracts a hydrogen from the 1 position of an internal glucose residue; the CDH yields a second electron to the PMO, triggering the release of water and the generation of a copper oxo radical; the PMO hydroxylates the 1 position, thereby triggering the elimination of the glycosidic linkage.

Though there are only two main cellulase classes, biomass-degrading bacteria and fungi produce multiple variants of each. The filamentous fungus *Neurospora crassa* is a representative example. Its genome contains nine predicted cellulase genes, which undergo different levels of transcription during growth on biomass (*66*). When cellulose is used as a carbon source, the *N. crassa* secretome is dominated by two cellobiohydrolases (53%), three endoglucanases (13%), and a β-glucosidase (3.8%) (*67*). Similar to other Ascomycetes, the dominant cellobiohydrolase (39.5%) is a member of GH family 7, which implies that it engages cellulose chains from the reducing end and employs a retaining mechanism (*68*); the minority cellobiohydrolase is a member of GH family 6, which implies that it engages cellulose chains from the non-reducing end and employs a non-retaining mechanism (*69*). The endoglucanase enzymes are from families 5, 6, and 7; the β-glucosidase, family 3. While the two varieties of cellobiohydrolases have been shown to act synergistically on cellulose (*70*), the advantage conferred by the secretion of multiple endoglucanase variants has remained poorly understood . We explore the possible advantages of cellulase variants within multi-enzyme mixtures later on in this work.

1.5.2 Oxidative Processes

Non-complexed cellulolytic mixtures are thought to facilitate oxidative cleavage of cellulose either directly through polysaccharide monooxgenases (PMOs) or indirectly through the generation of hydroxyl radicals via Fenton chemistry. PMOs, which were originally classified as family 61 glycosyl hydrolases, are metalloeznymes that work in concert with cellobiose dehydrogenases (CDHs) and/or other protein/small molecule electron donors to hydroxylate glucose carbons (C1 or C4) proximal to internal glycosidic linkages of cellulose chains, thus triggering β-1,4 bond cleavage via elimination (*71*). Again, the cellulolytic system of *N. crassa* contains representative examples. With cellulose as the carbon source, the secretome of this fungus has been shown to include a CDH (CDH-1, 2.4%), an oxidoreductase with an N-terminal heme domain and a C-terminal flavin domain, and four PMOs (15%), monooxygenases which contain N-terminal copper coordination sites (*67, 71*). These enzymes act through the following mechanism (*71, 72*) (Figs. 1.3A and 1.3C): (1) the catalytic flavin-containing domain of CDH oxidizes the reducing end of either a cellulose chain or a cello-oligomer, generating a lactone; (2) the heme domain of the CDH transfers one electron to a PMO, reducing its Cu(II) to Cu(I); (3) O_2 binds the Cu(I) of the PMO, thereby generating a copper superoxo intermediate; (4) the superoxo species generates a carbohydrate radical by abstracting a hydrogen from the 1 or 4 position of a glucose residue internal to a cellulose chain, thus oxidizing Cu(I) back to Cu(II); (5) the CDH yields a second electron to the PMO, triggering the release of water and the generation of a copper oxo radical; (6) the PMO hydroxylates the 1 or 4 position from which it previously abstracted a hydrogen, thereby triggering the elimination of the glycosidic linkage. Evidence for direct enzymatic oxidation of cellulose has only emerged over the last several years (*73, 74*); thus, the detailed mechanism is still being worked out (*75*). By contrast, indirect oxidative cleavage via the Fenton reaction has been discussed for several decades (*76, 77*). This mechanism relies on the Fe(II)-catalyzed decomposition of H_2O_2 into water and a hydroxyl radicals (HO•), which can depolymerize cellulose and other cell wall polymers via random oxidation. Many basidiomycetes, which are efficient fungal degraders of lignocellulosic biomass, secrete reductase and oxidase enzymes, sources of Fe(II) and H_2O_2, in addition to cellulases (*76, 78*). These enzymes have been proposed to enhance cellulose degradation rates by generating hydroxyl radicals near the substrate surface. Though the

generation of hydroxyl radicals during biomass depolymerization has been shown, however, the role of these radicals in microbial cellulose degradation is not well understood (*79*). In general, both of the direct (enzymatic) and indirect (hydroxyl radical) oxidative systems have a very clear advantage over cellulolytic systems: they obviate the need for a cellulose chain abstraction step, which is energetically costly (*80*). Future investigation of their mechanisms of targeting and synergy will allow for the development of more efficient cellulolytic mixtures.

1.6 Fuels Synthesis

Soluble sugars generated from the depolymerization of cellulose and hemicellulose can be reduced to fuels via microbial fermentation, chemical catalysis, or a combination of the two. In microbial fuel production, carbohydrates consumed by fungi or bacteria are converted to fuel products via endogenous or engineered (nonnative) metabolic pathways. Organisms such as *Saccharomyces cerevisiae, Escherichia coli*, *Clostridium acetobutylicum*, and *Zymomonas mobilis* have been engineered to produce ethanol, butanol, fatty esters, or fatty alcohols at substantial yields from plant-derived C5 and C6 sugars (*81-86*). Pathways for alkanes and cyclic isoprenoid (jet fuel) production within several of these organisms are also being pursued (*15*). In catalytic fuel production, solid catalysts are employed in multi-step reactions designed to convert carbohydrates to compounds with a lower oxygen content and greater density of C-C bonds (*20*). Several highlights include (1) the generation of alkanes via aqueous-phase dehydration/hydrogenation of C6 sugars (*87*), (2) the production of 2,5-dimehtylfuran from fructose dehydration to HMF followed by hydrogenolysis to DMF (*88*), and (3) aqueous phase reforming of carbohydrates to light alkanes, syngas, and monofunctional hydrocarbons (*89*). Combined fermentative/chemocatlalytic approaches involve the use of fermentation products in catalysis feed streams. Recently, a combined approach was demonstrated using *C. acetobutylicum* and palladium catalysts: acetone, butanol, and ethanol, which were selectively extracted from *C. thermocellum* fermentation broth via glycerol tributerate, were converted into ketones by palladium-catalyzed alkylation (*90*). With continual advances in the fields of metabolic engineering and catalyst design, the possibilities for fuel generation via microbial fermentation and chemical catalysis are expanding. All rely on reducing the carbohydrate sugars produced from biomass into compounds with higher energy densities and good combustion characteristics. Future work on fuel production will involve the design of synthesis routes that are highly selective and high yielding, and which require minimal energy inputs (high temperatures).

1.7 Cellulose Hydrolysis and the Future of Biofuels

Most life cycle assessments indicate that lignocellulosic biofuels can mitigate greenhouse gas emissions caused by transportation (*17*). Unfortunately, production prices (~$4.00) remain uncompetitive with those for conventional fossil fuels ($3.5 at the pump as of November 3, 2012 in the U.S.) (*91, 92*). To achieve cost parity, biofuels research currently underway is aimed at improving biomass-to-fuel yields, conversion rates, and product characteristics so that high quality lignocellulosic biofuels can become cost competitive with petrochemical alternatives (*24, 93*).

Inefficient cellulose depolymerization presents one of the greatest opportunities for reducing cellulosic biofuels production costs. By most estimates, high cellulolytic enzyme requirements are one of the largest barriers to successful competition between cellulosic fuels and fossil fuel alternatives (*93, 94*). Improvements in the stability and activity oft these enzymes can significantly lower production costs by lowering cellulase/cellulose loading requirements (*40*). In the present work, we describe mechanistic models and experimental approaches that we employed to investigate the kinetic details of cellulase-catalyzed hydrolysis of cellulose. Our results reveal important rate-limiting phenomena and demonstrate new strategies for improving and/or optimizing the activity of cellulolytic mixtures.

1.8 References

1. S. Chu, A. Majumdar, Opportunities and challenges for a sustainable energy future. *Nature* 488, 294 (Aug 16, 2012).
2. S. Pacala, R. Socolow, Stabilization wedges: Solving the climate problem for the next 50 years with current technologies. *Science* 305, 968 (Aug 13, 2004).
3. K. R. Smith, J. M. Samet, I. Romieu, N. Bruce, Indoor air pollution in developing countries and acute lower respiratory infections in children. *Thorax* 55, 518 (Jun, 2000).
4. P. Wilkinson, K. R. Smith, M. Joffe, A. Haines, A global perspective on energy: health effects and injustices. *Lancet* 370, 965 (Sep 15, 2007).
5. T. L. Friedman, The First Law of Petropolitics. *The Ecologist* 36, (2006).
6. E. Regnier, Oil and Energy Price Volatility. *Energy Economics* 29, 405 (2005).
7. S. Awerbuch, Portfolio-based Electricity Generation Planning: Poly Implications for Renewables and Energy Security. *Mitigation and Adaptation Strategies for Global Change* 11, 693 (2006).
8. D. Rajagopal, D. Zilberman, "Review of Environmental, Economic, and Policy Aspects of Biofuels" (The Wod Bank Deveopment Research Group, 2007).
9. P. Forster *et al.*, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I o the Fourth Assessment Report of the Intergovernmental Panel on limate Change*. (Cambridge University Press, Cambridge, United Kingdom, 2007).
10. .
11. B. Amigun, H. von Blottnitz, Investigation of Scale Economies for African Biogas Installations. *Environmental and Process Systems Engineering Research Group*, (2007).
12. D. Dai, Z. Hu, G. Pu, H. Li, C. Wang, Energy Efficiency and Potentials of Cassava Fuel Ethanol in Guangxi Region in China. *Energy Conversion and Management* 47, 1686 (2005).
13. M. S. Dresselhaus, I. L. Thomas, Alternative energy technologies. *Nature* 414, 332 (Nov 15, 2001).
14. R. Hahn, P. Passell, The Economics of Allowing More U.S. Oil Drilling. *Energy Economics* 32, 638 (2010).
15. J. L. Fortman *et al.*, Biofuel alternatives to ethanol: pumping the microbial well. *Trends Biotechnol* 26, 375 (Jul, 2008).
16. B. Strogen, A. Horvath, Greenhouse Gas Emissions from the Construction, Manufacturing, Operation and Maintenance of US Distribution Infrastructure for Petroleum and Biofuels. *Journal of Infrastructure Systems*, (2012).
17. S. C. Davis, K. J. Anderson-Teixeira, E. H. Delucia, Life-cycle analysis and the ecology of biofuels. *Trends Plant Sci* 14, 140 (Mar, 2009).
18. A. Carroll, C. Somerville, Cellulosic biofuels. *Annu Rev Plant Biol* 60, 165 (2009).
19. T. P. Vispute, H. Zhang, A. Sanna, R. Xiao, G. W. Huber, Renewable chemical commodity feedstocks from integrated catalytic processing of pyrolysis oils. *Science* 330, 1222 (Nov 26, 2010).

20. J. C. Serrano-Ruiz, R. M. West, J. A. Dumesic, Catalytic conversion of renewable biomass resources to fuels and chemicals. *Annu Rev Chem Biomol Eng* 1, 79 (2010).
21. C. Somerville, H. Youngs, C. Taylor, S. C. Davis, S. P. Long, Feedstocks for lignocellulosic biofuels. *Science* 329, 790 (Aug 13, 2010).
22. H. Youngs, C. Somerville, Development of feedstocks for cellulosic biofuels. *F1000 Biol Rep* 4, 10 (2012).
23. B. Strogen, A. Horvath, T. E. McKone, Fuel miles and the blend wall: costs and emissions from ethanol distribution in the United States. *Environ Sci Technol* 46, 5285 (May 15, 2012).
24. S. P. Chundawat, G. T. Beckham, M. E. Himmel, B. E. Dale, Deconstruction of lignocellulosic biomass to fuels and chemicals. *Annu Rev Chem Biomol Eng* 2, 121 (2011).
25. D. J. Cosgrove, Growth of the plant cell wall. *Nat Rev Mol Cell Biol* 6, 850 (Nov, 2005).
26. C. Somerville *et al.*, Toward a systems approach to understanding plant cell walls. *Science* 306, 2206 (Dec 24, 2004).
27. B. B. Buchanan, W. Gruissem, R. L. Jones, *Biochemistry & molecular biology of plants*. (American Society of Plant Physiologists, Rockville, Md., 2000), pp. xxxix, 1367 p.
28. A. S. Gross, J. W. Chu, On the Molecular Origins of Biomass Recalcitrance: The Interaction Network and Solvation Structures of Cellulose Microfibrils. *Journal of Physical Chemistry B* 114, 13333 (Oct 28, 2010).
29. Y. Nishiyama, P. Langan, H. Chanzy, Crystal structure and hydrogen-bonding system in cellulose Ibeta from synchrotron X-ray and neutron fiber diffraction. *Journal of the American Chemical Society* 124, 9074 (Aug 7, 2002).
30. Y. Nishiyama, J. Sugiyama, H. Chanzy, P. Langan, Crystal structure and hydrogen bonding system in cellulose I(alpha) from synchrotron X-ray and neutron fiber diffraction. *Journal of the American Chemical Society* 125, 14300 (Nov 26, 2003).
31. M. C. Mccann, B. Wells, K. Roberts, Direct Visualization of Cross-Links in the Primary Plant-Cell Wall. *J Cell Sci* 96, 323 (Jun, 1990).
32. W. Boerjan, J. Ralph, M. Baucher, Lignin biosynthesis. *Annu Rev Plant Biol* 54, 519 (2003).
33. S. K. Khanal, D. Takara, in *Green Energy and Technology*. (2012).
34. H. K. Murnen *et al.*, Optimization of ammonia fiber expansion (AFEX) pretreatment and enzymatic hydrolysis of Miscanthus x giganteus to fermentable sugars. *Biotechnol Prog* 23, 846 (Jul-Aug, 2007).
35. K. Shill *et al.*, Ionic Liquid Pretreatment of Cellulosic Biomass: Enzymatic Hydrolysis and Ionic Liquid Recycle. *Biotechnol Bioeng* 108, 511 (Mar, 2011).
36. S. H. Lee, T. V. Doherty, R. J. Linhardt, J. S. Dordick, Ionic liquid-mediated selective extraction of lignin from wood leading to enhanced enzymatic cellulose hydrolysis. *Biotechnol Bioeng* 102, 1368 (Apr 1, 2009).
37. R. Rinaldi, N. Meine, J. vom Stein, R. Palkovits, F. Schuth, Which controls the depolymerization of cellulose in ionic liquids: the solid acid catalyst or cellulose? *ChemSusChem* 3, 266 (Feb 22, 2010).
38. S. Larsson *et al.*, The Generation of Fermentation Inhibitors During Dilute Acid Hydrolysis of Softwood. *Enzyme and Microbial Technology* 24, 151 (1999).
39. D. Shallom, Y. Shoham, Microbial hemicellulases. *Curr Opin Microbiol* 6, 219 (Jun, 2003).
40. D. Klein-Marcuschamer, P. Oleskowicz-Popiel, B. A. Simmons, H. W. Blanch, The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnol Bioeng* 109, 1083 (Apr, 2012).
41. S. E. Levine, J. M. Fox, D. S. Clark, H. W. Blanch, A mechanistic model for rational design of optimal cellulase mixtures. *Biotechnol Bioeng* 108, 2561 (Nov, 2011).
42. A. Berlin, V. Maximenko, N. Gilkes, J. Saddler, Optimization of enzyme complexes for lignocellulose hydrolysis. *Biotechnol Bioeng* 97, 287 (Jun 1, 2007).
43. C. M. Dana *et al.*, Biased clique shuffling reveals stabilizing mutations in cellulase Cel7A. *Biotechnol Bioeng* 109, 2710 (Nov, 2012).

44. P. Heinzelman *et al.*, A family of thermostable fungal cellulases created by structure-guided recombination. *Proc Natl Acad Sci U S A* 106, 5610 (Apr 7, 2009).
45. C. Boisset *et al.*, Digestion of crystalline cellulose substrates by the clostridium thermocellum cellulosome: structural and morphological aspects. *Biochem J* 340 (Pt 3), 829 (Jun 15, 1999).
46. E. A. Bayer, E. Morag, R. Lamed, The cellulosome--a treasure-trove for biotechnology. *Trends Biotechnol* 12, 379 (Sep, 1994).
47. G. K. Sonan *et al.*, The linker region plays a key role in the adaptation to cold of the cellulase from an Antarctic bacterium. *Biochem J* 407, 293 (Oct 15, 2007).
48. N. S. Mosier, P. Hall, C. M. Ladisch, M. R. Ladisch, Reaction kinetics, molecular action, and mechanisms of cellulolytic proteins. *Adv Biochem Eng Biotechnol* 65, 23 (1999).
49. L. R. Lynd, P. J. Weimer, W. H. van Zyl, I. S. Pretorius, Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev* 66, 506 (Sep, 2002).
50. M. Saloheimo, T. M. Pakula, The cargo and the transport system: secreted proteins and protein secretion in Trichoderma reesei (Hypocrea jecorina). *Microbiology* 158, 46 (Jan, 2012).
51. C. Divne *et al.*, The 3-Dimensional Crystal-Structure of the Catalytic Core of Cellobiohydrolase-I from Trichoderma-Reesei. *Science* 265, 524 (Jul 22, 1994).
52. J. Rouvinen, T. Bergfors, T. Teeri, J. K. C. Knowles, T. A. Jones, 3-Dimensional Structure of Cellobiohydrolase-Ii from Trichoderma-Reesei. *Science* 249, 380 (Jul 27, 1990).
53. G. Sulzenbacher, F. Shareck, R. Morosoli, C. Dupont, G. J. Davies, The Streptomyces lividans family 12 endoglucanase: Construction of the catalytic core, expression, and X-ray structure at 1.75 angstrom resolution. *Biochemistry-Us* 36, 16032 (Dec 23, 1997).
54. K. Ohmiya, M. Shirai, Y. Kurachi, S. Shimizu, Isolation and properties of beta-glucosidase from Ruminococcus albus. *J Bacteriol* 161, 432 (Jan, 1985).
55. G. Davies, B. Henrissat, Structures and mechanisms of glycosyl hydrolases. *Structure* 3, 853 (Sep 15, 1995).
56. J. Lehtio *et al.*, The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. *Proc Natl Acad Sci U S A* 100, 484 (Jan 21, 2003).
57. N. Din *et al.*, C-1-C-X Revisited - Intramolecular Synergism in a Cellulase. *P Natl Acad Sci USA* 91, 11383 (Nov 22, 1994).
58. K. M. KlemanLeyer, M. SiikaAho, T. T. Teeri, T. K. Kirk, The cellulases endoglucanase I and cellobiohydrolase II of Trichoderma reesei act synergistically to solubilize native cotton cellulose but not to decrease its molecular size. *Appl Environ Microb* 62, 2883 (Aug, 1996).
59. E. J. Jervis, C. A. Haynes, D. G. Kilburn, Surface diffusion of cellulases and their isolated binding domains on cellulose. *J Biol Chem* 272, 24016 (Sep 19, 1997).
60. T. V. Vuong, D. B. Wilson, Processivity, Synergism, and Substrate Specificity of Thermobifida fusca Cel6B. *Appl Environ Microb* 75, 6655 (Nov 1, 2009).
61. B. Nidetzky, W. Zachariae, G. Gercken, M. Hayn, W. Steiner, Hydrolysis of Cellooligosaccharides by Trichoderma-Reesei Cellobiohydrolases - Experimental-Data and Kinetic Modeling. *Enzyme and Microbial Technology* 16, 43 (Jan, 1994).
62. M. Vrsanska, P. Biely, The Cellobiohydrolase-I from Trichoderma-Reesei Qm-9414 - Action on Cello-Oligosaccharides. *Carbohyd Res* 227, 19 (Apr 6, 1992).
63. K. Igarashi *et al.*, High Speed Atomic Force Microscopy Visualizes Processive Movement of Trichoderma reesei Cellobiohydrolase I on Crystalline Cellulose. *J Biol Chem* 284, 36186 (Dec 25, 2009).
64. G. T. Beckham *et al.*, Identification of Amino Acids Responsible for Processivity in a Family 1 Carbohydrate-Binding Module from a Fungal Cellulase. *J Phys Chem B* 114, 1447 (Jan 28, 2010).
65. M. Kurasin, P. Valjamae, Processivity of Cellobiohydrolases Is Limited by the Substrate. *J Biol Chem* 286, 169 (2011).
66. C. Tian *et al.*, Systems analysis of plant cell wall degradation by the model filamentous fungus Neurospora crassa. *Proc Natl Acad Sci U S A* 106, 22157 (Dec 29, 2009).

67. C. M. Phillips, A. T. Iavarone, M. A. Marletta, Quantitative proteomic approach for cellulose degradation by Neurospora crassa. *J Proteome Res* 10, 4177 (Sep 2, 2011).
68. C. Divne *et al.*, The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from Trichoderma reesei. *Science* 265, 524 (Jul 22, 1994).
69. J. Rouvinen, T. Bergfors, T. Teeri, J. K. Knowles, T. A. Jones, Three-dimensional structure of cellobiohydrolase II from Trichoderma reesei. *Science* 249, 380 (Jul 27, 1990).
70. K. Igarashi *et al.*, Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface. *Science* 333, 1279 (Sep 2, 2011).
71. C. M. Phillips, W. T. Beeson, J. H. Cate, M. A. Marletta, Cellobiose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by Neurospora crassa. *ACS Chem Biol* 6, 1399 (Dec 16, 2011).
72. W. T. Beeson, C. M. Phillips, J. H. Cate, M. A. Marletta, Oxidative cleavage of cellulose by fungal copper-dependent polysaccharide monooxygenases. *Journal of the American Chemical Society* 134, 890 (Jan 18, 2012).
73. P. V. Harris *et al.*, Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family. *Biochemistry* 49, 3305 (Apr 20, 2010).
74. R. J. Quinlan *et al.*, Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc Natl Acad Sci U S A* 108, 15079 (Sep 13, 2011).
75. J. A. Langston *et al.*, Oxidoreductive cellulose depolymerization by the enzymes cellobiose dehydrogenase and glycoside hydrolase 61. *Appl Environ Microbiol* 77, 7007 (Oct, 2011).
76. R. Cohen, K. A. Jensen, C. J. Houtman, K. E. Hammel, Significant levels of extracellular reactive oxygen species produced by brown rot basidiomycetes on cellulose. *FEBS Lett* 531, 483 (Nov 20, 2002).
77. J. W. Koenigs, Hydrogen Peroxide and Iron: a Proposed System for Decomposition of Wood by Brown-rot Basidiomycetes. *Wood and Fiber Science* 6, 66 (1974).
78. D. Martinez *et al.*, Genome, transcriptome, and secretome analysis of wood decay fungus Postia placenta supports unique mechanisms of lignocellulose conversion. *Proc Natl Acad Sci U S A* 106, 1954 (Feb 10, 2009).
79. S. Kaneko, K. Yoshitake, S. Itakura, H. Tanaka, A. Enoki, Relationship between production of hydroxyl radicals and degradation of wood, crystalline cellulose, and a lignin-related compound or accumulation of oxalic acid in cultures of brown-rot fungi. *Journal of Wood Science* 51, 262 (2005).
80. G. T. Beckham *et al.*, Molecular-level origins of biomass recalcitrance: decrystallization free energies for four common cellulose polymorphs. *J Phys Chem B* 115, 4118 (Apr 14, 2011).
81. E. J. Steen *et al.*, Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* 463, 559 (Jan 28, 2010).
82. J. Lee *et al.*, Metabolic engineering of Clostridium acetobutylicum ATCC 824 for isopropanol-butanol-ethanol fermentation. *Appl Environ Microbiol* 78, 1416 (Mar, 2012).
83. J. Y. Lee, Y. S. Jang, J. Lee, E. T. Papoutsakis, S. Y. Lee, Metabolic engineering of Clostridium acetobutylicum M5 for highly selective butanol production. *Biotechnol J* 4, 1432 (Oct, 2009).
84. S. Yang *et al.*, Paradigm for industrial strain improvement identifies sodium acetate tolerance loci in Zymomonas mobilis and Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* 107, 10395 (Jun 8, 2010).
85. H. Alper, J. Moxley, E. Nevoigt, G. R. Fink, G. Stephanopoulos, Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science* 314, 1565 (Dec 8, 2006).
86. L. O. Ingram, T. Conway, D. P. Clark, G. W. Sewell, J. F. Preston, Genetic engineering of ethanol production in Escherichia coli. *Appl Environ Microbiol* 53, 2420 (Oct, 1987).
87. G. W. Huber, J. N. Chheda, C. J. Barrett, J. A. Dumesic, Production of liquid alkanes by aqueous-phase processing of biomass-derived carbohydrates. *Science* 308, 1446 (Jun 3, 2005).

88. Y. Roman-Leshkov, C. J. Barrett, Z. Y. Liu, J. A. Dumesic, Production of dimethylfuran for liquid fuels from biomass-derived carbohydrates. *Nature* 447, 982 (Jun 21, 2007).
89. E. L. Kunkes *et al.*, Catalytic conversion of biomass to monofunctional hydrocarbons and targeted liquid-fuel classes. *Science* 322, 417 (Oct 17, 2008).
90. P. Anbarasan *et al.*, Integration of Chemical Catalysis with Extractive Fermentation to Produce Fuels. *Nature*, (2012).
91. D. Klein-Marcuschamer, P. Oleskowicz-Popiel, B. A. Simmons, H. W. Blanch, Technoeconomic analysis of biofuels: A wiki-based platform for lignocellulosic biorefineries. *Biomass and Bioenergy* 30, (2010).
92. U. E. I. Administration, in *Independent Statistics and Analysis*. (2012), vol. 2012.
93. M. E. Himmel *et al.*, Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* 315, 804 (Feb 9, 2007).
94. G. Stephanopoulos, Challenges in engineering microbes for biofuels production. *Science* 315, 801 (Feb 9, 2007).

Chapter 2: A Mechanistic Model of the Enzymatic Cellulose Hydrolysis*

2.1 Abstract

A detailed mechanistic model of enzymatic cellulose hydrolysis has been developed. The behavior of individual cellulase enzymes and parameters describing the cellulose surface properties are included. Results obtained for individual enzymes (*T. reesei* endoglucanase 2 and cellobiohydrolase I) and systems with both enzymes present are compared with experimental literature data. The model was sensitive to cellulase-accessible surface area; the EG2-CBHI synergy observed experimentally was only predicted at a sufficiently high cellulose surface area. Enzyme crowding, which is more apparent at low surface areas, resulted in differences between predicted and experimental rates of hydrolysis. Model predictions also indicated that the observed decrease in hydrolysis rates following the initial rate of rapid hydrolysis is not solely caused by product inhibition and/or thermal deactivation. Surface heterogeneities, which are not accounted for in this work, may play a role in decreasing the hydrolysis rate. The importance of separating the enzyme adsorption and complexation steps is illustrated by the model's sensitivity to the rate of formation of enzyme-substrate complexes on the cellulose surface.

2.2 Introduction

The mechanism of enzymatic cellulose hydrolysis is not well understood. The complexity of the system, which arises from the concerted action of several enzymes acting on a heterogeneous solid substrate, makes experimental kinetic and mechanistic studies difficult. As a result, the rate-determining features of the cellulase-cellulose system have yet to be identified, and the physical and chemical details of the enzyme-substrate interaction remain to be elucidated. Improved understanding of the overall system is crucial for the design of optimal enzymatic hydrolysis processes.

During enzymatic cellulose hydrolysis, saccharification rates typically decline dramatically after an initial "burst" phase (*1-5*). This decrease in hydrolysis rates makes high enzyme loadings and long reaction times necessary to achieve desired cellulose conversions. While many hypotheses for the cause or causes of this decline have been proposed (*6-13*), the origin of this phenomenon is still unresolved.

Detailed mechanistic models provide tools for studying the kinetics and mechanisms of complex reaction systems. Their importance in efforts to understand and improve enzymatic cellulose hydrolysis processes has been described earlier (*14, 15*). Mechanistic models allow investigators to rapidly probe a proposed reaction network through the use of computational predictions and sensitivity analysis. By understanding the underlying kinetics and mechanism of the cellulase reaction system more directed and rational approaches can be used for cellulase engineering and process optimization.

Many models of cellulase-catalyzed cellulose hydrolysis have been developed over the past 30 years. Most of these models contained simplified representations of the cellulases and/or

* This chapter, with few modifications from its original format, represents the following peer-reviewed publication: S. E. Levine, J. M. Fox, H. W. Blanch, D. S. Clark, A mechanistic model of the enzymatic hydrolysis of cellulose. *Biotechnol Bioeng* **107**, 37 (Sep 1, 2010)

the substrate. The activities of the different cellulase enzymes that work together during cellulose hydrolysis are commonly lumped together as a single enzyme concentration. The cellulose substrate has been represented using a variety of simplified representations: (1) considering the solid cellulose phase to be equivalent to a soluble polymeric substrate (*15*), (2) describing the cellulose by a single bulk concentration (*16-18*), (3) treating the cellulose as a mixture of digestible and inert substrate concentrations (*19, 20*), (4) using one or two adsorption sites (e.g., crystalline and amorphous) on a surface area that is then related to the bulk cellulose concentration (*21-25*), and (5) using a hypothetical periodic surface composed of random polymeric chains (*26*). These simplifications limit the ability of these models to explain enzyme-enzyme cooperativity or to predict evolution of the cellulose substrate surface morphology as hydrolysis occurs.

Most models that include an enzyme adsorption step rely on simple assumptions about the enzyme-substrate interaction. Cellulase adsorption onto cellulose has typically been described through the use of a Langmuir-type equilibrium equation (*16, 25, 27-29*). The Langmuir isotherm, while often able to be fit to experimental data, is based on assumptions that are not representative of the cellulase-cellulose system. In addition, inconsistencies between experimental adsorption data and the Langmuir equation predictions have been reported in the literature (*30*). Several models have avoided using an equilibrium adsorption model (*20, 22*), but their reliance on lumped enzyme and/or substrate concentrations has remained, rendering the physical meaning of the adsorption step unclear.

Representing adsorption in a mechanistic model is further complicated by the formation of the enzyme-substrate complex on the cellulose surface. Early models assumed that an enzyme adsorbs to the cellulose surface and complexes with a cellulose chain in one concerted reaction step (*15, 31, 32*); later models derived from these early models have also made this assumption (*28, 29, 33*). Most cellulases, however, have a structural architecture that suggests a two-step process. Cellulases contain a carbohydrate binding module (CBM) that is separated from the catalytic domain by a short linker; adsorption of the CBM domain to the surface and complexation of the catalytic domain to a chain may occur as two distinct events. A recent high-speed AFM study of cellulase enzymes on cellulose supports this two-step interaction (*34*). Mathematical representations must capture this complexation event. The utility of previous models in predicting enzymatic cellulose hydrolysis behavior has been severely limited by their reliance on lumped terms and inadequate representations of cellulase-cellulose interactions.

In this work a model has been developed which avoids some of the unrealistic assumptions that have limited earlier efforts. It is based on a mechanistic description that includes distinct enzyme adsorption and complexation steps. The equilibrium assumption is abandoned, thus permitting dynamic interactions between cellulases and the cellulose surface to be incorporated. Individual cellulases of a well-defined enzyme mixture are explicitly tracked; substrate concentration and the degree of cellulose polymerization are monitored; and surface concentrations of each cellulose chain length are individually described. Representation of the substrate is focused on capturing the time course of cellulose surface area; as the cellulose particles shrink, new chains are exposed, and the total cellulose surface area is reduced. This approach provides useful insights into the impact of substrate surface area on the hydrolysis rate, the roles of various mechanisms in the slowdown in the rate of hydrolysis, and the importance of an adequate representation of the enzyme-substrate interaction.

2.3 Methods

2.3.1 Representation of the Cellulose Substrate

The solid cellulose substrate is represented as an assembly of spherical particles. This representation reflects an established mapping of arbitrary shapes into an assembly of monodisperse spheres that has the same total surface area and volume as the original shape (*35*). This approach has been used to describe the radiative scattering and absorption of atmospheric ice crystals (*35, 36*). The rate of cellulose hydrolysis is directly dependent to the amount of substrate surface area; the manner in which surface area changes depends on the initial shape. In order to describe the change of surface area with time, we posit that the original shape can be represented by a polydisperse assembly of spheres (Box 1 provides the details of arbitrary shape mapping to spheres). In the present work, both monodisperse and bimodal distributions are considered.

BOX 1. Mapping Arbitrary Particle Shapes to Collections of Spheres

Monodisperse distributions of spheres can be used to model properties for static shapes. When the arbitrary shapes of interest change with time, a polydisperse distribution of spheres must be used to properly map the properties. Figure A shows the requirements and capabilities for matching static and dynamic properties of a given shape.

The polydisperse distribution of spheres required to map a given shape is difficult to determine. It requires knowledge of how the surface area changes with time. Thus a reverse approach may be used, where a given polydisperse distribution is used to model the dynamic change of a particle's properties. The form of these results is compared to measured results. Thus, instead of directly finding the sphere distribution to match an arbitrary shape, an arbitrary shape is found that matches a given sphere distribution.

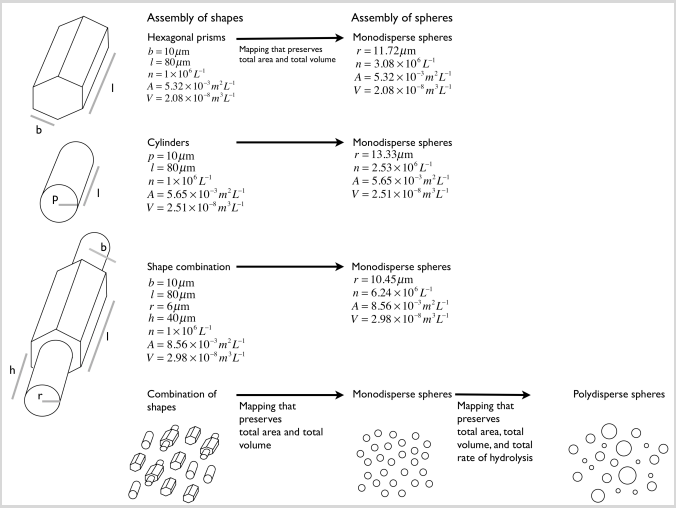


Figure A. Hexagonal columns, cylinders, and shape combinations represented as spheres: a monodisperse distribution of spheres has the same total volume and total surface area as the original shape, a polydisperse distribution of spheres has the same total volume, total surface area, and overall hydrolysis rate as the original shape

Particles are composed of cellulose chains of varying length described by a Poisson distribution based on the initial degree of polymerization. The cellulose particles shrink as soluble cello-oligosaccharides are released from the surface. A material balance describes the rate at which the radius changes with time (Eq. 2.1). The surface area is related to the density and the radius of a spherical particle by a simple relationship (Eq. 2.2).

$$\frac{dR}{dt} = R\frac{SA}{3}MW_1\sum_{i=1}^{6} ir_j^{'} \tag{2.1}$$

$$SA = \frac{3}{R\rho_{cellulose}} \tag{2.2}$$

As soluble cello-oligosaccharides are released from the solid cellulose surface, fresh cellulose chains become part of the surface. A term is required in the equations for cellulose chain surface concentrations to account for this new chain exposure. Particles are assumed to be composed of a continuum of cellulose chains rather than discrete layers, and their physical characteristics (e.g., sites per area and density) are assumed to remain constant as the particles shrinks. The rate at which new chains are exposed is equal to the rate at which old chains are lost, as the number of sites per area is assumed to be constant. Individual chains have an exposure term that depends on the rate of loss of old chains and the initial degree of polymerization probability distribution function (Eq. 2.3).

$$\frac{d[C_i^{'}]}{dt} = +pdf(DP_0, i)\frac{1}{DP_0}\sum_{j=1}^{6} jr_i^{'} \tag{2.3}$$

The cellulose chains on the surface are represented as a lattice of glucose units. In the case of lignocellulosic substrate, the representation of the surface can be modified to include non-hydrolyzable sites. Each cellulase enzyme that adsorbs onto this lattice occupies a certain number of sites, described by its footprint. Footprint estimates are based on small angle X-ray scattering measurements of the *Trichoderma reesei* cellobiohydrolase I (CBHI) (*37, 38*). If the size of each glucose unit is assumed to be 0.25 nm^2, CBHI occupies approximately 84 glucose sites. If all cellulases are assumed to have this footprint, regardless of shape, a monolayer of perfectly packed enzymes would cover all lattice sites. A more physically-realistic description of adsorption can be developed by considering random adsorption by cellulases of a defined shape to the surface. Consequently, random sequential adsorption (RSA) simulations of a cellulase-shaped adsorbate consisting of a large catalytic domain, a short linker, and a small CBM, adsorbing onto a periodic square lattice were used to determine an effective footprint. An adsorbed enzyme represented this way occupies 156 glucose units, nearly double the physical footprint. The details of the RSA simulation are provided in Appendix B of the supplemental material.

2.3.2 Hydrolysis Mechanism

The mechanism by which cellulases catalyze the hydrolysis of cellulose can be considered in three steps: (1) adsorption, (2) complexation, and (3) reaction. Adsorption and complexation were treated as reversible steps, while the reaction step was treated as irreversible. A schematic of this mechanism is given in Fig. 2.1, which illustrates these steps for an endoglucanase (EG) and a cellobiohydrolases (CBH).

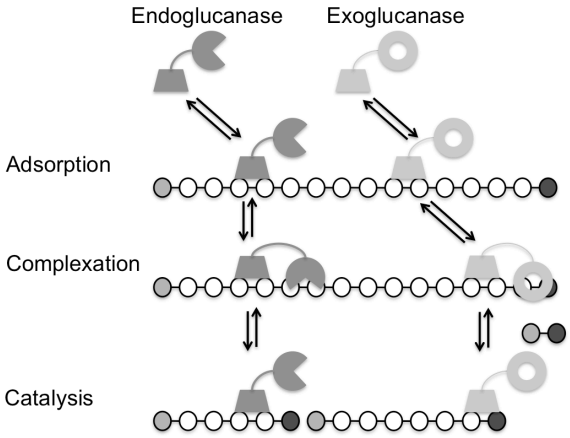


Figure 2.1 Schematic of the general mechanistic steps of cellulase-catalyzed hydrolysis of cellulose by an endoglucanase and an exoglucanase.

Adsorption and desorption of the cellulases were described using site and enzyme balances, which provide concentrations of cellulose surface sites and solution-phase enzymes. All cellulases are allowed to adsorb to identical free sites on the cellulose surface. Adsorption and desorption reactions are treated as elementary reactions, leading to balance equations for adsorbed, uncomplexed enzyme concentrations (Eq. 2.4).

$$\frac{d[E'_{x-ads}]}{dt} = k_{x-ads}[E_{x-free}][*'] - k_{x-des}[E'_{x-ads}] \tag{2.4}$$

The terms resulting from adsorbed enzyme complexing and decomplexing with a cellulose chain and catalysis are not included in Eq. 2.4. The catalysis step is assumed to be slow in comparison with the other reactions, allowing the complexation/decomplexation step to be considered at equilibrium.

The equilibrium assumption for complexation leads to a simple relationship between the surface concentrations of adsorbed complexed enzyme, adsorbed uncomplexed enzymes, and surface cellulose chains (Eq. 2.5). The difference in the equilibrium relationship for endoglucanase and cellobiohydrolases enzymes stems from the ability of endoglucanase to complex with any glycosidic bond on the cellulose chain while cellobiohydrolases can only complex with a specific chain end (reducing or non-reducing).

$$[E'_{EG-ads}C'_i] = \frac{[E'_{EG-ads}]\theta_{EG}(i-1)[C'_i]}{K_{M-EG-cel}} \qquad [E'_{CBH-ads}C'_i] = \frac{[E'_{CBH-ads}]\theta_{CBH}[C'_i]}{K_{M-CBH-cel}} \tag{2.5}$$

Using this equilibrium relationship and the mechanism depicted in Fig. 2.1, balance equations for the surface concentration of solid cellulose chains of length i can be derived (Eq. 2.6).

$$(i>6)\qquad \frac{d[C_i^{'}]}{dt}=\frac{k_{cat-EG-cel}}{K_{M-EG-cel}}[E_{EG-ads}^{'}]\theta_{EG}\left(2\sum_{j=i+1}^{\infty}[C_j^{'}]-(i-1)[C_i^{'}]\right)+\frac{k_{cat-CBH-cel}}{K_{M-CBH-cel}}[E_{CBH-ads}^{'}]\theta_{CBH}\left([C_{i+2}^{'}]-[C_i^{'}]\right) \tag{2.6}$$

Additional endoglucanase or cellobiohydrolases enzymes can be incorporated into the model; this will add terms to Eq. 2.6 that are similar to those above.

Soluble cello-oligosaccharides are formed by enzymes acting on both soluble, short-chain sugars and on insoluble chains within the solid substrate. The equations describing the generation of soluble short-chain sugars (DP≤6) assume a Michaelis-Menten mechanism for the soluble phase reaction terms, and rely on assumptions similar to those in Eq. 2.6 for the solid phase reaction terms. The equation for cellobiose concentration is shown as an example in Eq. 2.7. The full set of material balance equation can be found in Appendix A with detailed derivations in Appendix C of the supplemental material. While it is not shown here, the action of β-glucosidase can be integrated into the model equations with an additional Michaelis-Menten type term in the cellobiose and glucose equations. Both free and adsorbed cellulases are competitively inhibited by glucose and cellobiose. Terms for this inhibition are present in the enzyme and site balances. The full set of model equations is shown in Appendix A (with detailed derivations in Appendix C of the supplemental material).

$$\begin{aligned}\frac{d[C_2]}{dt}=&\left(2\frac{k_{cat-EG-cel}}{K_{M-EG-cel}}[E_{EG-ads}^{'}]\theta_{EG}\sum_{i=7}^{\infty}[C_i^{'}]\right.\\&+\frac{k_{cat-CBH-cel}}{K_{M-CBH-cel}}[E_{CBH-ads}^{'}]\theta_{CBH}\sum_{i=7}^{\infty}[C_i^{'}]\\&+2\frac{k_{cat-EG-sol}}{K_{M-EG-sol}}[E_{EG-ads}^{'}]\sum_{i=3}^{6}[C_i]\\&\left.+\frac{k_{cat-CBH-sol}}{K_{M-CBH-sol}}[E_{CBH-ads}^{'}]\left(\sum_{i=3}^{6}[C_i]+[C_4]\right)\right)\frac{A_{cel}}{V_{liq}}\\&+2\frac{k_{cat-EG-sol}}{K_{M-EG-sol}}[E_{EG-free}]\sum_{i=3}^{6}[C_i]\\&+\frac{k_{cat-CBH-sol}}{K_{M-CBH-sol}}[E_{CBH-free}]\left(\sum_{i=3}^{6}[C_i]+[C_4]\right)\end{aligned} \tag{2.7}$$

2.3.3 Model Parameters

The model requires a variety of adsorption, kinetic, and physical parameters. Sets of experimental values were determined in the present work, or values from the literature were used. In the absence of reliable experimentally determined parameters, estimates were made by fitting the model to experimental data.

Kinetic parameters for the activity of cellulase enzymes on a variety of substrates have been reported in the literature. The work detailed here relies on accurate determination of Michaelis-Menten constants for cellulase activity on soluble cello-oligosaccharides. For *Trichoderma reesei* parameters are available for CBH1 (*39*), CBH2 (*40*) and a range of endoglucanases (*41*). Recent results for the rate at which a CBH1 enzyme moves as it degrades the end of a cellulose chain (*34*) reveal that the turnover number of a cellulase complexed with a solid cellulose chain (~7 s^{-1}) is roughly equivalent to that of a cellulase acting on a soluble cello-oligosaccharides in solution (~9 s^{-1}). Accordingly, k_{cat} values determined on substrates like cellohexaose and cellopentaose can be used not only for solution-phase reactions but also for solid cellulose substrates. There are no comparable experimental values for the complexation constants of adsorbed enzyme with chains on the surface of a solid substrate. These parameters were thus optimized to best fit the experimental data.

Product inhibition equilibrium constants have been reported for cellulase hydrolysis of soluble substrates. Reported values range over several orders of magnitude (*39, 42-44*), making the determination of reliable values difficult. Literature values were used only to set reasonable bounds for the optimization of the product inhibition parameters.

The majority of cellulase adsorption studies have employed solution depletion measurements. A Langmuir model has typically been fit to such data to determine adsorption parameters. Because the model proposed here requires values of $k_{adsorption}$ and $k_{desorption}$ for each enzyme in the system (Eq. 2.4), fits to the Langmuir isotherm have not been employed. Values for these parameters have been estimated for a commercial cellulase mixture (Celluclast 1.5L) using data obtained from ellipsometry (*45*).

The parameters associated with the cellulose substrate are the initial surface area, the sites per area, and the average degree of polymerization of the cellulose chains. The substrate surface is assumed to consist of a homogeneous lattice of glucose units. While an experimental value for sites per area is not available an upper bound for this value can be calculated based on the size of a single glucose unit (0.25 $nm^{2)}$; accordingly a glucose lattice structure has a surface concentration of 6.64×10^{-5} $mmol^{-1}$ dm^{-2}. The degree of polymerization of several of cellulosic substrates (including Avicel, cotton, bacterial cellulose) has been determined using a variety of techniques (*46-50*). These values provide an estimate of a typical degree of polymerization for a given cellulosic substrate. The literature provides only a limited set of surface area measurements. Nitrogen BET measurements are available; however, nitrogen molecules are far smaller than proteins, and this method requires the use of dry substrates; therefore, these measurements will give inaccurate values for the surface area accessible to cellulases. Attempts to use surface depletion of fluorescent cellulase-like constructs have also been made (*51*), but these use only one type of CBM and thus likely underestimate the accessible surface area. Appendix D in the supplemental material provides more detail about the parameter values used in this study.

2.3.4 Model Assembly and Solution

The model consists of a system of differential and algebraic equations. Balance equations for the surface concentrations for chains of DP 7 to DP_{max} (set to $1.5xDP_0$); the concentrations of soluble cello-saccharides from glucose to cellohexaose; the adsorbed, uncomplexed enzyme surface concentrations; and the radius of the cellulose particles are included. Algebraic equations for enzyme and site balances are also part of the system of equations that comprise the model. The model was solved in Matlab using the *ode23t* function. For parameter optimization the *fmincon* function was used, which utilizes a sequential quadratic programming optimization algorithm. A normalized sum of square error between experimental values and model results was used as the objective function.

2.4 Results

2.4.1 Surface Area Requirements

The inadequacy of currently available surface area measurements is made clear by a comparison between measured values and the cellulose surface area required based on adsorbed enzyme footprint estimates. A typical enzyme loading of 0.15 μmol g^{-1} (approximately 9 mg enzyme per gram substrate) requires a surface area of 1.9 m^2 g^{-1} assuming an enzyme footprint based on the crystal structure enzyme dimensions, and a surface area of 3.5 m^2 g^{-1} is required based on the effective (RSA) enzyme footprint. The 0.15 μmol g^{-1} loading is usually considered to be industrially relevant, and the accessible surface area of Avicel has recently been reported as 2.38 m^2 g^{-1} (*51*). However, the 2.38 m^2 g^{-1} estimate for the cellulase-accessible surface area for Avicel is insufficient to accommodate a loading of 0.15 μmol g^{-1}. Despite being experimentally elusive, reliable measurements for enzyme accessible surface area are critical for understanding and modeling the enzyme-substrate interactions during hydrolysis.

2.4.2 Enzyme Activity with Low Surface Area (Monodisperse Spheres)

The model was tested using *T. reesei* CBHI and endoglucanase 2 (EG2) hydrolysis time-course data reported by Medve et al. (*5*). The hydrolysis conditions were as follows: 10 g L^{-1} Avicel; 0.16 μmol g^{-1} of either EG2 or CBHI (individual pure enzymes) (approximately 10 mg enzyme per g cellulose); 0.32 μmol g^{-1} total enzyme (1:1 EG2:CBHI mixed enzymes) (approximately 20 mg enzyme per g cellulose); pH 4.8; 50 mM sodium acetate buffer; 40 °C. The experiments and simulations conducted with a single pure enzyme (EG2 or CBHI) will be referred to as "single enzyme hydrolysis" in this work. The summation of the total effective glucose release rates of the EG2 and CBHI single enzyme hydrolysis experiments and simulations will be referred to as "theoretical mixed enzyme hydrolysis". The experiments and simulations that use a 1:1 EG2:CBHI mixture will be referred to as "mixed enzyme hydrolysis". The parameters used in the model are shown in Table 2.1 (more details can be found in Appendix C of the supplemental material).

Table 2.1. Kinetic parameters used in model simulations for comparison to experimental data from Medve et al. (1998).

Kinetic Parameter	*Value*[a]	*Units*
$k_{EG2\text{-}ads}$	8640[b]	L $mmol^{-1}$ hr^{-1}
$k_{CBH1\text{-}ads}$	8640[b]	L $mmol^{-1}$ hr^{-1}
$k_{EG2\text{-}des}$	19.3[b]	hr^{-1}
$k_{CBH1\text{-}des}$	164[b]	hr^{-1}
$k_{cat\text{-}EG2}$	65[c]	s^{-1}
$k_{cat\text{-}CBH1}$	9.4[c]	s^{-1}
$K_{M\text{-}EG2\text{-}cel}$	0.0067[d]	mmol dm^{-2}
$K_{M\text{-}CBH1\text{-}cel}$	0.000014[d]	mmol dm^{-2}
$K_{M\text{-}EG2\text{-}sol}$	0.053[c]	mmol L^{-1}
$K_{M\text{-}CBH1\text{-}sol}$	0.0032[c]	mmol L^{-1}
$K_{i\text{-}EG2\text{-}cellobiose}$	0.01[d]	mmol L^{-1}
$K_{i\text{-}CBH1\text{-}cellobiose}$	0.093[d]	mmol L^{-1}
$K_{i\text{-}EG2\text{-}glucose}$	16.9[d]	mmol L^{-1}
$K_{i\text{-}CBH1\text{-}glucose}$	31[c]	mmol L^{-1}

[a]For expanded details about the model parameters and their sources see Appendix C in the supplemental material

[b]Parameters based on values from elipsiometry experiments(unpublished data)

[c]Experimental parameters from various literature sources

[d]Parameters optimized to early time data (less than 3 hours) for single enzyme hydrolysis results from Medve et al. (1998)

An Avicel surface area of 8 m^2 g^{-1} was used as a low surface area condition. This value was chosen because it corresponds to a case where an enzyme loading of 0.16 μmol g^{-1} represents 50% surface coverage based on the effective (RSA) area occupied by an enzyme. The model results are compared to the experimental values in Fig. 2.2. The activities for single enzyme hydrolysis are shown in Fig. 2.2A. The model exhibits excellent quantitative agreement at early times (up to 3 hr) and qualitative agreement for extended times for single enzyme hydrolysis. The model fails to capture the decrease in the rate of hydrolysis. The model does not capture the experimental data for the hydrolysis rates with both enzymes present, as shown in Fig. 2.2B. While the experimental results show a synergistic relationship between EG2 and CBH1, the model predicts a competitive relationship.

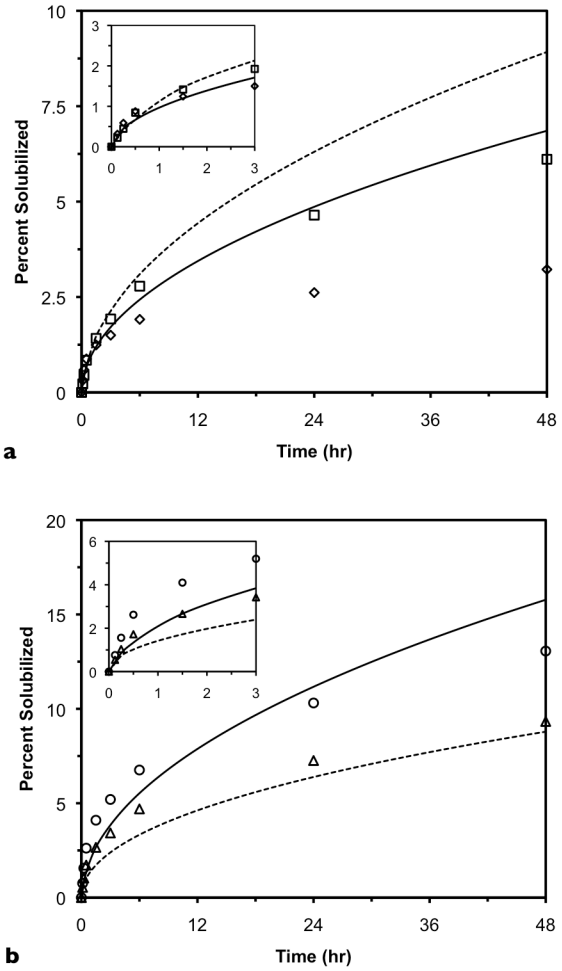


Figure 2.2 Model results using monodisperse spheres with the low initial surface area value (8 m^2 g^{-1}) compared to experimental data for a) single enzyme hydrolysis with CBH1 (□, experiment; --- ,model) and EG2 (◇, experiment; —, model) and b) theoretical mixed enzyme hydrolysis based on single enzyme experiments (△, experiment; —, model) and the actual mixture (○, experiment; ---, model). The experimental data are from Medve et al. (1998). The insets show an expanded view of the early time points (up to 3 hours).

The model allows the cause of this inconsistency to be examined. Fig. 2.3 shows the fraction of available sites on the substrate surface for each enzyme during single enzyme and mixed enzyme hydrolysis. The fraction of sites available to EG2 for hydrolysis is much lower in mixed cellulase hydrolysis compared to hydrolysis with EG2 alone. This decrease in substrate accessibility severely hinders the effectiveness of EG2 in hydrolyzing cellulose within a mixture. The CBH1 enzyme, on the other hand, suffers almost no change in the fraction of available sites (chain ends) on the surface. The difference between how a crowded surface affects EG2 and CBHI is expected. EG2 has far more sites to react with than CBHI because there are many more glycosidic bonds than chain ends. An enzyme on the surface therefore obstructs more EG2 sites than CBHI sites. In addition, EG2 activity converts glycosidic bonds into chain ends, thus adding to the number of CBHI sites. A low surface area under mixed enzyme hydrolysis conditions causes the synergistic relationship between EG2 and CBH1 to break down; EG2 is unable to effectively cleave cellulose chains to increase the number of chain ends available for CBH1 to hydrolyze to cellobiose.

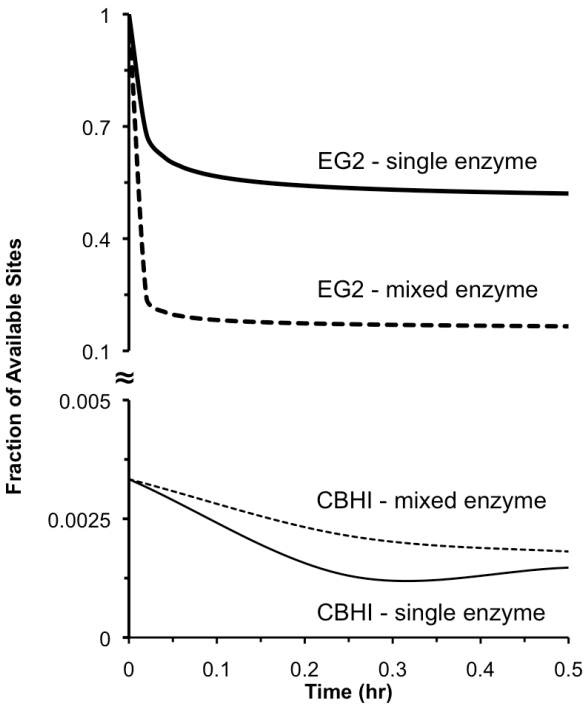


Figure 2.3 Fraction of sites on the cellulose surface available for catalysis by each enzyme type during the model simulation using monodisperse spheres with the low initial surface area value (8 m^2 g^{-1}). Only the first half hour is shown because the values have reached steady state at that point.

2.4.3 High Surface Area (Monodisperse spheres)

The effect of surface area on hydrolysis was further examined by employing an initial Avicel surface area of 47.6 m^2 g^{-1} (all other conditions unchanged). This value was chosen to minimize any effect a lack of available surface area would have on the cellulose and still be within the range of surface areas reported for other types of cellulose (e.g., PASC and bacterial cellulose) (*51*). For single enzyme hydrolysis (Fig. 2.4A), the model results and experimental data show good quantitative agreement at early times (up to 3 hours) and qualitative agreement at extended times. The mixed enzyme model results (Fig. 2.4B) are now also in good quantitative agreement with the experimental data at early times (up to 3 hours) and qualitative agreement at extended time. It is clear that the increased cellulose surface area is required to predict the observed synergistic interaction between the cellulase enzymes. However, the model fails to capture the full extent of the observed decrease in the rate of hydrolysis.

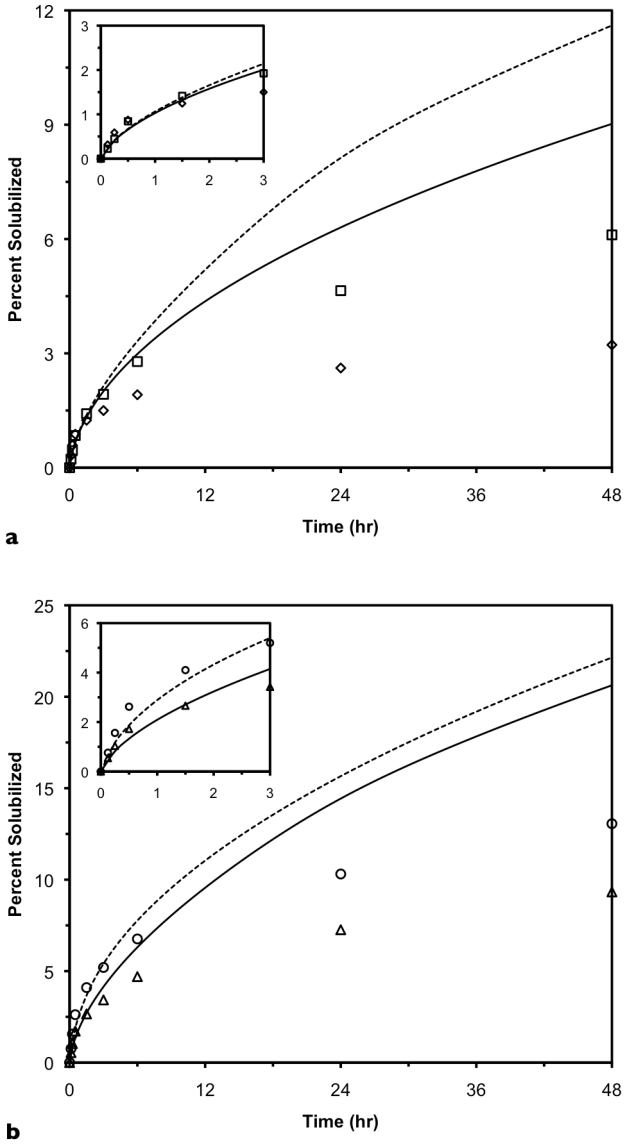


Figure 2.4 Model results using monodisperse spheres with the high initial surface area value (47.6 m^2 g^{-1}) compared to experimental data for a) single enzyme hydrolysis with CBH1 (□, experiment; ---, model) and EG2 (◇, experiment; —, model) and b) theoretical mixed enzyme hydrolysis based on single enzyme experiments (△, experiment; —, model) and the actual mixture (○, experiment; ---, model). The experimental data is from Medve et al. (1998). The insets show an expanded view of the early time points (up to 3 hours).

The role of increased surface area in allowing the synergistic interactions to occur can be explored by examining the change in the fraction of available sites for complexation for each enzyme in the model (Fig. 2.5). While EG2 still demonstrates a decreased fraction of accessible surface sites in the mixed enzyme hydrolysis case compared to the single enzyme hydrolysis case, this decline in available surface sites is much lower than that seen in the low surface area case. CBH1 has a large increase in the fraction of available sites (chain ends) on the substrate surface during mixed enzyme hydrolysis compared to the single enzyme hydrolysis case. This result is expected because EG2 cleaves cellulose chains on the surface into shorter chains, thereby increasing the number of chain ends available. The results of the high and low surface

area scenarios illustrate that in order for synergy to occur, the endoglucanase must be able to act almost as efficiently in the mixture as it does alone.

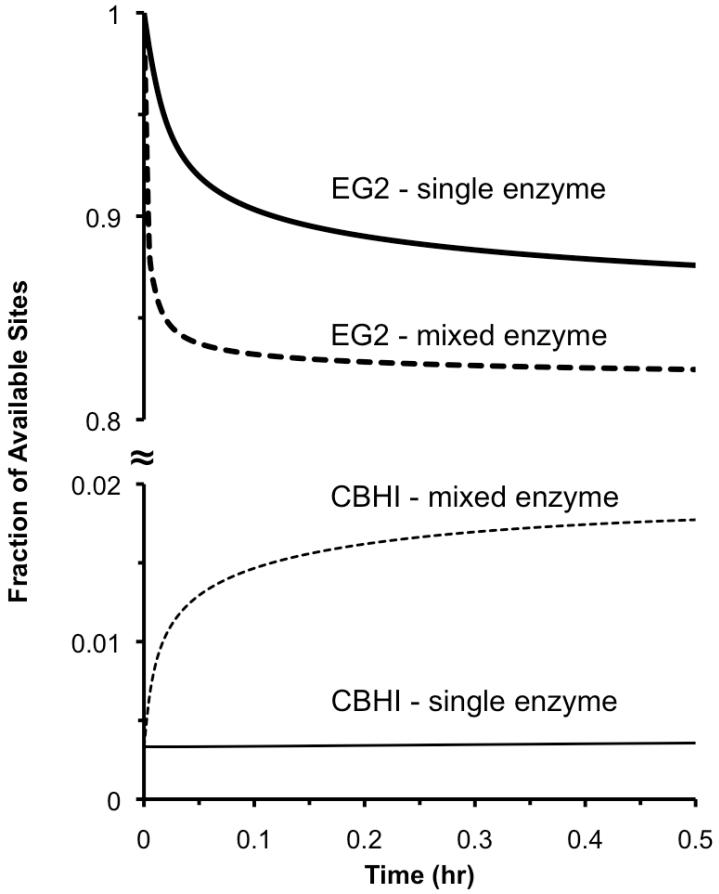


Figure 2.5. Fraction of sites on the cellulose surface available for catalysis by each enzyme type during the model simulation using monodisperse spheres with the high initial surface area value (47.6 m^2 g^{-1}). Only the first half hour is shown because the values have reached steady state at that point.

2.4.4 The Inclusion of Thermal Deactivation (Monodisperse Spheres)

In the model cases above only changes in the available surface area and product inhibition are capable of causing a decrease in the rate of hydrolysis. Clearly, even with strong product inhibition from cellobiose, the model is unable to capture the full extent of the observed decrease in the rate of hydrolysis. The extent to which thermal deactivation could contribute to the slower rate of hydrolysis was included through addition of first-order thermal decay of enzyme. The entire enzyme population was assumed to be identically susceptible to the thermal deactivation, and for simplicity thermal deactivation was assumed to occur only in solution. When the enzyme half-lives were reduced significantly, the model results achieved excellent agreement with the experimental data at all hydrolysis times (Fig. 2.6). The half-life of EG2 was set to 4.3 hours and the half-life of CBH1 was set to 10.6 hours. These are much shorter thermal half-life values than the value of the half-life for a commercial *T. reesei* cellulase mixture reported in the literature, 42.5 hours (*52*).

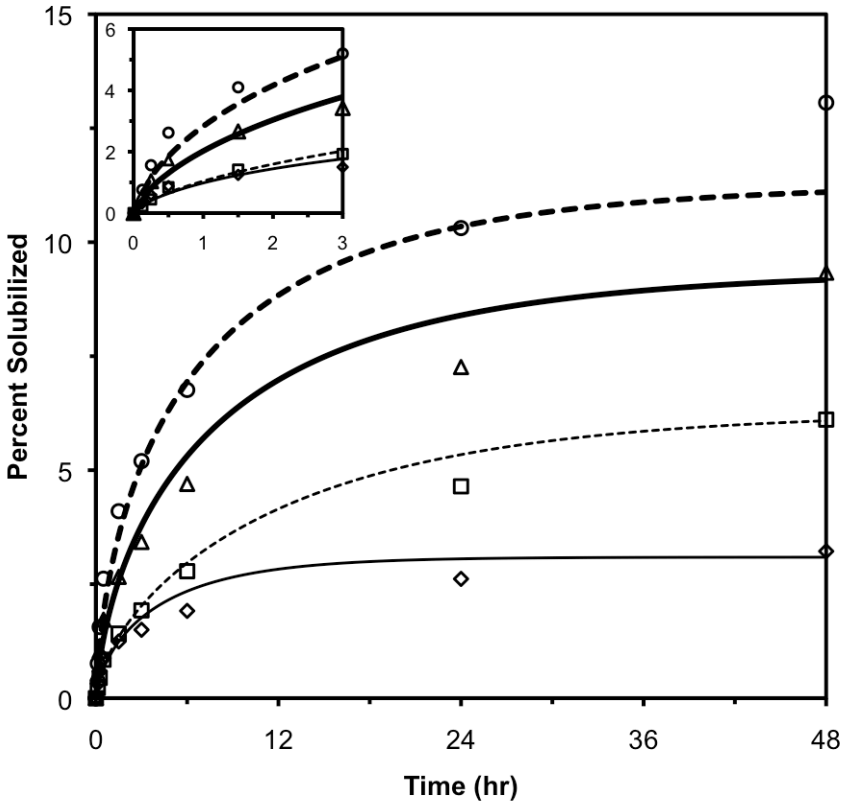


Figure 2.6. Comparison of model results with experimental data using monodisperse spheres with the high initial surface area (47.6 m^2 g^{-1}) case with the inclusion of first-order thermal enzyme deactivation in the model. Results are shown for hydrolysis with EG2 (◊, experiment; —, model), CBH1 (□, experiment; ---, model), theoretical mixture from single enzyme hydrolysis results (△, experiment; **▬**, model), and the actual EG2 and CBH1 mixture (○, experiment; **—**, model). The experimental data are from Medve et al. (1998). The insets show an expanded view of the early time points (up to 3 hours).

2.4.5 The Effect of a Bimodal Particle Size Distribution

All of the simulations discussed above used a monodisperse spherical particle distribution. As a demonstration of the use of polydisperse spherical particle distributions to capture the hydrolysis behavior of non-spherical cellulose particles, a range of bimodal distributions was employed under the same conditions as the high surface area case without thermal enzyme deactivation. The time course of cellulose surface area in the reactor with different bimodal distributions is shown in Fig. 2.7. In the cases shown in Fig. 2.7, the proportion and size of small spheres in the distribution has been varied; the fraction of initial surface area arising from small spheres was held constant at 83%. The decrease in cellulose area can also be altered by changing the fraction of the initial area resulting from each of he two spherical particle in the bimodal distribution (data not shown).

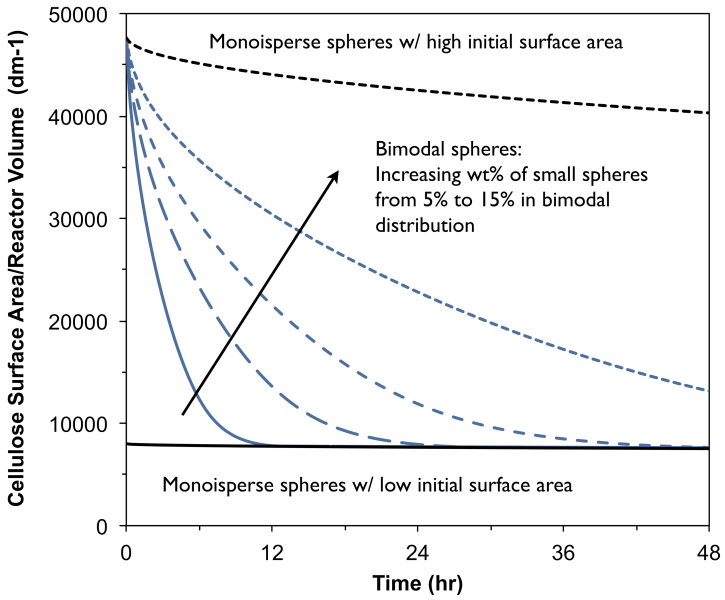


Figure 2.7. Model results for the accessible surface area per reactor volume under various sphere size distributions. The monodisperse cases are for high initial surface area (47.6 m^2 g^{-1}) (**▬**) and low initial surface area (8 m^2 g^{-1}) (—). The bimodal cases all have a total initial surface area that matches the high initial surface area value. The distribution is initially set to split the surface area with 83% on the small spheres and 17% large sphere. The cases presented vary weight % of the small spheres in the distribution (and the size to maintain the initial surface area ratios): 5% 1.5 µm spheres 95% 142.5 µm spheres (—), 7.5% 2.25 µm spheres 92.5% 139 µm spheres (— —), 10% 3 µm spheres 90% 135 µm spheres (-- --), 15% 4.5 µm spheres 85% 127.5 µm spheres (---).

The model results for the bimodal particle distribution 2.25 µM particles at a concentration of 0.75 g L^{-1} (83% of the initial surface area) and 139 µM particles at a concentration of 9.25 g L^{-1} (17% of the initial surface area) are shown in Fig. 2.8. This bimodal distribution was able to fit the experimental data for a mixture of EG2 and CBHI at all times but failed to match the experimental data for the single enzyme hydrolysis cases at longer times (Fig. 2.8). The decline in the hydrolysis rate is captured in the enzyme mixture but not in the single enzyme hydrolysis, resulting in the inability of the model to capture the synergy exhibited in the experimental data. By examining the predicted fraction of available sites (Fig. 2.9), it can be seen that the sites available to EG2 initially exhibit a slight decrease, similar to that seen in the high surface area case (Fig. 2.5). However the fraction of available sites for EG2 for complexation rapidly decreases as the particle size distribution transitions between being similar to the high surface area case to that of the low surface area case. The predicted fraction of available sites for CBHI exhibits a similar trend, where the fraction matches the behavior seen in the high surface area case (Fig. 2.5) but rapidly looses available sites as surface area declines.

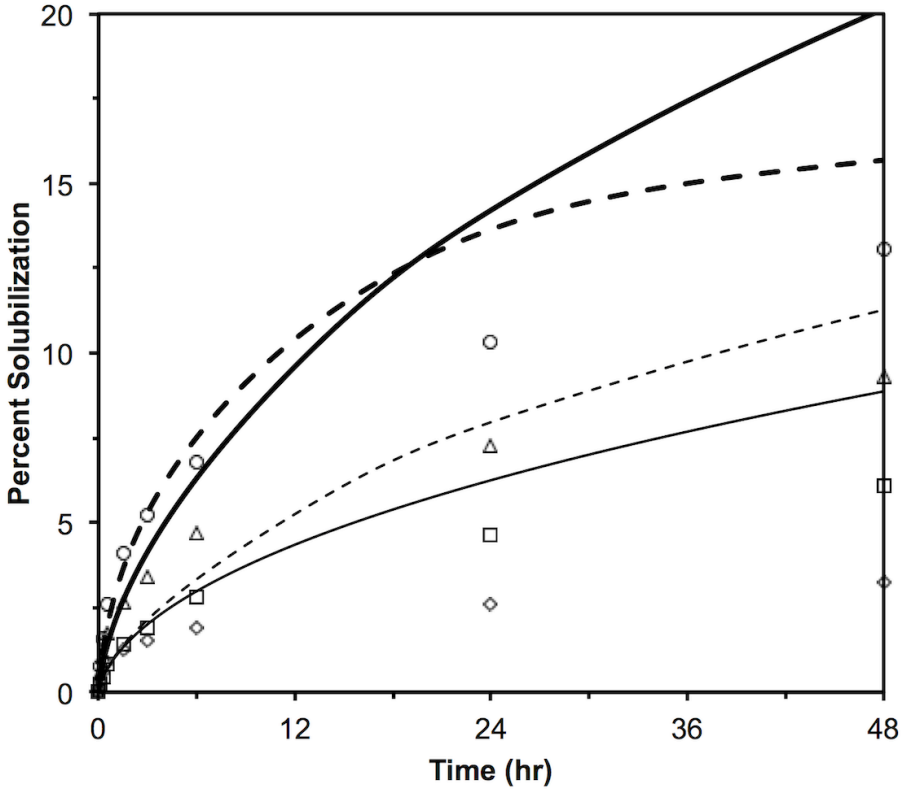


Figure 2.8 Comparison of model results with experimental data using a bimodal sphere distribution with 7.5 wt% 2.25 µm spheres and 92.5 wt% 139 µm spheres with an initial total surface area of 47.6 m^2 g^{-1}. Results are shown for hydrolysis with EG2 (◊, experiment; —, model), CBH1 (□, experiment; ---, model), theoretical mixture from single enzyme hydrolysis results (△, experiment; **▪▪▪**, model), and the actual EG2 and CBH1 mixture (○, experiment; **—**, model). The experimental data are from Medve et al. (1998). The insets show an expanded view of the early time points (up to 3 hours).

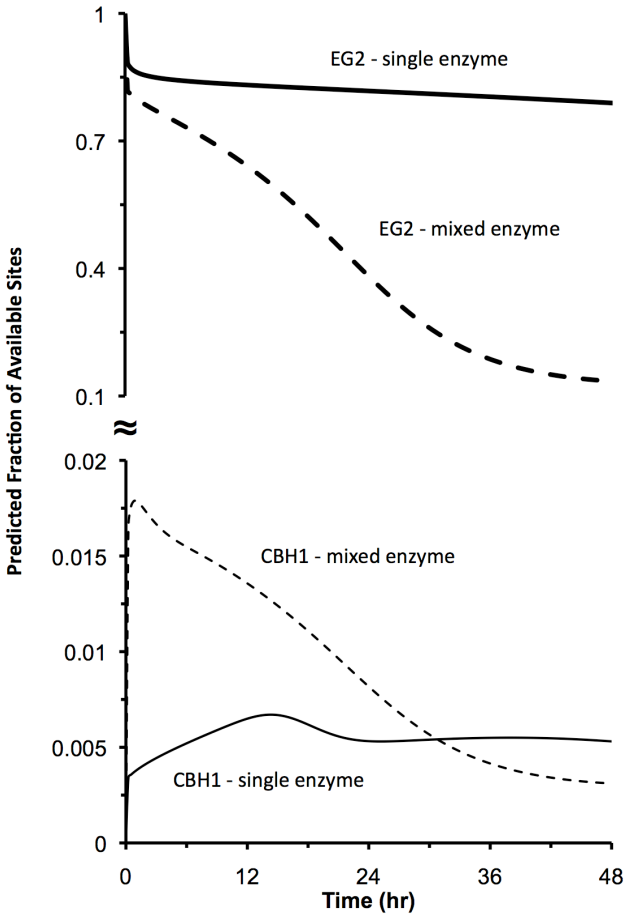


Figure 2.9 Fraction of sites on the cellulose surface available for catalysis by each enzyme type during the model simulation using the bimodal sphere distribution with 7.5 wt% 2.25 μm spheres and 92.5 wt% 139 μm spheres with an initial total surface area of 47.6 m^2 g^{-1}.

2.5 Discussion

2.5.1 The Role of Surface Area

The total cellulase-accessible cellulose surface area is a governing parameter for enzymatic hydrolysis. The model's ability to capture EG2-CBHI synergy depends strongly on surface area. The activities of enzymes acting alone or present in a mixture depend on the amount of surface that is accessible for cellulase complexation. The enzyme crowding effect is exacerbated as the amount of enzyme on the surface increases.

The extent to which enzyme crowding occurs may be influenced by the existence of enzyme-specific interactions between different cellulases and heterogeneities within the cellulose surface. When exposed to a low surface area cellulosic substrate, the adsorption of CBHI and EG2 results in high enzyme surface concentrations of both enzymes limiting the activity of EG2. Because the two enzymes are allowed to adsorb anywhere on the homogeneous surface, CBHI and EG2 compete for surface area. Lignocellulose substrates, however, are heterogeneous, containing crystalline and amorphous sub-sites. Furthermore, there is evidence that different cellulose binding domains exhibit specificity for particular morphological regions (*53-56*). By altering the

manner in which cellulases compete for sites on the surface, the distribution of subsites may impact the competitive behavior that leads to crowding effects.

2.5.2 The Decrease in Hydrolysis Rate

With significant enhancement in cellobiose inhibition and thermal inactivation parameters, the model was able to capture the slowdown in the hydrolysis rate exhibited by the experimental data (Fig. 2.7). Strong cellobiose inhibition was required: K_i = 0.01 mM for EG2 and K_i = 0.093 mM for CBHI. While cellobiose is known to be a strong inhibitor of cellulases, these values are low relative to reported values reported. Product inhibition alone did not enable the model to quantitatively capture the experimental data (Figs. 2.4 and 2.5). The sensitivity of the model to these parameters underscores the need to obtain reliable experimental values for the inhibition constants.

The inclusion of short thermal inactivation half-lives allowed the model to achieve quantitative agreement with the experimental data. These half-lives, however, were up to an order of magnitude smaller than those reported for a *T. reesei* cellulase mixture. The interaction of the cellulases at the solid-liquid interface may enhance their rate of thermal denaturation; alternatively, another rate-limiting mechanism may be at play. Regardless, the discrepancy between experimentally determined inactivation parameters and those required by the model indicates that the molecular picture is not complete. As with inhibition parameters, accurate half-life measurements for individual cellulases need to be obtained. Additional mechanisms that limit enzyme activity also need to be explored.

The model also indicates that enzyme crowding is not an exclusive explanation for the decrease in the hydrolysis rate. The kinetic slowdown has been observed under a wide range of hydrolysis conditions; however, enzyme crowding only arises with high enzyme loadings or low surface areas. The experimental data referenced in this work shows the decline in hydrolysis rate both with single enzymes acting alone and with enzymes acting within a mixture. Model results show crowding only with mixed enzymes and low substrate surface area. Again the need to explore additional rate-limiting mechanisms is illustrated by the incidence of a hydrolysis rate decrease.

2.5.3 Parameter Sensitivity

A sensitivity analysis was used to examine the importance of specific mechanistic steps in the performance of the model. The adsorption, catalytic, complexation, and cellobiose inhibition constants were increased and decreased by 20% under the high surface area, mixed enzyme, with thermal deactivation scenarios. These results were used to determine the fractional change in the amount of cellulose solubilized with respect to variations in each parameter. This is a standard method for sensitivity analysis of a kinetic model (*57*). The sensitivity analysis indicated that the model was highly responsive to changes in four parameters: $K_{M\text{-}CBH1\text{-}cel}$, $K_{M\text{-}EG2\text{-}cel}$, $K_{i\text{-}EG2\text{-}cellobiose}$, and $K_{i\text{-}CBH1\text{-}cellobiose}$, which are listed from most to least sensitive. These parameters relate to the complexation of the enzymes either with the solid cellulose surface for catalysis or with cellobiose to inhibit the enzymes. The model was relatively insensitive to changes in the adsorption, desorption, and catalytic parameters for both EG2 and CBHI. The sensitivity results suggest that complexation of adsorbed enzyme with the substrate is the

governing kinetic step in cellulose hydrolysis. The sensitivity of the model to the inhibition parameters reinforces the need for reliable experimentally determined values.

2.5.4 Capturing Particle Shape Effects

The bimodal distribution serves as a demonstration that more complicated rates of change in the total surface area of substrate can be represented using a distribution of spheres chosen to match the initial total area and volume of the particles. More complex particle size distributions would be required to match the rate of change in surface area for a specific particle shape. To determine these particle size distributions, detailed information about how a particle shape decreases during hydrolysis would be necessary. Box 1 details an approach for developing sphere size distributions for different particle shapes.

The model results for the bimodal case (Figs. 2.8 and 2.9) demonstrate the large effect more complex rates of surface area change can have on the predicted hydrolysis behavior. The inability of the model to capture synergy in the bimodal case reinforces the point that by itself, enzyme crowding related to a lack of accessible surface area cannot explain the decline in the rate of hydrolysis. If the decline in the rate of hydrolysis was caused simply by surface area and crowding there would be evidence available that lowering the enzyme loading can reduce, delay, or alleviate the effect; this has not been reported in the literature. In fact, the single enzyme hydrolysis results of Medve et al. (1998) show a decline in the rate of hydrolysis at a lower conversion than the mixed enzyme hydrolysis results, even though the single enzyme loading was only half the mixed enzyme loading.

2.5.5 Broader Implications of the Model Results

The model used k_{cat} values for soluble oligosaccharide (cellohexaose or cellopentaose) in representing catalysis steps performed by enzymes bound to soluble chains as well as to the solid substrate. The use of soluble-substrate k_{cat} values relies on the assumption that once the substrate is bound the rate of chemical catalysis does not change. Recently published experimental evidence supports this assumption: high-speed AFM results (*34*) showed the CBHI enzyme moving along a cellulose surface at 3.5 nm s^{-1}, which corresponds to a k_{cat} of 7 s^{-1}, similar to the soluble cello-oligosaccharide values.

If parameters describing the catalytic activity of enzymes on solid and soluble cellulose chains are identical, improvements in intrinsic cellulase kinetics (k_{cat}) on soluble substrates should improve the enzyme's activity on solid substrates. However, the model shows that the conversion of solid substrates to soluble sugars is not sensitive to the k_{cat} values of the enzymes. As discussed above, the complexation of adsorbed enzyme with solid substrate is the most significant parameter that influences the model predictions. Therefore, improvements in k_{cat} may not translate into improvements in solid-substrate cellulolytic activity.

The importance of accessible surface area is also highlighted by the model results. The discrepancy in activity and synergy behaviors between the low and high surface area cases indicates the broadly beneficial role of maximizing the enzyme accessible surface area before hydrolysis starts. A pretreatment process that enhances substrate surface area can reduce the likelihood that enzyme crowding will become rate limiting.

2.6 Conclusions

A detailed mechanistic model for cellulase-catalyzed hydrolysis of cellulose has been developed. The model explicitly tracks individual cellulases and key cellulose surface properties. Independent enzyme adsorption and complexation steps have been incorporated in an attempt to capture the most important details of the enzyme-substrate interaction.

Individual enzyme hydrolysis (EG2 or CBHI) and mixed enzyme hydrolysis scenarios were used to compare model results with experimental data from the literature. The model results were not consistent with all of the experimental data in the case of relatively low surface area. When the surface area was increased, limiting the effect of enzyme crowding during mixed enzyme hydrolysis, the model achieved good agreement with the experimental data, including EG2-CBHI synergy.

The model was not capable of capturing the full extent of the decrease in the rate of cellulose hydrolysis often reported in the literature. Strong product inhibition and short enzyme half-lives were required to match the slowdown apparent in experimental data. Neither of these effects is the primary cause of the observed slowdown. The ability to capture the rate of change in available surface area for shapes more complex than spheres, using a distribution of spherical particle sizes was demonstrated using a bimodal distribution. The available substrate surface area and how this area changes during hydrolysis was shown to be important, but these changes are unlikely to explain the decline in the rate of hydrolysis alone. The differential manner in which cellulases interact with structural heterogeneities within the substrate may play an important role in causing the rapid reduction in hydrolysis rate. These physical subtleties are not currently included in the model. The changes to the cellulose morphology that occur during hydrolysis may alter the character of the sub-sites on the surface leading to a reduction in the observed hydrolysis rate. The model results presented in this paper illustrate the importance of understanding the effect of relevant surface areas to enzyme hydrolysis activity. This work also highlights the utility of future investigations attempting to elucidate further details of cellulase-cellulose interaction.

2.7 Acknowledgements

This work was funded by the Energy Biosciences Institute. The authors wish to thank Clayton Radke and Sam Mauer for helpful discussions.

2.8 Nomenclature

A_{cel}	Total surface area of cellulose in the system (dm^2)
C_i'	Solid cellulose chain of length i (i>6)
C_i	Soluble cellooligosaccharide chain of length i (i<7)
DP	Degree of polymerization
DP_0	Initial degree of polymerization
DP_{max}	Maximum degree of polymerization tracked in the model
$E'_{CBH\text{-}ads}$	Uncomplexed cellobiohydrolase adsorbed to the surface
$E'_{CBH\text{-}ads}C_i$	Cellobiohydrolase adsorbed to the surface and complexed with an i length cellulose chain
$E_{CBH\text{-}free}$	Uncomplexed cellobiohydrolase in solution
$E'_{EG\text{-}ads}$	Uncomplexed Endoglucanase adsorbed to the surface
$E'_{CBH\text{-}ads}C_i$	Endoglucanase adsorbed to the surface and complexed with an i length cellulose chain
$E_{CBH\text{-}free}$	Uncomplexed Endoglucanase in solution
$E'_{x\text{-}ads}$	Uncomplexed cellulase of type x adsorbed to the surface
$E_{x\text{-}free}$	Uncomplexed cellulase of type x in solution
$k_{cat\text{-}CBH\text{-}cel}$	Catalytic constant for cellobiohydrolase acting on a solid cellulose chain (s^{-1})
$k_{cat\text{-}CBH\text{-}sol}$	Catalytic constant for cellobiohydrolase acting on a soluble cello-oligosaccharide (s^{-1})
$k_{cat\text{-}EG\text{-}cel}$	Catalytic constant for endoglucanase acting on a solid cellulose chain (s^{-1})
$k_{cat\text{-}EG\text{-}sol}$	Catalytic constant for endoglucanase acting on a soluble cello-oligosaccharide (s^{-1})
$K_{M\text{-}CBH\text{-}cel}$	Complexation equilibrium constant for cellobiohydrolase with a solid cellulose chain (mmol dm^{-2})
$K_{M\text{-}CBH\text{-}sol}$	Complexation equilibrium constant for cellobiohydrolase with a soluble cello-oligosaccharide (mmol L^{-1})
$K_{M\text{-}EG\text{-}cel}$	Complexation equilibrium constant for endoglucanase with a solid cellulose chain (mmol dm^{-2})
$K_{M\text{-}EG\text{-}sol}$	Complexation equilibrium constant for endoglucanase with a soluble cello-oligosaccharide (mmol L^{-1})
$k_{x\text{-}ads}$	Adsorption constant for a cellulase of type x onto a cellulose surface (L $mmol^{-1}$ s^{-1})
$k_{x\text{-}des}$	Desorption constant for a cellulase of type x from a cellulose surface (s^{-1})
MW_1	Cellulose monomer molecular weight (g $mmol^{-1}$)
n	Concentration of cellulose particles (mmol L^{-1})
R'	Cellulose particle radius (dm)
r_i'	Rate of formation of a length i soluble sugar from solid substrate (mmol dm^{-2} s^{-1})
SA	Cellulose surface area (dm^2 g^{-1})
t	Time (s)
V_{liq}	Volume of reactor (L)
θ_{CBH}	Fraction of free sites on the surface available to a cellobiohydrolase for complexation
θ_{EG}	Fraction of free sites on the surface available to an endoglucanase for complexation
$\rho_{cellulose}$	Cellulose density during hydrolysis (g L^{-1})
*	Free sites
[ ]	Concentration (mmol L^{-1}) if unprimed symbol or surface concentration (mmol dm^{-2}) if primed symbol

2.9 Appendix 1: Model Equations

In the following equations a prime (') indicates that the quantity is using a per area basis as opposed to a per volume basis.

Cellulose Particle Size Determination from a Material Balance:

$$\frac{dR}{dt} = R\frac{SA}{3}MW_1\sum_{i=1}^{6} ir_j^{'} \tag{2.1}$$

$$SA = \frac{3}{R\rho_{cellulose}} \tag{2.2}$$

Cellulose Area per Volume Equation:

$$\frac{A_{cel}}{V_{liq}} = 4\pi R^2 n \tag{2.A1}$$

Soluble Saccharide Mass Balances:

$$\begin{aligned}\frac{d[C_1]}{dt} = &\left(2\frac{k_{cat-EG-cel}}{K_{M-EG-cel}}[E^{'}_{EG-ads}]\theta_{EG}\sum_{i=7}^{\infty}[C_i^{'}] \right.\\ &+ 2\frac{k_{cat-EG-sol}}{K_{M-EG-sol}}[E^{'}_{EG-ads}]\sum_{i=3}^{6}[C_i]\\ &\left. + \frac{k_{cat-CBH-sol}}{K_{M-CBH-sol}}[E^{'}_{CBH-ads}][C_3] \right)\frac{A_{cel}}{V_{liq}}\\ &+ 2\frac{k_{cat-EG-sol}}{K_{M-EG-sol}}[E_{EG-free}]\sum_{i=3}^{6}[C_i]\\ &+ \frac{k_{cat-CBH-sol}}{K_{M-CBH-sol}}[E_{CBH-free}][C_3]\end{aligned} \tag{2.A2}$$

$$\frac{d[C_2]}{dt} = \left(2\frac{k_{cat-EG-cel}}{K_{M-EG-cel}}[E'_{EG-ads}]\theta_{EG}\sum_{i=7}^{\infty}[C'_i] + \frac{k_{cat-CBH-cel}}{K_{M-CBH-cel}}[E'_{CBH-ads}]\theta_{CBH}\sum_{i=7}^{\infty}[C'_i] + 2\frac{k_{cat-EG-sol}}{K_{M-EG-sol}}[E'_{EG-ads}]\sum_{i=3}^{6}[C_i] + \frac{k_{cat-CBH-sol}}{K_{M-CBH-sol}}[E'_{CBH-ads}]\left(\sum_{i=3}^{6}[C_i]+[C_4]\right)\right)\frac{A_{cel}}{V_{liq}} + 2\frac{k_{cat-EG-sol}}{K_{M-EG-sol}}[E_{EG-free}]\sum_{i=3}^{6}[C_i] + \frac{k_{cat-CBH-sol}}{K_{M-CBH-sol}}[E_{CBH-free}]\left(\sum_{i=3}^{6}[C_i]+[C_4]\right) \quad (2.7)$$

$$(i = 3,4) \quad \frac{d[C_i]}{dt} = \left(2\frac{k_{cat-EG-cel}}{K_{M-EG-cel}}[E'_{EG-ads}]\theta_{EG}\sum_{j=7}^{\infty}[C'_j] + 2\frac{k_{cat-EG-sol}}{K_{M-EG-sol}}[E'_{EG-ads}]\left(\sum_{j=3}^{6}[C_j]-(i-1)[C_i]\right) + \frac{k_{cat-CBH-sol}}{K_{M-CBH-sol}}[E'_{CBH-ads}]\left([C_{i+2}]-[C_i]\right)\right)\frac{A_{cel}}{V_{liq}} + 2\frac{k_{cat-EG-sol}}{K_{M-EG-sol}}[E_{EG-free}]\left(\sum_{j=3}^{6}[C_j]-(i-1)[C_i]\right) + \frac{k_{cat-CBH-sol}}{K_{M-CBH-sol}}[E_{CBH-free}]\left([C_{i+1}]-[C_i]\right) \quad (2.A3)$$

$$(i = 5,6) \qquad \frac{d[C_i]}{dt} = \left(2\frac{k_{cat-EG-cel}}{K_{M-EG-cel}}[E'_{EG-ads}]\theta_{EG}\sum_{j=7}^{\infty}[C'_j] \right.$$

$$+ \frac{k_{cat-CBH-cel}}{K_{M-CBH-cel}}[E'_{CBH-ads}]\theta_{CBH}[C'_{i+2}]$$

$$+ 2\frac{k_{cat-EG-sol}}{K_{M-EG-sol}}[E'_{EG-ads}]\left(\sum_{j=3}^{6}[C_j] - (i-1)[C_i] \right)$$

$$\left. - \frac{k_{cat-CBH-sol}}{K_{M-CBH-sol}}[E'_{CBH-ads}][C_i] \right) \frac{A_{cel}}{V_{liq}}$$

$$+ 2\frac{k_{cat-EG-sol}}{K_{M-EG-sol}}[E_{EG-free}]\left(\sum_{j=3}^{6}[C_j] - (i-1)[C_i] \right)$$

$$+ \frac{k_{cat-CBH-sol}}{K_{M-CBH-sol}}[E_{CBH-free}][C_i] \qquad (2.A4)$$

Solid Cellulose Mass Balances:

$$(i > 6) \qquad \frac{d[C'_i]}{dt} = \frac{k_{cat-EG-cel}}{K_{M-EG-cel}}[E'_{EG-ads}]\theta_{EG}\left(2\sum_{j=i+1}^{\infty}[C'_j] - (i-1)[C'_i] \right)$$

$$+ \frac{k_{cat-CBH-cel}}{K_{M-CBH-cel}}[E'_{CBH-ads}]\theta_{CBH}\left([C'_{i+2}] - [C'_i] \right) \qquad (2.6)$$

Cellulose Site Balance Equations:

$$\frac{[*']}{[*'_{\max}]} = \theta = 1 - \frac{\sigma_{EG}[E'_{EG-ads}]}{[*'_{\max}]}\left(1 + \sum_{i=7}^{\infty}\frac{\theta_{EG}(i-1)[C'_i]}{K_{M-EG-cel}} \right.$$

$$\left. + \sum_{i=3}^{6}\frac{(i-1)[C_i]}{K_{M-EG-sol}} + \frac{[C_1]}{K_{i-EG-glucose}} + \frac{[C_2]}{K_{i-EG-cellobiose}} \right)$$

$$- \frac{\sigma_{CBH}[E'_{CBH-ads}]}{[*_{\max}]'}\left(1 + \sum_{i=7}^{\infty}\frac{\theta_{CBH}[C'_i]}{K_{M-CBH-cel}} + \sum_{i=3}^{6}\frac{[C_i]}{K_{M-CBH-sol}} \right.$$

$$\left. + \frac{[C_1]}{K_{i-CBH-glucose}} + \frac{[C_2]}{K_{i-CBH-cellobiose}} \right) \qquad (2.A5)$$

$$\theta_{EG} = \theta + \frac{\sigma_{EG}[E'_{EG-ads}]}{[*'_{\max}]} \tag{2.A6}$$

$$\theta_{CBH} = \theta + \frac{\sigma_{CBH}[E'_{CBH-ads}]}{[*'_{\max}]} \tag{2.A7}$$

Enzyme Balances:

$$[E_{EG}] = \frac{\left([E'_{EG-ads}] + \frac{[E'_{EG-ads}]\theta_{EG}}{K_{M-EG-cel}} \sum_{i=7}^{\infty}(i-1)[C'_i]\right)\frac{A_{cel}}{V_{liq}}}{1 + \sum_{i=3}^{6}\frac{(i-1)[C_i]}{K_{M-EG-sol}} + \frac{[C_1]}{K_{i-EG-glucose}} + \frac{[C_2]}{K_{i-EG-cellobiose}}} + \frac{\left(\frac{[E'_{EG-ads}]}{K_{M-EG-sol}} \sum_{i=3}^{6}(i-1)[C_i] + \frac{[E'_{EG-ads}][C_1]}{K_{i-EG-glucose}} + \frac{[E'_{EG-ads}][C_2]}{K_{i-EG-cellobiose}}\right)\frac{A_{cel}}{V_{liq}}}{1 + \sum_{i=3}^{6}\frac{(i-1)[C_i]}{K_{M-EG-sol}} + \frac{[C_1]}{K_{i-EG-glucose}} + \frac{[C_2]}{K_{i-EG-cellobiose}}} \tag{2.A8}$$

$$[E_{CBH}] = \frac{\left([E'_{CBH-ads}] + \frac{[E'_{CBH-ads}]\theta_{CBH}}{K_{M-CBH-cel}} \sum_{i=7}^{\infty}[C'_i]\right)\frac{A_{cel}}{V_{liq}}}{1 + \sum_{i=3}^{6}\frac{[C_i]}{K_{M-CBH-sol}} + \frac{[C_1]}{K_{i-CBH-glucose}} + \frac{[C_2]}{K_{i-CBH-cellobiose}}} + \frac{\left(\frac{[E'_{CBH-ads}]}{K_{M-CBH-sol}} \sum_{i=3}^{6}[C_i] + \frac{[E'_{CBH-ads}][C_1]}{K_{i-CBH-glucose}} + \frac{[E'_{EG-ads}][C_2]}{K_{i-CBH-cellobiose}}\right)\frac{A_{cel}}{V_{liq}}}{1 + \sum_{i=3}^{6}\frac{[C_i]}{K_{M-CBH-sol}} + \frac{[C_1]}{K_{i-CBH-glucose}} + \frac{[C_2]}{K_{i-CBH-cellobiose}}} \tag{2.A9}$$

2.10 Appendix 2: Random Sequential Adsorption Simulations

Random sequential adsorption (RSA) describes a process where shapes are placed onto a surface at random locations subject only to the constraint that no shape is allowed to overlap with another shape. Cellulase enzymes adsorbing to a surface may be described by such a random sequential adsorption process. If an RSA simulation is performed to the limit where no unoccupied areas on the surface are large enough to accommodate a new cellulase, the saturation limit of cellulase on a model cellulose surface may be determined.

A stochastic simulation approach was used for the RSA simulations used in this work. The cellulose surface was represented by a periodic square lattice. A periodic lattice was used to account for the small size of the cellobiose units compared to the actual surface area of a cellulose particle. The adsorption area was assumed to be square lattice with minimum image boundary conditions; the adsorbate shape could extend fromone side of the lattice and appear on the other side. Adsorbate shapes were defined by the arrangements of sites they would occupy on the lattice when on the surface.

Shapes were placed using one of two methods. Early in the simulations, when little adsorption had occured, a lattice site was randomly chosen on the surface and that lattice site was then used as an anchor to test whether a shape with a random orientation could be placed there without overlapping previously placed shapes. If the shape did not overlap any occupied sites the free lattice sites it covered were then considered to be occupied; if the shape overlapped anywhere it was not placed. This process was repeated until a preset threshold of failed placements occurred in successive placement attempts. After this threshold was reached, a second shape placement algorithm was used. The second algorithm was designed to be more efficient on an already crowded surface. Each lattice site was examined to see if a shape in any allowed orientation could successfully be placed using that site as its anchor. One of these site and shape orientation combinations was then randomly chosen and the lattice was altered to reflect this new shape placement. This process was repeated, updating the list of possible adsorption sites based on the new lattice configuration each iteration until there were no possible lattice sites that would lead to the successful placement of a new shape. These two algorithms used together allow for efficient RSA simulations to determine the saturation coverage on a surface (*58*). The simulation program was written and run using Matlab. It was validated by determining the saturation coverage for shapes with available literature values: linear dimers, trimers, and square tetramers (*58*) (data no shown). The saturation coverage was determined by averaging the result of five separate stochastic simulations. As the size of the lattice was increased the standard deviation of these results declined.

To determine the saturation coverage for a typical cellulase on cellulose a 512-by-512 periodic square lattice was used for the cellulose surface. This gave a standard deviation for the saturation coverage of less than 0.5%. Each site on the lattice was assumed to be the size of a cellobiose unit. The adsorbing shape was designed to mimic a cellulase with a large coverage domain to represent the catalytic domain, a small coverage domain to represent the cellulose binding module, and a short, thin region to represent a short linker region. The cellulase shape was designed to occupy 42 cellobiose sites on the surface. Figure 2.A1 depicts the cellulase shape used for adsorption. Five RSA simulation runs conducted using this shape and surface gave saturation coverage of $53.5 \pm 0.1\%$. This saturation coverage gives an effective footprint of

78 cellobiose units (156 glucose units) for a cellulase that is 42 cellobiose units large (84 glucose units). This effective footprint accounts not just for the area that is physically blocked by an adsorbed cellulase but also for the area that is inaccessible because of the crowding of multiple adsorbed enzymes.

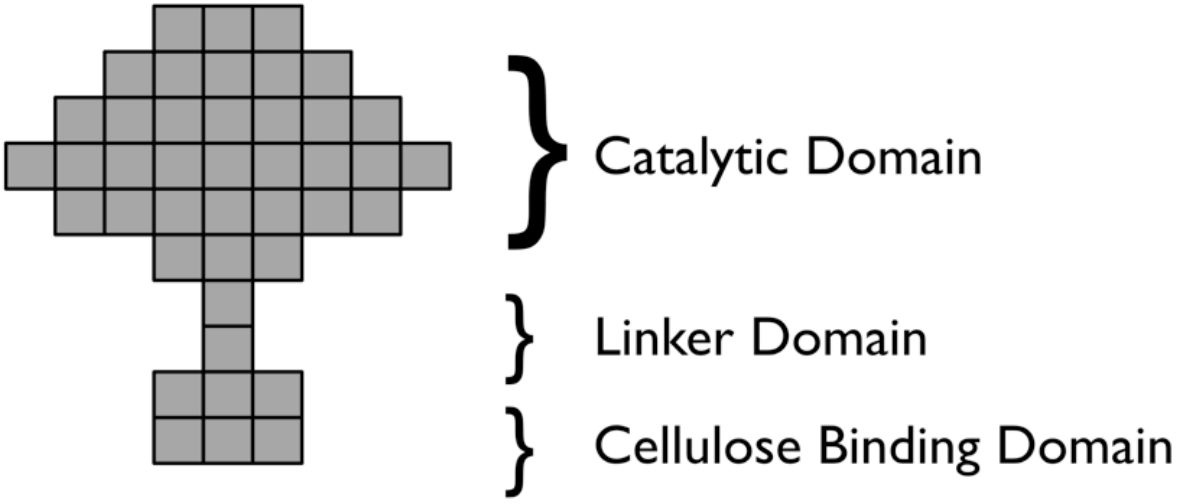


Figure 2.A1 The "cellulase" shape used in the random sequential adsorption simulations used to determine the effective size of the enzyme on a cellulose surface. The shape was designed to take the typical size of a cellulase of 42 cellobiose units (21 nm^2) and distribute it properly among the three basic cellulase domains. A square in the shape is a cellobiose unit.

2.11 Appendix 3: Derivation of Model Equations

2.11.1 Cellulose Chain Surface Concentrations

The equations describing the surface concentrations of cellulose chains incorporate the action of the adsorbed cellulase enzymes. A surface reaction mechanism for each cellulase can be developed from the overall hydrolysis mechanism (Fig. 2.1) as shown in Fig. 2.A2.

Endogluconase

$$E_{EG-ads} + C_i \underset{k_{decomp-EG-C_i}}{\overset{k_{comp-EG-C_i}}{\Leftrightarrow}} E_{EG-ads}C_i \overset{k_{cat-EG-cel}}{\rightarrow} E_{EG-ads} + C_{i-x} + C_x$$

$$E_{EG-ads} + C_1 \underset{k_{decomp-EG-C_1}}{\overset{k_{comp-EG-C_1}}{\Leftrightarrow}} E_{EG-ads}C_1$$

$$E_{EG-ads} + C_2 \underset{k_{decomp-EG-C_2}}{\overset{k_{comp-EG-C_2}}{\Leftrightarrow}} E_{EG-ads}C_2$$

Cellobiohydrolase

$$E_{CBH-ads} + C_i \underset{k_{decomp-CBH-C_i}}{\overset{k_{comp-CBH-C_i}}{\Leftrightarrow}} E_{CBH-ads}C_i \overset{k_{cat-CBH-cel}}{\rightarrow} E_{CBH-ads} + C_{i-2} + C_2$$

$$E_{CBH-ads} + C_1 \underset{k_{decomp-EG-C_1}}{\overset{k_{comp-EG-C1}}{\Leftrightarrow}} E_{CBH-ads}C_1$$

$$E_{CBH-ads} + C_2 \underset{k_{decomp-EG-C_2}}{\overset{k_{comp-EG-C_2}}{\Leftrightarrow}} E_{CBH-ads}C_2$$

Figure 2.A2 Mechanism for surface action of a typical endogluconase and a typical cellobiohydrolase on cellulose. $2<i\leq 6$ are soluble sugars and $6<i$ are solid cellulose chains.

Competitive inhibition by cellobiose and glucose have also been included in the surface reaction mechanism. Using this mechanism, balance equations for the surface concentration of solid cellulose chains (DP>6) and for adsorbed enzyme complexed with a solid cellulose chain with respect with time can be written (Eqs. 2.A10-2.A12)

$$\begin{aligned}\frac{d[C_i^{'}]}{dt} &= -k_{comp-EG-C_i}[E_{EG-ads}^{'}]\theta_{EG}(i-1)[C_i^{'}] + k_{decomp-EG-C_i}[E_{EG-ads}C_i^{'}] \\ &+2k_{cat-EG-cel}\sum_{j=i+1}^{\infty}\frac{[E_{EG-ads}C_j^{'}]}{(j-1)} \\ &-k_{comp-EG-C_1}[E_{EG-ads}^{'}][C_1] + k_{decomp-EG-C_1}[E_{EG-ads}C_1^{'}] \\ &-k_{comp-EG-C_2}[E_{EG-ads}^{'}][C_2] + k_{decomp-EG-C_2}[E_{EG-ads}C_2^{'}] \\ &-k_{comp-CBH-C_i}[E_{CBH-ads}^{'}]\theta_{CBH}[C_i^{'}] + k_{decomp-CBH-C_i}[E_{CBH-ads}C_i^{'}] \\ &+k_{cat-CBH-cel}[E_{EG-ads}C_{i+2}^{'}] \\ &-k_{comp-CBH-C_1}[E_{CBH-ads}^{'}][C_1] + k_{decomp-CBH-C_1}[E_{CBH-ads}C_1^{'}] \\ &-k_{comp-CBH-C_2}[E_{CBH-ads}^{'}][C_2] + k_{decomp-CBH-C_2}[E_{CBH-ads}C_2^{'}]\end{aligned} \tag{2.A10}$$

$$\begin{aligned}\frac{d[E_{EG-ads}C_i^{'}]}{dt} &= k_{comp-EG-C_i}[E_{EG-ads}^{'}]\theta_{EG}(i-1)[C_i^{'}] \\ &-(k_{decomp-EG-C_i}+k_{cat-EG-cel})[E_{EG-ads}C_i^{'}]\end{aligned} \tag{2.A11}$$

$$\begin{aligned}\frac{d[E_{CBH-ads}C_i^{'}]}{dt} &= k_{comp-CBH-C_i}[E_{CBH-ads}^{'}]\theta_{CBH}[C_i^{'}] \\ &-(k_{decomp-CBH-C_i}+k_{cat-CBH-cel})[E_{CBH-ads}C_i^{'}]\end{aligned} \tag{2.A12}$$

Cleavable glycosidic bonds are carefully tracked. The endogluconase complexation term includes an (i-1) multiplier to account for the number of glycosidic bonds within a cellulose chain of length i. Enzymes bound to any of these bonds are included in the complexed enzyme term, so proper accounting is required because only a fraction of the bond cleavages of an enzyme-cellulose chain complex result in the formation of a chain of a particular length. For a

chain of length j, where j is larger than i, there are two out of j-1 cleavable bonds that will result in the formation of a chain of length i. Accordingly, in Eq. 2.A10, the term for the formation of a chain of length i from an endoglucanase-cellulose complex is multiplied by 2/(j-1). This relies on the assumption that all have an equal probability of complexing with an endoglucanase. No such assumptions are required for the cellobiohydrolase terms, which depend only on cellulose chain surface concentrations; each chain has only one reducing or non-reducing end.

Equations 2.A11 and 2.A12 lead to Eq. 2.4 if two assumptions are made (1) the complexation reaction is in equilibrium and (2) the catalytic step is slow in comparison with the rate of complexation.

If the inhibition reactions are assumed to be in equilibrium, the terms involving cellobiose and glucose in Eq. A10 cancel each other out. Eq. 2.4 can then be substituted into Eq. 2.A10 to which can be simplified based on $K_{M\text{-}x\text{-}cel} = (k_{cat\text{-}x\text{-}cel}+k_{decomp\text{-}x\text{-}Ci})/k_{comp\text{-}x\text{-}Ci}$ to yield Eq. 2.5. This is the basic model differential equation for the surface concentration of a solid cellulose chain of length i with respect for time.

2.11.2 Soluble Cello-oligosaccharide Concentrations

Soluble cello-oligosaccharides are both substrates and products for cellulase enzymes that are either free in solution or adsorbed to solid cellulose. The balance equations for the soluble sugars must capture all formation and consumption processes.

The terms describing the formation of soluble sugars from solid cellulose are identical to the terms derived for the action of cellulases on long cellulose chains. Based on the mechanism shown in Fig. 2.A2, cellobiohydrolase catalytic activity will always form cellobiose. If the solid chain is a C_7 or C_8 oligomer, cellopentaose or cellohexaose will also be formed. An endoglucanase can produce soluble sugars on chains of any length if the bond that is cleaved is six or fewer glucose units from a chain end. The derivations discussed above can be adapted to develop the terms for the formation of soluble sugars from solid cellulose as shown in Eqs. 2.A13-2.A15.

$$(i = 1,3,4) \qquad \frac{d[C_i]}{dt} = \left(2\frac{k_{cat-EG-cel}}{K_{M-EG-cel}}[E'_{EG-ads}]\theta_{EG}\sum_{j=7}^{\infty}[C'_j] \right)\frac{A_{cel}}{V_{liq}} \tag{2.A13}$$

$$\frac{d[C_2]}{dt} = \left(2\frac{k_{cat-EG-cel}}{K_{M-EG-cel}}[E'_{EG-ads}]\theta_{EG}\sum_{i=7}^{\infty}[C'_i] \right)\frac{A_{cel}}{V_{liq}} + \left(\frac{k_{cat-CBH-cel}}{K_{M-CBH-cel}}[E'_{CBH-ads}]\theta_{CBH}\sum_{i=7}^{\infty}[C'_i] \right)\frac{A_{cel}}{V_{liq}} \tag{2.A14}$$

$$(i = 5,6) \qquad \frac{d[C_i]}{dt} = \left(2\frac{k_{cat-EG-cel}}{K_{M-EG-cel}}[E'_{EG-ads}]\theta_{EG}\sum_{i=7}^{\infty}[C'_i] \right)\frac{A_{cel}}{V_{liq}} + \left(\frac{k_{cat-CBH-cel}}{K_{M-CBH-cel}}[E'_{CBH-ads}]\theta_{CBH}[C'_{i+2}] \right)\frac{A_{cel}}{V_{liq}} \tag{2.A15}$$

The A_{cel}/V_{liq} multipliers are used to convert the surface reaction rates to volumetric reaction rates.

The remaining terms in the balance equations for soluble sugar concentrations describe the activity of enzymes on cellooligosaccharides that are free in solution (cellotriose up to cellohexaose). Adsorbed cellulases can react with soluble substrate according to the mechanism shown in Fig. 2.A2, where the degree of polymerization (i) will vary from 3 to 6. The action of soluble enzyme on a soluble sugar is mechanistically identical; the adsorbed enzyme term just needs to be replaced with a free enzyme term.

A derivation similar to that used above can be used for the terms involving enzyme action on soluble substrates. Two slight differences exist (1) the θ_{EG} or θ_{CBH} will no longer be used in the equilibrium relationships for enzyme-substrate complex concentrations and (2) soluble phase complexation constants will be needed in place of the solid phase constants. With these changes and careful attention to reaction stoichiometries the following additional term are obtained for the soluble sugar differential equations (Eqs. 2.A16-2.A18):

$$\frac{d[C_1]}{dt} = \left(2\frac{k_{cat-EG-sol}}{K_{M-EG-sol}}[E'_{EG-ads}]\sum_{i=3}^{6}[C_i] + \frac{k_{cat-CBH-sol}}{K_{M-CBH-sol}}[E'_{CBH-ads}][C_3] \right)\frac{A_{cel}}{V_{liq}} + 2\frac{k_{cat-EG-sol}}{K_{M-EG-sol}}[E_{EG-free}]\sum_{i=3}^{6}[C_i] + \frac{k_{cat-CBH-sol}}{K_{M-CBH-sol}}[E_{CBH-free}][C_3] \tag{2.A16}$$

$$\frac{d[C_2]}{dt} = \left(2\frac{k_{cat-EG-sol}}{K_{M-EG-sol}}[E'_{EG-ads}]\sum_{i=3}^{6}[C_i] + \frac{k_{cat-CBH-sol}}{K_{M-CBH-sol}}[E'_{CBH-ads}]\left(\sum_{i=3}^{6}[C_i] + [C_4] \right) \right)\frac{A_{cel}}{V_{liq}} + 2\frac{k_{cat-EG-sol}}{K_{M-EG-sol}}[E_{EG-free}]\sum_{i=3}^{6}[C_i] + \frac{k_{cat-CBH-sol}}{K_{M-CBH-sol}}[E_{CBH-free}]\left(\sum_{i=3}^{6}[C_i] + [C_4] \right) \tag{2.A17}$$

$$(i = 3,4) \qquad \frac{d[C_i]}{dt} = \left(2\frac{k_{cat-EG-sol}}{K_{M-EG-sol}}[E'_{EG-ads}]\left(\sum_{j=3}^{6}[C_j] - (i-1)[C_i] \right) + \frac{k_{cat-CBH-sol}}{K_{M-CBH-sol}}[E'_{CBH-ads}]\left([C_{i+2}] - [C_i] \right) \right)\frac{A_{cel}}{V_{liq}} + 2\frac{k_{cat-EG-sol}}{K_{M-EG-sol}}[E_{EG-free}]\left(\sum_{j=3}^{6}[C_j] - (i-1)[C_i] \right) + \frac{k_{cat-CBH-sol}}{K_{M-CBH-sol}}[E_{CBH-free}]\left([C_{i+2}] - [C_i] \right) \tag{2.A18}$$

$$(i = 5,6) \qquad \frac{d[C_i]}{dt} = \left(2\frac{k_{cat-EG-sol}}{K_{M-EG-sol}}[E'_{EG-ads}]\left(\sum_{j=3}^{6}[C_j] - (i-1)[C_i] \right) - \frac{k_{cat-CBH-sol}}{K_{M-CBH-sol}}[E'_{CBH-ads}][C_i] \right) \frac{A_{cel}}{V_{liq}}$$
$$+ 2\frac{k_{cat-EG-sol}}{K_{M-EG-sol}}[E_{EG-free}]\left(\sum_{j=3}^{6}[C_j] - (i-1)[C_i] \right) - \frac{k_{cat-CBH-sol}}{K_{M-CBH-sol}}[E_{CBH-free}][C_i] \tag{2.A19}$$

When Eq. 2.A17 is combined with Eq. 2.A14, Eq. 2.6 is obtained. The other equations for the soluble sugar concentrations can be formed by combining their terms from Eqs. 2.A13-2.A19.

2.11.3 Site Balance

The cellulose surface is occupied by a variety of species: adsorbed uncomplexed enzymes, adsorbed enzymes complexed with solid substrate, adsorbed enzymes complexed with soluble substrates, and adsorbed enzymes inhibited by cellobiose or glucose. If the footprints for these species are known, a term for surface concentration of free sites on the surface ([*']) can be used to develop a simple site balance. The surface concentration of total sites ([*'$_{max}$]) is assumed to remain constant (Eq. 2.A20).

$$[*'_{max}] = [*'] + \sigma_{EG}\left([E'_{EG-ads}] + \sum_{i=7}^{\infty}[E'_{EG-ads}C_i] + \sum_{i=3}^{6}[E_{EG-ads}C'_i] + [E_{EG-ads}C'_1] + [E_{EG-ads}C'_2] \right)$$
$$+ \sigma_{CBH}\left([E'_{CBH-ads}] + \sum_{i=7}^{\infty}[E_{CBH-ads}C'_i] + \sum_{i=3}^{6}[E_{CBH-ads}C'_i] + [E_{CBH-ads}C'_1] + [E_{CBH-ads}C'_2] \right) \tag{2.A20}$$

Equilibrium relationships for the inhibited and complexed enzyme terms can be used to simplify the site balance (Eq. 2.A21).

$$[*'_{max}] = [*'] + \sigma_{EG}[E'_{EG-ads}]\left(1 + \sum_{i=7}^{\infty}\frac{\theta_{EG}(i-1)[C'_i]}{K_{M-EG-cel}} + \sum_{i=3}^{6}\frac{(i-1)[C'_i]}{K_{M-EG-sol}} + \frac{[C_1]}{K_{i-EG-glucose}} + \frac{[C_2]}{K_{i-EG-cellobiose}} \right)$$
$$+ \sigma_{CBH}[E'_{CBH-ads}]\left(1 + \sum_{i=7}^{\infty}\frac{\theta_{CBH}[C'_i]}{K_{M-CBH-cel}} + \sum_{i=3}^{6}\frac{[C'_i]}{K_{M-CBH-sol}} + \frac{[C_1]}{K_{i-CBH-glucose}} + \frac{[C_2]}{K_{i-CBH-cellobiose}} \right) \tag{2.A21}$$

The free site fraction can be determined from the site balance by subtracting the terms related to the enzyme surface concentrations (in all forms) from both sides and by dividing both sides by $[*_{max}]'$ (Eq. 2.A5).

$$\frac{[*']}{[*'_{\max}]} = \theta = 1 - \frac{\sigma_{EG}[E'_{EG-ads}]}{[*'_{\max}]}\left(1 + \sum_{i=7}^{\infty}\frac{\theta_{EG}(i-1)[C'_i]}{K_{M-EG-cel}} + \sum_{i=3}^{6}\frac{(i-1)[C_i]}{K_{M-EG-sol}} + \frac{[C_1]}{K_{i-EG-glucose}} + \frac{[C_2]}{K_{i-EG-cellobiose}}\right) - \frac{\sigma_{CBH}[E'_{CBH-ads}]}{[*'_{\max}]}\left(1 + \sum_{i=7}^{\infty}\frac{\theta_{CBH}[C'_i]}{K_{M-CBH-cel}} + \sum_{i=3}^{6}\frac{[C_i]}{K_{M-CBH-sol}} + \frac{[C_1]}{K_{i-CBH-glucose}} + \frac{[C_2]}{K_{i-CBH-cellobiose}}\right) \quad (2.A5)$$

For a particular enzyme, the fraction of sites available for complexation can be found by adding the fraction of sites occupied by uncomplexed adsorbed enzymes of that specific type to θ, to yield θ_{EG} (Eq. 2.A6) and θ_{CBH} (Eq. 2.A7).

$$\theta_{EG} = \theta + \frac{\sigma_{EG}[E'_{EG-ads}]}{[*'_{\max}]} \quad (2.A6)$$

$$\theta_{CBH} = \theta + \frac{\sigma_{CBH}[E'_{CBH-ads}]}{[*'_{\max}]} \quad (2.A7)$$

The site balance is used for determining the surface concentration of free sites for adsorption while the enzyme specific available site fractions are used for scaling the concentrations of surface chains in the complexation relationships to account for enzyme crowding.

2.11.4 Enzyme Balances

Steady-state enzyme balances, which are used in the adsorption equations as well as in equations representing soluble-phase enzyme action, can be written for each type of cellulase in the system. The basic enzyme balances for a typical endoglucanase and cellobiohydrolase are shown in Eqs. 2.A22 and 2.A23.

$$[E_{EG-total}] = [E_{EG}] + \sum_{i=3}^{6}[E_{EG}C_i] + [E_{EG}C_1] + [E_{EG}C_2] + \left([E'_{EG-ads}] + \sum_{i=7}^{\infty}[E_{EG-ads}C'_i] + \sum_{i=3}^{6}[E_{EG-ads}C'_i] + [E_{EG-ads}C'_1] + [E_{EG-ads}C'_2]\right)\frac{A_{cel}}{V_{liq}} \quad (2.A22)$$

$$[E_{CBH-total}] = [E_{CBH}] + \sum_{i=3}^{6}[E_{CBH}C_i] + [E_{CBH}C_1] + [E_{CBH}C_2] + \left([E'_{CBH-ads}] + \sum_{i=7}^{\infty}[E_{CBH-ads}C'_i] + \sum_{i=3}^{6}[E_{CBH-ads}C'_i] + [E_{CBH-ads}C'_1] + [E_{CBH-ads}C'_2]\right)\frac{A_{cel}}{V_{liq}} \quad (2.A23)$$

If equilibrium relationships are substituted into Eqs. 2.A22 and 2.A23, these equations can be solved for the concentration of free enzyme in solution (Eqs. 2.A8 and 2.A9).

$$[E_{EG}] = \frac{\left([E'_{EG-ads}] + \frac{[E'_{EG-ads}]\theta_{EG}}{K_{M-EG-cel}} \sum_{i=7}^{\infty} (i-1)[C'_i]\right)\frac{A_{cel}}{V_{liq}}}{1 + \sum_{i=3}^{6} \frac{(i-1)[C_i]}{K_{M-EG-sol}} + \frac{[C_1]}{K_{i-EG-glucose}} + \frac{[C_2]}{K_{i-EG-cellobiose}}} + \frac{\left(\frac{[E'_{EG-ads}]}{K_{M-EG-sol}} \sum_{i=3}^{6} (i-1)[C_i] + \frac{[E'_{EG-ads}][C_1]}{K_{i-EG-glucose}} + \frac{[E'_{EG-ads}][C_2]}{K_{i-EG-cellobiose}}\right)\frac{A_{cel}}{V_{liq}}}{1 + \sum_{i=3}^{6} \frac{(i-1)[C_i]}{K_{M-EG-sol}} + \frac{[C_1]}{K_{i-EG-glucose}} + \frac{[C_2]}{K_{i-EG-cellobiose}}} \tag{2.A8}$$

$$[E_{CBH}] = \frac{\left([E'_{CBH-ads}] + \frac{[E'_{CBH-ads}]\theta_{CBH}}{K_{M-CBH-cel}} \sum_{i=7}^{\infty} [C'_i]\right)\frac{A_{cel}}{V_{liq}}}{1 + \sum_{i=3}^{6} \frac{[C_i]}{K_{M-CBH-sol}} + \frac{[C_1]}{K_{i-CBH-glucose}} + \frac{[C_2]}{K_{i-CBH-cellobiose}}} + \frac{\left(\frac{[E'_{CBH-ads}]}{K_{M-CBH-sol}} \sum_{i=3}^{6} [C_i] + \frac{[E'_{CBH-ads}][C_1]}{K_{i-CBH-glucose}} + \frac{[E'_{EG-ads}][C_2]}{K_{i-CBH-cellobiose}}\right)\frac{A_{cel}}{V_{liq}}}{1 + \sum_{i=3}^{6} \frac{[C_i]}{K_{M-CBH-sol}} + \frac{[C_1]}{K_{i-CBH-glucose}} + \frac{[C_2]}{K_{i-CBH-cellobiose}}} \tag{2.A9}$$

2.11.5 Cellulose Material Balance

The shrinking size of the cellulose particles is tracked using a material balance. If soluble sugars formed from the solid cellulose are assumed to instantly enter the soluble phase, then this is the only loss of mass from the cellulose particles (Eq. 2.A24).

$$\frac{dM_{particle}}{dt} = -MW_1 A_{particle} \sum_{i=1}^{6} ir'_i \tag{2.A24}$$

For a spherical particle, this mass balance can be simplified to Eq. 2.A25.

$$\frac{dR}{dt} = -\frac{MW_1}{\rho_{cellulose}} \sum_{i=1}^{6} ir'_i \tag{2.A25}$$

The density of the cellulose, the radius of the particle, and the surface area per mass of the cellulose (SA) are related by Eq. 2.A26 for a spherical particle.

$$\rho_{cellulose} = \frac{3}{R \cdot SA} \tag{2.A26}$$

When Eq. 2.A26 is substituted into Eq. 2.A25, Eq. 2.1 is obtained. Mass balances for other particle shapes can be developed using Equation 2.A24 as a starting point.

2.12 Appendix 4: Model Parameters

Parameter	Value	Units	Justification
DP_0	300	-	The degree of polymerization of a solid cellulosic substrate is difficult to determine because there are limited solvents that dissolve cellulose, a requirement for standard techniques such as viscometry and gel permeation chromatography. The value of 300 was chosen as a typical average DP for Avicel based on the reported values in the literature ranging from 181 to 327 (Ng and Zelkus 1980; Ryu et al. 1982).
$[E_{EG2\text{-}total}]$	0 or 1.6	$\mu mol\ L^{-1}$	Enzyme loadings set to match experimental conditions used in Medve et al. (1998)
$[E_{CBHI\text{-}total}]$	0 or 1.6	$\mu mol\ L^{-1}$	Enzyme loadings set to match experimental conditions used in Medve et al. (1998)
$k_{CBHI\text{-}ads}$	8640	$L\ mmol^{-1}\ hr^{-1}$	Value measured using elipsometry to determine the adsorption and desorption kinetics of a *T. reesei* cellulase mixture on a spin coated cellulose surface (Mauer and Radke 2009). While this value is not specific to an individual cellulase, it can be considered to be in an appropriate range for CBHI and the other major components of the *T. reesei* cellulase mixture.
$k_{EG2\text{-}ads}$	8640	$L\ mmol^{-1}\ hr^{-1}$	Value measured using elipsometry to determine the adsorption and desorption kinetics of a *T. reesei* cellulase mixture on a spin coated cellulose surface (Mauer and Radke 2009). While this value is not specific to an individual cellulase, it can be considered to be in an appropriate range for EG2 and the other major components of the *T. reesei* cellulase mixture.
$k_{cat\text{-}CBHI}$	34000	hr^{-1}	The same value of k_{cat} for CBHI was used for enzymatic catalysis of both soluble oligosaccharides as well as for solid cellulose. It was assumed that once a cellulose chain was bound in the cellulase active site the rate of catalysis would be minimally affected by the differences between soluble short chain oligosaccharides and solid long chain cellulose. The value reported by Nidetzky et al. (1994) for hydrolysis of cellohexaose by *T. reesei* CBHI was used in this work. The values were determined using a Lineweaver-Burk plot (conditions: 2.1 nM of CBHI, 1-100 µM of cellohexaose, 30 °C, acetate buffer, pH 4.8).
$k_{cat\text{-}EG2}$	234000	hr^{-1}	The same value of k_{cat} for EG2 was used for enzymatic catalysis of both soluble oligosaccharides and solid cellulose. It was assumed that once a cellulose chain was bound in the cellulase active site the rate of catalysis would be minimally affected by the differences between soluble short chain oligosaccharides and solid long chain cellulose. The value reported by Karlsson et al. (2002) for hydrolysis of cellopentaose by *T. reesei* EG2 was used in this work. The values were determined using a Lineweaver-Burk plot (conditions: 1.0 nM of CBHI, 10-150 µM of cellopentaose, 40 °C, Na-acetate buffer, pH 5.0).

$k_{CBHI\text{-}des}$	164	hr^{-1}	Value based on elipsometry experiments to determine the adsorption and desorption kinetics of a *T. reesei* cellulase mixture on a spin coated cellulose surface (Mauer and Radke 2009). The value increased for CBHI to give adsorption behavior similar to that reported by Medve et al. (1997) in the study used for comparison in this work, which was rapid achievement of a steady value of approximately 95% of the CBHI adsorbed in both the single enzyme and the mixed enzyme experiments.
$k_{EG2\text{-}des}$	19.3	hr^{-1}	Value based on elipsometry experiments to determine the adsorption and desorption kinetics of a *T. reesei* cellulase mixture on a spin coated cellulose surface (Mauer and Radke 2009). The value increased for EG2 to give adsorption behavior similar to that reported by Medve et al. (1997) in the study used for comparison in this work, which was rapid achievement of a steady value of approximately 95% of the EG2 adsorbed in both the single enzyme and the mixed enzyme experiments. In the mixed enzyme experiments Medve et al. showed that EG2 adsorbed more readily than CBHI, hence the use of a lower $k_{desorption}$ for EG2 than for CBHI.
$K_{i\text{-}CBHI\text{-}cellobiose}$	0.093	mmol L^{-1}	Cellobiose is known to strongly inhibit the action of cellulases but the reported values for K_i for *T. reesei* cellulases range from 0.021 mM to 1.4 mM for hydrolysis of soluble substrates (Hsu et al. 1980; Koivula et al. 1996; Nidetzky et al. 1994; von Ossowski et al. 2003). This range of was used to set sensible bounds (0.01 to 5 mM) to fit the value of the $K_{i\text{-}CBHI\text{-}cellobiose}$ in the model. The value was fit to the single enzyme hydrolysis data reported by Medve et al. (1997). It should be noted that the fit value of 0.093 mM is in fair agreement with the value reported by Nidetzky et al. (1994) of 0.034 mM for competitive inhibition of cellotetraose by cellobiose.
$K_{i\text{-}CBHI\text{-}glucose}$	31	mmol L^{-1}	Cellulases are inhibited by glucose, but to a lesser degree than by cellobiose. The range of reported values for glucose inhibition parameters for *T. reesei* cellulases is 2.1 to 31 (Hsu et al. 1980; Koivula et al. 1996). This range was used to fit the inhibition parameter for glucose to the single enzyme hydrolysis data reported by Medve et al. (1997). The value for CBHI did not change from 31 mM since glucose is not produced via the action of CBHI in the model. Time-course data for hydrolysis of cellulose by CBHI with β-glucosidase in the system would allow an improved value for this inhibition parameter to be fit. It is known, however, that glucose inhibition plays a less significant role in affecting the hydrolysis rate than cellobiose.

$K_{i\text{-}EG2\text{-}cellobiose}$	0.01	mmol L^{-1}	Cellobiose is known to strongly inhibit the action of cellulases but the reported values for K_i for *T. reesei* cellulases range from 0.021 mM to 1.4 mM for hydrolysis of soluble substrates (Hsu et al. 1980; Koivula et al.. 1996; Nidetzky et al. 1994; von Ossowski et al. 2003). This range of was used to set sensible bounds (0.01 to 5 mM) to fit the value of the $K_{i\text{-}EG2\text{-}cellobiose}$ in the model. The value was fit to the single enzyme hydrolysis data reported by Medve et al. (1997). The fit parameter fell at the lower bound set for the parameter optimization, which can indicate that the bounds chosen may need to be adjusted. Allowing a K_i value that makes EG2 more susceptible to cellobiose inhibition is not supported by the literature so the value of 0.01 was used for this work.
$K_{i\text{-}EG2\text{-}glucose}$	16.9	mmol L^{-1}	Cellulases are inhibited by glucose, but to a lesser degree than by cellobiose. The range of reported values for glucose inhibition parameters for *T. reesei* cellulases is 2.1 to 31 (Hsu et al. 1980; Koivula et al. 1996). This range was used to fit the inhibition parameter for glucose to the single enzyme hydrolysis data reported by Medve et al. (1997). The value for EG2 fell within this range. The large difference in value for inhibition of EG2 by glucose versus cellobiose demonstrates that glucose affects the rate of hydrolysis by EG2 much less than cellobiose.
$K_{M\text{-}CBHI\text{-}cel}$	0.000014	mmol dm^{-2}	Currently, there are no reliable values for K_M for the complexation of a cellulase with solid cellulose on the substrate surface. Standard initial rate experiments using Lineweaver-Burk analysis are not relevant for solid substrates because the adsorption step is not accounted for in the traditional kinetic analysis of enzymatic reactions. Without a reliable value or experimental technique to use to obtain a value for this important parameter, a value was fit to the single enzyme hydrolysis results for CBHI from Medve et al. (1997).
$K_{M\text{-}CBHI\text{-}sol}$	0.0032	mmol L^{-1}	The value reported by Nidetzky et al. (1994) for hydrolysis of cellohexaose by *T. reesei* CBHI was used in this work. The values were determined using a Lineweaver-Burk plot (conditions: 2.1 nM of CBHI, 1-100 µM of cellohexaose, 30 °C, acetate buffer, pH 4.8). There are limited reported values for purified cellulase kinetic constants on soluble cello-oligosaccharides. The effect of chain length on $K_{M\text{-}CBHI\text{-}sol}$ was ignored as the accumulation of soluble cello-oligosaccharides larger than cellobiose during hydrolysis is usually negligible.

$K_{M\text{-}EG2\text{-}cel}$	0.0067	mmol dm^{-2}	Currently, there are no reliable values for K_M for the complexation of a cellulase with solid cellulose on the cellulose surface. Standard initial rate experiments using Lineweaver-Burk analysis are not relevant for solid substrates because the adsorption step is not accounted for in the traditional kinetic analysis of enzymatic reactions. Without a reliable value or experimental technique to use to obtain a value for this important parameter, a value was fit to the single enzyme hydrolysis results for EG2 from Medve et al. (1997).
$K_{M\text{-}CBHI\text{-}sol}$	0.053	mmol L^{-1}	The value reported by Karlsson et al. (2002) for hydrolysis of cellopentaose by *T. reesei* EG2 was used in this work. The values were determined using a Lineweaver-Burk plot (conditions: 1.0 nM of CBHI, 10-150 μM of cellopentaose, 40 °C, Na-acetate buffer, pH 5.0). There are limited reported values for purified cellulase kinetic constants on soluble cello-oligosaccharides. The effect of chain length on $K_{M\text{-}EG2\text{-}sol}$ was ignored as the accumulation of soluble cello-oligosaccharides larger than cellobiose during hydrolysis is usually negligible.
MW_1	0.162	g $mmol^{-1}$	Molecular weight of anhydrous glucose
SA	8 or 47.6	m^2 g^{-1}	The lower value was used because it gives a coverage, based on the RSA enzyme size, of approximately 50% with an enzyme loading of 1.6 μmol L^{-1} if the enzymes readily adsorb. This ensures that the enzymes will be able to adsorb to the surface when there is twice as much enzyme. The upper value was used as it minimizes the effect of enzyme crowding. The upper value is 20 times the experimental value of 2.38 m^2 g^{-1}. The same source of the above experimental value had values of 41.9 and 33.5 for PASC and BMCC respectively which are similar to the upper value used in this study.
S_0	10	g L^{-1}	Substrate loadings set to match experimental conditions used in Medve et al. (1998)
σ_{CBHI}	156	-	Footprint size determined using RSA simulations of a "cellulase" shaped adsorbate. See Appendix B
σ_{EG2}	156	-	Footprint size determined using RSA simulations of a "cellulase" shaped adsorbate. See Appendix B
$*_{max}$	$6.64x10^{-5}$	mmol dm^{-2}	Based on the dimensions of a glucose unit (0.5 nm x 0.5 nm) for the area of a single site on the model surface. This area was used to determine the upper limit of sites per area by calculating the surface concentration in a perfectly packed square lattice.

2.13 References

1. A. A. Huang, Kinetic Studies on Insoluble Cellulose-Cellulase System. *Biotechnology and Bioengineering* **17**, 1421 (1975).
2. T. Jeoh *et al.*, Cellulase Digestibility of Pretreated Biomass is Limited by Cellulose Accessibility. *Biotechnology and Bioengineering* **98**, 112 (2007).
3. T. Jeoh, D. B. Wilson, L. P. Walker, Effect of Cellulase Mole Fraction and Cellulose Recalcitrance on Synergism in Cellulose Hydrolysis and Binding. *Biotechnology Progress* **22**, 270 (2006).
4. Y.-H. Lee, L. T. Fan, Kinetic Studies of Enzymatic Hydrolysis of Insoluble Cellulose: (II). Analysis of Extended Hydrolysis Times. *Biotechnology and Bioengineering* **25**, 939 (1983).
5. J. Medve, J. Karlsson, D. Lee, F. Tjerneld, Hydrolysis of Microcrystalline Cellulose by Cellobiohydrolysis I and Endoglucanase II from Trichoderma reesei: Adsorption, Sugar Production Pattern, and Synergism of the Enzymes. *Biotechnology and Bioengineering* **59**, 621 (1998).
6. A. S. Bommarius *et al.*, Cellulase Kinetics as a Function of Cellulose Pretreatment. *Metabolic Engineering* **10**, 370 (2008).
7. S. G. Desai, A. O. Converse, Substrate Reactivity as a Function of the Extent of Reaction in the Enzymatic Hydrolysis of Lignocellulose. *Biotechnology and Bioengineering* **56**, 650 (1997).
8. T. Erikkson, J. Karlsson, F. Tjerneld, A Model Explaining Declining Rate in Hydrolysis of Lignocellulose Substrates with Cellobiohydrolase I (Cel7A) and Endoglucanase I (Cel7B) of Trichoderma reesei. *Applied Biochemistry and Biotechnology* **101**, 41 (2002).
9. B. Nidetzky, W. Steiner, A New Approach for Modeling Cellulase-Cellulose Adsorption and the Kinetics of the Enzymatic Hydrolysis of Microcrystalline Cellulose. *Biotechnology and Bioengineering* **42**, 469 (1993).
10. K. Ohmine, H. Ooshima, Y. Harano, Kinetic Study on Enzymatic Hydrolysis of Cellulose by Cellulase from Trichoderma viride. *Biotechnology and Bioengineering* **25**, 2041 (1983).
11. P. Valjamae, V. Sild, G. Pettersson, G. Johansson, The Initial Kinetics of Hydrolysis by Cellobiohydrolases I and II is Consistent with a Cellulose Surface-Erosion Model. *European Journal of Biochemistry* **253**, 469 (1998).
12. B. Yang, D. M. Willies, C. E. Wyman, Changes in the Enzymatic Hydrolysis Rate of Avicel Cellulose with Conversion. *Biotechnology and Bioengineering* **94**, 1122 (2006).
13. S. Zhang, D. E. Wolfgang, D. B. Wilson, Substrate Heterogeneity Causes the Nonlinear Kinetics of Insoluble Cellulose Hydrolysis. *Biotechnology and Bioengineering* **66**, 35 (1999).
14. P. Bansal, M. Hall, M. J. Realff, J. H. Lee, A. S. Bommarius, Modelling Cellulase Kinetics on Lignocellulosic Substrates. *Biotechnology Advances*, (2009).
15. M. Okazaki, M. Moo-Young, KInetics of Enzymatic Hydrolysis of Cellulose: Analytical Description of a Mechanistic Model. *Biotechnology and Bioengineering* **20**, 637 (1978).
16. M. T. Holtzapple, H. S. Caram, A. E. Humphrey, The HCH-1 Model of Enzymatic Cellulose Hydrolysis. *Biotechnology and Bioengineering* **26**, 775 (1984).
17. K. L. Kadam, E. C. Rydholm, J. D. McMillan, Development and Validation of a Kinetic Model for Enzymatic Saccharification of Lignocellulosic Biomass. *Biotechnology Progress* **20**, 698 (2004).
18. Y. Zheng, Z. Pan, R. Zhang, B. M. Jenkins, Kinetic Modeling for Enzymatic Hydrolysis of Pretreated Creeping Wild Ryegrass. *Biotechnology and Bioengineering*, (2009).
19. L. T. Fan, Y.-H. Lee, Kinetic Studies of Enzymatic Hydrolysis of Insoluble Cellulose: Derivation of a Mechanistic Kinetic Model. *Biotechnology and Bioengineering* **25**, 2707 (1983).
20. Q. Gan, S. J. Allen, G. Taylot, Kinetic Dynamics in Heterogeneous Enzymatic Hydrolysis of Cellulose: and Overview, an Experimental Study and Mathematical Modelling. *Process Biochemistry* **38**, 1003 (2003).

21. A. O. Converse, R. Matsuno, M. Tanaka, M. Taniguchi, A Model of Enzyme Adsorption and Hydrolysis of Microcrystalline Cellulose with Slow Deactivation of the Adsorbed Enzyme. *Biotechnology and Bioengineering* **32**, 38 (1988).
22. A. O. Converse, J. D. Optekar, A Synergistic Kinetics Model for Enzymatic Cellulose Hydrolysis Compared to Degree-of-Synergism Experimental Results. *Biotechnology and Bioengineering* **42**, 145 (1993).
23. K. Movagarnejad, M. Sohrabi, T. Kaghazchi, F. Vahabzadeh, A Model for the Rate of Enzymatic Hydrolysis of Cellulose in Heterogeneous Solid-Liquid Systems. *Biochemical Engineering Journal* **4**, 197 (2000).
24. K. Movagharnejad, M. Sohrabi, A Model for the Rate of Enzymatic Hydrolysis of Some Cellulosic Waste Materials in Heterogeneous Solid-Liquid Systems. *Biochemical Engineering Journal* **14**, 1 (2003).
25. S. Wald, C. R. Wilke, H. W. Blanch, Kinetics of the Enzymatic Hydrolysis of Cellulose. *Biotechnology and Bioengineering* **26**, 221 (1984).
26. J. J. Fenske, M. H. Penner, J. P. Bolte, A Simple Individual-based Model of Insoluble Polysaccharide Hydrolysis: the Potential for Autosynergism with Dual-activity Glycosidases. *Journal of Theoretical Biology* **199**, 113 (1999).
27. S. Peri, S. Karra, Y. Y. Lee, M. N. Karim, Modeling Intrinsic Kinetics of Enzymatic Cellulose Hydrolysis. *Biotechnology Progress* **23**, 626 (2007).
28. W. Zhou, Z. Hao, Y. Xu, H.-B. Schuttler, Cellulose Hydrolysis in Evolving Morphologies II: Numerical Results and Analysis. *Biotechnology and Bioengineering*, (2009).
29. W. Zhou, H.-B. Schuttler, Z. Hao, Y. Xu, Cellulose Hydrolysis in Evolving Morphologies I: A General Modeling Formalism. *Biotechnology and Bioengineering*, (2009).
30. J. Medve, J. Stahlberg, F. Tjerneld, Isotherms for Adsorption of Cellobiohydrolase I and II from Trichoderma reesei on Microcrystalline Cellulose. *Applied Biochemistry and Biotechnology* **66**, 39 (1997).
31. K. Suga, G. van Dedem, M. Moo-Young, Enzymatic Breakdown of Water Insoluble Substrates. *Biotechnology and Bioengineering* **17**, 185 (1975).
32. K. Suga, G. van Dedem, M. Moo-Young, Degradation of Polysaccharides by Endo and Exo Enzymes: A Theoretical Analysis. *Biotechnology and Bioengineering* **17**, 433 (1975).
33. Y.-H. P. Zhang, L. R. Lynd, A Functionally Based Model for Hydrolysis of Cellulose by Fungal Cellulase. *Biotechnology and Bioengineering* **94**, 888 (2006).
34. K. Igarashi *et al.*, High-Speed Atomic Force Microscopy Visualizes Processive Movement of Trichoderma Reesei Cellobiohydrolase I on Crystalline Cellulose. *Journal of Biological Chemistry*, (2009).
35. J. E. Hansen, L. D. Travis, Light Scattering in Planetary Atmospheres. *Space Science Review* **16**, 527 (1974).
36. S. G. Warren, C. G. Thomas, Representation of a Nonsperical Ice Particle by a Collection of Independent Spheres for Scattering and Absorption of Radiation. *Journal of Geophysical Research* **104**, 697 (1999).
37. P. M. Abuja, I. Pilz, M. Claeyssens, P. Tomme, Domain Structure of Cellobiohydrolase II as Studied by Small Angle X-Ray Scattering: Close Resemblance to Cellobiohydrolase I. *Biochemical and Biophysical Research Communications* **156**, 180 (1988).
38. P. M. Abuja *et al.*, Structural and Functional Domains of Cellobiohydrolase I from Trichoderma reesei A Small Angle X-Ray Scattering Study of the Intact Enzyme and its Core. *European Biophysics Journal* **15**, 339 (1988).
39. B. Nidetzky, W. Zachariae, G. Gercken, M. Hayn, W. Steiner, Hydrolysis of Cellooligosaccharides by Trichoderma reesei Cellobiohydrolases: Experimental Data and Kinetic Modeling. *Enzyme and Microbial Technology* **16**, 43 (1994).

40. A. Koivula *et al.*, The Active Site of Cellobiohydrolase Cel6A from Trichoderma reesei: The Roles of Aspartic Acids D221 and D175. *Journal of the American Chemical Society* **124**, 10015 (2002).
41. J. Karlsson, M. Siika-aho, M. Tenkanen, F. Tjerneld, Enzymatic Properties of the Low Molecular Mass Endoglucanases Cel12A (EG III) and Cel45A (EG V) of Trichoderma reesei. *Journal of Biotechnology* **99**, 63 (2002).
42. T.-A. Hsu, C.-S. Gong, G. T. Tsao, Kinetic Studies of Cellodextrins Hydrolyses by Exocellulase from Trichoderma reesei. *Biotechnology and Bioengineering* **22**, 2305 (1980).
43. A. Koivula *et al.*, The Active Site of Trichoderma reesei Cellobiohydrolase II: the Role of Tyrosine 169. *Protein Engineering* **9**, 691 (1996).
44. I. von Ossowski *et al.*, Engineering the Exo-loop of Trichoderma reesei Cellobiohydrolase, Cel7A. A Comparison with Phanerochaete chrysosporium Cel7D. *Journal of Molecular Biology* **333**, 817 (2003).
45. S. Mauer, C. J. Radke. (2009).
46. T. K. Ng, J. G. Zelkus, A Continuous Spectrophotometric Assay for the Determination of Cellulase Solubilizing Activity. *Analytical Biochemistry* **103**, 42 (1980).
47. D. D. Y. Ryu, S. B. Lee, T. Tassinari, C. Macy, Effect of Compression Milling on Cellulose Structure and on Enzymatic Hydrolysis Kinetics. *Biotechnology and Bioengineering* **24**, 1047 (1982).
48. T. M. Wood, Preparation of Crystalline, Amorphous and Dyed Cellulase Substrates. *Methods in Enzymology* **160**, 19 (1988).
49. P. Valjamae, V. Sild, A. Nutt, G. Pettersson, G. Johansson, Acid Hydrolysis of Bacterial Cellulose Reveals Differenct Modes of Synergistic Action Between Cellobiohydrolase I and Endoglucanase I. *European Journal of Biochemistry* **266**, 327 (1999).
50. K. Kleman-Leyer, E. Agosin, A. H. Conner, T. K. Kirk, Changes in Molecular Size Distribution of Cellulose during Attack by White Rot and Brown Rot Fungi. *Applied and Environmental Microbiology* **58**, 1266 (1992).
51. J. Hong, X. Ye, Y.-H. P. Zhang, Quantitative Determination of Cellulose Accessibility to Cellulase Based on Adsorption of a Nonhydrolytic Fusion Protein Containing CBM and GFP with Its Applications. *Langmuir* **23**, 12535 (2007).
52. R. E. T. Drissen *et al.*, A Generic Model for Glucose Production from Various Cellulose Sources by a Commerical Cellulase Complex. *Biocatalysis and Biotransformation* **25**, 419 (November-December, 2007).
53. A. B. Boraston, E. Kwan, P. Chiu, R. A. J. Warren, D. G. Kilburn, Recognition and Hydrolysis of Noncrystalline Cellulose. *Journal of Biological Chemistry* **278**, 6120 (2003).
54. G. Carrard, A. Koivula, H. Soderlund, P. Beguin, Cellulose-Binding Domains Promote Hydrolysis of Different Sites on Crystalline Cellulose. *Proceedings of the National Academemy of Science* **97**, 10342 (2000).
55. A. L. Creagh, E. Ong, E. Jervis, D. G. Kilburn, C. A. Haynes, Binding of the Cellulose-Binding Domain of Exoglucanase Cex from Cellulomonas fimi to Insoluble Microcrystalline Cellulose is Entropically Driven. *Proceedings of the National Academemy of Science* **93**, 12229 (1996).
56. S. P. Voutilainene *et al.*, Cloning, Expression and Characterization of Novel Thermostable Family 7 Cellobiohydrolases. *Biotechnology and Bioengineering* **101**, 515 (2008).
57. T. Turanyi, Sensitivity Analysis of Complex Kinetic Systems. Tools and Applications. *Journal of Mathematical Chemistry* **5**, 203 (1990).
58. J. W. Evans, Random and Cooperative Sequential Adsorption. *Rev Mod Phys* **65**, 1281 (Oct, 1993).

3. A Mechanistic Model for Rational Design of Optimal Cellulase Mixtures*

3.1 Abstract

A model-based framework is described that permits the optimal composition of cellulase enzyme mixtures to be found for lignocellulose hydrolysis. The rates of hydrolysis are shown to be dependent on the nature of the substrate. For bacterial microcrystalline cellulose (BMCC) hydrolyzed by a ternary cellulase mixture of EG2, CBHI, and CBHII, the optimal predicted mixture was 1:0:1 EG2:CBHI:CBHII at 24 hours and 1:1:0 at 72 hours, at loadings of 10 mg enzyme per g substrate. The model was validated with measurements of soluble cello-oligosaccharide production from BMCC during both single enzyme and mixed enzyme hydrolysis. Three-dimensional diagrams illustrating cellulose conversion were developed for mixtures of EG2, CBHI, CBHII acting on BMCC and predicted for other substrates with a range of substrate properties. Model predictions agreed well with experimental values of conversion after 24 hours for a variety of enzyme mixtures. The predicted mixture performances for substrates with varying properties demonstrated the effects of initial degree of polymerization (DP) and surface area on the performance of cellulase mixtures. For substrates with a higher initial DP, endoglucanase enzymes accounted for a larger fraction of the optimal mixture. Substrates with low surface areas showed significantly reduced hydrolysis rates regardless of mixture composition. These insights, along with the quantitative predictions, demonstrate the utility of this model-based framework for optimizing cellulase mixtures.

3.2 Introduction

Cost-effective enzymatic hydrolysis of cellulose is required for the conversion of biomass to renewable transportation fuels. Nature employs a mixture of enzymes acting in concert to breakdown cellulosic biomass (*1*). Efficient deconstruction of the various lignocellulosic substrates proposed as feedstocks for biofuels production requires cellulolytic enzyme mixtures tailored to individual biomass compositions and pretreatment methods. To date, design of optimal cellulase mixtures has been based primarily on statistically designed experiments (*2-6*). Most studies have focused on a single lignocellulosic substrate without testing the optimality of the cellulase mixture on additional substrates with different characteristics (for example, surface area, degree of polymerization, and lignin content). An individual cellulase mixture, even if optimized for a specific biomass substrate, would not be expected to perform optimally on lignocellulosic materials with different characteristics. Commercially available cellulase mixtures exhibit different rates and extents of hydrolysis when applied to different substrates (*7-9*), demonstrating the importance of the cellulase-substrate relationship.

Recently, Banerjee and coworkers (*4*) applied a high-throughput statistical scheme to optimize mixtures for a variety of substrates and pretreatment methods. Their results show that the composition of the optimal mixture is different for each feedstock/pretreatment combination. Designing cellulase mixtures using statistical techniques is useful, but it requires a large amount of experimental resources and time for each mixture. Understanding the specific features and

* This chapter, with few modifications from its original format, represents the following peer-reviewed publication: S. E. Levine, J. M. Fox, D. S. Clark, H. W. Blanch, A mechanistic model for rational design of optimal cellulase mixtures. *Biotechnol Bioeng* **108**, 2561 (Nov, 2011).

mechanisms of the cellulase-substrate system that impact cellulase mixture performance would allow the rational development of optimal enzyme mixtures.

Mechanistic modeling is a powerful tool to provide insights into specific substrate features and enzymatic reaction steps that affect enzymatic hydrolysis (*10*). Even the most comprehensive cellulose hydrolysis models to date lack the mechanistic detail necessary to capture important features that impact how specific soluble products are formed, and instead focus on determining overall hydrolytic activity (*11-15*). For example, enzyme processivity is central to the action of cellobiohydrolases (CBHs) but has not been described in models to date. In the current work, we have built upon our previous model (*13*) to include mechanistic details that account for individual product formation by both CBH and endoglucanase (EG) enzymes. This model is used to study the activity of cellulase mixtures in the degradation of cellulosic substrates with different physical properties. This framework forms the basis for a rapid, rational approach to design and optimize enzyme mixtures, and requires basic information about the cellulase-substrate system of interest. This approach also provides an understanding of why a particular mixture works more or less efficiently on a given substrate. This information is useful for both development of enzyme "cocktails" and in assessing the effectiveness of pretreatment methods.

3.3 Methods

3.3.1 Mechanistic Model Development

The model employed in this study was adapted from the mechanistic model of enzymatic cellulose hydrolysis described by Levine et al. (*13*). To improve the model's ability to predict not only overall cellulose conversion to sugars, but also the time-course of specific products, the model was expanded to incorporate additional information on the physicochemical processes involved in cellulose hydrolysis. For example, by incorporating an explicit processive mechanism for CBH enzymes, we removed the previous assumption that CBH enzyme-substrate complexation was in equilibrium. This assumption, while commonly made in cellulase kinetic models (*14, 16-19*), has no experimental support. The removal of this assumption results in the mechanism shown in Fig. 3.1A. Adsorbed CBH enzymes can complex with a chain end on the cellulose surface and then either decomplex from that cellulose chain or catalyze the release of cellobiose, resulting in a CBH-cellulose complex on a chain that is two glucose units shorter. The new enzyme-substrate complex can then react further or decomplex.

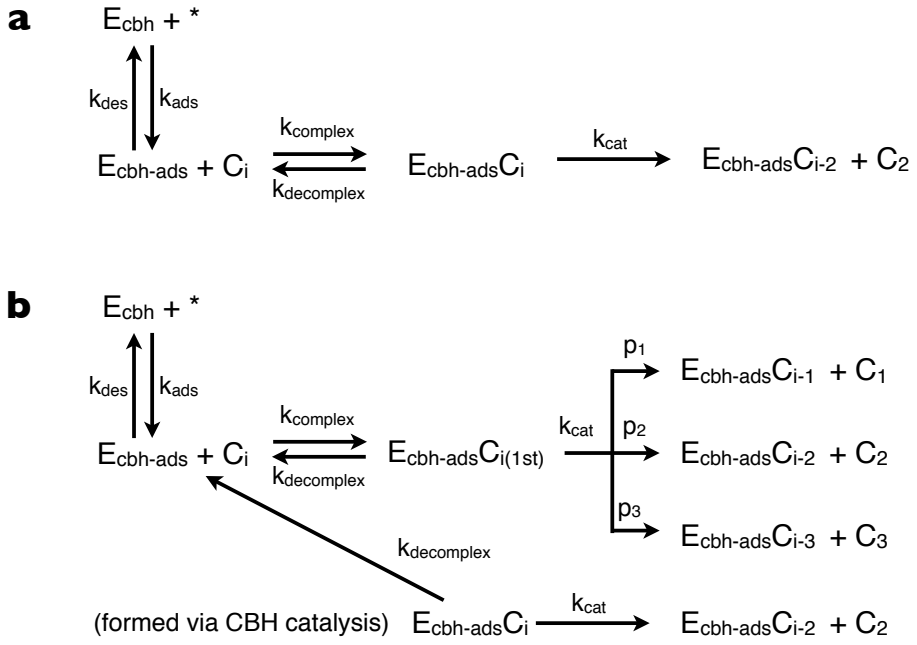


Figure 3.1 Processivity mechanisms for CBH enzymes: (a) assumes that CBH enzymes can only produce cellobiose; (b) assumes that glucose, cellobiose, and cellotriose can be formed in the first cut after complexation, with subsequent cuts forming only cellobiose.

Removal of the equilibrium assumption for CBH complexation with cellulose requires that the enzyme-substrate complexes be tracked with separate material balances. An example of one of these material balances is shown in Eq. 3.1. In addition, it was necessary to modify the CBH-related terms in the material balance equations for surface cellulose chains of a given length, as shown in Eq. 3.2.

$$\begin{aligned}\frac{d[E_{CBH-ads}C_i^{'}]}{dt} &= +k_{comp-CBH-cel}\theta_{CBH}[E_{CBH-ads}^{'}][C_i^{'}] \\ &\quad -(k_{decomp-CBH-cel}+k_{cat-CBH})[E_{CBH-ads}C_i^{'}] \\ &\quad +k_{cat-CBH}[E_{CBH-ads}C_{i+2}^{'}]\end{aligned} \tag{3.1}$$

$$\begin{aligned}\frac{d[C_i^{'}]}{dt} &= -k_{comp-CBH-cel}\theta_{CBH}[E_{CBH-ads}^{'}][C_i^{'}] \\ &\quad +k_{decomp-CBH-cel}[E_{CBH-ads}C_i^{'}]\end{aligned} \tag{3.2}$$

Cellobiose is the primary product of CBH activity on cellulose, but glucose and cellotriose are formed in significant quantities as well (*20*). Experimental studies of CBH hydrolysis of oligosaccharides suggest that glucose, cellobiose, and cellotriose can be formed from the first cut the enzyme makes after complexing with a chain (*21, 22*). Subsequent processive cuts form cellobiose. To incorporate this mechanism into the model, the CBH-cellulose material balances were divided into those pertaining to enzyme-substrate complexes

formed via complexation and those where complexes resulted from processive catalytic action. The mechanism used in the model is shown in Fig. 3.1B. The soluble product of the first hydrolytic cut by a CBH enzyme after complexation was determined by multiplying the k_{cat} for hydrolysis by a partition constant based on experimental data (*21, 22*). Balance equations for an uncomplexed cellulose chain, a CBH-cellulose complex formed via complexation, and a CBH-cellulose complex formed via processive catalysis are shown in Eqs. 3.3-3.5.

$$\frac{d[C_i^{'}]}{dt} = -k_{comp-CBH-cel}\theta_{CBH}[E_{CBH-ads}^{'}][C_i^{'}] + k_{decomp-CBH-cel}([E_{CBH-ads}C_{i-1st}^{'}] + [E_{CBH-ads}C_i^{'}]) \tag{3.3}$$

$$\frac{d[E_{CBH-ads}C_{i-1st}^{'}]}{dt} = +k_{comp-CBH-cel}\theta_{CBH}[E_{CBH-ads}^{'}][C_i^{'}] - (k_{decomp-CBH-cel} + k_{cat-CBH})[E_{CBH-ads}C_{i-1st}^{'}] \tag{3.4}$$

$$\begin{aligned}\frac{d[E_{CBH-ads}C_i^{'}]}{dt} = {} & +k_{cat-CBH}([E_{CBH-ads}C_{i+2}^{'}] - [E_{CBH-ads}C_i^{'}]) \\ & + p_1 k_{cat-CBH}[E_{CBH-ads}C_{i+1-1st}^{'}] + p_2 k_{cat-CBH}[E_{CBH-ads}C_{i+2-1st}^{'}] \\ & + p_3 k_{cat-CBH}[E_{CBH-ads}C_{i+3-1st}^{'}] - k_{decomp-CBH-cel}[E_{CBH-ads}C_i^{'}]\end{aligned} \tag{3.5}$$

The mechanism typically used to describe EG activity assumes that all bonds in a cellulose chain are equally likely to be hydrolyzed by the enzyme. This mechanism would consequently lead to an equal distribution of all soluble products, producing near equimolar amounts of glucose and cellobiose. The product distributions observed experimentally are far from this assumed product distribution (*20*). One mechanism that explains the observed product distribution is reduced activity of EG on the terminal glycosidic bond of a cellulose chain (*23*). This reduced activity can be incorporated into the model by separating the cuts by EG on the terminal glycosidic bonds from general EG cuts. An example of the resulting EG terms in a cellulose chain material balance that includes this distinction is given by Eq. 3.6 (this equation was derived using the same method detailed in our previous modeling work (*13*)).

$$\begin{aligned}\frac{d[C_i^{'}]}{dt} = {} & +\frac{k_{cat-EG}}{K_{comp-EG-cel}}[E_{EG-ads}^{'}]\theta_{EG}\left(2\sum_{j=i+2}^{\infty}[C_j^{'}] - (i-3)[C_i^{'}]\right) \\ & + \frac{k_{cat-EG-term}}{K_{comp-EG-cel-term}}[E_{EG-ads}^{'}]\theta_{EG}2\left([C_{i+1}^{'}] - [C_i^{'}]\right)\end{aligned} \tag{3.6}$$

The previous model (*13*) was also altered to include the observed reduction in activity of cellulases on cellotriose. Both EG and CBH enzymes have been shown to have reduced activity on cellotriose compared to longer oligosaccharides (*21-25*). This was included by separating the rates of complexation and catalysis of a cellulase with cellotriose from the general action of these enzymes on longer soluble oligosaccharides.

3.3.2 Model Parameters

The model parameters were determined using an approach similar to that outlined previously (*13*). Experimental parameters were used when available. In the absence of experimental values, parameters were estimated using literature values for related parameters or simple calculations for dimensional parameters (for example, glucose sites per area). When neither of the above sources was suitable, parameters were estimated by fitting to a defined data set. Most of the new parameters required for the improvements to the model were based on experimental data or relevant literature. For a small number of parameters, primarily those associated with enzyme-solid substrate complexation, it was necessary to estimate parameters from best fits of experimental hydrolysis data. The IMSL nonlinear optimizer NNLPF was used for optimization against time-course hydrolysis data to estimate parameters for individual cellulase enzymes. Table 3.1 lists the kinetic parameters used in the model. As shown in the previous modeling study this work builds upon (*13*), the sensitivity to surface area in model performance is only significant at longer hydrolysis times (>12 hrs) and has minimal effect on the fit parameters.

3.3.3 Model Assembly and Solution

The model was constructed in FORTRAN 95 using the Absoft Pro Fortran v11 compiler. The model was solved using the DASPG solver from IMSL, which uses the Petzold-Gear method to solve the differential and algebraic equations that comprises the model. To aid in model construction for a variety of enzyme mixtures and initial cellulose degrees of polymerization, a PERL script was developed that builds the model equations from basic input related to the number and type of cellulases and the initial average degree of polymerization of cellulose in the system. The number of equations in the model depends on the number of enzymes included in the system as well as the maximum cellulose chain length tracked. The primary model used in this work had 5635 differential equations and 15 algebraic equations, includes one EG enzyme and two CBH enzymes and tracks cellulose chains up to 1125 (average chain length of 750) glucose units long.

Table 3.1 Parameters used in the model of enzymatic cellulose hydrolysis

Parameter	Value	Units	Source
$k_{EG2\text{-}ads}$	2.4	L $mmol^{-1}$ s^{-1}	(*13*)[a]
$k_{CBHI\text{-}ads}$	2.4	L $mmol^{-1}$ s^{-1}	(*13*)[a]
$k_{CBHII\text{-}ads}$	2.4	L $mmol^{-1}$ s^{-1}	(*13*)[a]
$k_{EG2\text{-}des}$	5.36 x 10^{-3}	s^{-1}	(*13*)[a]
$k_{CBHI\text{-}des}$	6.10 x 10^{-3}	s^{-1}	This study[b]
$k_{CBHII\text{-}des}$	2.08 x 10^{-4}	s^{-1}	This study[b]
$k_{cat\text{-}EG2}$	65	s^{-1}	(*13*)[a]
$k_{cat\text{-}CBHI}$	4.42	s^{-1}	This study[b]
$k_{cat\text{-}CBHII}$	14	s^{-1}	(*25*)[c]
$K_{M\text{-}EG2\text{-}cel}$	2	mmol m^{-2}	This study[b]
$k_{comp\text{-}CBHI\text{-}cel}$	1990	m^2 $mmol^{-1}$ s^{-1}	This study[b]
$k_{comp\text{-}CBHII\text{-}cel}$	7780	m^2 $mmol^{-1}$ s^{-1}	This study[b]
$k_{decom\text{-}CBHI\text{-}cel}$	1.37	s^{-1}	This study[b]
$k_{decom\text{-}CBHII\text{-}cel}$	2.78	s^{-1}	This study[b]
$K_{M\text{-}EG2\text{-}sol}$	0.053	mM	(*13*)[a]
$k_{comp\text{-}CBHI\text{-}sol}$	483	L $mmol^{-1}$ s^{-1}	(*22*)[d]
$k_{comp\text{-}CBHII\text{-}sol}$	1200	L $mmol^{-1}$ s^{-1}	(*25*)[d]
$k_{decom\text{-}CBHI\text{-}sol}$	2.78	s^{-1}	(*22*)[d]
$k_{decom\text{-}CBHII\text{-}sol}$	2.78	s^{-1}	(*25*)[d]
$k_{cat\text{-}EG2\text{-}C3}$	5.44	s^{-1}	(*34*)[d]
$k_{cat\text{-}CBHI\text{-}C3}$	0.016	s^{-1}	This study[e]
$k_{cat\text{-}CBHII\text{-}C3}$	0.061	s^{-1}	(*25*)
$K_{M\text{-}EG2\text{-}C3}$	19.5	mM	This study[a]
$k_{comp\text{-}CBHI\text{-}C3}$	26	L $mmol^{-1}$ s^{-1}	This study[e]
$k_{comp\text{-}CBHII\text{-}C3}$	142	L $mmol^{-1}$ s^{-1}	(*25*)
$k_{decom\text{-}CBHI\text{-}C3}$	2.78	s^{-1}	This study[e]
$k_{decom\text{-}CBHII\text{-}C3}$	2.78	s^{-1}	(*25*)
P1:P2:P3 CBHI	4:4:3	NA	(*21*)[c]
P1:P2:P3 CBHII	1:1:1	NA	(*21*)[d]
$k_{cat\text{-}EG2\text{-}C1}$	5.44	s^{-1}	(*34*)
$K_{M\text{-}EG2\text{-}C1\text{-}cel}$	900	mmol m^{-2}	This study[b]
$K_{M\text{-}EG2\text{-}C1\text{-}sol}$	90	mM	This study[f]
$K_{i\text{-}EG2\text{-}cellobiose}$	0.03	mM	This study[b]
$K_{i\text{-}CBHI\text{-}cellobiose}$	0.19	mM	This study[b]
$K_{i\text{-}CBHII\text{-}cellobiose}$	0.06	mM	This study[b]
$K_{i\text{-}EG2\text{-}glucose}$	16.0	mM	This study[b]
$K_{i\text{-}CBHI\text{-}glucose}$	31	mM	(*13*)[a]
$K_{i\text{-}CBHII\text{-}glucose}$	29	mM	This study[b]
$k_{d\text{-}EG2}$	3.89 x 10^{-6}	s^{-1}	(*35*)[d]
$k_{d\text{-}CBHI}$	3.89 x 10^{-6}	s^{-1}	(*35*)[d]
$k_{d\text{-}CBHII}$	8.06 x 10^{-6}	s^{-1}	(*35*)[d]

[a]Parameters used were taken from the earlier version of the model; see Levine et al. 2010 for details
[b]Parameters optimized against single enzyme hydrolysis data
[c]Parameters taken directly from published values
[d]Parameter set based on published data
[e]Parameters obtained via kinetic experiments using soluble cello-oligosaccharides in our laboratory (data not shown)
[f]Parameter set for soluble substrate set based on optimized value for solid substrate reaction

3.3.4 Enzymes

Both *Trichoderma longibrachiatum* Cellobiohydrolase I (CBHI) and *Taleromyces emersonii* Endoglucanase 2 (EG2) were purchased from Megazyme. Previous experiments by our group revealed minor endoglucanase and β-glucosidase contamination of these products. The analysis in this work relied on accurate measurements of cellooligosaccharide concentrations in cellulase/cellulose reaction mixtures over 72-hour periods, so careful measures were taken to ensure high enzyme purity. A GE Healthcare ATKA Explorer system was used for this purpose. Approximately 20 mg of the *Tr*CBH1 enzyme, which was purchased in 3.2 M ammonium sulfate solution, was dissolved into 15 ml of 25 mM HEPES pH 7.35 and desalted on a GE HiPrep 26/10 desalting column with a 4 ml/min flow of identical buffer. The protein-containing fraction was purified further via exchange on a HiLoad 16/10 Q Sepharose HP anion exchange column. The main peak from this step was then concentrated and run through a Superdex 75 10/300 gel filtration column (15% 1.0M HEPES solution). The leading peak from this step was desalted again and purified on a Mono Q 10/100 GL anion exchange column. Single-band purity was ensured with a 10% Tris-HCl gel (Bio-Rad). The *Tem*EG2 enzyme was purified in an identical way, but without the Mono Q step. The final *Tem*EG2 enzyme had a slight lower molecular weight impurity with no known enzymatic activity (Megazyme E-CELTE Product Information). In both cases, beta-glucosidase contamination was tested for by incubation for 4 hours at 50°C with 1.1 mM of cellobiose (50 mM sodium acetate buffer, pH 4.8). Samples were quenched at this time with equal volumes of 0.100 M NaOH and run on a Dionex PA-200 column (method described below). No β-glucosidase activity was detected.

Trichoderma reesei CBH2 (CBHII) was purified from Celluclast (Sigma-Aldrich), a lyophilized mixture of enzymes secreted by *T. reesei* when grown on cellulose. The CBHII enzyme was purified in a fashion identical to that of TrCBH1. The final product was analyzed by SDS-PAGE (7.5% Tris-HCl gel) to ensure the absence of *Tr*CBH1.

3.3.5 Cellulose Preparation

Many previously reported BMCC preparations have been shown to modify substrate morphology and enhance digestibility. Pretreatment effects are minimized in the preparation employed here by the exclusion of steps involving strongly acidic or basic solutions.

A 20 μL freezer stock of *Acetobacter xylinum* (ATCC 53582) was used to inoculate 1 L of medium as described previously (*26*). This inoculum was grown in a 2-L Erlenmeyer flask for 7 days at 25°C in a shaker at 250 RPM). The resulting mixture was spun down at 6000 RPM in a Beckman Coulter Avanti J-26 XP Centrifuge. The lower ¼ of the resulting heterogeneous suspension, which contained most of the cellulose, was then resuspended in 600 ml of water, and spun down again. This process was repeated three times with 600 ml of 50% ethanol solution. After being incubated at 10°C in 50% ethanol for three weeks, the resulting suspension was spun down, and blended for 3 minutes in a Waring blender. The absence of live cells was confirmed with a Leica Epifluorescence Microscope and the subsequent suspension was washed once with 50% ethanol. The final BMCC preparation was suspended in 30% ethanol, and its cellulose concentration was measured with the phenol sulfuric acid method as described previously (*27*).

3.3.6 Measurement of Cello-oligosacchride Production in Various Mixtures

Cello-oligosaccharide production by defined mixtures of *Tl*CBH1, *Tr*CBH2, and *Tem*EG2 acting on BMCC was monitored over the course of 72 hours. Reactions were carried out in -50 ml Erlenmeyer flasks containing 25.25 ml of 1 g/L BMCC, 50 mM sodium acetate buffer (pH 4.8), and the enzyme concentrations described in Table 3.2.

Table 3.2 Hydrolysis reaction mixtures

Reaction #	CBHI (μM)	CBHII (μM)	EG2 (μM)
1	0	0	0
2A	0.094	0	0
2B	0	0.031	0
2C	0	0	0.022
3A	0.094	0.031	0.022
3B	0.076	0	0.019
3C	0.152	0	0

The reactions fell into three groups: (1) enzyme-free controls, (2) reactions used to obtain fit parameters, (3) reactions used for model validation. After reagent addition, the flasks were sealed with foil and placed in humidified shakers at 200 RPM. Samples were taken at the following times: 0, 0.83, 1.83, 3.08, 4.53, 5.53, 7.05, 8.83, 10.62, 17.5, 20.83, 26.8, 31.9, 41.78, 49.15, 57.5, 66.4, and 74.57 hours. At all time points, two 20 μL samples were withdrawn from each reaction mixture and added to 180 μL of water (each) within 1.5 ml Eppendorf tubes; these were subsequently dipped into an ethanol and dry ice bath for fast freezing. This small sample volume minimized reduction in the reaction volume (1.5% by the end of the experiment). For Dionex HPLC preparation, samples were thawed with the addition of 200 μL of 0.1 M NaOH solution, and spun down for 20 minutes at 1400 rpm. Supernatant collected from each Eppendorf tube was diluted 2-200 fold with 0.05 M NaOH before Dionex analysis. All reactions were carried out in duplicate.

3.4 Results

3.4.1 Single Enzyme Hydrolysis

Certain kinetic parameters in the model were determined from fits of the time-course of all detectable soluble products (glucose, cellobiose, and cellotriose) measured in the hydrolysis of BMCC by purified, single enzymes. Fig. 3.2 shows the optimized model results compared to the experimental data for EG2 (Fig. 3.2A), CBHI (Fig. 3.2B), and CBHII (Fig. 3.2C).

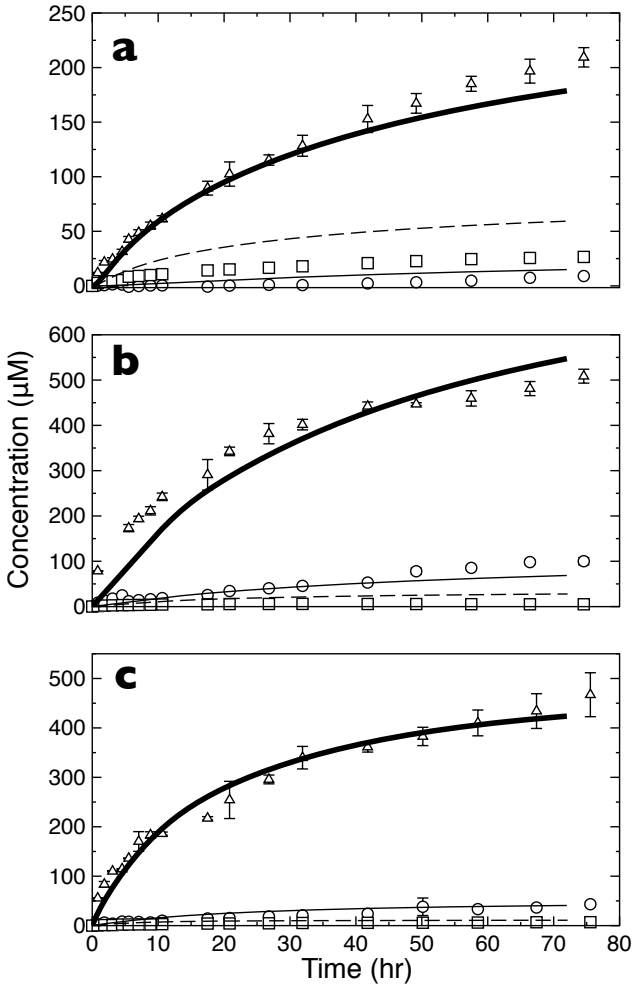


Figure 3.2 Model results compared to experimental time-course data for (a) EG2, (b) CBHI, and (c) CBHII single-enzyme hydrolysis. All detectable soluble products are shown: glucose (○, experiment; −, model), cellobiose (△, experiment; —, model), and cellotriose (□, experiment; ---, model).

The substrate was modeled with an initial degree of polymerization (DP) of 750 and an initial surface area of 47.6 m^2 g^{-1}. The enzymes loadings were 0.022 μmol g^{-1} for EG2, 0.031 μmol g^{-1} CBHII, and 0.094 μmol g^{-1} for CBHI (approximately 10 mg cellulase per g substrate). As shown in Table 3.1, 14 of the 41 enzyme parameters in the model were determined from the optimization of the model to the single-enzyme hydrolysis data. All model results exhibit good agreement with the experimental data. That the model is able to capture a large data set with a minimal requirement for parameter optimization provides strong support for the accuracy of the assumed mechanisms.

3.4.2 Model Validation

To demonstrate the accuracy and predictive ability of the model, comparisons were made to three additional hydrolysis time-courses (Fig. 3.3).

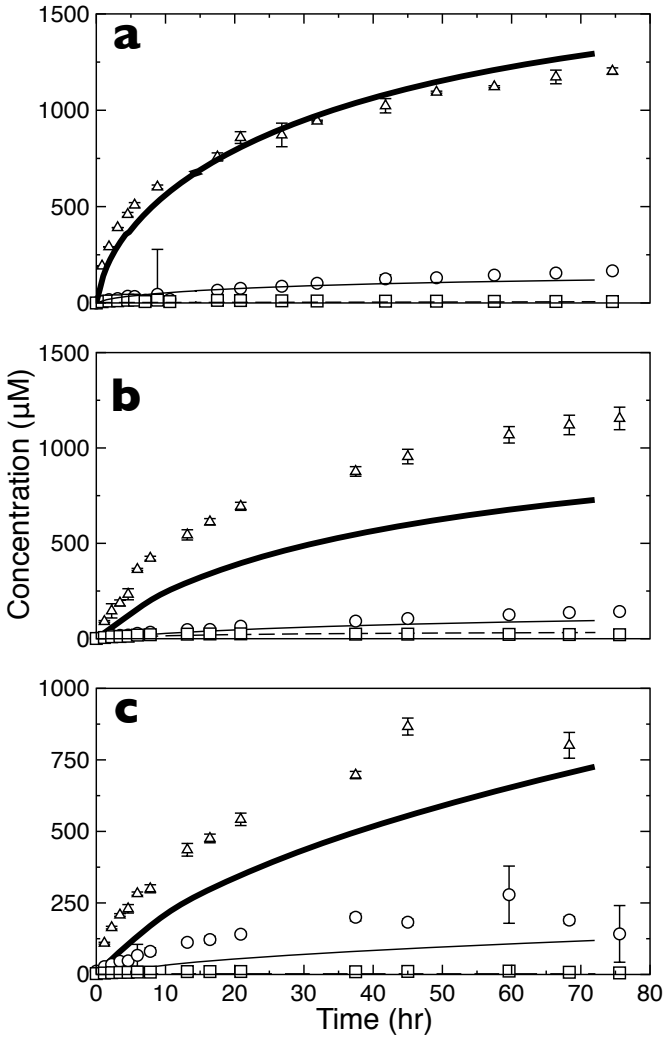


Figure 3.3 Model results compared to experimental time-course data for (a) EG2-CBHI-CBHII ternary mixture, (b) EG2-CBHI binary mixture, and (c) CBHI single-enzyme (different loading than Fig. 3.2B) hydrolysis. All detectable soluble products are shown: glucose (○, experiment; −, model), cellobiose (△, experiment; —, model), and cellotriose (□, experiment; ---, model).

The model accurately predicted the hydrolysis behavior for an EG2-CBHI-CBHII ternary mixture (Fig. 3.3A), an EG2-CBHI binary mixture (Fig. 3.3B), and a different loading of CBHI alone (Fig. 3.3C). The substrate parameters were the same as above. The enzyme loadings were 0.0218 μmol g^{-1} EG2, 0.0938 μmol g^{-1} CBHI, and 0.0313 μmol g^{-1} CBHII for the ternary mixture; 0.0117 μmol g^{-1} EG2 and 0.0762 μmol g^{-1} CBHI for the binary mixture, and 0.152 μmol g^{-1} CBHI for the single enzyme system. The ternary mixture had the same individual enzyme loadings as those employed in the single enzyme hydrolysis reactions discussed above. Equivalent loadings were used to demonstrate that the model is able to capture the synergy

observed in the ternary mixture. This is an important effect primarily related to EG enzymes shortening the average cellulose chain length on the surface, creating additional free chain ends with which the CBH enzymes can react. The ability to describe hydrolysis by different cellulase mixtures composed of EG and CBH enzymes enables use of the model to determine which enzyme mixtures will efficiently degrade a simple cellulosic substrate.

3.4.3 Mixture Effects on Substrates

To demonstrate how the model can be used to design efficient enzyme mixtures, three-dimensional plots of the total cellulose conversion at a specific hydrolysis time (24 and 72 hours) as it varies with ternary cellulase mixture composition were constructed for BMCC (substrate variables were the same as above) with a constant total enzyme loading of 10 mg enzyme per g of substrate (Fig. 3.4). Four 24-hour end-point assays for total cellulose conversion were used to validate the model predictions further. As can be seen in Fig. 3.4A the experimental 24-hour conversion values are in good agreement with the model predictions. Mixtures B and C have very close conversion values similar to the model predictions, but mixture C is larger than mixture B contradicting the model. This discrepancy may have resulted from experimental errors and/or uncertainty inherent to the model. The model predicts the composition of the optimal cellulase mixture for this system. The optimal cellulase mixture shifts from1:0:1 EG2:CBHI:CBHII for 24-hour conversion to 1:1:0 for 72-hour conversion. This shift is consistent with the differences in thermostability between CBHI and CBHII.

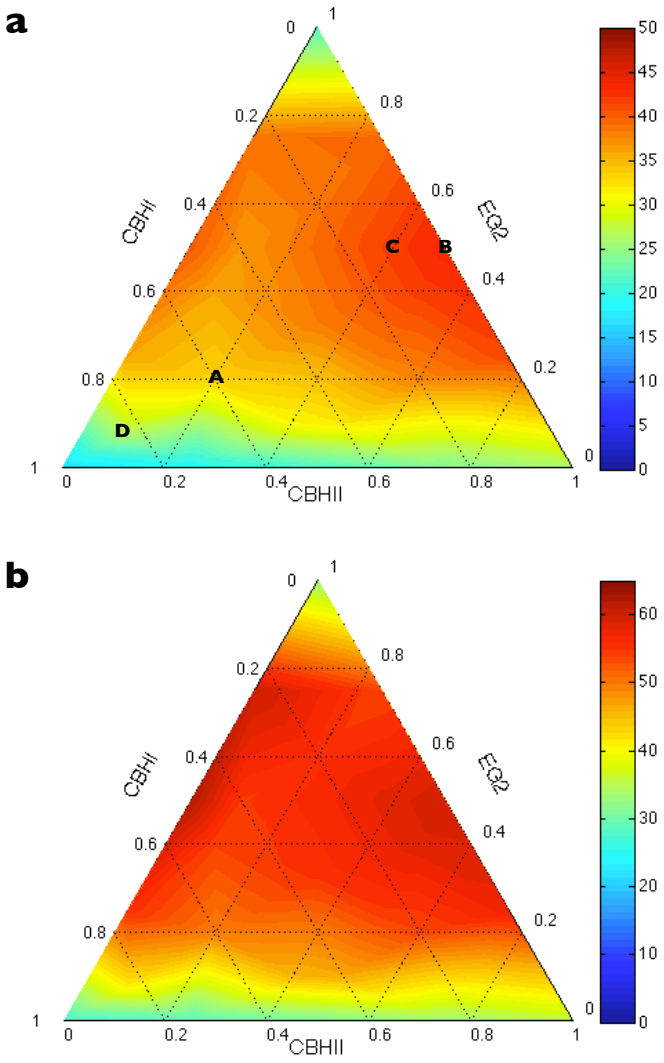


Figure 3.4 Three-dimensional plots of percent conversion of cellulose at (a) 24 hours and (b) 72 hours of hydrolysis with mixtures of EG2-CBHI-CBHII. The vertical axis (colored) is total cellulose conversion (%). The substrate has an initial DP of 750 and an initial surface area of 47.6 m^2 g^{-1}. Total cellulase loading is 10 mg g^{-1}. Model predictions were compared to experimental values at the marked points in (a) A: model 33.3%, experiment 29.9%; B: model 43.3%, experiment 40.4%; C: model 41.8%, experiment 46.0%; D: model 27.0%, experiment 28.7%.

The effect of substrate characteristics on hydrolysis by cellulase mixtures was examined for a range of model substrates. Figure 3.5 shows ternary diagrams for varying initial DPs and initial cellulose surface areas. Comparing the different substrate conditions shown in Fig. 3.5 clearly reveals that the effectiveness of different mixtures varies substantially. Accommodating different substrate characteristics is important for enzyme mixture design because lignocellulosic materials undergo physical or chemical pretreatment before saccharification. Different pretreatments have different effects on the substrate properties. These range from dilute acid pretreatment, which reduces substrate DP significantly but has less of an effect on surface area,

to biomass dissolution in ionic liquids, which leads to minimal changes in substrate DP but significantly increases surface area. It should be noted that pretreatments have other important effects relating to substrate parameters that are not incorporated into the current model (for example, crystallinity and the presence of lignin and other non-cellulosic species).

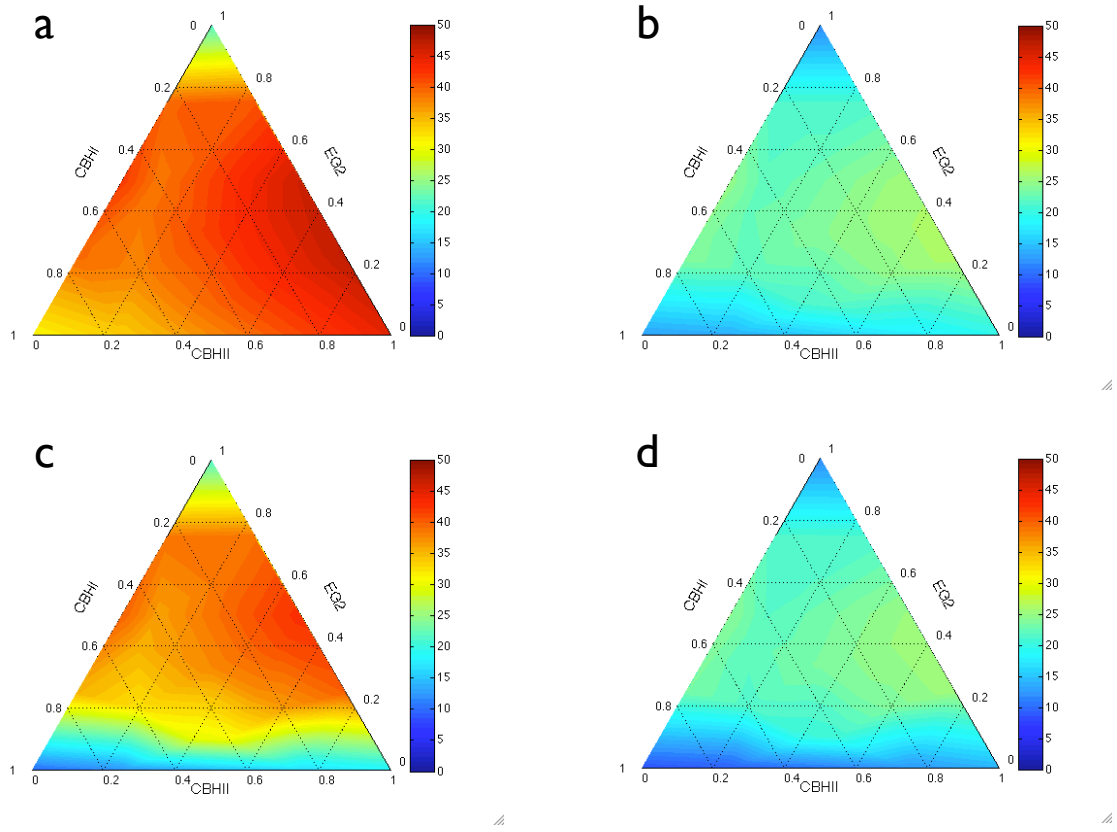


Figure 3.5 Three dimensional plots of percent conversion of cellulose at 24 hours of hydrolysis with mixtures of EG2-CBHI-CBHII. The vertical axis (colored) is total cellulose conversion (%). Total cellulase loading is 10 mg g^{-1}. The substrate has initial properties of (a) DP = 200 and surface area = 47.6 m^2 g^{-1}, (b) DP = 750 and surface area = 5.95 m^2 g^{-1}, (c) DP = 1500 and surface area = 47.6 m^2 g^{-1}, and (d) DP = 1500 and surface area = 5.95 m^2 g^{-1}.

3.5 Discussion

The rational design of cellulase mixtures tailored to a specific substrate requires a detailed understanding of the underlying mechanism of interaction between the enzymes and the substrate. As shown in Fig. 3.2 and Fig. 3.3 the model successfully predicts the evolution of the major soluble oligosaccharide products, validating the assumed mechanisms from which the model is derived. The validated model can be used as a tool to study enzyme-substrate interactions and to develop a better understanding of how both individual and mixed cellulases degrade cellulose. This understanding, as well as the model itself, forms the basis of a rational approach to design and optimize cellulase mixtures.

The model's prediction of how cellulase mixtures perform on substrates with different properties demonstrates interesting trends. Comparing Fig. 3.4, Fig. 3.5A, and Fig. 3.5C indicates that the higher the initial DP of the cellulose, the more important EG becomes to the

mixture. This relationship is expected based on the accepted endo-exo synergy mechanism whereby EG lowers the surface DP of the substrate, increasing the number of chain ends available for CBH enzymes. The model results show that the optimal level of EG in the mixtures increases as initial DP increases. However the increase is limited, indicating that an upper bound exists for the optimal EG fraction as DP increases. This is likely caused by a balance between increasing the number of chain ends available by increasing the EG fraction in the mixture and maintaining enough CBH enzymes in the system to use those chain ends efficiently. Furthermore, it is clear that the overall conversion achieved decreases with increasing DP, in agreement with published results on a range of cellulose substrates (*28, 29*).

An additional trend revealed by the related to the performance of enzyme mixtures is shown in Figs. 3.5B and 3.5D. These 3-dimensional plots were developed for model substrates with low initial surface area and show a large decline in the conversion achieved by all cellulase mixtures at 24 hours. The lower surface area was chosen to ensure it was a limiting factor. Comparing Fig. 3.5B to Fig. 3.4A and Fig. 3.5C to Fig. 5D reveals that the relative topology of the three-dimensional plot is unaltered as the surface area is reduced. These results demonstrate that surface area is more important than DP for the overall conversion of a hydrolyzed substrate, while the relative effectiveness of a specific cellulase mixture is primarily controlled by DP. A major focus of all biomass pretreatment methods should thus be to increase the available surface area of the biomass before saccharification, in agreement with our previous modeling study indicating that adequate surface area was critical for efficient cellulase action and synergy (*13*). These general trends of cellulase-mixture performance can be used to understand how various pretreatments should impact the optimal mixture composition. Harsher pretreatments (such as dilute-acid pretreatment), which decrease the DP of the cellulose drastically, would be expected to reduce the amount of endoglucanase required in the optimal mixture compared to less severe pretreatments (such as AFEX), which have a smaller effect on the cellulose DP. Furthermore, a pretreatment's overall effectiveness in improving cellulase activity towards a substrate will be controlled by how effectively the process "opens up" the biomass, increasing the accessible surface area.

By including sufficient mechanistic details, a description of the solid substrate, and a reliable parameter set for each enzyme in the mixture being designed, a straightforward optimization of the mixture composition can be performed to determine an efficient, tailored mixture for a given substrate. The method for determining efficient cellulose mixtures is shown in Fig. 3.6. The three steps in this rational mixture design can be summarized as (1) system definition, (2) model assembly and validation (if necessary), (3) and mixture optimization. Basing mixture optimization on a well-characterized model of the system of interest provides a clear link between the enzyme mixture and the substrate characteristics. These connections guide approaches to further improve hydrolysis rates, including modifying the substrate by new pretreatment methods, and altering the activities of individual cellulases through mutagenesis.

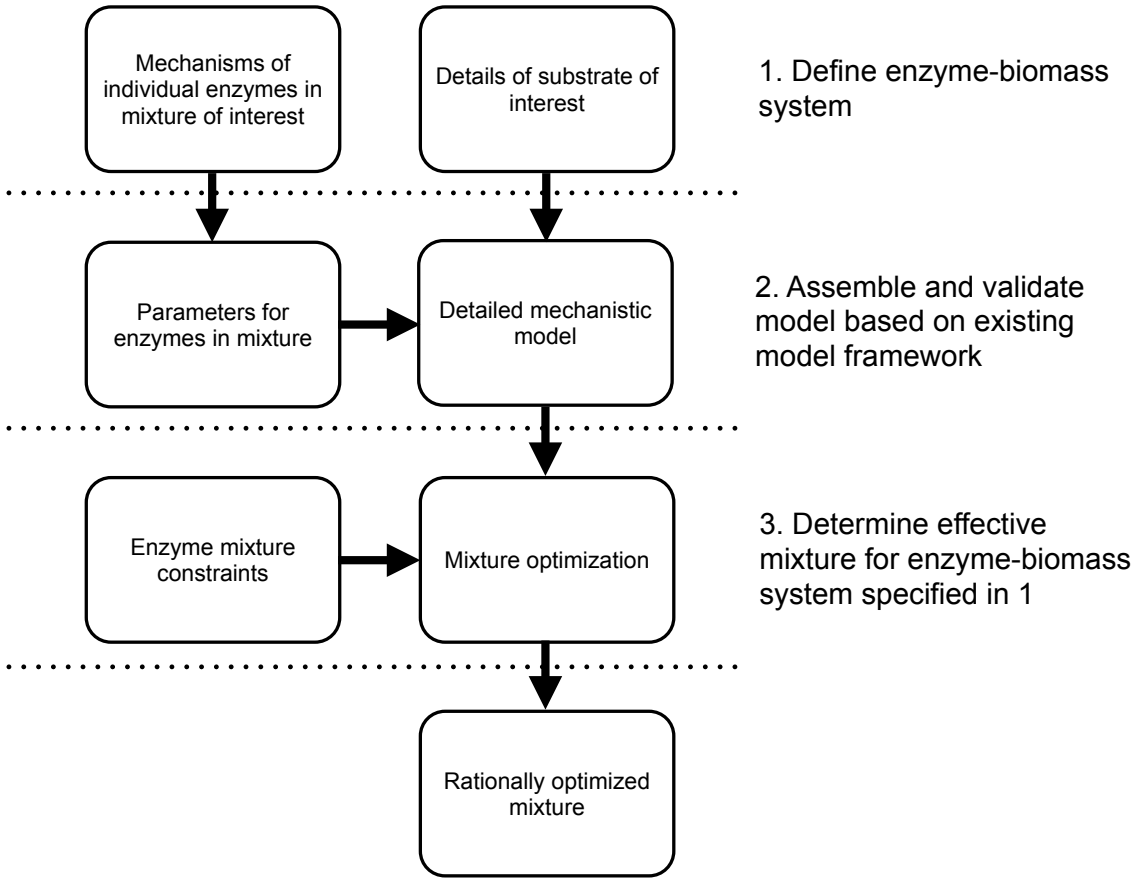


Figure 3.6 Schematic of the model-based rational approach to design and optimize a cellulase mixture for a specific substrate.

To date, the model-based mixture design method has only been applied to the proof-of-concept system of a ternary enzyme mixture acting on BMCC. BMCC is one of the most homogeneous cellulosic substrates available with a moderately high degree of polymerization (*30, 31*) and an open ribbon structure leading to high surface area (*32*). This system was chosen to demonstrate the rational design approach because it minimizes the impact of substrate surface heterogeneity on cellulase activity. In addition to determining the optimal ternary mixture for BMCC, qualitative trends of how substrate properties influence cellulase mixture performance have been revealed.

As the details of lignocellulosic substrates and biomass-degrading enzymes become better understood, it is critical to combine this information in a detailed mechanistic model to translate this knowledge into a practical tool. Understanding enzyme-substrate interactions of the various helper enzymes, such as Cel61A (*33*) as well as different hemicellulases, is important to expand the range of enzymes that can be included in the modeled enzyme mixtures. This requires knowledge of enzyme-specific kinetic and inhibition parameters. Equally important is the manner in which individual enzymes interact with different cellulosic surfaces. Understanding the range of adsorption-, complexation-, and mobility-related rates for the various enzymes on crystalline cellulose (including different crystal types and faces) as well as amorphous cellulose is important for discerning the different roles played by each enzyme in the mixture when degrading complex substrates. In order for the model to be used to optimize enzyme mixtures

toward real lignocellulosic substrates (e.g., corn stover, sugarcane bagasse, or miscanthus) enzyme interactions with lignin and hemicellulose need to be included.

3.6 Conclusions

We have developed a detailed modeling framework for cellulase hydrolysis of cellulose that includes mechanistic details for both CBH and EG enzymes that account for specific product formation. These include explicit inclusion of CBH processivity, an important mechanism usually neglected in kinetic models of enzymatic cellulose hydrolysis. The model was validated using a large set of time-course data for soluble oligosaccharide evolution during hydrolysis by both single and mixed cellulases. The model showed excellent predictive agreement with experimental data.

Three-dimensional plots of the performance of cellulase mixtures consisting of EG2, CBHI, and CBHII were generated for cellulosic substrates having a range of initial DPs and surface areas including values representing BMCC. These diagrams demonstrated the ability of the model to predict efficient cellulase mixtures. For BMCC, the substrate used to validate the model, the most efficient mixture was 1:0:1 EG2:CBHI:CBHII at 24 hours and 1:1:0 at 72 hours. Moreover, the results reveal general trends for the impact of specific cellulose characteristics on cellulase mixture performance. These include the increasing importance of EG in the mixture as the initial DP of the substrate increases, and the dominant role of initial surface area in controlling enzyme behavior. Additional substrate details can be incorporated into the model framework and studied for their impact on cellulase mixture performance.

The model described in this work forms the basis of a rational approach for designing effective cellulase mixtures. It relies on properly defining the important features of the system of interest, assembling a detailed model, and using that model for mixture optimization. This method illuminates crucial relationships between the substrate and the enzymes, and enables further refinement of the overall saccharification process. Furthermore, this method is based on a computational approach, thus avoiding the large amount of materials and experimental effort required for statistically designed experimental mixture optimization.

3.7 Acknowledgements

This work was supported by the Energy Biosciences Institute. JMF is grateful to the National Science Foundation for a Graduate Research Fellowship.

3.8 Nomenclature

$C_i^{'}$	Solid cellulose chain of length i (i>6)
DP	Degree of polymerization
$E^{'}_{CBH\text{-}ads}$	Uncomplexed cellobiohydrolase adsorbed to the surface
$E_{CBH\text{-}ads}C_i^{'}$	Cellobiohydrolase adsorbed to the surface and complexed with an i length cellulose chain formed via processive catalytic action
$E^{'}_{CBH\text{-}ads}C^{'}_{i\text{-}1at}$	Cellobiohydrolase adsorbed to the surface and complexed with an i length cellulose chain formed via complexation with a cellulose chain
$E^{'}_{EG\text{-}ads}$	Uncomplexed Endoglucanase adsorbed to the surface
$E_{CBH\text{-}ads}C_i^{'}$	Endoglucanase adsorbed to the surface and complexed with an i length cellulose chain
$k_{cat\text{-}CBH}$	Catalytic constant for cellobiohydrolase acting on cellulose (s^{-1})
$k_{cat\text{-}EG}$	Catalytic constant for endoglucanase acting on cellulose (s^{-1})
$k_{cat\text{-}EG\text{-}term}$	Catalytic constant for endoglucanase acting on cellulose at the terminal glycosidic bond (s^{-1})
$k_{comp\text{-}CBH\text{-}cel}$	Kinetic rate constant for complexation of a cellobiohydrolase with the end of a solid cellulose chain (m^2 $mmol^{-1}$ s^{-1})
$K_{comp\text{-}EG}$	Complexation equilibrium constant for endoglucanase with a cellulose chain (mmol m^{-2})
$K_{comp\text{-}EG\text{-}term}$	Complexation equilibrium constant for endoglucanase with a cellulose chain at the terminal glycosidic bond (mmol m^{-2})
p_i	Probability the initial catalytic event of a cellobiohydrolase after complexation will form and i length cello-oligosaccharide (i=1,2,3)
t	Time (s)
θ_{CBH}	Fraction of free sites on the surface available to a cellobiohydrolase for complexation
θ_{EG}	Fraction of free sites on the surface available to an endoglucanase for complexation
[ ]	surface concentration (mmol m^{-2}) when used with a primed symbol

3.9 References

1. M. Mandels, E. T. Reese, Fungal Cellulases and the Microbial Decomposition of Cellulosic Frabric. *Developments in Industrial Microbiology* **5**, 5 (1964).
2. J. O. Baker, C. I. Ehrman, W. S. Adney, S. R. Thomas, M. E. Himmel, Hydrolysis of Cellulose Using Ternary Mixtures of Purified Celluloses. *Applied Biochemistry and Biotechnology* **70-2**, 395 (1998).
3. G. Banerjee *et al.*, Synthetic Enzyme Mixtures for Biomass DecronstructionL Production and Optimization of a Core Set. *Biotechnology and Bioengineering* **106**, 707 (2010).
4. G. Banerjee, S. Car, J. S. Scott-Craig, M. S. Borrusch, J. D. Walton, Rapid Optimization of Enzyme Mixtures for Deconstruction of Diverse Pretreatment/Biomass Feedstock Combinations. *Biotechnology for Biofuels* **3**, (October 12, 2010, 2010).
5. D. Gao, S. P. S. Chundawat, C. Krishnan, V. Balan, B. E. Dale, Mixture Optimization of Six Core Glycosyl Hydrolases for Maximizing Saccharification of Ammonia Fiber Expansion (AFEX) Pretreated Corn Stover. *Bioresource Technology* **101**, 2770 (2010).
6. X. Zhao, T. R. Rignall, C. McCabe, W. S. Adney, M. E. Himmel, Molecular Simulation Evidence for Processive Motion of Trichoderma reesei. *Chemical Physics Letters* **460**, 284 (2008).
7. F. M. Gama, M. Mota, Enzymatic Hydrolysis of Cellulose .1. Relationship Between Kinetics and Physico-Chemical Parameters. *Biocatalysis and Biotransformation* **15**, 221 (1997).
8. R. Gupta, Y. Y. Lee, Mechanism of Cellulase Reaction on Pure Cellulosic Substrates. *Biotechnology and Bioengineering* **102**, 1570 (2009).
9. C. E. Wyman *et al.*, Comparitive Sugar Recovery and Fermentation Data Following Pretreatment of Poplar Wood by Leading Technologies. *Biotechnology Progress* **25**, 333 (2009).
10. P. Bansal, M. Hall, M. J. Realff, J. H. Lee, A. S. Bommarius, Modelling Cellulase Kinetics on Lignocellulosic Substrates. *Biotechnology Advances* **27**, 833 (2009).
11. A. O. Converse, J. D. Optekar, A Synergistic Kinetics Model for Enzymatic Cellulose Hydrolysis Compared to Degree-of-Synergism Experimental Results. *Biotechnology and Bioengineering* **42**, 145 (1993).
12. J. J. Fenske, M. H. Penner, J. P. Bolte, A Simple Individual-based Model of Insoluble Polysaccharide Hydrolysis: the Potential for Autosynergism with Dual-activity Glycosidases. *Journal of Theoretical Biology* **199**, 113 (1999).
13. S. E. Levine, J. M. Fox, H. W. Blanch, D. S. Clark, A Mechanistic Model of the Enzymatic Hydrolysis of Cellulose. *Biotechnology and Bioengineering* **107**, 37 (2010).
14. M. Okazaki, M. Moo-Young, KInetics of Enzymatic Hydrolysis of Cellulose: Analytical Description of a Mechanistic Model. *Biotechnology and Bioengineering* **20**, 637 (1978).
15. W. Zhou, H.-B. Schuttler, Z. Hao, Y. Xu, Cellulose Hydrolysis in Evolving Morphologies I: A General Modeling Formalism. *Biotechnology and Bioengineering* **104**, 261 (2009).
16. M. T. Holtzapple, H. S. Caram, A. E. Humphrey, The HCH-1 Model of Enzymatic Cellulose Hydrolysis. *Biotechnology and Bioengineering* **26**, 775 (1984).
17. S. Peri, S. Karra, Y. Y. Lee, M. N. Karim, Modeling Intrinsic Kinetics of Enzymatic Cellulose Hydrolysis. *Biotechnology Progress* **23**, 626 (2007).
18. S. Wald, C. R. Wilke, H. W. Blanch, Kinetics of the Enzymatic Hydrolysis of Cellulose. *Biotechnology and Bioengineering* **26**, 221 (1984).
19. Y.-H. P. Zhang, L. R. Lynd, A Functionally Based Model for Hydrolysis of Cellulose by Fungal Cellulase. *Biotechnology and Bioengineering* **94**, 888 (2006).
20. J. Medve, J. Karlsson, D. Lee, F. Tjerneld, Hydrolysis of Microcrystalline Cellulose by Cellobiohydrolysis I and Endoglucanase II from Trichoderma reesei: Adsorption, Sugar Production Pattern, and Synergism of the Enzymes. *Biotechnology and Bioengineering* **59**, 621 (1998).
21. P. Biely, M. Vrsanska, The Cellobiohydrolase I from Trichoderma reesei QM 9414: Action on Cello-Oligosaccharides. *Carbohydrate Research* **227**, 19 (1992).

22. B. Nidetzky, W. Zachariae, G. Gercken, M. Hayn, W. Steiner, Hydrolysis of Cellooligosaccharides by Trichoderma reesei Cellobiohydrolases: Experimental Data and Kinetic Modeling. *Enzyme and Microbial Technology* **16**, 43 (1994).
23. P. Biely, M. Vrsanska, M. Claeyssens, The Endo-1,4-Beta-Glucanase I from Trichoderma reesei Action on Beta-1,4-oligomers and Polymers Derived from D-glucose and D-xylose. *European Journal of Biochemistry* **200**, 157 (1991).
24. A. Koivula *et al.*, The Active Site of Trichoderma reesei Cellobiohydrolase II: the Role of Tyrosine 169. *Protein Engineering* **9**, 691 (1996).
25. A. Koivula *et al.*, The Active Site of Cellobiohydrolase Cel6A from Trichoderma reesei: The Roles of Aspartic Acids D221 and D175. *Journal of the American Chemical Society* **124**, 10015 (2002).
26. K. Zaar, Biogenesis of Cellulose by Acetobacter Xylinum. *Cytobiologie* **16**, 1 (1977).
27. M. Dubois, K. A. Gilles, J. K. Hamilton, P. A. Rebers, F. Smith, Colorimetric Method for Determination of Sugars and Related Substances. *Analytical Chemistry* **28**, 350 (1956).
28. G. Beldman, A. G. J. Voragen, F. M. Rombouts, M. F. Searle-van Leeuwen, W. Pilnik, Adsorption and Kinetic Behavior of Purified Endoglucanases and Exoglucanases from Trichoderma viride. *Biotechnology and Bioengineering* **30**, 251 (1987).
29. T. M. Wood, Properties and Mod of Action of Cellulases. *biotechnology and Bioengineering Symposium*, 111 (1975).
30. K. M. Kleman-Leyer, N. R. Gilkes, R. C. Miller Jr., T. K. Kirk, Changes in the Molecular-Size Distribution of Insoluble Celluloses by the Action of Recombinant Cellulomonas fimi Cellulases. *Biochemical Journal* **302**, 463 (1994).
31. P. Valjamae, V. Sild, A. Nutt, G. Pettersson, G. Johansson, Acid Hydrolysis of Bacterial Cellulose Reveals Differenct Modes of Synergistic Action Between Cellobiohydrolase I and Endoglucanase I. *European Journal of Biochemistry* **266**, 327 (1999).
32. J. Hong, X. Ye, Y.-H. P. Zhang, Quantitative Determination of Cellulose Accessibility to Cellulase Based on Adsorption of a Nonhydrolytic Fusion Protein Containing CBM and GFP with Its Applications. *Langmuir* **23**, 12535 (2007).
33. P. V. Harris *et al.*, Stimulation of Lignocellulosic Biomass Hydrolysis by Proteins of Glycoside Hydrolase Family 61: Structure and Function of a Large Enigmatic Family. *Biochemistry* **49**, 3305 (2010).
34. J. Karlsson, M. Siika-aho, M. Tenkanen, F. Tjerneld, Enzymatic Properties of the Low Molecular Mass Endoglucanases Cel12A (EG III) and Cel45A (EG V) of Trichoderma reesei. *Journal of Biotechnology* **99**, 63 (2002).
35. R. E. T. Drissen *et al.*, A Generic Model for Glucose Production from Various Cellulose Sources by a Commerical Cellulase Complex. *Biocatalysis and Biotransformation* **25**, 419 (November-December, 2007).

Chapter 4: An Evaluation of Cellulose Saccharification and Fermentation with an Engineered *Saccharomyces cerevisiae* Capable of Cellobiose and Xylose Utilization[*]

4.1 Abstract

Commercial-scale cellulosic ethanol production has been hindered by high costs associated with cellulose-to-glucose conversion and hexose and pentose co-fermentation. Simultaneous saccharification and fermentation with an *S.* cerevisiae strain capable of cellobiose and xylose co-utilization has been proposed as a possible avenue to reduce these costs. Cellobiose is the primary product of cellulase-catalyzed hydrolysis of cellulose, but as a potent inhibitor of cellulase enzymes, it is usually hydrolyzed by exogenously added β-glucosidase. A yeast strain capable of cellobiose consumption has the potential to alleviate cellobiose inhibition of cellulase enzymes while preventing the generation of extracellular glucose, a known inhibitor of xylose transport in yeast. The recently developed DA24-16 strain of *S. cerevisiae* incorporates a xylose assimilation pathway and a cellodextrin transporter (CDT) that permit rapid growth on xylose and cellobiose. In the present work, a mechanistic kinetic model of cellulase-catalyzed hydrolysis of cellulose was combined with a multi-substrate model of microbial growth to investigate the ability of cellobiose consumption by this strain to obviate the need for exogenously added β-glucosidase and to assess the impact of cellobiose utilization on simultaneous saccharification and fermentation (SSF) and separate hydrolysis and fermentation (SHF). Results indicate that with SSF, the strain is not as effective at enhancing cellulose hydrolysis or ethanol production as a non-CDT-containing strain supplemented with β-glucosidase. Under SSF conditions, reducing the Monod constant for growth on cellobiose enabled DA24-16 to produce ethanol nearly as rapidly as non-CDT-containing yeast. In producing 75 g/L ethanol, SSF with any strain did not result in shorter residence times than SHF with a 12-hour saccharification step. Strains with improved cellobiose utilization are therefore unlikely to allow higher titers to be reached more quickly in SSF than in SHF.

4.2 Introduction

Cellulosic biofuels may provide a viable low-carbon alternative to traditional fossil fuels, but high cellulose-to-sugar conversion costs and inefficient fermentation of both hexose and pentose sugars have made commercial production economically unattractive. Simultaneous saccharification and fermentation (SSF) has been proposed as a possible avenue to improve process economics. In this one-reactor scenario, cellulase enzymes catalyze the hydrolysis of cellulose to glucose, which is simultaneously fermented into alcohols or other liquid fuels, thus preventing monomeric sugars and soluble cello-oligosaccharides from inhibiting the enzymatic de-polymerization process. Alcohols produced during SSF have the potential to reduce microbial contamination that might occur in non-aseptic saccharification reactors, and the use of one reactor has been proposed as a possible strategy to lower equipment costs (*1-3*).

Yeast strains capable of simultaneous cellodextrin and xylose fermentation may be particularly advantageous in SSF processes. The plant cell wall of biofuel-candidate grasses and trees contains xylose and glucose polymers that must be broken down into their constituent

[*] This chapter, with few modifications from its original format, represents the following peer-reviewed publication: J. M. Fox, S. E. Levine, H. W. Blanch, D. S. Clark, An evaluation of cellulose saccharification and fermentation with an engineered Saccharomyces cerevisiae capable of cellobiose and xylose utilization. *Biotechnol J*, (Jan 9, 2012).

sugars prior to being metabolized into useful fuels (*4*). Xylose is present primarily as the hemicellulose xylan, which is released as soluble xylose oligomers during pretreatment of lignocellulosic substrates (*5, 6*). Glucose is present in the β-1,4-linked cellobiose polymer cellulose, which undergoes saccharification through an enzyme-catalyzed hydrolysis process that takes place after pretreatment. The cellulase enzymes responsible for saccharification are inhibited by cellobiose; β-glucosidase is often added to alleviate this inhibition by converting cellobiose to glucose (*7*). Ideally, fuel-producing microbes would be capable of using both xylose and glucose; the additional ability to consume cellobiose may obviate the need for the addition of β-glucosidase.

The engineered DA24-16 strain of *Saccharomyces cerevisiae* may improve the economics of SSF. In this strain, a xylose assimilation pathway, a *Neurospora crassa* cellodextrin transporter (CDT), and an intracellular β-glucosidase were introduced to facilitate rapid consumption of xylose and cellobiose (*8*). Previous attempts to engineer xylose assimilation pathways into *S. cerevisiae* have been successful, but DA24-16 is able to grow on xylose at rates nearly equivalent to those of *Pichia stipitis*, which natively consumes xylose more rapidly than any other wild type yeast strain yet discovered (*8-11*). *S. cerevisiae* strains capable of growth on cellobiose have also been engineered previously, but all have relied on secreted β-glucosidase enzymes to hydrolyze cellobiose to glucose (*12-14*). The advantage of the CDT-containing DA24-16 strain relates to its ability to avoid extracellular production of glucose, which inhibits xylose consumption (*15*). By obviating the need for exogenous β-glucosidase and by facilitating the uninhibited metabolism of xylose, the DA24-16 *S. cerevisiae* strain and yeast with similar capabilities may reduce enzyme costs and increase the sugar conversion efficiencies associated with SSF processes.

In the present work, we explore the potential advantages of cellobiose utilization by employing a mechanistic kinetic model of cellulase action and a multi-substrate model of microbial growth to compare the performance of a CDT-containing strain in SSF to the performance of a non-CDT-containing strain in SSF and SHF. Our analysis places particular emphasis on assessing the capacity of cellobiose transporters acting in concert with an intracellular β-glucosidase to replace exogenously added β-glucosidase enzymes in the hydrolysis process, and on evaluating the impact of cellobiose utilization on the possible advantages of SSF over SHF. Xylose growth parameters for both strains and cellobiose growth parameters for the cellodextrin-containing strain were equated to those of DA24-16; however, our results are generally applicable to other CDT-containing *S. cerevisiae* strains that are able to consume multiple substrates. The sensitivity of our results to these parameters and the likelihood of altered growth behavior in response to changes in these parameters are also discussed. This work marks the first time that an independently validated mechanistic model of cellulase action capable of tracking cello-oligosaccharide chain length has been used alongside a microbial growth model to assess the potential benefits of cellobiose consumption in yeast.

4.3 Methods

4.3.1 Choice of Enzymatic Hydrolysis Conditions

The process economics of commercial-scale cellulosic biofuels production are enhanced by high solids loadings (20% or higher), which lower energy costs associated with substrate flow and product separation (*16*). In the present work, commercial operation requirements were

accommodated through the use of cellulose and xylose concentrations of 129 g/L and 71 g/L, respectively. These match the average cellulose:xylan ratio (15:9) of candidate biofuels crops and amount to a 200 g/L (20%) solids loading in reactions where both substrates are present (*4*). This solids loading refers to the pretreated slurry where the lignin has been removed (*16*).

The cellulase enzymes chosen for this study were *Trichoderma sp.* Cel7A (*Tr*Cel7A), *Trichoderma sp.* Cel6A (*Tr*Cel6A), and *Talaromyces emersonii* GH5 (*Tem*GH5*)*. These fungal cellulases are similar to the main constituents of commercial cellulase mixtures. The *Tr*Cel7A and *Tr*Cel6A cellobiohydrolases catalyze the hydrolysis of insoluble cellulose by adsorbing to the substrate, complexing with reducing ends (*Tr*Cl7A) or non-reducing ends (*Tr*Cel6A) of cellulose chains, engaging in processive catalysis, decomplexing from chains, and recomplexing or desorbing from the surface. The *Tem*GH5 endoglucanase enzyme engages in a similar manner, but complexes anywhere within chains and catalyzes only one hydrolysis event before decomplexing. In solution, all three enzymes can directly hydrolyze soluble cello-oligosaccharides (DP<7). The kinetics of these steps are described by Levine et al., and parameter values are included in Table 4.1. Cellulose was assumed to have the properties of Avicel, as described by Levine et al (*17*).

In our studies, xylose was used as a surrogate for the xylan hydrolyzate produced during pretreatment of lignocellulosic substrates. Pretreatments such as dilute acid hydrolysis, alkaline hydrolysis, steam explosion, liquid hot water pretreatment, and ionic liquid pretreatment can remove the majority of xylan from cellulosic substrates, producing xylose and a variety of soluble xylo-oligomers (*5, 6, 18, 19*). These xylo-oligomers, which can make up over 50% of xylan hydrolyzate, have been shown to inhibit cellulase activity as strongly as cellobiose, which is about 100-fold more inhibitory than glucose. Xylo-oligomer concentrations, however, can be significantly reduced and converted to xylose through optimized pretreatment processes or through the addition of β-D-xylosidase to hydrolysis reactions (*20-22*). Accordingly, xylan hydrolyzate is represented as xylose in the present work. Pretreatments that incompletely hydrolyze xylan, creating a requirement for significant xylanase action during the cellulose hydrolysis step, are not addressed here.

Table 4.1 Parameter Values

Parameter	Value	Units	Source	Description
$k_{EG2\text{-}ads}$	2.4	L $mmol^{-1}$ s^{-1}	(*17*)[a]	Adsorption of EG2 onto the cellulose surface
$k_{CBHI\text{-}ads}$	2.4	L $mmol^{-1}$ s^{-1}	(*17*)[a]	Adsorption of CBH1 onto the cellulose surface
$k_{CBHII\text{-}ads}$	2.4	L $mmol^{-1}$ s^{-1}	(*17*)[a]	Adsorption of CBH2 onto the cellulose surface
$k_{EG2\text{-}des}$	5.36×10^{-3}	s^{-1}	(*17*)[a]	Desorption of EG2 from the cellulose surface
$k_{CBHI\text{-}des}$	6.10×10^{-3}	s^{-1}	(*23*)[b]	Desorption of CBH1 from the cellulose surface
$k_{CBHII\text{-}des}$	2.08×10^{-4}	s^{-1}	This study[b]	Desorption of CBH2 from the cellulose surface
$k_{cat\text{-}EG2}$	65	s^{-1}	(*17*)[a]	Activity of EG2 on cellulose chains
$k_{cat\text{-}CBHI}$	4.42	s^{-1}	This study[b]	Activity of CBH1 on cellulose chains
$k_{cat\text{-}CBHII}$	14	s^{-1}	(*24*)[c]	Activity of CBH2 on C_i cellulose chains

$K_{M\text{-}EG2\text{-}cel}$	2	mmol m^{-2}	This study[b]	K_M of EG2 for insoluble cellulose chains
$k_{comp\text{-}CBHI\text{-}cel}$	1990	m^2 $mmol^{-1}$ s^{-1}	(*23*)[b]	Complexation of CBH1 / insoluble chain ends
$k_{comp\text{-}CBHII\text{-}cel}$	7780	m^2 $mmol^{-1}$ s^{-1}	(*23*)[b]	Complexation of CBH2 / insoluble chain ends
$k_{decom\text{-}CBHI\text{-}cel}$	1.37	s^{-1}	(*23*)[b]	Decomplexation of CBH1 / insol chain ends
$k_{decom\text{-}CBHII\text{-}cel}$	2.78	s^{-1}	(*23*)[b]	Decomplexation of CBH2 / insol chain ends
$K_{M\text{-}EG2\text{-}sol}$	0.053	mM	(*17*)[a]	K_M for EG2 and soluble cellulose chains
$k_{comp\text{-}CBHI\text{-}sol}$	483	L $mmol^{-1}$ s^{-1}	(*25*)[d]	Complexation of CBH1 / soluble chain ends
$k_{comp\text{-}CBHII\text{-}sol}$	1200	L $mmol^{-1}$ s^{-1}	(*24*)[d]	Complexation of CBH2 / soluble chain ends
$k_{decom\text{-}CBHI\text{-}sol}$	2.78	s^{-1}	(*25*)[d]	Decomplexation of CBH1 / sol chain ends
$k_{decom\text{-}CBHII\text{-}sol}$	2.78	s^{-1}	(*24*)[d]	Decomplexation of CBH2 / sol chain ends
$k_{cat\text{-}EG2\text{-}C3}$	0.253	s^{-1}	(*26*)[d]	Activity of EG2 on cellotriose
$k_{cat\text{-}CBHI\text{-}C3}$	0.016	s^{-1}	(*26*)[e]	Activity of CBH1 on cellotriose
$k_{cat\text{-}CBHII\text{-}C3}$	0.051	s^{-1}	(*26*)[d]	Activity of CBH2 on cellotriose
$K_{M\text{-}EG2\text{-}C3}$	0.96	mM	(*26*)[a]	K_M for EG2 on cellotriose
$k_{comp\text{-}CBHI\text{-}C3}$	26	L $mmol^{-1}$ s^{-1}	(*23*)[e]	Complexation of CBH1 with cellotriose
$k_{comp\text{-}CBHII\text{-}C3}$	142	L $mmol^{-1}$ s^{-1}	(*24*)[d]	Complexation of CBH2 with cellotriose
$k_{decom\text{-}CBHI\text{-}C3}$	2.78	s^{-1}	(*23*)[e]	Decomplexation of CBH1 with cellotriose
$k_{decom\text{-}CBHII\text{-}C3}$	2.78	s^{-1}	(*24*)[d]	Decomplexation of CBH2 with cellotriose
C1:C2:C3 CBH1	6:1.5:2.5	NA	(*26*)c	Ratio of C1:C2:C3 for CBH1 initial cuts
C1:C2:C3 CBH2	6:1.5:2.5	NA	(*26*)[d]	Ratio of C1:C2:C3 for CBH2 initial cuts
$k_{cat\text{-}EG2\text{-}C1}$	5.44	s^{-1}	(*27*)[d]	Activity of EG2 at chain ends
$K_{M\text{-}EG2\text{-}C1\text{-}cel}$	900	mmol m^{-2}	(*23*)[b]	K_M for EG2 on insoluble chain ends
$K_{M\text{-}EG2\text{-}C1\text{-}sol}$	90	mM	(*23*)[b]	K_M for EG2 on soluble chain ends
$K_{i\text{-}EG2\text{-}C2}$	0.03	mM	(*23*)[b]	Competitive inhibition of EG2 by C2
$K_{i\text{-}CBHI\text{-}C2}$	0.19	mM	(*23*)[b]	Competitive inhibition of CBH1 by C2
$K_{i\text{-}CBHII\text{-}C2}$	0.06	mM	(*23*)[b]	Competitive inhibition of CBH2 by C2
$K_{i\text{-}EG2\text{-}C1}$	16.0	mM	(*23*)[b]	Competitive inhibition of EG2 by C1
$K_{i\text{-}CBHI\text{-}C1}$	31	mM	(*17*)[a]	Competitive inhibition of CBH1 by C1
$K_{i\text{-}CBHII\text{-}C1}$	29	mM	(*23*)[b]	Competitive inhibition of CBH2 by C1
$k_{d\text{-}EG2}$	3.89×10^{-6}	s^{-1}	(*28*)[d]	Desorption of EG2 from cellulose
$k_{d\text{-}CBHI}$	3.89×10^{-6}	s^{-1}	(*28*)[d]	Desorption of CBH1 from cellulose
$k_{d\text{-}CBHII}$	8.06×10^{-6}	s^{-1}	(*28*)[d]	Desorption of CBH2 from cellulose
$\mu_{max\text{-}C1}$	0.5	h^{-1}	(*29*)[c]	Maximum specific growth rate on C1
K_{C1}	1.7	mM	(*30*)[c]	Monod constant for growth on C1
$Y_{X/C1}$	0.19	g/g	(*30*)[c]	Yield coefficient for dry cell weight from C1
$Y_{Eth/C1}$	0.43	g/g	(*8*)[c]	Yield coefficient for ethanol from C1
$\mu_{max\text{-}C2}$	0.39	h^{-1}	This study[g]	Maximum specific growth rate on C2
K_{C2}	350	mM	This study[g]	Monod constant for growth on C2
$B_{i\text{-}C1\text{-}C2}$	9.0	mM^{-1}	(*30*)[c]	Inhibition of C2 uptake by C1

$Y_{X/C2}$	0.15	g/g	(*30*)[c]	Yield coefficient for dry cell weight from C2
$Y_{Eth/C2}$	0.42	g/g	(*8*)[c]	Yield coefficient for ethanol from C2
$\mu_{max\text{-}Xy}$	0.19	h^{-1}	This study[g]	Maximum specific growth rate on xylose
K_{Xy}	20	mM	This study[g]	Monod constant for growth on xylose
$q_{max\text{-}Xy}$	4.73	mM/g/h	(*8*)[c]	Maximum specific uptake rate for xylose
$B_{i\text{-}C1\text{-}Xy}$	9.0	mM^{-1}	(*30*)[c]	Inhibition of xylose uptake by glucose
$Y_{Eth/Xy}$	0.35	g/g	(*8*)[c]	Yield coefficient for ethanol form xylose
$K_{i\text{-}Xy\text{-}EG2}$	16.0	mM	This study[d]	Competitive inhibition of EG2 by xylose
$K_{i\text{-}Xy\text{-}CBH1}$	31	mM	This study[d]	Competitive inhibition of CBH1 by xylose
$K_{i\text{-}Xy\text{-}CBH2}$	29	mM	This study[d]	Competitive inhibition of CBH2 by xylose
$K_{i\text{-}X\text{-}Eth}$	87	g/L	(*31*)[d]	Threshold inhibition of cell growth by ethanol
$K_{i\text{-}Eth\text{-}EG2}$	50.35	g/L	(*32*)[d]	Noncompetitive inhibition of EG2 by ethanol
$K_{i\text{-}Eth\text{-}CBH1}$	50.35	g/L	(*32*)[d]	Noncompetitive inhibition of CBH1 by ethanol
$K_{i\text{-}Eth\text{-}CBH2}$	50.35	g/L	(*32*)[d]	Noncompetitive inhibition of CBH2 by ethanol
$k_{cat\text{-}BG\text{-}C2}$	518	s^{-1}	(*33*)	Activity of β-glucosidase on cellobiose at 30°C
$k_{cat\text{-}BG\text{-}C2}$	1810	s^{-1}	(*33*)	Activity of β-glucosidase on cellobiose at 50°C
$K_{M\text{-}BG\text{-}C2}$	0.88	mM	(*33*)	K_M for BG on cellobiose
$K_{i\text{-}BG\text{-}C1}$	3.4	mM	(*33*)	Competitive inhibition of BG by glucose

[a]Parameters used were taken from the hydrolysis model developed by Levine et al (*17*).
[b]Parameters optimized against single enzyme hydrolysis data in the latest version of the hydrolysis model developed by Levine et al (*23*).
[c]Parameters taken directly from the literature.
[d]Parameter set based on published data.
[e]Parameters obtained via kinetic experiments using soluble cello-oligosaccharides as reported in Levine et al (*23*).
[f]Parameters for soluble substrate based on optimized values for solid substrate reactions as in Levine et al (*23*).
[g]Parameters optimized against batch fermentation data from Ha et al (*8*).

Xylose, which is significantly less inhibitory than longer DP xylo-oligomers, is thought to be a competitive inhibitor of cellulase enzymes (*5, 34*); however, there are no directly measured xylose inhibition parameters available in the literature. In the present work, these parameters were equated with the glucose inhibition parameters for each enzyme ($K_{i\text{-}EG\text{-}G} = K_{i\text{-}EG\text{-}Xy}$).

Pretreatments that hydrolyze the majority of the hemicellulose produce xylose feed streams that that can be routed either to glucose/xylose co-fermentation reactors or to separate xylose fermentation units (*18, 35*). Process scenarios in which xylose undergoes a separate fermentation step are not addressed in the present work. In our model, xylan hydrolyzate is

included in the combined hydrolysis and fermentation of SSF, whereas in SHF, it is included in the fermentation but not necessarily in the hydrolysis reaction.

In the present work, xylose inhibition of cellulase enzymes is explored through modeling runs where its inhibitory effect is assumed to be as weak as that of glucose or as strong as that of cellobiose. The implications of this range of inhibition on the present assessment of multi-substrate utilizing yeasts in SSF processes and on SHF vs. SSF comparisons are addressed below.

Cellulase loadings of 25 mg enzyme / g glucan were used in each simulation. This loading is similar to that reported for proposed commercial operations and to those employed in recently-published SSF studies (*36-39*). The cellulase mixture contained cellobiohydrolase I (CBHI), cellobiohydrolase II (CBHII), and endoglucanase (EG1) in a 1:3:1 molar ratio, which is close to the optimal composition reported for these enzymes (*40*). When β-glucosidase was present, it represented 10% of the total enzyme concentration (by mass), a commonly reported β-glucosidase concentration (*21*).

The optimal conditions for the fungal enzymes considered are pH of 4.85 and a temperature of 50°C. The growth and ethanol production rates of *S. cerevisiae* are highest at a pH of 5.5 and a temperature of 30°C. Previous studies have shown the activity of *Trichoderma sp.* enzymes to be 50% lower at 30°C than at 50°C. We therefore accounted for the reduction in enzyme activity at 30°C by reducing the cellulase enzyme k_{cat} values by 50%. The effect of pH 5.5 on optimal enzyme activity is small (<5% reduction in activity) (*41*), and was ignored in this work. The kinetic parameters for β-glucosidase activity at 30 and 50°C were taken directly from Chauve et al (*33*). All other parameters were obtained from the literature and were assumed to be independent of temperature in the 30-50°C range (Table 4.1).

4.3.2 Hydrolysis Model Development

Cellulose saccharification was modeled with the mechanistic kinetic model of cellulase-catalyzed hydrolysis recently developed by Levine et al (*23*). Cellulase action on insoluble cellulosic substrates is modeled as a heterogeneous biocatalytic reaction involving endo- and exocellulases that either interact directly with soluble cello-oligosaccharides (DP<7) or form complexes with insoluble cellulose chains after first adsorbing to available sites on the substrate surface. Unlike previously developed models of cellulase catalysis, this model tracks all chain lengths and products, and it does not assume equilibrium for cellobiohydrolase-substrate complexes (*23, 42-45*). This model also explicitly accommodates reduced activity by endoglucanase enzymes at chain ends, and processive catalysis by cellobiohydrolase enzymes along chains.

The action of beta-glucosidase on cellobiose was modeled with Michaelis-Menten kinetics modified to include competitive inhibition by glucose:

$$\frac{d[C_2]}{dt} = \frac{k_{cat-BG-C2} BG [C_2]}{[C_2] + K_{M-BG-C2}\left(1 + \frac{[C_1]}{K_{i-BG-C1}}\right)} \tag{4.1}$$

where BG is the concentration of β-glucosidase and $K_{i\text{-}BG\text{-}C1}$ is the equilibrium constant for competitive inhibition by glucose.

For SSF, the additional influences of ethanol and xylose on enzyme behavior were incorporated through the addition of non-competitive and competitive inhibition terms, which were obtained from reports on the effect of these compounds on cellulose hydrolysis rates (*5, 46*). These terms appear in equations for the site and enzyme balances. Inhibition effects of ethanol and xylose were described by the following relationships:

$$k_{cat}^{app} = \frac{k_{cat}}{1 + \frac{[Eth]}{K_{i-Eth}}} \tag{4.2}$$

$$K_M^{app} = K_M \left(1 + \sum_{i=7}^{\infty} \frac{q_{CBH1}\left[C_i^{'}\right]}{K_{M\text{-}CBH1\text{-}cel}} + \sum_{i=3}^{6} \frac{[C_i]}{K_{M\text{-}CBH1\text{-}sol}} + \frac{[C_1]}{K_{i\text{-}CBH1\text{-}C1}} + \frac{[C_2]}{K_{M\text{-}CBH1\text{-}C2}} + \frac{[Xy]}{K_{i\text{-}CBH1\text{-}Xy}} \right) \tag{4.3}$$

where k_{cat} represents an arbitrary kinetic constant, and $K_{i\text{-}Eth}$ represents the noncompetitive equilibrium constant between ethanol and a given enzyme. The other parameters are defined as in Levine et al. (*17, 23*); their values are reported in Table 4.1.

4.3.3 Choice of SSF and SHF Conditions

Our comparisons of SSF and SHF processes are based on the total residence times required to reach particular ethanol concentrations. Estimates of the combined residence time for hydrolysis and fermentation processes range from 3-6 days, but 2-3 day targets have been proposed for commercial operations (*47*). For SHF, we considered three different hydrolysis times spanning the period over which the hydrolysis rate varied significantly (Fig. 4.1): 12, 24, and 36-hr.

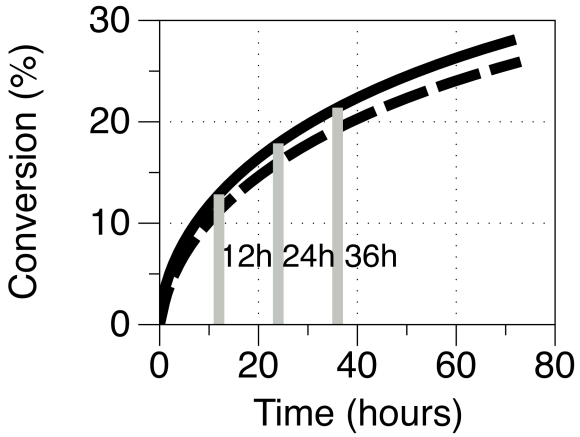


Figure 4.1 Profiles of cellulose conversion during saccharification reactions. Runs were carried out for 144 hours at 50°C with a combined total enzyme loading of 2.5 g/L and substrate concentrations of 128.57g/L cellulose with or without 71.43 g/L xylose. Cellulases were present in a 1:3:1 EG2:CBH1:CBH2 (molar) enzyme ratio; β-glucosidase made up 10% of the enzyme mixture by mass. Lines are as follows: solid black line (cellulose present); dashed black line (cellulose and xylose present).

Fermentations were simulated at the optimal pH and temperature for growth and ethanol production by *S. cerevisiae* (pH = 5.5; T = 30°C). When fermentation and hydrolysis were conducted separately, they were done so under pH and temperature conditions that were optimal for each step: pH 4.85 and 50°C for saccharification and pH 5.5 and 30°C for fermentation.

Fermentations were modeled with and without xylose to assess the impact of this sugar on the results.

All fermentations had an inoculum of 0.252 g cells/L (OD_{600}~1.0) so that growth could be compared with published batch fermentation processes. The impact of larger inocula on residence times is addressed below.

4.3.4 Model Development for SSF and SHF

Yeast growth was modeled using Monod kinetics with growth-associated ethanol production and threshold-based ethanol inhibition. Simultaneous utilization of multiple carbon sources (glucose, cellobiose, and xylose) was incorporated through additive growth kinetics modified to include catabolic repression of cellobiose and xylose utilization by glucose. Previous reports on the DA24-16 strain showed no evidence that xylose inhibits cellobiose uptake or vice versa. The resulting equation for batch growth in the absence of cellulases is identical in form to that proposed for the glucose, cellobiose, and xylose fermenting yeast *Candida lusitaniae* (*30*):

$$\frac{dX}{dt}=\left(\frac{\mu_{\max-C_1}C_1}{K_{C_1}+C_1}+\frac{\mu_{\max-C_2}C_2}{K_{C_2}+C_2}\frac{1}{K_{i-C_1-C_2}C_1+1}+\frac{\mu_{\max-Xy}Xy}{K_{Xy}+Xy}\frac{1}{K_{i-C_1-Xy}C_1+1}\right)\left(1-\frac{[Ethanol]}{K_{i-Eth-X}}\right)X \tag{4.4}$$

Ethanol production by each substrate was modeled in an analogous fashion:

$$\frac{dEth}{dt}=\left(\begin{array}{l}\frac{Y_{Eth/C_1}}{Y_{X/C_1}}\frac{\mu_{\max-C_1}C_1}{K_{C_1}+C_1}+\frac{Y_{Eth/C_2}}{Y_{X/C_2}}\frac{\mu_{\max-C_2}C_2}{K_{C_2}+C_2}\frac{1}{B_{i-C_1-C_2}C_1+1}\\ +\frac{Y_{Eth/Xy}}{Y_{X/Xy}}\frac{\mu_{\max-Xy}Xy}{K_{Xy}+Xy}\frac{1}{B_{i-C_1-Xy}C_1+1}\end{array}\right)\left(1-\frac{[Ethanol]}{K_{i-Eth-X}}\right)X \tag{4.5}$$

Uptake equations for each carbon source were based on Monod kinetics:

$$\text{rate of glucose uptake}=-\left(\frac{1}{Y_{X/C_1}}\right)\left(\frac{\mu_{\max-C_1}C_1}{K_{C_1}+C_1}\right)\left(1-\frac{[Ethanol]}{K_{i-X-Eth}}\right)X \tag{4.6}$$

$$\text{rate of cellobiose uptake}=-\left(\frac{1}{Y_{X/C_2}}\right)\left(\frac{\mu_{\max-C_2}C_2}{K_{C_2}+C_2}\right)\left(\frac{1}{B_{i-C_1-C_2}C_1+1}\right)\left(1-\frac{[Ethanol]}{K_{i-Eth-X}}\right)X \tag{4.7}$$

$$\text{rate of xylose uptake}=-\left(\frac{1}{Y_{X/Xy}}\right)\left(\frac{\mu_{\max-Xy}Xy}{K_{Xy}+Xy}\right)\left(\frac{1}{B_{i-C_1-Xy}C_1+1}\right)\left(1-\frac{[Ethanol]}{K_{i-Eth-X}}\right)X \tag{4.8}$$

where Y, μ_{max}, and K represent the specific yield coefficient, maximum specific growth rate, and Monod constant for growth on each substrate; $B_{i\text{-}C1}$ and $K_{i\text{-}Eth}$ represent parameters describing glucose inhibition of substrate uptake and ethanol inhibition of cell growth.

Equations 4.6 and 4.7 were included in the material balance for glucose and cellobiose, effectively linking the fermentation model to the cellulose hydrolysis model. Equation 4.8 was included in the xylose material balance.

4.3.5 Model parameters

To estimate values of K_{C2}, $\mu_{max\text{-}C2}$, K_{Xy}, and $\mu_{max\text{-}Xy}$ for the DA24-16 strain, the IMSL nonlinear optimizer NNLPF was used to fit batch reactor material balance equations consistent with the conditions employed by Ha et al. (*8*):

$$\frac{dC_2}{dt} = -\left(\frac{1}{Y_{X/C_2}}\right)\left(\frac{\mu_{\max-C_2} C_2}{K_{C_2} + C_2}\right)\left(1 - \frac{[Ethanol]}{K_{i-Eth-X}}\right)X \tag{4.9}$$

$$\frac{dX_y}{dt} = -\left(\frac{1}{Y_{X/Xy}}\right)\left(\frac{\mu_{\max-Xy} Xy}{K_{Xy} + Xy}\right)\left(1 - \frac{[Ethanol]}{K_{i-Eth-X}}\right)X \tag{4.10}$$

Initial cellobiose and xylose concentrations were set at their reported experimental values of 40 g/L. Inhibition of *S. cerevisiae* growth by ethanol was assumed to be independent of substrate, so the previously measured value of $K_{i\text{-}Eth\text{-}X}$ = 87 g/L was employed (*46*); the Monod constants (K_{C1} and K_{C2}) and maximum specific growth rates ($\mu_{max\text{-}C1}$ and $\mu_{max\text{-}C2}$) were treated as adjustable parameters and optimized. The number of available data points was low (7 total), which increased the uncertainty associated with the fitted values.

After $\mu_{max\text{-}Xy}$ was determined, the xylose yield coefficient was estimated from the $q_{max\text{-}Xy}$ value reported for the DA24-16 strain according to the following equation (*8*):

$$q_{\max-Xy} = \frac{\mu_{\max-Xy}}{Y_{X/Xy}} \tag{4.11}$$

4.3.6 Model Assembly and Solution

The model was constructed in Fortran 95 using the Absoft Pro Fortran v11 compiler. The model used the IMSL function DASPG differential and algebraic equation solver, which employs the Petzold-Gear method. To capture hydrolysis and fermentation, 5,365 differential equations and 15 algebraic equations from the original hydrolysis model, modified to include xylose and ethanol inhibition, were supplemented with three differential equations to track cell density, ethanol concentration, and xylose concentration.

4.4. Results and Discussion

4.4.1 The Potential of CDT Yeast Strains to Reduce β-glucosidase Requirements

Previous reports on cellobiose-consuming yeast have asserted that such strains may be able to alleviate cellobiose inhibition of cellulase and thereby obviate the need for exogenously added β-glucosidase enzymes, which account for up to 10% of the total enzyme cost (*8, 48*). The capacity of the cellodextrin transporter (CDT) to remove cellobiose as quickly as extracellular β-glucosidase hydrolyzes it to glucose was investigated in separate SSF reactions (Table 4.2; reactions 1-4).

Table 4.2 Total residence times to achieve target ethanol concentrations within various process scenarios

Run	Process Scenarios	τ_{res} (hr) 40 g/L Ethanol	τ_{res} (hr) 75 g/L Ethanol
1	SSF with CDT-containing yeast; cellulose present	>144	>144
2	SSF with CDT-containing yeast; cellulose and xylose present	>144	>144
2'	SSF with CDT-containing yeast with K_{C1} = 0.14 mM; cellulose and xylose present	>144	>144
3	SSF with non-CDT-containing yeast supplemented with β-glucosidase; cellulose present	21.25	89
4	SSF with non-CDT-containing yeast supplemented with β-glucosidase; cellulose and xylose present	37.5	71
4'	SSF with non-CDT-containing yeast supplemented with β-glucosidase; K_{C1} = 0.14 mM; cellulose and xylose present	26.5	56.75
5	SSF with CDT-containing yeast with no glucose inhibition of xylose and cellobiose transport; cellulose present	>144	>144
6	SSF with CDT-containing yeast with no glucose inhibition of xylose and cellobiose transport; cellulose and xylose present	>144	>144
7	SSF with CDT-containing yeast where K_{C1}=K_{C2}; cellulose present	70.5	>144
8	SSF with CDT-containing yeast where K_{C1}=K_{C2}; cellulose and xylose present	35.5	102.75
9	SSF CDT-containing yeast where $\mu_{max\text{-}C2} = \mu_{max\text{-}C1}$; cellulose present	>144	>144
10	SSF CDT-containing yeast where $\mu_{max\text{-}C2} = \mu_{max\text{-}C1}$; cellulose and xylose present	>144	>144
11	SSF CDT-containing yeast where $K_{C2} = K_{C1}$ and $\mu_{max\text{-}C2} = \mu_{max\text{-}C1}$; cellulose present	63.5	>144
12	SSF CDT-containing yeast where $K_{C2} = K_{C1}$ and $\mu_{max\text{-}C2} = \mu_{max\text{-}C1}$; cellulose and xylose present	34.5	96.75
13	SHF with 12-hour hydrolysis step; xylose added in fermentation	31.5	59.5
13'	SHF with 12-hour hydrolysis step; K_{C1} = 0.14 mM; xylose added in fermentation	28	54.75
14	SHF with 12-hour hydrolysis step; K_{C1} = 0.14 mM; xylose added in hydrolysis	32.75	61.25
15	SHF with 24-hour hydrolysis step; xylose added in fermentation	40.5	67.25
16	SHF with 24-hour hydrolysis step; xylose added in hydrolysis	41.5	68.5
17	SHF with 36-hour hydrolysis step; xylose added in fermentation	50.75	76.5
18	SHF with 36-hour hydrolysis step; xylose added in hydrolysis	51.5	77.75
19	SSF CDT-containing yeast where K_{C2} = 0 mM, $\mu_{max\text{-}C2} = \mu_{max\text{-}C1}$, and B_{C1} = 0; cellulose and xylose	28.5	62.25
19'	SSF CDT-containing yeast where K_{C2} = 0 mM, $\mu_{max\text{-}C2} = \mu_{max\text{-}C1}$, and $B_{C1\text{-}C2}$ = 0; K_{C1} = 0.14 mM ; cellulose and xylose	25	58

Time-course profiles of cellulose conversion into products for each of these conditions are plotted in Fig. 4.2A. With cellulose alone, SSF with the CDT-containing strain resulted in 10% conversion of cellulose in 140 hr. In contrast, SSF with the non-CDT-containing strain supplemented with β–glucosidase achieved nearly 100% conversion in 90 hr. Interestingly, when xylose was added, the CDT-containing strain achieved 25% cellulose conversion in 140 hr; growth on xylose led to higher concentrations of yeast, which facilitated faster overall cellobiose uptake (see Eq. 4.6), reducing its inhibitory effect. The DA24-16 strain does not remove cellobiose as quickly as a non-CDT-containing strain that is supplemented with β-glucosidase, and is therefore less effective in alleviating cellobiose inhibition of cellulase-catalyzed hydrolysis.

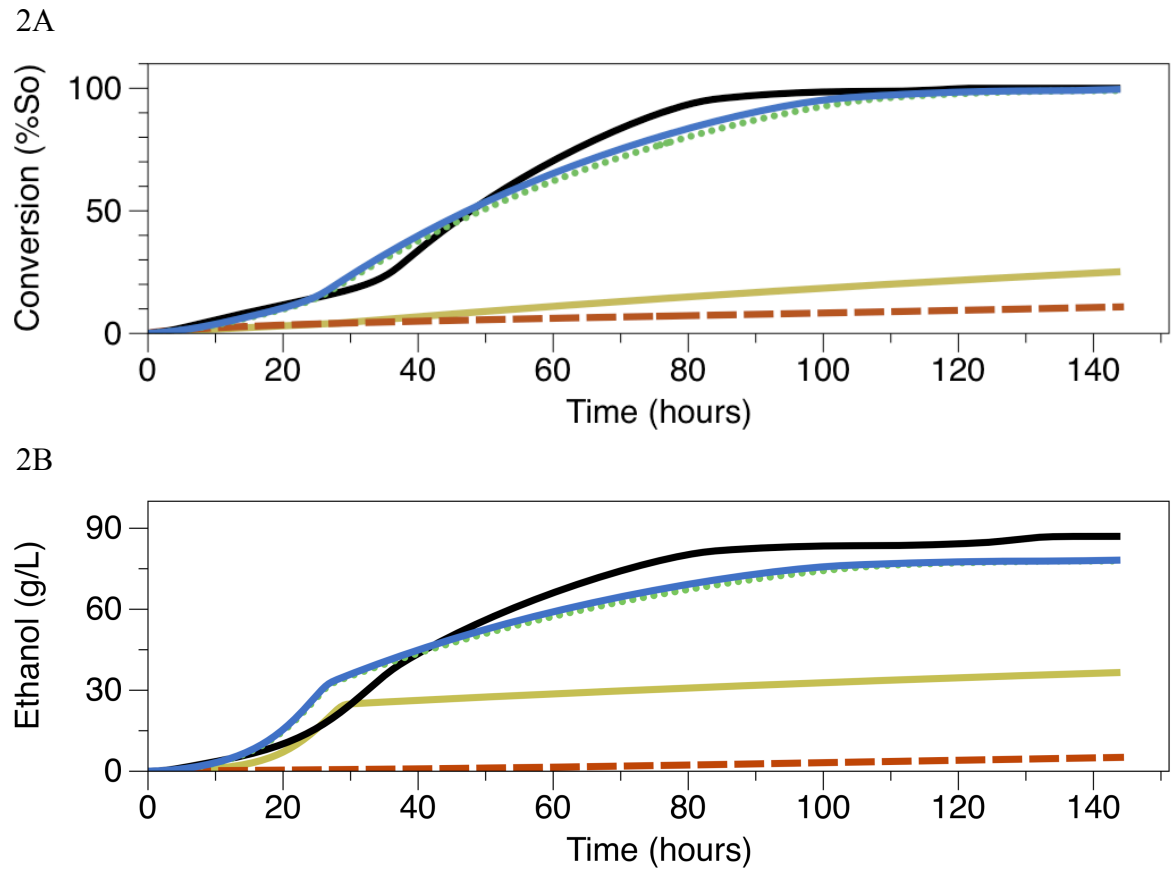


Figure 4.2 Profiles of cellulose hydrolysis and ethanol production in SSF scenarios. Runs were carried out for 144 hours at 30°C with a 1:3:1 EG2:CBH1:CBH2 (molar) enzyme ratio, a combined total enzyme loading of 2.5 g/L, and substrate concentrations of 129 g/L cellulose with or without 71 g/L xylose. When β-glucosidase was present, it comprised 10% of the total enzyme mass. Cultures were started with a yeast inoculum of 0.252 g/L, corresponding to an OD_{600} of approximately 1. Runs are annotated with the following abbreviations: simultaneous saccharification and fermentation (SSF); cellodextrin transporter and intracellular β-glucosidase present (CDT); cellulose present (C); xylose present (X). Lines show time course profiles of (2A) cellulose conversion to products and (2B) ethanol production for the following conditions: SSF CDT C (dashed red line, run 1); SSF CDT C X (solid yellow line, run 2); SSF CDT w/ $K_{C2} = K_{C1}$ C X (dotted green line, run 8); SSF CDT w/ $K_{C2} = K_{C1}$, $\mu_{ma\text{-}C2} = \mu_{ma\text{-}C1}$ C X (solid blue line, run 12); SSF C X BG (solid black line, run 4). Note: lines for runs 9 and 10 (SSF CDT without glucose inhibition of growth cellobiose) are coincident with those from runs 1 and 2, respectively.

Both xylose and cellobiose transporters are inhibited by glucose (*8, 30*). To address the inhibitory effect of glucose on the cellobiose consumption rates in SSF, we examined a situation in which glucose inhibition of cellobiose and xylose uptake by the CDT-containing yeast was removed (runs 5 and 6, Table 4.2). These scenarios brought about only a slight increase in cellulose conversion rates. After 144 hours, SSF with CDT strains uninhibited by glucose inhibition of cellobiose uptake allowed cellulose conversion to reach 11.4% and 25.9% for runs 5 and 6, respectively; this is a 5 and 3 % increase over runs 1 and 2. Clearly, in these hydrolysis scenarios, glucose was consumed nearly as rapidly as it was generated, failing to reach strongly inhibitory levels.

S. cerevisiae strains capable of cellobiose assimilation could be engineered to have faster rates of cellobiose uptake through improvements to the transporter protein or through modifications to the assimilation pathway. Enhancements in cellobiose uptake rates likely to result from such efforts were investigated by varying the kinetic parameters associated with growth of *S. cerevisiae* on cellobiose.

For the DA24-16 strain, the Monod constant for cellobiose (K_{C2}=350 mM) is about 200-fold larger than the analogous constant for glucose (K_{C1}=1.7 mM). The influence of K_{C2} on the ability of cellobiose-utilizing *S. cerevisiae* to alleviate cellobiose inhibition of cellulase enzymes was investigated with SSF runs in which it was set equal to K_{C1} (runs 7 and 8, Table 4.2). This lower value of K_{C2} enabled cellobiose consumption to increase rates of cellulose hydrolysis almost as effectively as the addition of β-glucosidase (Fig. 4.2A).

The impact of increasing the specific growth rate of the CDT-containing yeast in SSF was assessed with simulations where $\mu_{max\text{-}C2}$ was increased to $\mu_{max\text{-}C1}$ (runs 9 and 10, Table 4.2). This change did not allow for faster conversion rates in SSF; conversion profiles for runs 9 and 10 were coincident with those of runs 1 and 2. This maximum specific growth rate is the highest measured for *S. cerevisiae*; it is likely limited by the maximal rate of rRNA synthesis in this yeast, not glucose uptake capacity, and is therefore independent of the nature of the substrate (*49-52*). Higher values of $\mu_{max\text{-}C2}$ were not investigated. As illustrated in Figs. 4.2A and 4.2B, cellulose conversion and ethanol production results for the DA24-16 strain were insensitive to the increase in $\mu_{max\text{-}C2}$.

The influence of a faster specific growth rate was also investigated in simulations where K_{C2} and $\mu_{max\text{-}C2}$ were set equal to K_{C1} and $\mu_{max,C1}$ (runs 11 and 12, Table 4.2). At 72 hours, when compared to run 8 (with the CDT-containing yeast with K_{C2} =K_{C1}), run 12 showed a 4.6% improvement in conversion (Fig. 4.2A) and a 3% improvement in ethanol concentration (Fig. 4.2B). Thus, only when K_{C2} was set to K_{C1} did an increase in the maximum specific growth rate lead to slight increases in hydrolysis rates and ethanol titers.

In practice, it may be possible to reduce K_{C2} by enhancing the affinity of the transporter for cellobiose, by increasing the number of transporters per cell, or by raising V_{max} for the transporter enzyme (*53, 54*). The *N. crassa* transporter incorporated into the DA24-16 strain has high affinity for cellobiose (apparent K_M equal to ~3 μM), which is low compared to other substrate transporters; hence, improving affinity of the transporter for cellobiose is not likely to lower K_{C2} (*8*). Transporter abundance in the membrane or activity are more likely to be effective targets for improvement.

Improvements to downstream enzymes may also lower K_{C2}. After an intracellular hydrolysis step, the cellobiose assimilation pathway is identical to that of glucose assimilation. Thus, the activity of intracellular β-glucosidase could also be a limiting factor in cellobiose consumption.

4.4.2 Impact of Co-utilizing Cellobiose, Xylose, and Glucose on SSF vs. SHF

The impact of cellobiose utilization on SSF and SHF was explored by comparing residence times required to achieve 40 g/L and 75 g/L ethanol titers (Table 4.2). Kinetic parameters for xylose utilization were held constant in all runs (K_{xyl} and $\mu_{max\text{-}xyl}$ of the DA24-16 strain) and growth parameters for CDT-containing and non-CDT-containing yeast were identical. For all SHF runs, the residence times included in Table 4.2 represent the sum of the times required for hydrolysis and fermentation.

Regardless of ethanol titer, the SHF process with the 12-hour hydrolysis step led to the shortest residence times: 31.5 hours for 40 g/L and 59.5 hours for 75 g/L. These times are 8 and 18% shorter than the shortest SSF residence times for the processes considered in Table 4.2.

The rankings of total residence times for the SSF scenarios depended on the ethanol titers that were targeted (Table 4.2). For 40 g/L ethanol, the shortest SSF residence time (34.5 hours) corresponded to the scenario with the best-case CDT-containing strain ($K_{C2} = K_{C1}$ and $\mu_{max\text{-}C2} = \mu_{max\text{-}C1}$). For 75 g/L, however, the SSF process where the non-CDT containing strain was supplemented with β-glucosidase resulted in a residence time of 71 hours compared to 96.75 hours for the best-case CDT-containing strain. These results indicate that the introduction of a CDT transporter into SSF scenarios provides no advantage over SHF.

The residence time required for a fermentation to achieve a particular ethanol titer depends on the concentration of cells used for inoculation (0.252 g/L in this study). We studied the influence of this dependence on the relative performance of different SSF and SHF processes by plotting the residence time required to reach 75 g/L ethanol against inoculant concentration for several scenarios (Fig. 4.3): SSF with the non-CDT yeast supplemented with β-glucosidase (Run 4), SSF with the best-case CDT yeast (Run 12), and SHF with the 12-hour hydrolysis step. The relative performance of the SSF runs was not affected by the inoculum size. This is consistent with the kinetics of the fermentation reaction; the rate of ethanol production is first order in yeast concentration. Thus, rate enhancements resulting from higher inoculant concentrations are identical within different SSF scenarios.

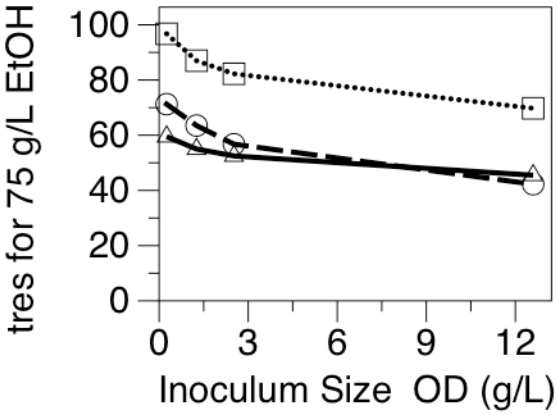


Figure 4.3 Sensitivity of time required to reach 75 g/L ethanol to inoculation size. The influence of inoculation size on the residence time required to produce 75 g/L ethanol was explored with scenarios where the fermentation inoculum was varied from 0.252 g/L (OD_{600} ~ 1) to 12.6 g/L (OD_{600} ~ 50). The following scenarios were investigated: SSF CDT w/ $K_{C2} = K_{C1}$, $\mu_{ma\text{-}C2} = \mu_{ma\text{-}C1}$ C X (open squares, run 12); SSF C X BG (open circles, run 4); SHF with a 12-hour hydrolysis step (open triangles, run 13).

By comparison, for an SHF process with a hydrolysis step of predetermined length (e.g., 12 hours), a larger inoculum size shortens only the duration of the fermentation step; the resulting reduction in overall residence time is less significant than it is within SSF. This effect is apparent in Fig. 4.3. As inoculum size is increased, the residence times required to reach 75 g/L ethanol decrease more rapidly in the SSF processes than in the SHF alternative. In fact, at inoculant concentrations of 12.6 g/L cells (OD_{600} ~ 50), the residence time for the SSF process conducted with a non-CDT yeast supplemented with β-glucosidase was shorter than that required by SHF with a 12-hour hydrolysis step: 42.25 hours compared to 45.5 hours. At this high inoculum size, however, shorter residence times can likely be achieved within SHF if the hydrolysis step is shortened; the duration of this step was not optimized in the present work. Thus, while the relative performance of particular SHF and SSF scenarios is sensitive to the inoculum size, the general result remains: cellobiose-consuming yeast are unlikely to make SSF more efficient than SHF.

4.4.3 The Influence of the Cellodextrin Transporter on Xylose Uptake Rates

The cellodextrin transporter's role in enhancing xylose uptake by preventing glucose inhibition of xylose uptake was investigated through a comparison of the times required for yeast to fully consume xylose in three scenarios: SSF with the best-case CDT-containing strain, SSF with the non-CDT-containing strain supplemented with β-glucosidase, and SHF with a 12-hour hydrolysis step (Table 4.3). For the SHF process, the xylose consumption time was calculated by adding the time required for complete xylose consumption in the fermentation step to the 12 hours required for hydrolysis. In SSF with the best CDT-containing strain, xylose was completely consumed in 27.25 hours; this was significantly shorter than the 38 hours required for SSF with the non-CDT-containing strain supplemented with β-glucosidase and the 37.5 hours required for SHF with a 12-hour hydrolysis step. As discussed above, however, the production of 75 g/L ethanol titers required residence times longer than 50 hours, and all processes allowed for complete xylose utilization within this time. Moreover, we found no correlation between rapid xylose uptake and shortest residence time. Thus, the enhanced xylose consumption rates of

the best-case CDT-containing strain are not beneficial when high concentrations of ethanol are desired.

Table 4.3 Time Taken to Consume Xylose

Run	Conditions	$B_{i\text{-}G\text{-}Xy}$	$t_{cons.}$ for xylose	τ_{res} (hr) 75 g/L Ethanol
13	SHF with 12-hour hydrolysis step; xylose added in fermentation	9 mM	37.5	59.5
4	SSF with non-CDT-containing yeast supplemented with β-glucosidase; cellulose and xylose present	9 mM	38	71
12	SSF CDT-containing yeast where $K_{C1}=K_{C2}$ and $\mu_{max\text{-}C2} = \mu_{max\text{-}C1}$; cellulose and xylose present	9 mM	27.25	96.75
13	SHF with 12-hour hydrolysis step; xylose added in fermentation	90 mM	93.5	94.5
4	SSF with non-CDT-containing yeast supplemented with β-glucosidase; cellulose and xylose present	90 mM	98.75	122
12	SSF CDT-containing yeast where $K_{C1}=K_{C2}$ and $\mu_{max\text{-}C2} = \mu_{max\text{-}C1}$; cellulose and xylose present	90 mM	50	119

Parameters describing glucose inhibition of xylose uptake by the DA24-16 strain have yet to be measured experimentally. The present work employs experimental parameters describing glucose inhibition of xylose transport in the yeast *Candida lusitaniae*, which exhibits glucose growth kinetics similar to those of *S. cerevisiae* (*30*). These parameters may be incorrect for the DA24-16 strain. Thus, the cellodextrin transporter's ability to increase xylose uptake rates was further investigated with simulations where the glucose inhibition parameter ($B_{C1\text{-}Xy}$) was increased 10-fold. In such a scenario, glucose inhibition of xylose uptake is stronger than reported in most yeast strains capable of xylose assimilation (*30, 55, 56*). In the SSF run with this stronger inhibition, the optimized CDT-containing strain consumed all xylose within 50 hours, a much shorter time than the 93.5 and 98.75 hours required for the SHF and SSF alternative processes, respectively. Within SSF, strong glucose inhibition of xylose transport led to conditions where rapid xylose uptake correlated with short residence times; SSF with the best-case CDT yeast allowed for 75 g/L ethanol to be produced more quickly than with the non-CDT yeast. However, this correlative relationship did not hold when SHF and SSF were compared. The SHF scenario with the 12-hour hydrolysis step still had the shortest residence time for a 75 g/L ethanol titer despite the longer time for complete xylose consumption.

The cellodextrin transporter facilitates rapid xylose consumption by alleviating glucose inhibition of xylose transport, but for high ethanol titers, this enhanced xylose utilization rate is only beneficial within SSF if glucose inhibition of xylose uptake is stronger in the DA24-16 strain than in other reported xylose-utilizing strains (*30, 55, 57*). Additionally, when SSF and SHF are compared, the overall residence time for high ethanol titers is not sensitive to rapid xylose uptake, and the SHF scenario allows for shorter residence times despite leading to conditions in which xylose is consumed more slowly.

4.4.4 Sensitivity of the SSF/SHF Comparison to Values of K_{C1} and K_{C2}

In the present work, the Monod constant for growth on glucose (K_{C1}) was assumed to be the lowest achievable value of the Monod constant for growth on cellobiose in an engineered CDT-containing yeast. However, reported values of K_{C1} for yeast growth on glucose range from 0.1 to 3 mM (*29, 30, 56, 58*). In our study, we used an intermediate value of 1.7 mM for K_{C1}. To more completely explore the sensitivity of our results to K_{C2} values that are even lower than 1.7 mM, we examined the residence times required to achieve 75 g/L ethanol in SSF with a CDT-containing strain where K_{C2} was varied from 0.035 mM to 3.5 mM (Fig. 4.4). The lowest estimates of K_{C1} for yeast are around 0.14 mM; the residence time for this value of K_{C2} is illustrated in Fig. 4.4 (starred point). SSF with such a dramatically improved CDT yeast allowed for a shorter residence time than SSF with a non-CDT yeast supplemented with β-glucosidase (71 hours; dashed gray line in Fig. 4.4). In all runs, however, the residence times were longer than the corresponding time for the SHF process with a 12-hour hydrolysis step (59.5 hours; black line in Fig. 4.4). Reductions in residence time from improvements to K_{C2} beyond 1.7 mM were minimal and appeared to reach a limit near K_{C2} = 0.035 mM (65 hours). Thus, while performance of the CDT-containing yeast in SSF scenarios is highly sensitive to K_{C2}, the result that SHF with a 12-hour hydrolysis step leads to shorter residence times at high ethanol titers is not.

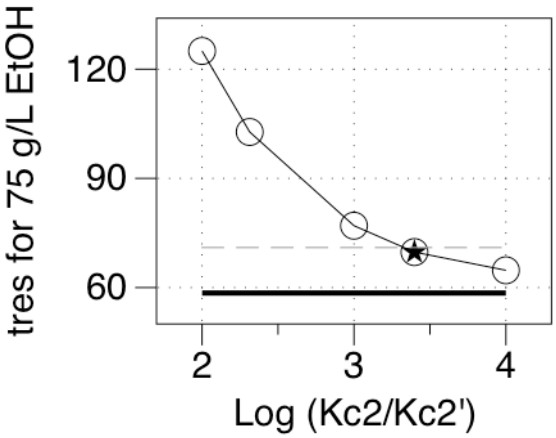


Figure 4.4 Sensitivity of time required to reach 75 g/L ethanol in SSF to K_{C2}. The influence of K_{C2} values on the residence time required to produce 75 g/L ethanol in SSF with a CDT yeast were explored. Points represent residence times for various reductions in the Monod constant associated with growth on cellobiose, which are plotted along the x-axis as $\log(K_{C2}/K_{C2}')$ where K_{C2} = 350 mM and K_{C2}' is an improved (lower) value of K_{C2}. The lines correspond to the residence times required to reach 75 g/L of ethanol in SHF with a 12-hour hydrolysis step (run 13; solid black line) and SSF with a non-CDT yeast supplemented with β-glucosidase (dashed gray line).

The possible influence of a lower K_{C1} on our results was also explored (Table 4.2). When this K_{C1} was lowered to 0.14 mM in runs 2, 4, and 13, the relative ordering of residence times was unchanged (2', 4', and 13' in Table 4.2). When K_{C2} was subsequently lowered (as above), the same result was observed. The SHF process with a 12-hour hydrolysis step still allowed 75 g/L ethanol to be produced more quickly than all other processes.

4.4.5 Sensitivity of the SSF/SHF Comparison to Glucose, Cellobiose, and Xylose Inhibition

The potential advantages of SSF and SHF may be sensitive to inhibition of cellulase enzymes by glucose, cellobiose, and xylose. Xylose inhibition in particular is not well studied; there are no directly measured inhibition parameters describing the impact of this sugar on cellulase enzymes. The influence of xylose inhibition was explored through a comparison of cellulose conversion profiles generated by three processes simulated with low and high xylose inhibition parameters: SSF with the CDT-containing yeast, SSF with the best CDT-containing yeast, and SHF with a 12-hour hydrolysis step (runs 2, 12, and 13 in Table 4.2).

Parameter values describing the competitive inhibition of cellulase enzymes by xylose ($K_{i\text{-}Xy}$) were estimated from published measurements of cellulase inhibition by other sugars. In the simulations discussed in 3.1-3.5, equilibrium constants for competitive xylose inhibition were set equal to measured equilibrium constants for glucose inhibition. Reported measurements of glucose inhibition, however, are inconsistent. For example, the values of $K_{i\text{-}C1}$ employed in the simulations discussed in 3.1 and 3.2 ranged from 17-31 mM, but a study by Phillippidis et al. reported a $K_{i\text{-}C1}$ value of 295 mM for a similar mixture of *Trichoderma sp.* enzymes (*32*). The implications of this range of values on the results of this work were investigated with runs where glucose and xylose inhibition parameters were set equal to this higher value ($K_{i\text{-}Xy} = K_{i\text{-}C1} = 295$ mM). These runs represent a low xylose inhibition scenario. A high xylose inhibition scenario was also simulated. Xylan hydrolyzate is not necessarily composed entirely of xylose, but can be dominated by xylo-oligomers. These longer DP sugars have been shown to be as inhibitory of cellulase enzymes as cellobiose (*5*). To investigate the implications of xylan hydrolyzate that is as inhibitory of cellulase enzymes as cellobiose, competitive equilibrium constants for xylose inhibition were equated with the analogous constants for cellobiose inhibition ($K_{i\text{-}Xy} = K_{i\text{-}C2}$ for each enzyme). Cellulose conversion profiles over 72 hours are reported in Fig. 4.5.

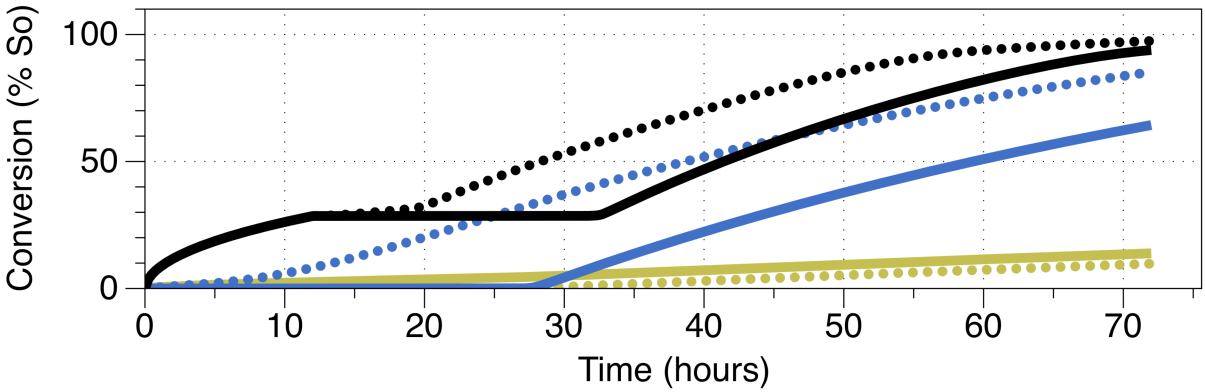


Figure 4.5 Sensitivity of the SSF vs. SHF comparison to product inhibition. The influence of higher or lower levels of xylose inhibition was explored through a comparison of SSF with the CDT-containing yeast (run 2), SSF with the best-case CDT-containing yeast (run 12), and SHF with a 12-hr hydrolysis step (run 13). Low inhibition (low Xi) corresponds to runs in which $K_{I=xyl}=K_{i\text{-}C1}=295$ mM. High inhibition (high Xi) corresponds to runs in which $K_{i\text{-}Xy}=K_{i\text{-}C2}$ for each enzyme. Lines correspond to conversion profiles for the following conditions: SSF CDT C X high Xi (solid yellow line); SSF CDT w/ $K_{C2} = K_{C1}$, $\mu_{ma\text{-}C2}= \mu_{ma\text{-}C1}$ high Xi (solid blue line); 12-H SHF C high Xi (solid black line); SSF CDT C X low Xi (dotted yellow line); SSF CDT w/ $K_{C2} = K_{C1}$, $\mu_{ma\text{-}C2}= \mu_{ma\text{-}C1}$ low Xi (dotted blue line); 12-H SHF C low Xi (dotted black line).

The cellulose conversion profile for the CDT-containing yeast in the SSF scenario was not significantly affected by xylose inhibition parameters. At reduced levels of xylose and glucose

inhibition, the SHF process resulted in 15% higher conversion at 72 hours than the SSF process with the best CDT-containing yeast; at the higher level of xylose inhibition, the SHF process led to a 25% higher conversion. Thus, SHF outperformed SSF at low and high levels of xylose inhibition, indicating that the results of this work are insensitive to the values of the parameters used to describe xylose inhibition.

4.4.6 A Best-case Scenario for Cellobiose Consumption

To assess the ability of a CDT-containing yeast to achieve high ethanol titers, we ran the model with $K_{C2} = 0$, $B_{C1\text{-}C2} = 0$, and $\mu_{max\text{-}C2} = 0.5\ hr^{-1}$ (Run 19 in Table 4.2). Under such circumstances, cellobiose uptake can be represented by the following equation:

$$\text{rate of cellobiose uptake} = \left(\frac{1}{Y_{X/C_2}}\right)\left(\mu_{\max-C2}\right)\left(1-\frac{[Ethanol]}{K_{i-Eth-X}}\right)X \qquad (12)$$

Here, uptake of cellobiose (and growth on cellobiose) is governed only by the maximum specific growth rate of *S. cerevisiae* and inhibition by ethanol, and is zeroth order in cellobiose. For such a strain, the residence time required to achieve 75 g/L of ethanol was 62.25 hours, which is still 4.6% higher than the residence time for SHF with a 12-hour hydrolysis step. The same result was obtained in the set of runs where $K_{C1} = 0.14$ mM (run 19' in Table 4.2). Accordingly, improvements to the cellobiose assimilation pathway are not likely to lead to a strain of yeast that allows high ethanol titers to be reached more quickly in SSF than in SHF.

4.4.7 Collection of Parameter Values

Some of the parameter values for the microbial growth model employed in the present analysis were not available in the literature and had to be estimated from limited datasets and non-*Saccharomyces* yeast strains. As discussed above, the conclusions of the present work are not sensitive to these parameters, but the model-predicted performance of any particular strain is. Future work to engineer yeast strains capable of xylose and cellobiose assimilation should involve attempts to accurately measure kinetic parameters for growth on these substrates ($K_{substrate}$, $\mu_{max\text{-}substrate}$, $B_{\text{-}C1\text{-}substrate}$).

4.5 Concluding Remarks

We combined a detailed mechanistic kinetic model of cellulase-catalyzed hydrolysis with a multi-substrate model of microbial growth to investigate the ability of cellobiose utilizing *S. cerevisiae* to facilitate cellulase hydrolysis and improve ethanol titers using SSF. The model showed that an existing strain with this capability (DA24-16) is not as effective at enhancing rates of cellulose hydrolysis or ethanol production as a non-cellobiose-utilizing alternative supplemented with β-glucosidase. However, lowering the Monod constant for growth on cellobiose would enable DA24-16 to produce ethanol from cellulose and xylose nearly as rapidly as the non-cellobiose-utilizing alternative. Strategies to lower the Monod constant include increasing transporter expression and/or increasing the rate of intracellular cellobiose hydrolysis. Such a best-case CDT strain would likely reduce, but not eliminate, the β-glucosidase requirement.

By relieving glucose inhibition of xylose uptake, the cellodextrin transporter allowed xylose to be consumed more quickly in CDT yeast than in non-CDT yeast with the same xylose assimilation pathway machinery. Rapid xylose uptake, however, did not result in short residence times for producing 75 g/L ethanol unless glucose inhibition of xylose uptake was assumed to be higher in the DA24-16 strain than for most other xylose-consuming yeast strains; and, under such circumstances, this correlation existed only in SSF scenarios. The SHF scenario with the 12-hour hydrolysis step required more time for complete xylose utilization but allowed high ethanol titers to be reached more quickly than in SSF processes, regardless of the level of glucose inhibition of xylose uptake.

Improvements to the cellobiose utilization system of the DA24-16 strain are unlikely to result in SSF residence times that are shorter than those of separate hydrolysis and fermentation (SHF) processes. Our simulations showed that SHF with a 12-hour hydrolysis step allowed 75 g/L ethanol titers to be reached more quickly than all SSF processes addressed in this work. A technoeconomic analysis is necessary to determine whether or not the longer residence times required in SSF with optimized CDT-containing strains of *S. cerevisiae* could allow for a net cost savings resulting from the elimination or reduction in β–glucosidase requirements.

In an SHF process, improved cellobiose-utilizing strains could still be useful. If cellulases that are insensitive to cellobiose are developed, there would be no need for β – glucosidase in the hydrolysis step; and enzyme cost would be reduced. In this scenario, any organism capable of rapid cellobiose consumption would be advantageous in the fermentation step, which would contain a high concentration of cellobiose.

4.6 References

1. L. R. Lynd, J. H. Cushman, R. J. Nichols, C. E. Wyman, Fuel Ethanol from Cellulosic Biomass. *Science* 251, 1318 (Mar 15, 1991).
2. L. R. Lynd, W. H. van Zyl, J. E. McBride, M. Laser, Consolidated bioprocessing of cellulosic biomass: an update. *Current Opinion in Biotechnology* 16, 577 (Oct, 2005).
3. C. E. Wyman, N. D. Hinman, Ethanol - Fundamentals of Production from Renewable Feedstocks and Use as a Transportation Fuel. *Appl Biochem Biotech* 24-5, 735 (Spr-Sum, 1990).
4. A. Carroll, C. Somerville, Cellulosic biofuels. *Annual review of plant biology* 60, 165 (2009).
5. Q. Qing, B. Yang, C. E. Wyman, Xylooligomers are strong inhibitors of cellulose hydrolysis by enzymes. *Bioresource Technol* 101, 9624 (Dec, 2010).
6. I. C. Roberto, S. I. Mussatto, R. C. L. B. Rodrigues, Dilute-acid hydrolysis for optimization of xylose recovery from rice straw in a semi-pilot reactor. *Ind Crop Prod* 17, 171 (May, 2003).
7. M. Holtzapple, M. Cognata, Y. Shu, C. Hendrickson, Inhibition of Trichoderma-Reesei Cellulase by Sugars and Solvents. *Biotechnol Bioeng* 36, 275 (Jul, 1990).
8. S. J. Ha *et al.*, Engineered Saccharomyces cerevisiae capable of simultaneous cellobiose and xylose fermentation. *P Natl Acad Sci USA* 108, 504 (Jan 11, 2011).
9. C. F. Wahlbom, R. R. C. Otero, W. H. van Zyl, B. Hahn-Hagerdal, L. J. Jonsson, Molecular analysis of a Saccharomyces cerevisiae mutant with improved ability to utilize xylose shows enhanced expression of proteins involved in transport, initial xylose metabolism, and the pentose phosphate pathway. *Appl Environ Microb* 69, 740 (Feb, 2003).
10. D. R. J. Grootjen, R. G. J. M. Vanderlans, K. C. A. M. Luyben, Conversion of Glucose Xylose Mixtures by Pichia-Stipitis under Oxygen-Limited Conditions. *Enzyme Microb Tech* 13, 648 (Aug, 1991).

11. M. Gardonyi, M. Jeppsson, G. Liden, M. F. Gorwa-Grausland, B. Hahn-Hagerdal, Control of xylose consumption by xylose transport in recombinant Saccharomyces cerevisiae. *Biotechnol Bioeng* 82, 818 (Jun 30, 2003).
12. S. Saitoh, T. Tanaka, A. Kondo, Breeding of Industrial Diploid Yeast Strain with Chromosomal Integration of Multiple beta-Glucosidase Genes. *J Biosci Bioeng* 106, 594 (Dec, 2008).
13. Y. Shen *et al.*, Simultaneous saccharification and fermentation of acid-pretreated corncobs with a recombinant Saccharomyces cerevisiae expressing beta-glucosidase. *Bioresource Technol* 99, 5099 (Jul, 2008).
14. M. I. Rajoka *et al.*, Cloning and expression of beta-glucosidase genes in Escherichia coli and Saccharomyces cerevisiae using shuttle vector pYES 2.0. *Folia Microbiol* 43, 129 (1998).
15. J. M. Galazka *et al.*, Cellodextrin Transport in Yeast for Improved Biofuel Production. *Science* 330, 84 (Oct 1, 2010).
16. D. Humbird, A. Mohagheghi, N. Dowe, D. J. Schell, Economic impact of total solids loading on enzymatic hydrolysis of dilute acid pretreated corn stover. *Biotechnol Prog* 26, 1245 (Sep-Oct, 2010).
17. S. E. Levine, J. M. Fox, H. W. Blanch, D. S. Clark, A Mechanistic Model of the Enzymatic Hydrolysis of Cellulose. *Biotechnology and Bioengineering* 107, 37 (2010).
18. C. N. Hamelinck, B. van Hooijdonk, A. P. Faaij, Ethanol from lignocellulosic biomass: techno-economic performance in short-, middle-, and long-term. *Biomass & Bioenergy* 28, 384 (2005).
19. W. Lan, C. F. Liu, R. C. Sun, Fractionation of bagasse into cellulose, hemicelluloses, and lignin with ionic liquid treatment followed by alkaline extraction. *J Agric Food Chem* 59, 8691 (Aug 24, 2011).
20. J. E. Morinelly, J. R. Jensen, M. Browne, T. B. Co, D. R. Shonnard, Kinetic Characterization of Xylose Monomer and Oligomer Concentrations during Dilute Acid Pretreatment of Lignocellulosic Biomass from Forests and Switchgrass. *Ind Eng Chem Res* 48, 9877 (Nov 18, 2009).
21. G. Banerjee *et al.*, Synthetic enzyme mixtures for biomass deconstruction: production and optimization of a core set. *Biotechnol Bioeng* 106, 707 (Aug 1, 2010).
22. O. Akpinar, O. Levent, S. Bostanci, U. Bakir, L. Yilmaz, The optimization of dilute acid hydrolysis of cotton stalk in xylose production. *Appl Biochem Biotechnol* 163, 313 (Jan, 2011).
23. S. E. Levine, J. M. Fox, H. W. Blanch, D. S. Clark, A Mechanistic Model for Rational Design of Optimal Cellulase Mixtures. *Biotechnol Bioeng*, (2011).
24. A. Koivula *et al.*, The Active Site of Cellobiohydrolase Cel6A from Trichoderma reesei: The Roles of Aspartic Acids D221 and D175. *Journal of the American Chemical Society* 124, 10015 (2002).
25. B. Nidetzky, W. Zachariae, G. Gercken, M. Hayn, W. Steiner, Hydrolysis of Cellooligosaccharides by Trichoderma reesei Cellobiohydrolases: Experimental Data and Kinetic Modeling. *Enzyme and Microbial Technology* 16, 43 (1994).
26. J. M. Fox, S. E. Levine, D. S. Clark, H. W. Blanch, Initial- and Processive-Cut Products Reveal Cellobiohydrolase Rate Limitations and the Role of Companion Enzymes. *Biochemistry-Us*, (Dec 14, 2011).
27. J. Karlsson, M. Siika-aho, M. Tenkanen, F. Tjerneld, Enzymatic Properties of the Low Molecular Mass Endoglucanases Cel12A (EG III) and Cel45A (EG V) of Trichoderma reesei. *Journal of Biotechnology* 99, 63 (2002).
28. R. E. T. Drissen *et al.*, A Generic Model for Glucose Production from Various Cellulose Sources by a Commerical Cellulase Complex. *Biocatalysis and Biotransformation* 25, 419 (November-December, 2007).
29. H. W. Blanch, D. S. Clark, *Biochemical engineering*. (M. Dekker, New York, 1997), pp. xii, 702 p.
30. Y. J. Kwon, C. R. Engler, Kinetic Models for Growth and Product Formation on Multiple Substrates. *Biotechnology and Bioprocess Engineering* 10, 6 (2005).

31. R. D. Tyagi, T. K. Ghose, Batch and Multistage Continuous Ethanol Fermentation of Cellulose Hydrolysate and Optimum Design of Fermenter by Graphical Analysis. *Biotechnol Bioeng* 22, 1907 (1980).
32. G. P. Philippidis, T. K. Smith, C. E. Wyman, Study of the Enzymatic-Hydrolysis of Cellulose for Production of Fuel Ethanol by the Simultaneous Saccharification and Fermentation Process. *Biotechnol Bioeng* 41, 846 (Apr 15, 1993).
33. M. Chauve *et al.*, Comparative kinetic analysis of two fungal beta-glucosidases. *Biotechnology for Biofuels* 3, (Feb 11, 2010).
34. K. L. Kadam, E. C. Rydholm, J. D. McMillan, Development and validation of a kinetic model for enzymatic saccharification of lignocellulosic biomass. *Biotechnol Progr* 20, 698 (May-Jun, 2004).
35. Y. Chen, B. Dong, W. Qin, D. Xiao, Xylose and cellulose fractionation from corncob with three different strategies and separate fermentation of them to bioethanol. *Bioresour Technol* 101, 7005 (Sep, 2010).
36. K. K. Podkaminer, X. J. Shao, D. A. Hogsett, L. R. Lynd, Enzyme Inactivation by Ethanol and Development of a Kinetic Model for Thermophilic Simultaneous Saccharification and Fermentation at 50 degrees C with Thermoanaerobacterium saccharolyticum ALK2. *Biotechnol Bioeng* 108, 1268 (Jun, 2011).
37. J. M. van Zyl, E. van Rensburg, W. H. van Zyl, T. M. Harms, L. R. Lynd, A Kinetic Model for Simultaneous Saccharification and Fermentation of Avicel With Saccharomyces cerevisiae. *Biotechnol Bioeng* 108, 924 (Apr, 2011).
38. Y. Sun, J. Y. Cheng, Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresource Technol* 83, 1 (May, 2002).
39. J. Houghton, S. Weatherwax, J. Ferrel, Breaking the Biological Barriers to Cellulosic Ethanol: A Joint Research Agenda. *US Dep Energy*, (Dec 7-9, 2005, 2006).
40. J. O. Baker, C. I. Ehrman, W. S. Adney, S. R. Thomas, M. E. Himmel, Hydrolysis of cellulose using ternary mixtures of purified celluloses. *Appl Biochem Biotech* 70-2, 395 (Spr, 1998).
41. in *Genencor A Danisco Division* G. A. D. Division, Ed. (2011).
42. M. T. Holtzapple, H. S. Caram, A. E. Humphrey, The Hch-1 Model of Enzymatic Cellulose Hydrolysis. *Biotechnol Bioeng* 26, 775 (1984).
43. M. Okazaki, M. Mooyoung, Kinetics of Enzymatic-Hydrolysis of Cellulose - Analytical Description of a Mechanistic Model. *Biotechnol Bioeng* 20, 637 (1978).
44. S. Peri, S. Karra, Y. Y. Lee, M. N. Karim, Modeling intrinsic kinetics of enzymatic cellulose hydrolysis. *Biotechnol Progr* 23, 626 (May-Jun, 2007).
45. S. Wald, C. R. Wilke, H. W. Blanch, Kinetics of the Enzymatic-Hydrolysis of Cellulose. *Biotechnol Bioeng* 26, 221 (1984).
46. T. K. Ghose, R. D. Tyagi, Rapid Ethanol Fermentation of Cellulose Hydrolysate .1. Batch Versus Continuous Systems. *Biotechnol Bioeng* 21, 1387 (1979).
47. G. Stephanopoulos, Challenges in engineering microbes for biofuels production. *Science* 315, 801 (Feb 9, 2007).
48. J. M. Galazka, J. H. Cate, A new diet for yeast to improve biofuel production. *Bioeng Bugs* 2, (Jul 1, 2011).
49. J. R. Warner, The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci* 24, 437 (Nov, 1999).
50. S. L. French, Y. N. Osheim, F. Cioci, M. Nomura, A. L. Beyer, In exponentially growing Saccharomyces cerevisiae cells, rRNA synthesis is determined by the summed RNA polymerase I loading rate rather than by the number of active genes. *Mol Cell Biol* 23, 1558 (Mar, 2003).
51. D. R. Kief, J. R. Warner, Coordinate Control of Syntheses of Ribosomal Ribonucleic-Acid and Ribosomal-Proteins during Nutritional Shift-up in Saccharomyces-Cerevisiae. *Mol Cell Biol* 1, 1007 (1981).
52. K. Elbing *et al.*, Role of hexose transport in control of glycolytic flux in Saccharomyces cerevisiae. *Appl Environ Microbiol* 70, 5323 (Sep, 2004).

53. D. K. Button, B. Robertson, E. Gustafson, X. Zhao, Experimental and theoretical bases of specific affinity, a cytoarchitecture-based formulation of nutrient collection proposed to supercede the Michaelis-Menten paradigm of microbial kinetics. *Appl Environ Microbiol* 70, 5511 (Sep, 2004).
54. D. K. Button, Kinetics of nutrient-limited transport and microbial growth. *Microbiol Rev* 49, 270 (Sep, 1985).
55. B. Hahn-Hagerdal *et al.*, Metabolic engineering of Saccharomyces cerevisiae for xylose utilization. *Adv Biochem Eng Biotechnol* 73, 53 (2001).
56. R. Morales-Rodriguez, K. V. Gernaey, A. S. Meyer, G. Sin, A Mathematical Model for Simultaneous Saccharification and Co-ferentation (SSCF) of C6 and C5 Sugars. *Chinese Journal of Chemical Engineering* 19, 185 (2010).
57. A. Saloheimo *et al.*, Xylose transport studies with xylose-utilizing Saccharomyces cerevisiae strains expressing heterologous and homologous permeases. *Appl Microbiol Biotechnol* 74, 1041 (Apr, 2007).
58. E. Postma, W. A. Scheffers, J. P. van Dijken, Kinetics of growth and glucose transport in glucose-limited chemostat cultures of Saccharomyces cerevisiae CBS 8066. *Yeast* 5, 159 (May-Jun, 1989).

Chapter 5: Initial- and Processive-cut Products Reveal Cellobiohydrolase Rate Limitations and Role of Companion Enzymes

5.1 Abstract

Efforts to improve the activity of cellulases, which catalyze the hydrolysis of insoluble cellulose, have been hindered by uncertainty surrounding the mechanistic origins of rate-limiting phenomena and by an incomplete understanding of complementary enzyme function. In particular, direct kinetic measurements of individual steps occurring after enzymes adsorb to the cellulose surface have proven to be experimentally elusive. The present work describes an experimental and analytical approach, derived from a detailed mechanistic model of cellobiohydrolase action, to determine rates of initial- and processive-cut product generation by *Trichoderma longibrachiatum* cellobiohydrolase I (*Tl*Cel7A) as it catalyzes the hydrolysis of bacterial microcrystalline cellulose (BMCC) alone and in the presence of *Talaromyces emersonii* endoglucanase II (*Tem*GH5). This analysis revealed that the rate of *Tl*Cel7A-catalyzed hydrolysis of crystalline cellulose is limited by the rate of enzyme complexation with glycan chains, which is shown to be equivalent to the rate of initial-cut product generation. This rate is enhanced in the presence of endoglucanase enzymes. The results confirm recent reports of the role of morphological obstacles on enzyme processivity, and also provide the first direct evidence that processive length may be increased by the presence of companion enzymes, including small amounts of *Tem*GH5. The findings of this work indicate that efforts to improve cellobiohydrolase activity should focus on enhancing the enzyme's ability to complex with cellulose chains, and the analysis employed provides a new technique for investigating the mechanism by which companion enzymes influence cellobiohydrolase activity.

5.2 Introductory Notes

5.2.1 Abbreviations

*Tem*GH5, *Talaromyces emersonii* GH5 endoglucanase; *Tl*Cel7A, *Trichoderma longibrachiatum* cellobiohydrolase Cel7A; G_i, cello-oligosaccharide of length i.

5.2.2 Textual Footnotes

***** This chapter, with few modifications from its original format, represents the following peer-reviewed publication: J. M. Fox, S. E. Levine, D. S. Clark, H. W. Blanch, Initial- and processive-cut products reveal cellobiohydrolase rate limitations and the role of companion enzymes. *Biochemistry* **51**, 442 (Jan 10, 2012).

**Initial rates (averages of the first three hours of hydrolysis) are normalized by *Tl*Cl7A concentration to facilitate comparison with the initial-cut experiments of *Tl*Cel7A on cellohexaose, where a lower concentration of enzyme had to be used to facilitate initial-rate measurements. Rates reported here correspond to reactions 4A-4E in Table 5.1, where [*Tl*Cel7A] is held constant but [*Tem*GH5] is increased. These rates are 30-300 times slower than initial rates of initial-cut product generation on cellohexaose (3 μM/ sec / μM *Tl*Cel7A).

***Estimates of processive length (n_{pr}) for each enzyme mixture. These estimates are based on the ratio of concentrations of processive-cut products to initial-cut products for a processive enzyme; this ratio levels off to the processive length of that enzyme. Figure 5.2 confirms this behavior. Values of processive length (n_{pr}) are estimated at 110 hours (the 110 hour time points from Fig. 5.2). The dependence of processive length on concentrations of *Tl*Cel7A and *Tem*GH5 enzymes is addressed in the discussion.

5.3 Introduction

Cellobiohydrolases comprise a major class of cellulase enzymes, which catalyze the hydrolysis of insoluble cellulose into soluble sugars. In Nature, they are the largest single component of the multi-cellulase mixtures secreted by cellulolytic fungi (*1*); in industry, they are a common target of enzyme engineering strategies to develop more efficient saccharification systems for cellulosic biofuels production (*2*). Efforts to understand and improve cellobiohydrolase activity in cellulolytic mixtures have been hindered by the experimental difficulties presented by the heterogeneous biocatalytic systems in which they participate (*3-5*).

Once adsorbed to the cellulose surface, cellobiohydrolases participate in a series of reaction steps that are difficult to study in isolation (*6-8*). Experimental evidence of a particular rate-limiting kinetic step has therefore remained elusive. In this work, we employed an experimental approach, based on a detailed mechanistic model of cellobiohydrolase action, to identify the rate-limitations imposed upon *Trichoderma longibrachiatum* cellobiohydrolase I (*Tl*Cel7A) as it catalyzes the hydrolysis of bacterial microcrystalline cellulose (BMCC) alone and in the presence of *Talaromyces emersonii* endoglucanase II (*Tem*GH5).

Cellobiohydrolases (CBH; EC. 3.2.1.91) and endoglucanases (EG; EC. 3.2.1.4) represent the two broad cellulase classes. Both have active sites consisting of two glutamate and/or aspartate residues positioned to promote general acid catalyzed hydrolysis of the β-1,4-bonds that link the repeating cellobiose subunits of cellulose chains (*9-11*). The distance between these catalytic residues (5.5 or 10 Å) determines whether hydrolysis results in a retention or inversion of the anomeric carbon's configuration. In cellobiohydrolases, which hydrolyze cellulose processively from either the reducing or non-reducing chain ends, the active site is located within a tunnel that can accommodate 6-10 glucosyl units (*12, 13*). In endoglucanases, which catalyze hydrolysis within cellulose chains, it is located within a 3-5 residue binding site cleft (*14*). In both cases, the catalytic domain is linked via a 6-109 residue polypeptide linker to a smaller cellulose binding module (CBM), which facilitates adsorption to insoluble cellulosic substrates (*15*). Endoglucanases sometimes lack a CBM, but cellobiohydrolases with such architectures have not been found.

The *Tl*Cel7A and *Tem*GH5 enzymes used in this study are thermostable examples of GH7 and GH5 cellulases. *Tl*Cel7A is a retaining cellobiohydrolase with a 10-residue binding site tunnel that facilitates processive hydrolysis from the reducing ends of cellulose chains (*12*). *Tem*GH5, which has been less extensively studied, is a retaining endoglucanase that appears to lack a CBM domain and binds cellulose with a binding site cleft that allows it to catalyze hydrolysis from within cellulose chains (*16, 17*).

Figure 1

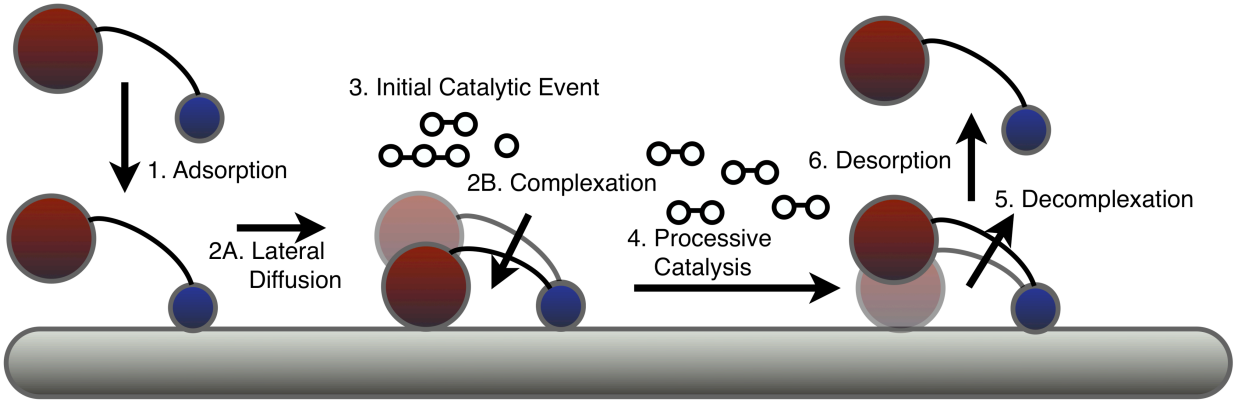


Figure 5.1 A general conceptual picture of the reaction path undertaken by cellobiohydrolases as they hydrolyze insoluble cellulose. Steps for endoglucanases are similar, but they do not engage in processive catalysis.

Possible reaction paths undertaken by these enzymes as they hydrolyze insoluble, heterogeneous cellulosic materials have been proposed in the literature (Fig. 5.1) (*3*). A cellobiohydrolase (1) adsorbs to the crystalline regions of the cellulose surface via its CBM domain (*18-20*); (2) it diffuses along the surface and complexes with a cellulose chain end via its catalytic domain (*21*); (3) it catalyzes an initial hydrolysis event, generating glucose, cellobiose, or cellotriose (G_1, G_2, and G_3) (*22-24*); (4) it catalyzes subsequent hydrolysis events, generating only G_2, as it processively moves along the chain (*25, 26*); (5) it decomplexes from the cellulose chain (*27*); (6) it recomplexes with a cellulose chain or desorbs from the surface. Endoglucanases undergo similar steps, but they are generally thought to engage in a non-processive random attack of less crystalline regions with single chains exposed. These enzymes generate both soluble and insoluble products, depending on the length the polymer they are hydrolyzing and the position of the hydrolyzed glycosidic linkage being broken.

Over the past 30 years, numerous mechanistic models describing the kinetics of cellulase-catalyzed hydrolysis of cellulose have been developed to determine the rate limitations of enzymatic hydrolysis (*3, 4, 28-33*). Often relying on simplified representations of cellulose (including those of a soluble polymer, a bulk phase substrate, a multi-component mixture, a periodic flat surface), these studies have been limited in their ability to determine kinetic limitations arising form the nature of the cellulose surface (surface diffusion, chain abstraction, processivity, and decomplexation) (*29-31, 33*). Almost all of these studies conclude that more experimental insights into surface reactions are required. Experimental efforts have been made to measure the rates of steps 1, 4, and 5 for Cel7A; however, conclusions about the sensitivity of overall hydrolysis kinetics to these steps have been difficult to draw in the absence of rate data for steps 2 and 3 (*25, 27, 34, 35*).

Initial- and processive-cut *Tl*Cel7A product profiles can be employed to reveal the kinetic limitations imposed by complexation of an adsorbed cellobiohydrolase with chain ends and by subsequent processive catalysis. When compared to initial-rate measurements on soluble cello-oligosaccharide chain ends, initial-cut product profiles from crystalline hydrolysis can reveal kinetic impediments associated with adsorption and complexation. Together with initial-cut time courses, processive-cut profiles can reveal limitations associated with processivity.

Endoglucanase enzymes engage in complementary hydrolysis activities that reveal the limitations of initial and processive modes of cellobiohydrolase catalysis. Endoglucanases increase the concentration of free chain ends on the substrate surface and, by acting preferentially on partially-exposed cellulose chains, may decrease the amount of solution-exposed pseudo-amorphous regions (*36*). The response of initial- and processive-cut *Tl*Cel7A product profiles to the presence of increasing concentrations of endoglucanase enzymes can reveal not only the surface features that limit *Tl*Cel7A complexation and processivity, but also the kinetic limitations imposed by them.

We used measurements of cello-oligosaccharide production from bacterial microcrystalline cellulose (BMCC) obtained with either (1) identical concentrations of *Tl*Cel7A and different concentrations of *Tem*GH5 or (2) different concentrations of *Tl*Cel7A acting alone to calculate time-course profiles for the rates of initial and processive cellobiohydrolase hydrolysis events. We show that initial-cut products are reporters of the complexation of adsorbed cellobiohydrolases with glycan chains and that ratios of processive- to initial-cut products at later times serve as a measure of processive length. Profiles of initial-cut products, processive-cut products, and the ratio of the two permit assessment of the extent to which the overall hydrolysis process is limited by the reaction steps associated with an adsorbed cellobiohydrolase's engagement with cellulose chain ends and by those associated with its subsequent processive hydrolysis. This represents the first experimental attempt to directly study (1) cellobiohydrolase complexation kinetics and their impact on overall rates of cellulose hydrolysis and (2) the influence of endoglucanase enzymes on cellobiohydrolase complexation kinetics and on processive length.

5.4 Materials and Methods

5.4.1 Enzyme Purification

Both *Trichoderma longibrachiatum* Cellobiohydrolase I (*Tl*Cel7A) and *Talaromyces emersonii* EG2 (*Tem*GH5) were purchased from Megazyme and purified to single band purity as described in SI Materials and Methods. A 5-hour assay of each enzyme against 10 g/L carboxymethyl cellulose ensured *Tl*Cel7A purity. At 0.27 μM, *Tem*GH5 hydrolyzed ~ 50% of the substrate in 5 hours, but *Tl*Cel7A had no detectable activity against it, indicating no endoglucanase contamination of *Tl*Cel7A (SI Materials and Methods).

5.4.2 Cellulose Preparation

Pretreatment effects were minimized in the preparation employed here by the exclusion of steps involving strongly acidic or basic solutions.

A 20 μL freezer stock of *Acetobacter xylinum* (ATCC 53582) was used to inoculate 1 L of media as described previously (*37*). This inoculum was grown in a 2 L Erlenmeyer flask for 7 days at 25°C in a shaker at 250 RPM); the resulting cellulose was purified as described in SI Materials and Methods. The final cellulose concentration was measured with the phenol sulfuric acid method (*37, 38*).

5.4.3 Measurement of *Tl*Cel7A Initial-cut Product Distribution

The first-cut product distribution generated by *Tl*Cel7A after binding to cellohexaose was measured with initial rate experiments. Sugar production was monitored for two minutes in 100 µL reactions containing the following: 0.01 µM *Tl*Cel7A, 1390 µM cellohexaose (G_6), and 50 mM sodium acetate buffer (pH 4.8). The G_6 solution (with buffer) was incubated for 5 minutes in a 96-Well PCR plate at 50°C before 4.2 µL of the enzyme was added to bring the total volume to 100 µL. Reactions were stopped at 0, 20, 40, 60, 80, 100, and 120 seconds with 100 µL of 0.100 M NaOH solution. Concentrations of glucose (G_1), cellobiose (G_2), cellotriose (G_3), cellotetraose (G_4), and cellopentaose (G_5) were measured with a Dionex ICS-3000 equipped with a CarboPac PA200 column and a Dionex Electrochemical Cell detection unit. All reactions were carried out in triplicate.

5.4.4 Measurement of Cellotriose Hydrolysis Rates

The Michaelis-Menten kinetic parameters for the activity of *Tl*Cel7A and *Tem*GH5 on cellotriose were calculated with nonlinear regression of initial rate data (Datagraph, Figs. A2*A* and S2*B*). Hydrolysis was monitored in 100 µL reactions with 0.051 µM *Tl*Cel7A or 0.074 µM *Tem*GH5 in 50 mM sodium acetate (pH 4.8) buffer (50°C). Initial cellotriose concentrations of 10.13, 50.65, 202.6, 506.5, and 919.804 µM were used for *Tl*Cel7A and 7.61, 50.03, 194.25, 476.02, 865.07, and 3323.6 µM for *Tem*GH5. Reactions were stopped after 1, 2, 3, and 4 hours with 100 µL of 0.10 M NaOH, and glucose production was measured on the Dionex HPLC.

5.4.5 Measurement of Cello-oligosaccharide Production in Various Mixtures

Cello-oligosaccharide production by defined mixtures of *Tl*Cel7A and *Tem*GH5 acting on BMCC was monitored discretely over the course of 120 hours. In humidified shakers at 50°C (200 RPM), reactions were carried out in 250 ml Erlenmeyer flasks containing 25.25 ml of 1 g/L BMCC, 50 mM sodium acetate buffer (pH 4.8), 0.01 % sodium azide, and enzyme concentrations as described in Table 5.1. Samples were taken from 1, 3A-3C, and 4A-4E at the following times: 0, 1.17, 2.22, 3.42, 4.6, 5.93, 7.8, 13.15, 16.42, 20.87, 37.47, 45, 59.63, 68.33, 75.6, 86.08, 97.73, 110.4, and 120.6 hours. Samples were taken from 2A-2C at the following times: 0, 1.75, 6.3, 16.5, 21.4, 26.9, 32.3, 49, 98.1, and 121.1 hours. At all time points, two 20 µL aliquots were withdrawn individually frozen and diluted with 0.05 M NaOH before being run on a Dionex. All reactions were carried out in duplicate.

Table 5.1 Enzyme Loadings in Various Mixtures

Reaction #	*[Ce7A] (µM)*	*[GH5] (µM)*	*Reaction #*	*[Cel7A] (µM)*	*[GH5] (µM)*
1	*0*	*0*	*3B*	*0.152*	*0*
2A	*0*	*0.022*	*3C*	*0.305*	*0*
2B	*0*	*0.037*	*4A*	*0.076*	*0*
2C	*0*	*0.19*	*4B*	*0.076*	*0.0187*
2D	*0*	*0.37*	*4C*	*0.076*	*0.0935*
2E	*0*	*0.74*	*4D*	*0.076*	*0.187*
3A	*0.015*	*0*	*4E*	*0.076*	*0.374*

5.4.6 Analysis of Kinetic Data

Prior to detailed kinetic analysis, measurements likely to be erroneous were detected with a MATLAB program written for this study. When a single measurement brought the standard error associated with an average of four measurements of the same phenomenon above 10%, that measurement was removed from the data analyzed. Of the total dataset, 19 measurements were removed; these represented ~2% of the total measurements. Bias resulting from the removal of this small percentage of data is unlikely to affect the results of this work.

5.4.7 Derivation of Equations for η and f

The development of equations for η, the fraction of enzymes that are not directly engaged in hydrolyzing the substrate, is described in Section 5.9. Derivations of equations for *f*, the fraction of cellotriose generated by *Tem*GH5 within *Tem*GH5-*Tl*Cel7A mixtures, are detailed in Section 5.11.

5.5 Results

5.5.1 Determination of *Tl*Cel7A Initial- and Processive-cut Product Generation Rates

The differences in product distributions generated by cellobiohydrolases hydrolyzing cellulose processively and non-processively permit the experimental assessment of the kinetic limitations of binding to, processing along, and dissociating from cellulose chains. Upon complexing with a glycan chain, *Tl*Cel7A generates G_1, G_2, and G_3 in a fixed ratio. The rate of *Tl*Cel7A initial-cut product (icut product) formation can therefore be written in terms of the rate of G_3 generation from chain ends:

$$\frac{d\left[\text{icut product}\right]_{Tl\text{Cel7A}}}{dt} = \left(\frac{d\left[G_3\right]_{Tl\text{Cel7A icut}}}{dt}\right)\cdot\left(1+\alpha+\beta\right) \tag{5.1}$$

where

$$\alpha = \frac{\dfrac{d\left[G_1\right]_{Tl\text{Cel7A icut}}}{dt}}{\dfrac{d\left[G_3\right]_{Tl\text{Cel7A icut}}}{dt}}, \quad \beta = \frac{\dfrac{d\left[G_2\right]_{Tl\text{Cel7A icut}}}{dt}}{\dfrac{d\left[G_3\right]_{Tl\text{Cel7A icut}}}{dt}}$$

A material balance for G_3 can be used to express the right side of Eq. 5.1 in terms of measureable quantities. Both the initial cuts of *Tl*Cel7A and the "random attack" of *Tem*GH5 on BMCC generate G_3, and both enzymes can catalyze its hydrolysis. The following equation for G_3 production takes these processes into account:

$$\frac{d\left[G_3\right]_{\text{observed}}}{dt} = \left(\frac{d\left[G_3\right]_{Tl\text{Cel7A icut}}}{dt} + \frac{d\left[G_3\right]_{Tem\text{GH5 BMCC}}}{dt} - \frac{d\left[G_3\right]_{G3\text{ hyd}}}{dt}\right) \tag{5.2}$$

Eqs. 5.1 and 5.2 can be combined to yield

$$\frac{d\left[\text{icut product}\right]_{Tl\text{Cel7A}}}{dt} = (1+\alpha+\beta)\cdot\left(\frac{d\left[G_3\right]_{\text{obs}}}{dt}+\frac{d\left[G_3\right]_{G\,3\text{ hyd}}}{dt}-\frac{d\left[G_3\right]_{Tem\text{GH5 BMCC}}}{dt}\right) \quad (5.3)$$

In Eqs. 5.2 and 5.3, soluble cello-oligosaccharide chains consisting of more than three glucosyl units are not considered as separate sources of G_3 production. While cellotetraose (G_4), cellopentaose (G_5), and cellohexaose (G_6) are slowly generated by the action of *Tem*GH5 on cellulose, they are quickly hydrolyzed by the same enzyme; thus, their production is included in the last term of either equation. Cello-oligosaccharide chains with more than 6 glucose subunits are insoluble.

While retaining glycosyl hydrolase enzymes have been shown to be capable of catalyzing transglycosylation, this activity has been measured to be minor in comparison with their hydrolytic activities by Harjunpaa et al, Gusakov et al, and Vřanská et al (*24, 39, 40*). Thus, as in kinetic models presented by Okazaki et al, Bezzera et al, Praestgaard et al, and Nidetzky et al, transglycosylation by *Tl*Cel7A and *Tem*GH5 was neglected in this analysis (*23, 29, 41, 42*)

If the initial concentrations of all species are assumed to be zero, Eq. 5.3 can be integrated to obtain an expression for the concentration of initial-cut products that have been generated at any given time:

$$\left[\text{icut product}\right]_{Tl\text{Cel7A}} = (1+\alpha+\beta)\cdot\left(\left[G_3\right]_{obs}+\left[G_3\right]_{G\,3\text{ hyd}}-\left[G_3\right]_{Tem\text{GH5 BMCC}}\right) \quad (5.4)$$

Each term on the right hand side represents some discretely measureable or estimable quantity.

An expression for the concentration of processive-cut products at any given time can be derived in a fashion identical to that described above (Eq. 5.5):

$$\left[\text{pcut product}\right]_{\text{Tr CBH1}} = \left(\left[G_2\right]_{\text{obs}}-\left[G_2\right]_{G\,3\text{hyd}}-\left[G_2\right]_{Tem\text{GH5 BMCC}}-\left[G_2\right]_{Tl\text{Cel7A icuts}}\right) \quad (5.5)$$

5.5.2 Measurement of Initial-cut Product Distribution by *Tl*Cel7A on Chain Ends

The kinetic analysis employed here relies on previous reports that cellobiohydrolase enzymes generate glucose, cellobiose, and cellotriose as initial-cut products and that the initial-cut product distribution is the same for both soluble and solid substrates. The first assumption is well supported by existing literature (*22-24, 43*). The second is based on a hypothesis that the factors influencing cellobiohydrolase product formation patterns result from substrate binding within the active site tunnel, rather than outside of it. Previous efforts to investigate processivity of cellobiohydrolases and chitinases suggest that initial-cut product distributions result from the tendency of the terminal glucosyl unit to occupy the +1, +2, and +3 sites located after the catalytic residues (*8, 43*). For *Trichoderma reesei* Cel7A, crystallographic data have revealed that the interaction between enzyme and bound glycan is determined (1) by an extensive network of direct and indirect hydrogen bonds lining the entire tunnel and (2) by four tryptophan residues responsible for

the -7, -4, -2, and +1 binding sites (*9, 12*). These data reveal an active site located seven binding sites into the enzyme, meaning that the dominant environment of a cello-oligosaccharide chain prior to being cleaved is the enzyme's interior. As the molecular determinants of binding are located inside the enzyme, the initial-cut product distribution that they influence should not depend on whether the substrate was soluble or abstracted from a surface. In this work, we have assumed this to be the case, and that substrate morphology is more likely to affect the rate at which initial cuts are generated and the rate with which chains are fed into the tunnel than the pattern of their cleavage once they get there. The limitations of this assumption are addressed below.

The initial cut (G_1-G_3) product distribution for *Tl*Cel7A acting on insoluble substrate chain ends was determined from the initial rates of cellohexaose hydrolysis (G_6). A high cellohexaose concentration (~100xK_m), a low enzyme/substrate loading, and short reaction times were used to ensure that *Tl*Cel7A operated at a V_{max} corresponding to its rate of catalysis at chain ends. Measured rates of G_1 and G_2 production were observed to be almost equivalent to the respective rates of G_5 and G_4 production (Table 5.2; Fig. 5.A2), suggesting that these products correspond only to initial cuts (*23, 24*).

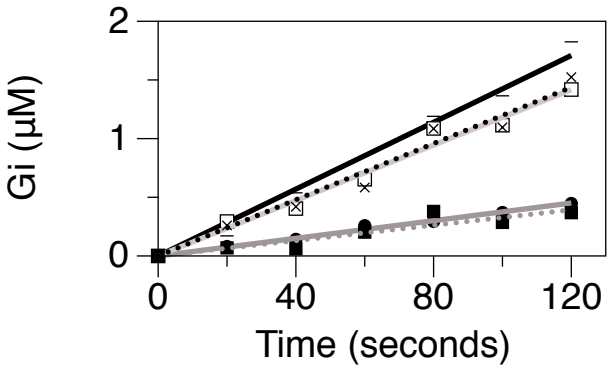


Fig. 5.A2. Initial cut products generated by *TlCel7A* Activity on G_6. Cellooligosaccharides with degrees of polymerization that add to 6 are produced at nearly identical rates. In this figure, sugars are labeled as follows: G_1 (--), G_2 (□), G_3 (□)G_4 (n) G_5

Table 5.2 Kinetic Parameters

Enzyme	*Substrate*	*Parameter*	*Value*	*Source*
TlCel7A	G_6	$(d[G_1]/dt)_i$	*0.0142 μM/s*	*Measured*
TlCel7A	G_6	$(d[G_2]/dt)_i$	*0.0038 μM/s*	*Measured*
TlCel7A	G_6	$(d[G_3]/dt)_i$	*0.0118 μM/s*	*Measured*
TlCel7A	G_6	$(d[G_4]/dt)_i$	*0.0033 μM/s*	*Measured*
TlCel7A	G_6	$(d[G_5]/dt)_i$	*0.012 μM/s*	*Measured*
TlCel7A	G_6	$k_{catTlCel7A}$	*4.5 1/s*	*Measured**
TlCel7A	G_6	$K_{MTlCel7A}$	*14.9 μM*	*Measured**
TlCel7A	G_3	$k_{catTlCel7A}$	*58.1* h^{-1}	*Measured*
TlCel7A	G_3	$K_{MTlCel7A}$	*107.3 μM*	*Measured*
TlCel7A	G_{3-6}	K_{IG1}	*10,000 μM*	*[3]*
TlCel7A	G_{2-6}	K_{IG2}	*100 μM*	*[31-33]*
TlCel7A	G_6	α	*2.41*	*Measured*
TlCel7A	G_6	β	*0.64*	*Measured*
TlCel7A	*BMCC*	ρ	*2.7*	*Estimated*
TlCel7A	*BMCC*	K'_{eq}	*0.592 μM*	*Measured*
TemGH5	G_3	$k_{catTemGH5}$	*910* h^{-1}	*Measured*
TemGH5	G_3	$K_{MTemGH5}$	*960 μM*	*Measured*
TemGH5	G_{3-6}	K_{IG1}	*10,000*	*[3]*
TemGH5	G_{3-6}	K_{IG2}	*500*	*[30]*
TemGH5	*BMCC*	ρ	*1*	*Estimated*
TemGH5	*BMCC*	K'_{eq}	*0.088 μM*	*Measured*

*These values were measured from initial rate experiments as described in Section 5.13.4.

5.5.3 Cellotriose Hydrolysis by *Tl*Cel7A and *Tem*GH5

Cellotriose hydrolysis by both enzymes was observed to follow Michaelis-Menten behavior modified to include competitive inhibition by G_1 and G_2 (Figs. 5.A2A and 5.A2B) (*44*):

$$\frac{d[G_3]_{hyd}}{dt} = \frac{-k_{catTlCel7A} \cdot \eta_{TlCel7A} \cdot [Tl\text{Cel7A}]_{total} \cdot [G_3]}{[G_3] + K_{MTlCel7A} \cdot \left(1 + \frac{[G_1]}{K_{IG1TlCel7A}} + \frac{[G_2]}{K_{IG2TlCel7A}}\right)}$$

$$\frac{d[G_3]_{hyd}}{dt} = \frac{-k_{catTemGH5} \cdot \eta_{TemGH5} \cdot [Tem\text{EG2}]_{total} \cdot [G_3]}{[G_3] + K_{MTemGH5} \cdot \left(1 + \frac{[G_1]}{K_{IG1TemGH5}} + \frac{[G_2]}{K_{IG2TemGH5}}\right)} \qquad (5.6 \text{ and } 5.7)$$

where η_{CBH} and η_{EG} represent the fractions of each enzyme that are free to interact with soluble substrates (η= [enzyme free in solution or adsorbed to the cellulose but not complexed with cellulose chains]/[total enzyme]), i.e., uncomplexed with cellulose chains. Previous work has shown that such enzyme speciation exists (*3, 27, 45*). All relevant kinetic constants from Eqs. 5.6 and 5.7 were measured from initial rate experiments or estimated from the literature (*3, 46-49*) (Table 5.2).

Values for η_{CBH} and η_{EG} were estimated from constitutive equations based on three assumptions: (1) cellulases can move to complexation sites only after becoming adsorbed to cellulose (complexation and adsorption are not coincident); (2) cellulase adsorption and complexation are in equilibrium (*50*); and (3) the total enzyme concentration and total number of surface

sites (adsorption sites plus complexation sites) are assumed to be within an order of magnitude of each other (*27, 34, 51*). An equation for η_{CBH} based on these assumptions takes the form (Appendix 1, Section 5.9):

$$\eta_{Tl\text{Cel7A}} = \rho_{Tl\text{Cel7A}} \cdot \frac{[Tl\text{Cel7A}]_{total}}{K'_{eqTl\text{Cel7A}} + [Tl\text{Cel7A}]_{total}} \qquad (5.8)$$

where ρ_{CBH} is a proportionality constant that can be estimated from data provided in Kurasin et al (*27*), and K'_{eqCBH} is the apparent equilibrium dissociation constant for complexed enzymes (Appendices 1 and 2, Sections 5.9 and 5.10). The equation for η_{EG} is identical in form. Values of ρ and K'_{eq} for each enzyme are included in Table 5.2.

For each mixture described in Table 5.1, rates of G_3 hydrolysis at 18 different time points were calculated from soluble cello-oligosaccharide measurements and the parameters described above using Eqs. 5.6 and 5.7 (Fig. 5.A3A). Fits to $d[G_3]_{G3hyd}/dt$ vs. t data were then integrated to yield time course profiles for G_3 hydrolysis (Fig. 5.A3B).

5.5.4 Estimation of Cellotriose Production by *Tem*GH5

Cellotriose resulting from the action of *Tem*GH5 on BMCC was estimated as a fraction *f* of the total G_3 produced in each reaction:

$$[G_3]_{Tem\text{GH5 BMCC}} = f \cdot ([G_{3\ total}] + [G_{3\ hyd}])$$
$$[G_2]_{Tem\text{GH5 BMCC}} = \lambda \cdot f \cdot ([G_{3\ total}] + [G_{3\ hyd}]) \qquad (5.9 \text{ and } 5.10)$$

where

$$\lambda = \frac{[G_2]_{Tem\text{GH5 BMCC}}}{[G_3]_{Tem\text{GH5 BMCC}}}$$

In solutions containing *Tem*GH5 alone, λ did not vary significantly with time (Fig. 5.A5). This observation is consistent with the "random attack" mechanism proposed for this enzyme, where it cleaves randomly within chains, generating soluble products either by cleaving near chain ends or by generating multiple cuts near the initial cut. Even if the degree of polymerization (DP) on the surface decreases throughout hydrolysis, a change in the pattern of glucose, cellobiose, or cellotriose generated from terminal glycosidic linkages is unlikely to result from a change in the number of such linkages on the surface. If these products are generated by multiple cuts close together, their ratio will also be independent of DP.

Values of *f* are difficult to measure, but upper and lower bounds can be estimated from Eqs. 5.11 and 5.12:

$$f(t) = \frac{\text{Total } G_3 \text{ by } \textit{Tem}\text{GH5 alone}}{\text{Total } G_3 \text{ by } \textit{Tem}\text{GH5 and } \textit{Tl}\text{Cel7A in mix}}$$

$$= \frac{\left[G_3\right]_{\text{tot } \textit{Tem}\text{GH5 i}}}{\left[G_3\right]_{\text{total } \textit{Tem}\text{GH5 i}} + \left[G_3\right]_{\text{tot } \textit{Tr}\text{Cel7A i}}}$$

$$f(t) = \frac{\left[G_3\right]_{\text{total by } \textit{Tem}\text{GH5 i}}}{1.4 \cdot \left(\left[G_3\right]_{\text{total } \textit{Tem}\text{GH5 i}} + \left[G_3\right]_{\text{total } \textit{Tl}\text{Cel7A i}}\right)} \qquad (5.11 \text{ and } 5.12)$$

Equation 5.11 overestimates *f* by relying on an underestimate of the total amount of G_3 generated by a two-enzyme mixture (the denominator); *Tl*Cel7A and *Tem*GH5 are assumed to act in concert as they would in isolation (subscript i = isolation; Appendix 3, Section 5.11). Equation 5.12 underestimates *f*; in this equation, the concerted action of both enzymes is assumed to lead to an increase in G_3 production that is identical to the average increase in overall hydrolysis commonly reported when these enzymes are combined (Appendix 3, Section 5.11) (*52*). This increase is often described by the ratio defined by Eq. 13:

$$R = \frac{\left[G_{eq}\right]_{\text{mix}}}{\left[G_{eq}\right]_{\textit{Tl}\text{Cel7A}} + \left[G_{eq}\right]_{\textit{Tem}\text{GH5}}} \qquad (5.13)$$

where G_{eq} is the glucose equivalent of soluble products formed. In the reactions employed in this work, the average value of this ratio was 1.4 (Appendix 3, Section 5.11). Averages of upper and lower bounds of *f*(t) were used to evaluate Eqs. 5.9 and 5.10 at each time point.

5.6 Discussion

5.6.1 Complexation Limits Hydrolysis

The significance of rates of initial-cut product generation from solid substrates can be explored through a comparison to rates on soluble substrates (Table 5.3). With 1.39 mM cellohexaose, a soluble substrate, *Tl*Cel7A generated initial-cut products at a rate of 3 μM/sec/μM *Tl*Cel7A. This cellohexaose concentration is about 100 times the K_M of *Tl*Cel7A for this substrate; the rate observed, therefore, is dependent only on the enzyme concentration and the intrinsic rate of catalysis (k_{cat}). On 1 g/L BMCC, *Tl*Cel7A generated initial-cut products at initial rates of 0.01 to 0.1 μM/sec/μM *Tl*Cel7A enzyme (Table 5.3); these rates are 30-300 times slower than those on the soluble substrate. Unlike rates with cellohexaose as substrate, however, the insoluble substrate rates are not necessarily measured where [chain ends] >> $K_{M\text{-surface}}$ (mole/surface area). To determine the maximum rates of hydrolysis where the chain end concentration was not limiting, *Tem*GH5 endoglucanase was added to provide additional cellulose chain ends for *Tl*Cel7A action. In Table 5.3, the rate at which initial-cut *Tl*Cel7A products are generated from cellulose appears to reach a point at which it is independent of chain end surface concentration. The last two data points indicate that when the endoglucanase concentration is doubled from 0.185 μM to 0.37 μM, the initial rate of initial-cut product generation by *Tl*Cel7A does not double, increasing instead from 0.093 to 0.107 μM/ sec / μM *Tl*Cel7A. While more endoglucanse might enhance that rate, this small rate increase suggests that, in the presence of excess surface chains ([chain ends] >> K_M), the rate is no longer dependent on the availability of chain ends.

The observed values on solid substrates, even with excess surface chains available, are still an order of magnitude lower than the analogous rate on soluble substrates.

Table 5.3 Initial Rates of Initial-cut Product Generation**

Step	*Substrate Conc.*	*[GH5]*	*icut Product Rate µM/sec/µM TlCel7A*
icut on BMCC	*1 g/L BMCC*	*0*	*0.015*
icut on BMCC	*1 g/L BMCC*	*0.019*	*0.021*
icut on BMCC	*1 g/L BMCC*	*0.094*	*0.044*
icut on BMCC	*1 g/L BMCC*	*0.187*	*0.093*
icut on BMCC	*1 g/L BMCC*	*0.374*	*0.108*
icut on C_6	*1.39 mM C_6*	*n/a*	*3.0*

Initial-cut products appear to be reporters for the complexation of *Tl*Cel7A enzymes with chain ends in the solid substrate. In accordance with the aforementioned rate disparities, the intrinsic rate of hydrolysis, which is at least as fast as the rate of initial-cut product generation from soluble cello-oligosaccharides, does not govern the rate at which initial-cut products are formed from crystalline cellulose. Instead, the cellobiohydrolases seem to be impeded by the physical processes associated with their ability to adsorb to the substrate, diffuse along the cellulose surface, and form a complex with polymer chain ends. As adsorption has been shown to reach equilibrium within one hour, the last two steps, referred to as complexation for this discussion, appear to determine the rate of initial-cut product formation (*53*).

When compared with recently measured rates of *Trichoderma* Cel7A processivity, slow cellulose chain association rates appear to limit overall hydrolysis. *In situ* observations of *Tl*Cel7A cellobiohydrolases made with fast-scan AFM show processive speeds of ~3.5 +/- 1 nm/sec Igarashi et al. (*25*). This speed is approximately equal to the k_{cat} reported for this cellobiohydrolase multiplied by the length of a cellobiose unit (~1 nm), and it amounts to a specific activity of ~ 3.5 µM/sec /µM processive *Tl*Cel7A when cellobiose is the dominant product. Rates of processive catalysis measured by Igarashi et al. are nearly identical to the rates of initial-cut product generation from soluble substrates measured in the present work (3 µM/sec /µM *Tl*Cel7A); both appear to be governed by the intrinsic kinetics of the enzyme (k_{cat}). Thus, the kinetics of cellulose hydrolysis are primarily governed by the rates at which cellobiohydrolases become processive, not by the rates at which they process.

These results shed some light on the kinetic limitations imposed by the rate with which processive enzymes decomplex from cellulose chains. If enzymes that become "unstuck" quickly re-engage in processive hydrolysis, then the rate at which they disengage with chain ends could limit overall hydrolysis kinetics. This argument has been put forth in the literature (*27*). We find, however, that, early on, the ability of *Tl*Cel7A enzymes to initiate catalysis at chain ends is limited by the availability of those ends. Faster rates of decomplexation are only certain to increase rates of overall hydrolysis if complexation is fast, and we find that it is slow.

5.6.2 Surface Morphology Limits Processivity

The presence of additional physical and kinetic impediments to processive cellobiohydrolase action was evidenced in time course profiles showing the ratios of processive to initial cuts (p/i ratio, Fig. 5.2*C*). The average number of processive events resulting from a single initial catalytic event is defined as the processive length (n_{pr}). For each cellulase mixture, the p/i ratio approaches a value of n_{pr} at later times. The p/i ratios at 110 hours ($\sim n_{pr}$) can be used as an estimate of processive length (Table 5.4); processive length values reported here are on the same order of magnitude as those reported in other studies of *Tl*Cel7A on BMCC (*27, 54*). As these values are two orders of magnitude lower than estimates of the intrinsic processive length of the *Trichoderma* Cel7A enzyme on this substrate (1000-4000), processive cellobiohydrolases are likely hindered by physical obstructions or decomplexation-encouraging surface features far more frequently than they are stopped by the intrinsic kinetics of disengagement from cellulose chains. These findings are consistent with previous work showing that processive length depends on the nature of the substrate (*27*).

Table 5.4 Estimates of Processive Length (n_{pr}) for Each Mixture[***]

[Cel7A] (μM)	*[GH5] (μM)*	*R (p/i) 110hr ~n_{pr}*	*X*	*[Cel7A] (μM)*	*[GH5] (μM)*	*R (p/i) 110hr ~n_{pr}*	*X*
0.076	*0*	*13*	*18*	*0.015*	*0*	*8*	*2*
0.076	*0.0187*	*16*	*26*	*0.076*	*0*	*13*	*32*
0.076	*0.0935*	*12*	*47*	*0.152*	*0*	*18*	*57*
0.076	*0.187*	*9*	*53*	*0.305*	*0*	*21*	
0.076	*0.374*	*4*	*60*				

R (p/i) = p/i ratio
X = % conversion at 110hr

The present results also evidence the influence of endoglucanase enzymes on processive length (Table 5.4). Mixtures with 0.076 μM of *Tl*Cel7A and 0.0187 of *Tem*GH5 permit longer or equivalent processive lengths than the reaction with 0.076 μM of *Tl*Cel7A alone. The failure of low *Tem*GH5 concentrations to reduce processive length by creating cuts in the surface is counterintuitive. By preferentially hydrolyzing exposed, pseudo-amorphous cellulose chains on the cellulose surface, the regions upon which endoglucanases are most active, low concentrations of these enzymes may remove impediments, increasing the length over which processive cellobiohydrolase catalysis can occur. Other authors have provided evidence of the influence of surface morphology on cellobiohydrolase processive length, but the current data provide evidence that morphological changes generated by low concentrations of endoglucanase enzymes may influence processive length (*27, 45, 55*). The mechanism of these endoglucanase-influenced structural changes is not known.

At concentrations above 0.0935 μM, *Tem*GH5 began to restrict the processive length of *Tl*Cel7A (Fig. 5.2*C*). This may result from a shortening of the average length of exposed cellulose strands or an enzyme crowding effect at the surface. As the latter has not previously been observed at the cellulase/cellulose loadings used in this study, the former is more likely (*3, 56, 57*). At high concentrations, *Tem*GH5 may generate cuts within the crystalline regions to which *Tl*Cel7A binds, thus reducing the length over which *Tl*Cel7A can freely process. Mechanisms

proposed here, however, are only speculative. In the end, the failure of low endoglucanase concentrations to rapidly decrease processive length deserves further experimental investigation.

5.6.3 Cellobiohydrolase Enzymes Potentiate Their Own Activity by Influencing Surface Morphology

In reactions with *Tl*Cel7A acting alone, the processive length increased with *Tl*Cel7A loading (Fig. 5.2D; Table 5.4), suggesting that these enzymes may affect the morphology of the cellulose in a manner that increases its susceptibility to processive catalysis. Bacterial microcrystalline cellulose is predominantly in the I_α crystalline phase, which consists of cellulose chains aligned in parallel (*58*). Previous work has shown that the *Tl*Cel7A cellobiohydrolase enzymes engage in unidirectional processive catalysis on the hydrophobic face of the crystal (*25, 58, 59*). We speculate that cellobiohydrolases may be able to potentiate their own activity by clearing obstructions associated with imperfections in the packing of intertwined or closely aligned chains.

The change in processive length with cellobiohydrolase loading may also evidence the minor endo-like activity that has been reported for cellobiohydrolases (27). The present work reveals that small amounts of endoglucanase can increase processive length. The minor endo-like activity of cellobiohydrolase enzymes may function in a similar fashion; as the cellobiohydrolase concentration is increased, this endo-like activity, while still minor, may increase and lead to increased processive length in a manner similar to that of small amounts of endoglucanase. Previous work has shown that modification or removal of the four surface loops that form the active site tunnel of Cel7A can open up its active site up, increasing its endo-like character(*49, 60*). A processive- and initial-cut analysis similar to that used in the present work but carried out with endo-like variants of Cel7A may reveal the influence of endo activity on the ability of cellobiohydrolase enzymes to carry out processive catalysis.

5.6.4 The Sensitivity of This Approach to Estimates Employed

The present results showing that complexation is slow are unlikely to change if the initial-cut product distribution for *Tl*Cel7A is different on soluble and insoluble substrates. The product distribution is embedded in the $(1+\alpha+\beta)$ coefficient of Eq. 5.3, which describes initial-cut products generated from insoluble cellulose. The $(1+\alpha+\beta)$ term represents the total number of initial-cut products generated each time cellotriose is generated from an initial cut. For cellohexaose, the value of $(1+\alpha+\beta)$ was found to be 4. Initial rates of initial-cut product generation on solid and soluble substrates were shown to differ by an order of magnitude when the initial-cut product distribution was assumed to be the same ($(1+\alpha+\beta) = 4$) on solid and soluble substrates. For initial rates of initial-cut product generation to be equivalent on solid and soluble substrates, the value of $(1+\alpha+\beta)$ would have to be ~40 for cellulose. In such a case, the processive length corresponding to the reaction with 0.305 μM *Tl*Cel7A acting alone, the longest in this study, would be 2. This is an order of magnitude smaller than any processive length reported for this enzyme, indicating that a value of 40 for $(1+\alpha+\beta)$ on solid cellulose is unrealistic(*27*). Thus, even if initial-cut product distributions on solid and soluble substrates are different, they are unlikely to be different enough to invalidate the result that complexation is rate limiting.

The slow complexation result is similarly unaffected by the processive-cut product distribution. Previous studies have shown that cellobiose is the dominant product of processive catalysis (*22, 23*). In the present work, cellobiose is treated as the only product of processive cataly-

sis. If glucose and cellotriose are generated by processing cellobiohydrolase enzymes, the analysis employed within this work will lead to an overestimate for initial rates of initial-cut product generation from soluble substrates. That is, they will be even slower than reported in Table 5.4.

Processive length trends are insensitive to the initial-cut product distribution. The product distribution is incorporated in the $(1+\alpha+\beta)$ coefficient in the $[G_2]_{Tl\text{Cel7A icuts}}$ term of Eqn. 5.5. Different initial-cut product distributions can increase or decrease the value of these terms, leading to higher or lower estimates of initial- and processive-cuts within all reactions. Processive length trends within and between reactions, however, are unaffected.

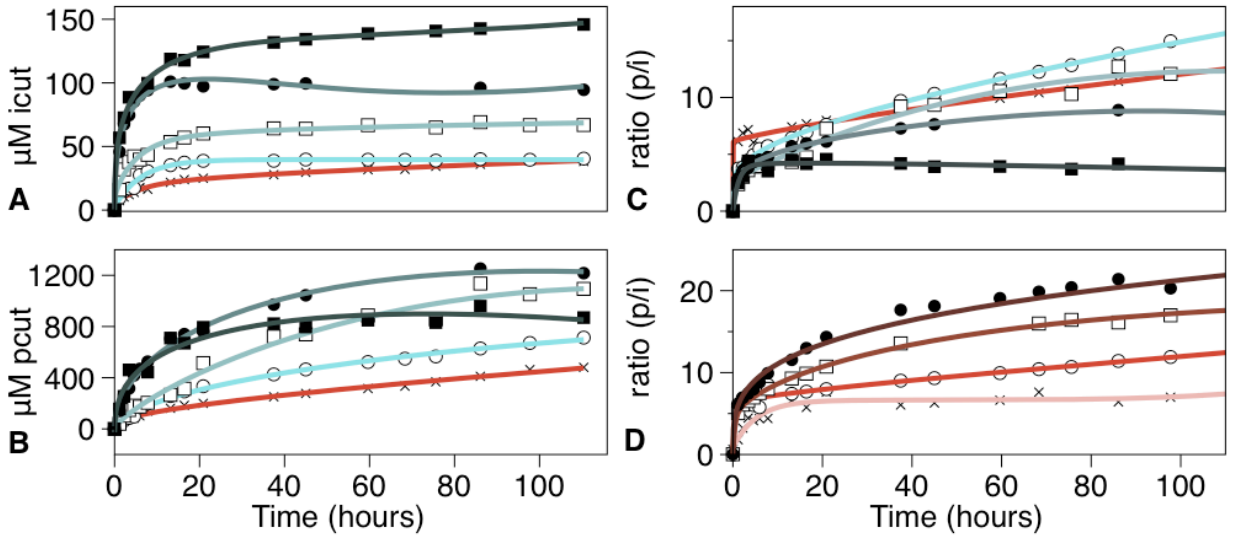


Figure 5.2 *Tl*Cel7A initial- and processive-cut product generation within various mixtures. For A, B, and C, lines represent arbitrary fits for the following discretely measured quantities within each *Tl*Cel7A-*Tem*GH5 mixture (1 g/L BMCC, pH 4.85, 50°C): (A) initial-cut products, (B) processive-cut products, and (C) the ratio between the two (p/i ratio = [pcut products]/[icut products]). The red line (X's) corresponds to the mixture with 0.076 μM *Tl*Cel7A and no endoglucanase. All other mixtures have 0.076 μM *Tl*Cel7A and concentrations of *Tem*GH5 that are proportional to their shades of blue: 0.019 μM *Tem*GH5 (-, ¢), 0.09 *Tem*GH5 (-, open squares), 0.19 μM *Tem*GH5 (-, □), 0.37 μM *Tem*GH5 (-, n). For the reactions with *Tl*Cel7A alone, values of the p/i ratio are given in D. *Tl*Cel7A concentrations increase from light red to dark red: 0.0152 μM (-, ¢), 0.076 μM (-, X), 0.152 μM (-, open squares), and 0.305 μM (-,n). Lines represent arbitrary fits to guide the eye.

The qualitative trends illustrated by the initial- and processive-cut *Tl*Cel7A product profiles are insensitive to values of η and *f* (Section 5.12). Plots of Eq. 5.8 show that 10-90% of *Tem*GH5 or *Tl*Cel7A is available to hydrolyze soluble sugars at any given time, depending on the enzyme loading (Figs. 5.A4C and 5.A4D). Values of ρ and K'_{eq} affect the value of η in any given mixture, but they do not affect the proportionality between η and total enzyme concentration or the time-invariant nature of η. The trends illustrated in Fig. 5.2 are most affected by these properties of η (Section 5.12) and, thus, they are insensitive to the manner in which η is estimated. Estimates of *f* have a similar influence. The method employed to estimate *f* affects the *Tem*GH5 concentration at which further increases in *Tem*GH5 lead to a reduction in processive length, rather than an increase, but this general trend, evident in Table 5.4, occurs with any *f* estimate (high, low, average).

5.6.5 Enzyme Denaturation

The results of this work are not sensitive to the slow denaturation of the cellulase enzymes that were used. When grown on Avicel, *Trichoderma reesei* secretes a mixture of cellulases which is dominated by Cel7A; the half-life of this mixture has been measured to be 43 hours at 50°C within a cellulose hydrolysis reaction (*61*). *T. emersonii* grows at higher temperatures than *T. reesei*, and many of its enzymes are more thermostable (*2*). The half-life of *Tem*GH5 has not been explicitly measured, but we conservatively equate it to that of *Tl*Cel7A for the current discussion. Assuming an exponential model of decay, 95% of the enzymes will remain active in solution at 50°C after 3 hours, the period over which initial-cut product generation rates from insoluble cellulose were measured. This 5% inactivation in enzyme is not enough to account for the initial-cut product generation rates being 3,000 to 30,000% slower on insoluble cellulose than on soluble cello-oligosaccharides.

Trends observed between the different reaction mixtures are similarly unaffected by enzyme denaturation. Each term in the kinetic equations used to calculate initial- and processive-cut products exhibits a first order dependence on enzyme concentration. Reductions in enzyme concentration due to denaturation would affect both icut and pcut product concentrations in proportion to the concentration of enzyme present in each mixture; the trends observed between different reaction mixtures would remain unaffected.

5.6.6 The Influence of Product Inhibition

The topic of product inhibition deserves special attention in our discussion of cellobiohydrolase kinetics. Glucose and cellobiose inhibition of cellobiohydrolase enzymes is well known (*3, 4, 62*). In a very recent study, Bu et al. calculated the absolute binding free energies of cellobiose to the Cel7A catalytic domain of *T. reesei* and found that this product is more stable in the enzyme tunnel than in free solution (*63*). The work of Bu et al. is significant in that it identifies the residues responsible for the enzyme's strong interaction with cellobiose, which acts as a competitive inhibitor and thereby reduces the effective enzyme concentration. If inhibition by glucose and cellobiose is reduced, regardless of the rate-limiting step, rates of Cel7A-catalyzed cellulose hydrolysis will be increased.

5.6.7 Evidence of a Morphological Element to Enzyme Synergy

The sensitivity of cellobiohydrolase enzymes to the nature of the cellulose substrate may provide a rationale for the evolution of multi-enzyme mixtures produced by fungi. Cellulases are often grouped into two main classes (EG and CBH), but fungi such as *Trichoderma reesei* and *Neurospora crassa* produce multiple variants of each (*1, 64*). Several studies have demonstrated that carbohydrate-active enzymes have binding site preferences for different surface structures, which focus their activities onto proximal regions (*65, 66*). Such preferences may allow different cellulases to interact with and hydrolyze a wider variety of surfaces, thus relaxing the fidelity and sensitivities that might otherwise limit two-enzyme mixtures. Recent investigations into the role of GH61 enzymes, which are produced by *N. crassa* and other filamentous fungi, reinforce this hypothesis. This enzyme, which falls into neither cellulase class, has been shown to have a stimulatory effect on cellulases that likely results from an influence on surface morphology or an ability to oxidatively generate cuts within crystalline cellulose regions (*67, 68*).

5.6.8 Future Work with CBM-containing Endoglucanase Enzymes

The results of this work are likely dependent on the morphological specificity of the enzymes used. The CBM-less *Tem*GH5 endoglucanase may have a more relaxed fidelity for particular surface structures than CBM-containing alternatives. There are 63 different families of CBM domains in the Carbohydrate Active Enzyme Database(*69*), and different adsorption behaviors and cellulose morphology specificities have been observed between them (*70, 71*). Future work includes attaching CBM's with demonstrated specificities to the catalytic domain of *Tem*GH5 and assessing the different influence of these variants on cellobiohydrolase behavior. A crystalline-specific CBM may make the enzyme less likely to act on amorphous or paracrystalline cellulose, and therefore less effective at increasing the processive length of cellobiohydrolase enzymes. This is a speculated result, however, and the experiments still need to be performed.

5.7 Conclusions

Complexation has been previously assumed to be rate-limiting, and endoglucanases have been asserted to influence complexation rates. The present work provides direct experimental measurement of complexation rates to validate these assumptions. These results complement recently published MD simulations evidencing high energy barriers associated with glycan chain abstraction with kinetic measurements showing slow rates of cellobiohydrolase complexation, a process in which cellulose chain decrystallization is a key step(*63, 72, 73*).

This study has several major implications for the design of better saccharification systems. First, cellobiohydrolase engineering efforts should focus on the development of enzymes with enhanced abilities to find and complex with chain ends. Results presented here show cellobiohydrolase complexation rates to be slow relative to other kinetic steps; enzymes that can more quickly extract chains from the cellulose surface are therefore likely to catalyze the hydrolysis of insoluble cellulose at higher rates. Second, this work presents the first experimental evidence that cellulases may remove the physical obstructions that cause cellobiohydrolases to become stuck. Endoglucanases are shown to potentiate the activity of cellobiohydrolases, not only be generating free chain ends, but also by altering the surface in a way that increases that enzyme's processive length. Cellobiohydrolases are shown to potentiate their own activity in an identical processive-length-enhancing fashion. Investigations that elucidate the mechanisms of these morphology-influenced enzyme synergies will facilitate the design of multi-cellulase mixtures. Such efforts would be complementary to recent studies showing that pretreatment-derived structural changes in the cellulose surface can make it more susceptible to cellobiohydrolase-catalyzed hydrolysis (*74, 75*). Ultimately, pretreatment methods, engineered cellulases, and novel helper enzymes (e.g., GH61 family enzymes) that facilitate the access of cellobiohydrolases to chain ends and reduce the incidence of surface obstructions will provide the most effective means of improving the rate and cost-effectiveness of biomass conversion to sugars.

5.8 Acknowledgements

We thank Dr. Mara Bryan for her comments on the manuscript and Dr. Stefan Bauer for his help with Dionex instrumentation.

5.9 Appendix 1: Derivation of Equations for η

This derivation of the η equation ($\eta_{Tl\text{Cel7A}}$ and $\eta_{Tem\text{GH5.}}$) applies to both *Tl*Cel7A and *Tem*GH5. Three assumptions were made in deriving expressions for η. First, enzymes were assumed to move to complexation sites only after becoming adsorbed to cellulose (Fig. 5.A1).

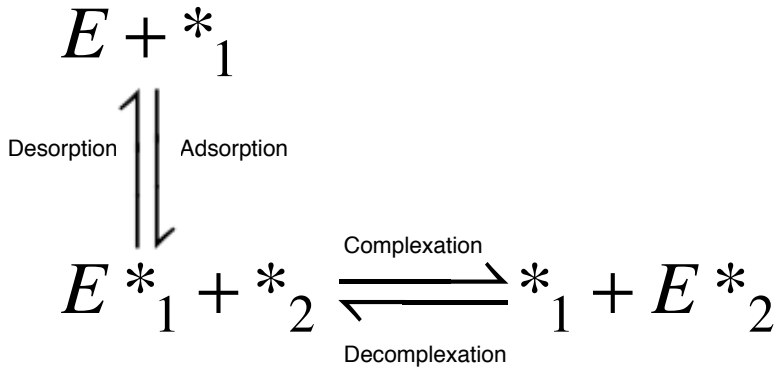


Fig. 5.A1 Depiction of the equilibrium between different states of *Tl*Cel7A. Cellulases are assumed to either be (1) free in solution, (2) adsorbed to the substrate but not complexed with a cellulose chain, and (3) complexed with a cellulose chain. Consistent with previous work, cellulases are assumed not to form productive complexes immediately upon adsorbing to the cellulose. That is, complexation is a separate step.

In accordance with previous work suggesting that cellulases rarely form productive complexes immediately upon adsorbing to the cellulose, direct adsorption of free enzymes to complexation sites was not considered in in this analysis. Secondly, equilibrium was assumed between the three different states in which the enzymes can be found: (1) free in solution, (2) adsorbed to the surface, but not complexed, and (3) complexed (Fig. 5.A1). This assumption is supported by previous work indicating that cellulase adsorption reaches a constant value within 1 hour. Dissociation constants can be subsequently defined as follows:

$$K_{eq,1} = \frac{[E][*_1]}{[E*_1]}$$
$$K_{eq,2} = \frac{[E*_1][*_2]}{[E*_2][*_1]} \qquad \text{(5.A1 and 5.A2)}$$

where $*_1$ represents the volumetric molar concentration of available adsorption sites to which an enzyme can bind before finding a complexation site; $E*_1$ represents the volumetric molar concentration of enzyme that is adsorbed, but no complexed; $*_2$ represents the volumetric molar concentration of available complexation sites to which an enzyme must bind to release a soluble product; $E*_2$ represents the volumetric molar concentration of enzyme that is complexed; and E represents the molar concentration of enzyme that is free in solution.

Site balances on adsorption and complexation sites can be combined with Eqs. 5.A1 and 5.A2 to produce an equation for $E*_2$, the volumetric concentration of complexed enzyme:

$$\left[E^{*}_{2}\right] = \frac{{}^{*}_{T,2}\left[E\right]}{K'_{eq} + \left[E\right]} \tag{5.A3}$$

where

$$K'_{eq} = K_{eq,1} \cdot K_{eq,2} = \frac{\left[E\right]\left[{}^{*}_{2}\right]}{\left[E^{*}_{2}\right]} \tag{5.A4}$$

Here, K'_{eq} represents an apparent dissociation constant for complexed enzymes,

${}^{*}_{T,1}$ is the volumetric concentration of total adsorption sites to which enzymes must bind before finding complextion sites, and ${}^{*}_{T,2}$ is the volumetric concentration of total complexation sites.

Eqs. 5.A3 and 5.A4 can be combined with $E_T=[E]+[E^*_1]+[E^*_2]$ to yield a cubic equation that can be solved for E as a function of E_T, ${}^{*}_{1,T}$, ${}^{*}_{2,T}$, K_{eq1}, and K'_{eq}. When the total enzyme concentration ($E_T=[E]+[E^*_1]+[E^*_2]$) and the total number of surface sites (adsorption+complexation; ${}^{*}_{T,1}+{}^{*}_{T,2}$) are within an order of magnitude of each other, E can be approximated as proportional to E_{tot} and the following proportionality follows:

$$\left[E^{*}_{2}\right] \propto \left[{}^{*}_{T,2}\right] \cdot \frac{E_T}{K_{eq} + E_T} \tag{5.A5}$$

This assumption can be tested experimentally. At early times, values for $[E^*_2]$ can be approximated by $[G_{eq}]$, the glucose equivalent concentration of product released from productive enzyme-cellulose interactions:

$$G_{eq} = G_1 + 2 \cdot G_2 + 3 \cdot G_3 \tag{5.A6}$$

Eq. 5.A3 can then be rewritten in terms of glucose equivalents:

$$\left[G_{eq}\right] = \kappa \cdot \left[G_{eq\ \max}\right] \cdot \frac{E_T}{K_{eq} + E_T} \tag{5.A7}$$

where κ is a proportionality constant and $[G_{eq\ max}]$ is the glucose equivalent soluble product generated if all complexation sites are occupied by enzyme. Mixtures with different concentrations of *Tem*GH5 or *Tl*Cel7A acting alone on BMCC can then be used to verify the relationship given in Eq. 5.A7. Total glucose equivalents were measured at 20, 40, 60, and 80 hours, and plots of $[G_{eq}]$ vs. E_{tot} were generated (Figs. 5.A3A and 5.ASB). Eq. 5.A4 adequately describes the manner in which G_{eq} changes with protein loading, validating the order of magnitude assumption about the values of E_T and ${}^{*}_{T,2}$.

An equation for θ, the fraction of total complexation sites occupied by enzyme, can be written by rearranging Eq. 5.A7:

$$\theta = \frac{\left[G_{eq}\right]}{\left[G_{eq\ \max}\right]} = \kappa \cdot \frac{E_T}{K'_{eq} + E_T} \tag{5.A8}$$

Plots of θ/κ as a function of E_T (Figs. 5.A3C and 5.A3D) were insensitive to the time at which $[G_{eq}]$ was measured, suggesting that θ could be approximated as a constant over the course of cellulose hydrolysis. Values of K_{eq} at 40 hours were used for calculations within this work.

When the cellulose concentration is constant (1 g/L in all reactions addressed in this work), both the fraction of complexation sites occupied by enzyme ($\theta=E^*_2/^*_{T,2}$) and the fraction of enzyme available to hydrolyze soluble substrate ($\eta = ([E] + [E^*_1])/ [E_T]$) are functions of E_T, and they are proportional:

$$\eta(E_T) \propto \theta(E_T) \tag{5.A9}$$

This proportionality allows for an equation for η to be written in terms of E_{tot}:

$$\eta = \frac{E}{E_T} = \rho \cdot \frac{E_T}{K'_{eq} + E_T} \tag{5.A10}$$

where ρ is a constant arising from both κ and the proportionality between η and θ. Plots of η vs. E_{tot} are show in Figs. 5.A4C and 5.A4D; as with θ/κ, they do not exhibit a strong time-dependence.

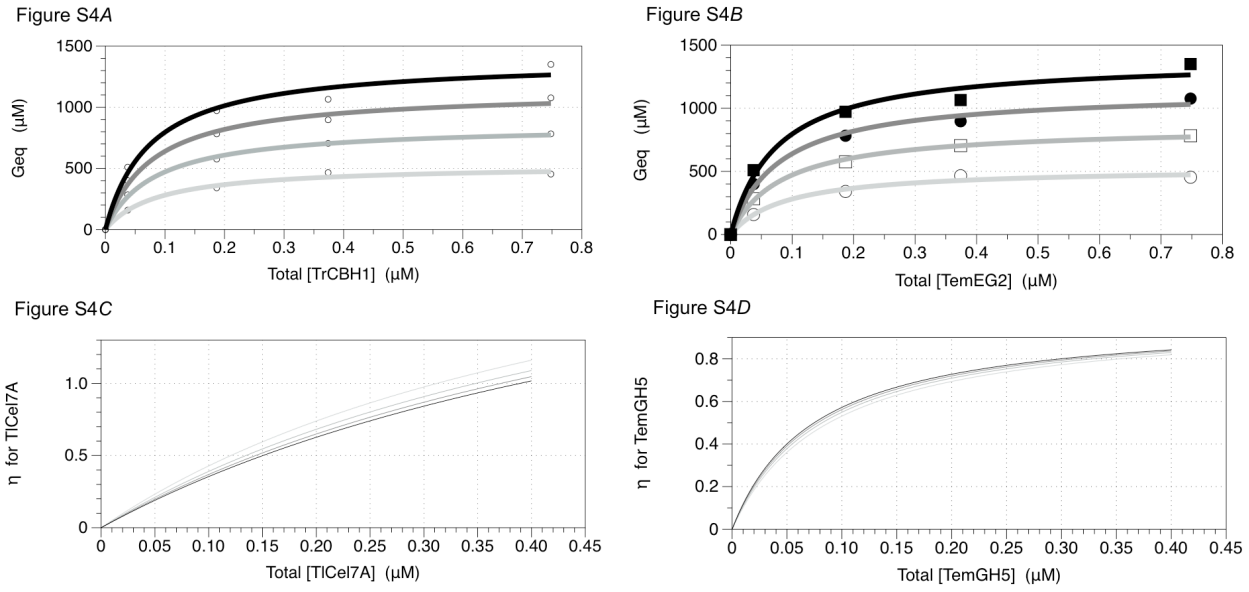


Fig. 5.A4. Estimation of enzyme fractions free to interact with soluble substrates. In S4A, glucose equivalents of soluble products ($G_1+2G_2+3G_3$) generated from 1 g/L BMCC are plotted for reactions with 0.015, 0.076, 0.15, 0.30 μM of *Tl*Cel7A at 20 (-), 40 (-), 60 (-), and 80 hours (-). Reactions were carried out at pH 4.85 and 50°C. In S4B, glucose equivalents of soluble products generated from 1 g/L BMCC are plotted for reactions with 0.037, 0.187, 0.374, and 0.748 μM of *Tem*GH5 at 20 (-), 40 (-), 60 (-), and 80 hours (-). Reactions were also carried out at pH 4.85 and 50°C. In S4C, equations for $\eta_{Tl\text{Cel7A}}$ developed from $\rho_{Tl\text{Cel7A}} = 2.7$ and fits to the points in Fig. 5.A4A are plotted against $[Tl\text{Cel7A}]_{tot}$. The four curves correspond to equations for $\eta_{Tl\text{Cel7A}}$ based on a $\rho_{Tl\text{Cel7A}}$ estimate of 2.7 and G_{eq} measurements at 20 (-), 40 (-), 60 (-), and 80 hours (-). In S4D, equations for η_{TemGH5} developed from $\rho_{EG}=1$ and fits to the points in Fig. 5.A4B are plotted against $[Tem\text{GH5}]_{tot}$. The four curves correspond to equations for η_{TemGH5} based on G_{eq} measurements at 20 (-), 40 (-), 60 (-), and 80 hours (-).

5.10 Appendix 2: Estimation of ρ

Values for ρ for *Tl*Cel7A can be estimated from the data of Kurasin et al [34]. In that work, when 2.5μM *Tl*Cel7A was mixed with 10 g/L of cellulose, approximately 80% of the enzyme was available to hydrolyze soluble substrate ($E = 0.8*E_T$). With this data and the K'_{eq} value reported in Table 5.3, Eq. 5.A7 was solved to yield a value of $\rho_{Tl\text{Cel7A}} = 2.7$. As no analogous data was available for *Tem*GH5, we used $\rho_{EG} = 1$ for this enzyme. As discussed below, the results of this study proved to be insensitive to ρ values.

5.11 Appendix 3: Derivation of Eqs. 5.11 and 5.12

In order to estimate the contribution of *Tem*GH5 to G3 production, we estimated upper and lower bounds of *f* with low and high estimates of the amount of cellotriose produced by a mixture of *Tl*Cel7A and *Tem*GH5. In Eq. 5.11, the enzymes are assumed to generate the same amount of cellotriose in a mixture as they would in isolation.

$$f(t) = \frac{\text{Total } G_3 \text{ by } \textit{Tem}\text{EG2 alone}}{\text{Total } G_3 \text{ by theoretical mix}} = \frac{\left[G_3\right]_{\text{tot } \textit{Tem}\text{EG2 i}}}{\left[G_3\right]_{\text{total } \textit{Tem}\text{EG2 i}} + \left[G_3\right]_{\text{tot } \textit{Tr}\text{CBH1 i}}} \tag{5.A11}$$

Experimental evidence, however, shows that these two enzymes generate more cellotriose when they act in concert than they do when they act alone; thus, the value of *f* in Eq. 5.11 represents an upper limit.

In Eq. 5.12, the enzymes are assumed to generate more cellotriose in a mixture than they would in isolation. In the mixtures investigated in this work, there was an average of 1.4 fold more soluble product ($[G_{eq}]$) generated when the enzymes worked in concert:

$$\frac{\left[G_{eq}\right]_{mix}}{\left[G_{eq}\right]_{Tl\text{Cel7A}} + \left[G_{eq}\right]_{\textit{Tem}\text{GH5}}} = 1.4 \tag{5.13}$$

This 1.4 fold increase in hydrolysis can be employed to create an equation for a lower bound of *f*:

$$f(t) = \frac{\left[G_3\right]_{\text{total by } \textit{Tem}\text{EG2 i}}}{1.4 \cdot \left(\left[G_3\right]_{\text{total } \textit{Tem}\text{EG2 i}} + \left[G_3\right]_{\text{total } \textit{Tr}\text{CBH1 i}}\right)} \tag{5.A12}$$

In Eq. 5.12, the enzymes are assumed to generate 1.4 fold more cellotriose in a mixture than they would in isolation. Because most of the increased sugar production in two-enzyme mixtures comes from increased production of G_2, not G_3, the denominator in Eq. 5.12 represents an overestimate of G_3 generation. The corresponding value of *f* is therefore a lower bound of the fraction of cellotriose generated by *Tem*GH5, in this study.

5.12 Appendix 4: Sensitivity of the Results to η and *f* Estimates

The results of this study are relatively insensitive to values of η. This constant is affected by estimates of ρ and K'_{eq}. As cellotriose hydrolysis rates are proportional to ρ in all enzyme mixtures examined, the value of this constant does not affect the trends observed from one mixture to the next. A higher value of ρ results in higher estimates of processive and initial cuts, but trends between initial and processive cut time course profiles remain unchanged; the processive length plots remain unaffected. Values of K'_{eq} have a similar effect; they alter the values of the results, but they have no affect on trends. The primary function of Eq. 8 is to illustrate the non-linear increase in the fraction of *Tem*GH5 or *Tl*Cel7A that becomes available to hydrolyze soluble sugars. At high *Tem*GH5 and *Tl*Cel7A loadings, more enzymes are available to interact with soluble substrates than would be available at lower loadings, where much of the enzymes would be engaged in hydrolyzing the substrate (Fig. 5.A4D).

High estimates of *f* (Eq. 5.10) lead to high estimated rates of production of processive- and initial-cut product; low estimates of *f* (Eq. 5.11) lead to lower estimated rates. In both cases, however, the rate at which initial-cut product generation slows down is greater than the rate at which processive-cut product generation slows. Similarly, for both estimates, the *Tl*Cel7A processive length increases with the addition of low concentrations of *Tem*GH5 and then begins to decrease as that value gets higher. The *Tem*GH5 concentration at which this happens is sensitive to the values used to estimate *f*; however the results of this work do not depend on the value of *f*.

5.13 Additional Materials and Methods

5.13.1 Chemicals

Cellotriose and cellohexaose were purchased from Seikagaku Biobusiness; sodium acetate, HEPES, and sodium azide, from Sigma; glycerol and 50% sodium hydroxide solution, from Fisher Scientific. Criterion 10% Tris-HCl gels and GelCode Blue Reagent were supplied by Bio-Rad and Thermo Scientific, respectively.

5.13.2 Enzymes

Fungal cellulase enzymes have numerous disulfide bonds and distinct glycosylation patterns that make them difficult to express recombinantly in non-cellulolytic hosts. The *Trichoderma sp* Cel7A cellobiohydrolase, which has twelve disulfide bonds, short O-linked glycoylsation on its linker, and N-linked glycosylation on its CBM and catalytic domain, has proven to be a difficult target for recombinant expression. Most attempts to produce Cel7A heterologously in other hosts have led to *Cel*7A variants that are less active or less stable than the native form; this reduced activity has been attributed to incomplete folding or incorrect glycosylation. For this reason, both *Trichoderma sp.* Cel7A and *Taleromyces emersonii* GH5 were purified from commercially available forms, in which they had already been brought to near purity.

Both *Trichoderma longibrachiatum* Cellobiohydrolase I (*Tl*Cel7A) and *Talaromyces emersonii* EG2 (*Tem*GH5) were purchased from Megazyme (E-CBH1 and E-CELTE). Previous experiments by our group revealed minor endoglucanase and β-glucosidase contamination within these products. The analysis in this work relied on accurate measurements of cellooligosaccharide concentrations in cellulase/cellulose reaction mixtures over 120 hour time periods, so great lengths were taken to ensure high enzyme purity. A GE Healthcare ATKA Explorer system was

used for this purpose. Approximately 20 mg of the *Tl*Cel7A enzyme, which arrived in 3.2 M ammonium sulfate solution, was dissolved into 15 ml of 25 mM HEPES pH 7.35 and desalted on a GE HiPrep 26/10 desalting column with a 4 ml/min flow of identical buffer. The protein-containing fraction was purified further via exchange on a HiLoad 16/10 Q Sepharose HP anion exchange column. The main peak from this step was then concentrated and run through a Superdex 75 10/300 gel filtration column (15% 1.0M HEPES solution). The leading peak from this step was desalted again and purified on a Mono Q 10/100 GL anion exchange column. Single band purity was ensured with a 10% Tris-HCl gel (Bio-Rad). The *Tem*GH5 enzyme was purified in an identical way, but without the Mono Q step. The final *Tem*GH5 enzyme (37 kDa) had a slight higher molecular weight impurity (47 kDa) with no known enzymatic activity (Megazyme E-CELTE Product Information). In both cases, beta-glucosidase contamination was checked by incubation for 4 hours at 50°C with 1.1 mM of cellobiose (50 mM sodium acetate buffer, pH 4.8). Samples were stopped at this time with equal volumes of 0.100 M NaOH run on a Dionex PA-200 column (method described below). No β-glucosidase activity was detected.

The endoglucanase activity of the purified enzymes was tested with the carboxymethyl cellulose (CMC) assay, which is the most commonly used assay for endoglucanase activity. For this assay, 0.27 μM of either enzyme was mixed with 10 g/L carboxymethyl cellulose sodium salt (Sigma 419273-1kg) at a pH of 4.85 (50 mM sodium acetate buffer) and a temperature of 50°C. Samples were agitated at 150 rpm in an incubator shaker. At 1, 3.03, and 5.16 hours, 30 μL samples were withdrawn from each reaction mixture and mixed with 60 μL DNS reagent as described in Ghose, T. K. Measurement of Cellulase Activities. *Pure and Appl. Chem 59*, 157-168. For the sample taken from the *Tem*GH5 reaction after 1 hour, a dilution step was required prior to the addition of DNS reagent. After thorough mixing, these 2:1 mixtures of DNS reagent to sample were heated at 99°C for 5 minutes in a thermocycler. The absorbance of the resulting solutions at 540 nm was then measured and converted into glucose equivalents with a glucose standard curve generated in an identical fashion. The *Tl*Cel7A cellobiohydrolase had no detectable activity on CMC, indicating no endoglucanase contamination (Fig 5.A7). The *Tem*GH5 endoglucanase, by comparison, degraded ~50% of the CMC in 5 hours (Fig. 5.A7).

5.13.3 Cellulose

Many previously reported BMCC preparations have been shown to modify substrate morphology and enhance digestibility. Pretreatment effects are minimized in the preparation employed here by the exclusion of steps involving strongly acidic or basic solutions.

A 20 μL freezer stock of *Acetobacter xylinum* (ATCC 53582) was used to inoculate 1 L of media as described previously [48]. This inoculum was grown in a 2 L Erlenmeyer flask for 7 days at 25°C in a shaker at 250 RPM). The resulting mixture was spun down at 6000 RPM in a Beckman Coulter Avanti J-26 XP Centrifuge. The lower 25% of the resulting heterogeneous suspension, which contained most of the cellulose, was then resuspended in 600 ml of water, and spun down again. This process was repeated three times with 600 ml of 50% ethanol solution. After being incubated at 10°C in 50% ethanol for three weeks, the resulting suspension was spun down, and blended at 3 minutes in a Waring blender. The absence of live cells was confirmed with a Leica Epifluorescence Microscope and the subsequent suspension was washed once with 50% ethanol. The final BMCC preparation was suspended in 30% ethanol, and its cellulose concentration was measured with the phenol sulfuric acid method as described previously [49].

5.13.4 Determination of k_{cat} and K_M for the activity of *Tl*Cel7A on G6

The Michaelis-Menten kinetic parameters pertaining to the activity of *Tl*Cel7A on cellohexaose (Table 5.2) were calculated via a Lineweaver-Burk plot of initial rate data. Hydrolysis was monitored in 80 μL reactions with 0.022 μM *Tl*Cel7A in 50 mM sodium acetate (pH 4.8) buffer (50°C). Initial cellohexaose concentrations of 1, 1.6, 2.1, 5.1, 20.1, 100, and 149.5 μM were used. Reactions were stopped after 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 2.25, and 2.5 minutes with 20 μL of 1.0 M NaOH on ice, and cellohexaose depletion was monitored on the Dionex HPLC. A plot of 1/V vs. 1/S was made, and the following line of best fit was calculated: 1/V=2.556*(1/S)+0.1721. This corresponds to a $k_{cat\,Tl\text{Cel7A}}$ of 4.5 s^{-1} and a $K_{M\,Tl\text{Cel7A}}$ of 14.9 μM (as reported in Table 2). These values are similar to the analogous parameters describing the activity of *Trichoderma reesei* Cel7A on cellohexaose ($k_{cat\,Tl\text{Cel7A}}$ of 9.4 s^{-1} and $K_{M\,Tl\text{Cel7A}}$ of 3.3 μM) calculated by Nidetzky et al in (*23*).

5.13.5 Determination of k_{cat} and K_M for the activity of *Tem*GH5 and *Tlong*Cel7A on G3

Figure S3*A*

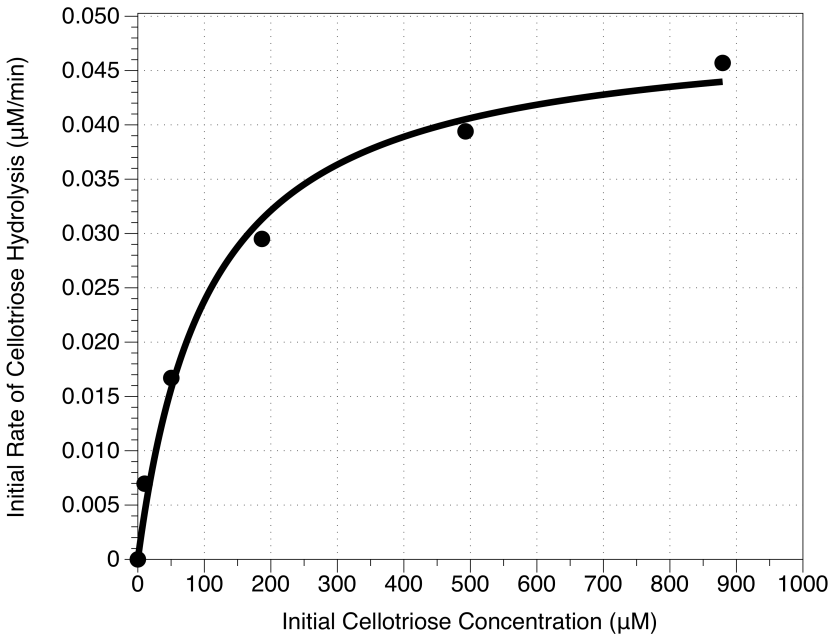


Figure S3*B*

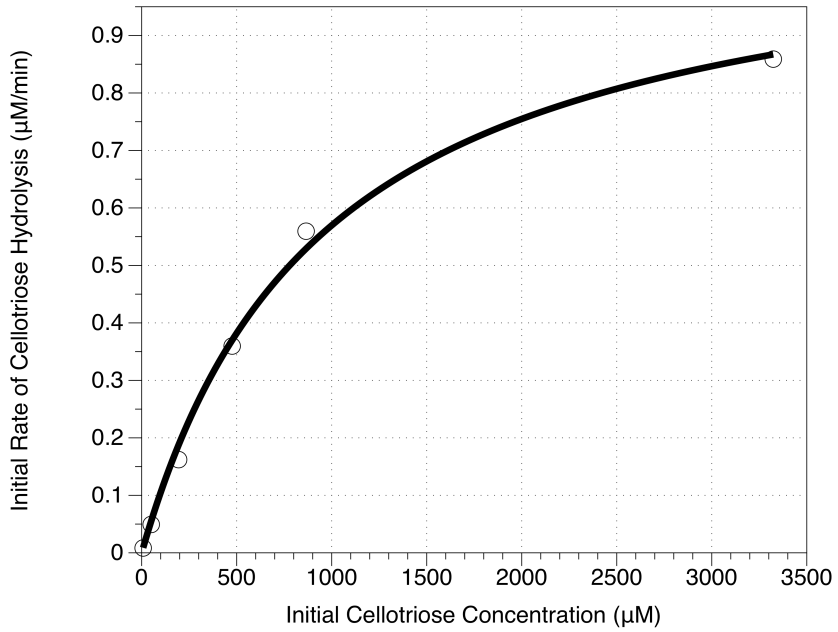


Fig. 5.A3 Michaelis-Menten plots of enzyme activity on G3 at pH 4.85 and 50°C. In S3A, initial rates of G_3 hydrolysis by 0.051 µM *Tl*Cel7A are plotted against corresponding initial substrate concentrations of 10.13, 50.65, 202.6, 506.5, and 919.804 µM (□). In S3B, initial rates of G_3 hydrolysis by 0.074 µM *Tem*GH5 are plotted against corresponding initial substrate concentrations of 7.61, 50.03, 194.25, 476.02, 865.07, and 3323.6 µM (¢). In both plots, the line (-) represents a nonlinear regression fit to the Michaelis-Menten equation.

5.13.6 Investigation of *Tem*GH5-produced Soluble Sugar Ratios during Saccharification

Figure S5

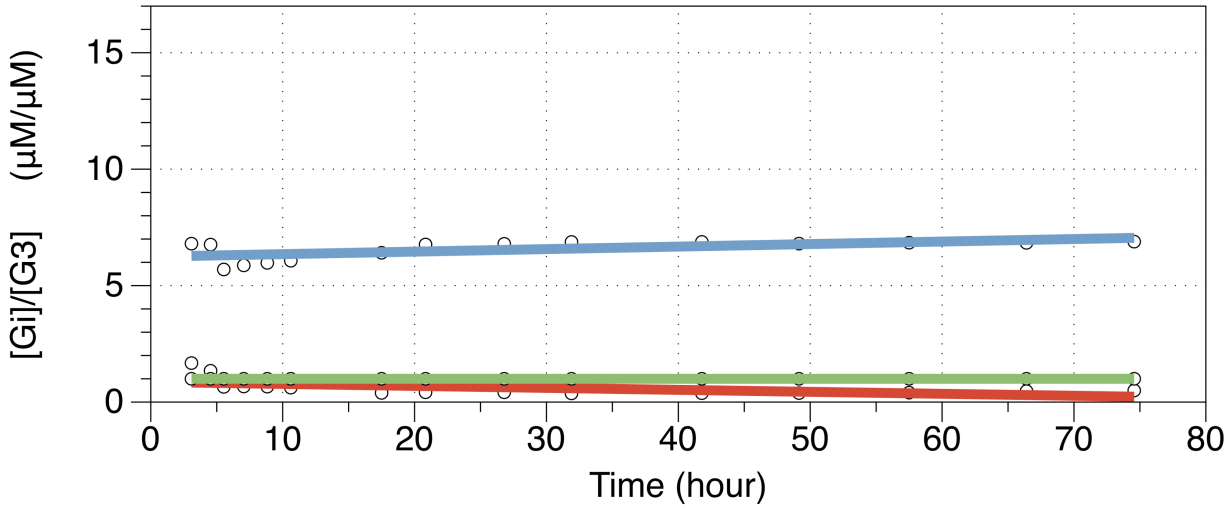


Fig. 5.A5 Time course profiles for cellotriose-normalized sugar concentrations in *Tem*GH5-catalyzed hydrolysis of BMCC. Cellooligosaccharide production from 1 g/L BMCC and 0.022 μM *Tem*GH5 alone on BMCC (pH 4.85, 50°C). The cellotriose concentration was corrected to account for *Tem*GH5-catalyzed cellotriose hydrolysis (Eqs. 5.6 and 5.7) and then used as a basis for normalization: G_1/G_3 (-), G_2/G_3(-), and G_3/G_3 (-). The *Te*mGH5 product distribution is relatively constant throughout hydrolysis; this is consistent with the random attack mechanism proposed for this enzyme and indicates that transglycosylation activity is very low.

5.13.7 Estimation of Rats of *Tem*GH5- and *Tlong*Cel7A-catalyzed Hydrolysis During Saccharification

Figure S6*A*

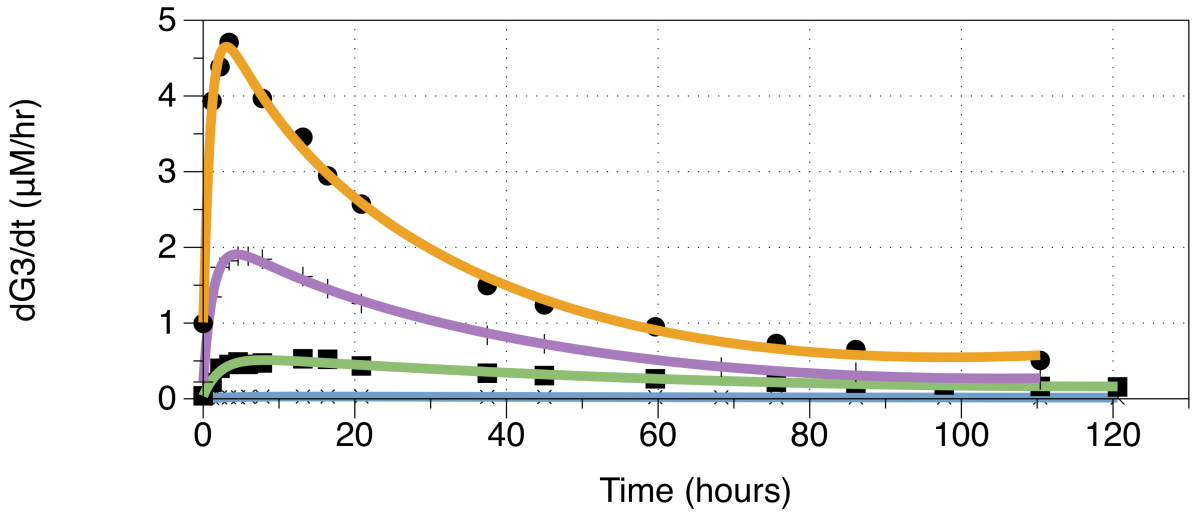


Figure S6*B*

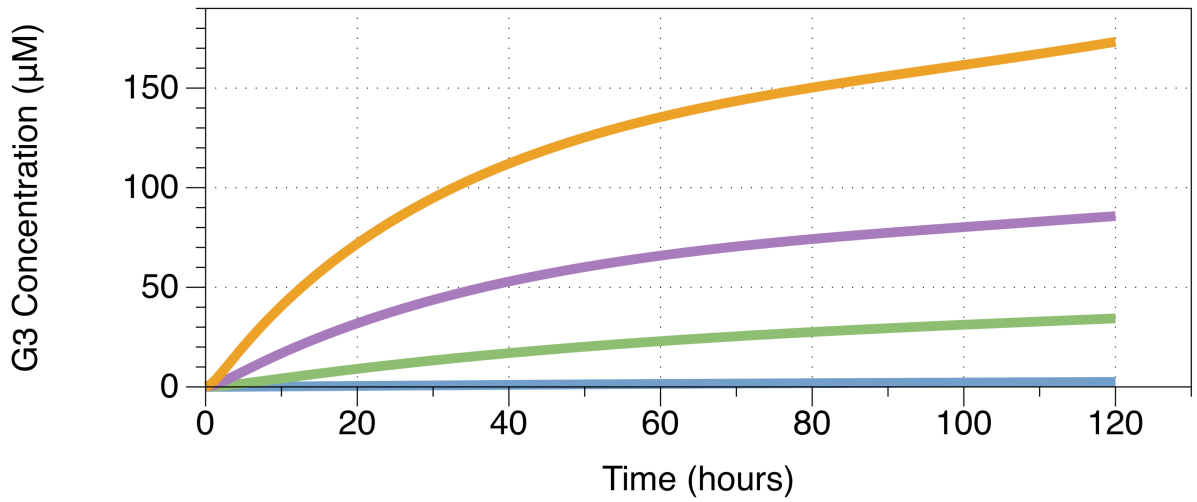


Fig 5.A6 *Tem*GH5-catalyzed hydrolysis of G3. In S6A, rates of cellotriose hydrolysis by *Tem*GH5 in various mixtures are plotted against time. Mixtures consisted of 1 g/L BMCC, 0.076 µM *Tl*Cel7A, and *Tem*GH5 concentrations as follows 0.019 µM (x, -), 0.09 µM (n,-), 0.19 µM (+,-), 0.37 µM (□,-). Plots of *Tl*Cel7A catalyzed hydrolysis in these mixtures and in separate mixtures without *Tem*GH5 looked similar, but rates were generally lower. Lines represent fits of theses points to equations of the form $a+b*x^c+d*x^f + m*n^x$, where x is an enzyme concentration. In S6B, total cellotriose hydrolyzed by *Tem*GH5 was determined by integrating lines of fit to the $d[G_3]/dt$ points described in Fig. 5.A6A. Total cellotriose hydrolyzed by this enzyme is plotted against time for mixtures consisting of 1 g/L BMCC, 0.076 µM *Tl*Cel7A, and *Tem*GH5 concentrations as follows 0.019 µM (-), 0.09 µM (-), 0.19 µM (-), 0.37 µM (-).

5.13.8 Confirmation of the Purity of *Tem*GH5

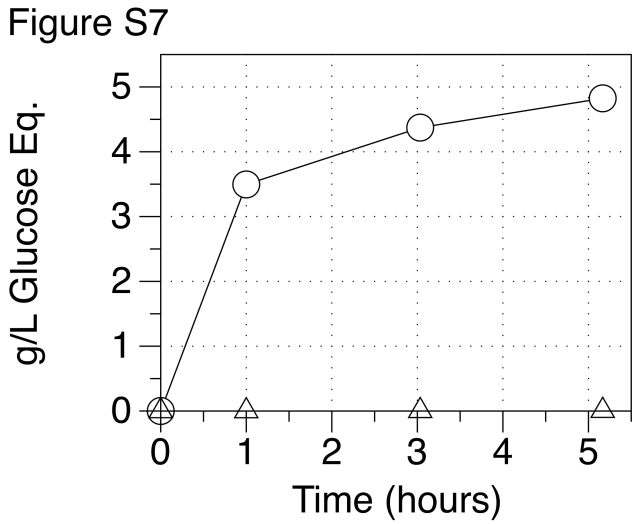


Figure 5.A7 Assay for Endoglucanase Activities of *Tem*GH5 and *Tl*CelA. In S7, the glucose equivalent released by *Tem*GH5 (open circle) and *Tl*Cel7A (open triangle) on carboxymethyl cellulose is plotted against time. For the assay, 0.27 μM of enzyme was incubated with 10 g/L carboxymethyl cellulose sodium salt at 50°C (1 ml total volume), and reducing sugar concentrations were measured by the DNS assay (detailed below). Absorbance at 540 nm was converted into glucose equivalent concentrations with a glucose standard curve. The *Tl*Cel7A cellobiohydrolase had no detectable activity on CMC.

5.13.9 Equations of Fit Used in Figure 5.2

In 2A, 2B, 2C, and 2D the following equations of fit are used to generate lines to guide the eye:

$$\frac{d[product]}{dt} = a + b \cdot t^{c} + d \cdot t^{f} + m \cdot n^{t} \qquad [S9]$$

As these curves are intended for visualization, they do not influence the results of this work.

5.14 References

References

1. Martinez, D., Berka, R. M., Henrissat, B., Saloheimo, M., Arvas, M., Baker, S. E., Chapman, J., Chertkov, O., Coutinho, P. M., Cullen, D., Danchin, E. G. J., Grigoriev, I. V., Harris, P., Jackson, M., Kubicek, C. P., Han, C. S., Ho, I., Larrondo, L. F., de Leon, A. L., Magnuson, J. K., Merino, S., Misra, M., Nelson, B., Putnam, N., Robbertse, B., Salamov, A. A., Schmoll, M., Terry, A., Thayer, N., Westerholm-Parvinen, A., Schoch, C. L., Yao, J., Barabote, R., Nelson, M. A., Detter, C., Bruce, D., Kuske, C. R., Xie, G., Richardson, P., Rokhsar, D. S., Lucas, S. M., Rubin, E. M., Dunn-Coleman, N., Ward, M., and Brettin, T. S. (2008) Genome sequencing and analysis of the biomass-degrading fungus Trichoderma reesei (syn. Hypocrea jecorina) (vol 26, pg 553, 2008), *Nat Biotechnol 26*, 1193-1193.
2. Voutilainen, S. P., Murray, P. G., Tuohy, M. G., and Koivula, A. (2010) Expression of Talaromyces emersonii cellobiohydrolase Cel7A in Saccharomyces cerevisiae and rational

mutagenesis to improve its thermostability and activity, *Protein Engineering Design & Selection 23*, 69-79.
3. Levine, S. E., Fox, J. M., Blanch, H. W., and Clark, D. S. (2010) A mechanistic model of the enzymatic hydrolysis of cellulose, *Biotechnol Bioeng 107*, 37-51.
4. Zhang, Y. H. P., and Lynd, L. R. (2004) Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems, *Biotechnol Bioeng 88*, 797-824.
5. Himmel, M. E. (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production (vol 315, pg 804, 2007), *Science 316*, 982-982.
6. Ting, C. L., Makarov, D. E., and Wang, Z. G. (2009) A Kinetic Model for the Enzymatic Action of Cellulase, *J Phys Chem B 113*, 4970-4977.
7. Crowley, M. F., Uberbacher, E. C., Brooks III, C. L., Walker, R. C., Nimlos, M. R., and Himmel, M. E. (2008) Developing improved MD codes for understanding processive cellulases, *Journal of Physics: Conference Series 125*, 1.
8. Horn, S. J., Sikorski, P., Cederkvist, J. B., Vaaje-Kolstad, G., Sorlie, M., Synstad, B., Vriend, G., Varum, K. M., and Eijsink, V. G. H. (2006) Costs and benefits of processivity in enzymatic degradation of recalcitrant polysaccharides, *P Natl Acad Sci USA 103*, 18089-18094.
9. Divne, C., Stahlberg, J., Teeri, T. T., and Jones, T. A. (1998) High-resolution crystal structures reveal how a cellulose chain is bound in the 50 angstrom long tunnel of cellobiohydrolase I from Trichoderma reesei, *Journal of Molecular Biology 275*, 309-325.
10. Koivula, A., Ruohonen, L., Wohlfahrt, G., Reinikainen, T., Teeri, T. T., Piens, K., Claeyssens, M., Weber, M., Vasella, A., Becker, D., Sinnott, M. L., Zou, J. Y., Kleywegt, G. J., Szardenings, M., Stahlberg, J., and Jones, T. A. (2002) The active site of cellobiohydrolase Cel6A from Trichoderma reesei: The roles of aspartic acids D221 and D175, *J Am Chem Soc 124*, 10015-10024.
11. Pereira, J. H., Chen, Z. W., McAndrew, R. P., Sapra, R., Chhabra, S. R., Sale, K. L., Simmons, B. A., and Adams, P. D. (2010) Biochemical characterization and crystal structure of endoglucanase Cel5A from the hyperthermophilic Thermotoga maritima, *J Struct Biol 172*, 372-379.
12. Divne, C., Stahlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J. K. C., Teeri, T. T., and Jones, T. A. (1994) The 3-Dimensional Crystal-Structure of the Catalytic Core of Cellobiohydrolase-I from Trichoderma-Reesei, *Science 265*, 524-528.
13. Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K. C., and Jones, T. A. (1990) 3-Dimensional Structure of Cellobiohydrolase-Ii from Trichoderma-Reesei, *Science 249*, 380-386.
14. Sulzenbacher, G., Shareck, F., Morosoli, R., Dupont, C., and Davies, G. J. (1997) The Streptomyces lividans family 12 endoglucanase: Construction of the catalytic core, expression, and X-ray structure at 1.75 angstrom resolution, *Biochemistry-Us 36*, 16032-16039.
15. Sonan, G. K., Receveur-Brechot, V., Duez, C., Aghajari, N., Czjzek, M., Haser, R., and Gerday, C. (2007) The linker region plays a key role in the adaptation to cold of the cellulase from an Antarctic bacterium, *Biochem J 407*, 293-302.
16. (2007) Cellulase (endo-beta-glucanase) from Talaromyces emersonii (Lot 30602), In *Data Booklet* (Megazyme, Ed.).
17. Murray, P., Collins, C., and Tuohy, M. (2002) Q8WZDZ_TALEM, In *Uniprot*.
18. Lehtio, J., Sugiyama, J., Gustavsson, M., Fransson, L., Linder, M., and Teeri, T. T. (2003) The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules, *Proc Natl Acad Sci U S A 100*, 484-489.
19. Din, N., Damude, H. G., Gilkes, N. R., Miller, R. C., Warren, R. A. J., and Kilburn, D. G. (1994) C-1-C-X Revisited - Intramolecular Synergism in a Cellulase, *P Natl Acad Sci USA 91*, 11383-11387.
20. KlemanLeyer, K. M., SiikaAho, M., Teeri, T. T., and Kirk, T. K. (1996) The cellulases endoglucanase I and cellobiohydrolase II of Trichoderma reesei act synergistically to solubilize native cotton cellulose but not to decrease its molecular size, *Appl Environ Microb 62*, 2883-2887.
21. Jervis, E. J., Haynes, C. A., and Kilburn, D. G. (1997) Surface diffusion of cellulases and their isolated binding domains on cellulose, *J Biol Chem 272*, 24016-24023.

22. Vuong, T. V., and Wilson, D. B. (2009) Processivity, Synergism, and Substrate Specificity of Thermobifida fusca Cel6B, *Appl Environ Microb 75*, 6655-6661.
23. Nidetzky, B., Zachariae, W., Gercken, G., Hayn, M., and Steiner, W. (1994) Hydrolysis of Cellooligosaccharides by Trichoderma-Reesei Cellobiohydrolases - Experimental-Data and Kinetic Modeling, *Enzyme and Microbial Technology 16*, 43-52.
24. Vrsanska, M., and Biely, P. (1992) The Cellobiohydrolase-I from Trichoderma-Reesei Qm-9414 - Action on Cello-Oligosaccharides, *Carbohyd Res 227*, 19-27.
25. Igarashi, K., Koivula, A., Wada, M., Kimura, S., Penttila, M., and Samejima, M. (2009) High Speed Atomic Force Microscopy Visualizes Processive Movement of Trichoderma reesei Cellobiohydrolase I on Crystalline Cellulose, *J Biol Chem 284*, 36186-36190.
26. Beckham, G. T., Matthews, J. F., Bomble, Y. J., Bu, L. T., Adney, W. S., Himmel, M. E., Nimlos, M. R., and Crowley, M. F. (2010) Identification of Amino Acids Responsible for Processivity in a Family 1 Carbohydrate-Binding Module from a Fungal Cellulase, *J Phys Chem B 114*, 1447-1453.
27. Kurasin, M., and Valjamae, P. (2011) Processivity of Cellobiohydrolases Is Limited by the Substrate, *J Biol Chem 286*, 169-177.
28. Converse, A. O., and Optekar, J. D. (1993) A Synergistic Kinetics Model for Enzymatic Cellulose Hydrolysis Compared to Degree-of-Synergism Experimental Results, *Biotechnol Bioeng 42*, 145-148.
29. Okazaki, M., and Mooyoung, M. (1978) Kinetics of Enzymatic-Hydrolysis of Cellulose - Analytical Description of a Mechanistic Model, *Biotechnol Bioeng 20*, 637-663.
30. Fan, L. T., Lee, Y. H., and Beardmore, D. H. (1980) Mechanism of the Enzymatic-Hydrolysis of Cellulose - Effects of Major Structural Features of Cellulose on Enzymatic-Hydrolysis, *Biotechnol Bioeng 22*, 177-199.
31. Fenske, J. J., Penner, M. H., and Bolte, J. P. (1999) A simple individual-based model of insoluble polysaccharide hydrolysis: the potential for autosynergism with dual-activity glycosidases, *Journal of Theoretical Biology 199*, 113-118.
32. Valjamae, P., Sild, V., Pettersson, G., and Johansson, G. (1998) The initial kinetics of hydrolysis by cellobiohydrolases I and II is consistent with a cellulose surface - erosion model, *Eur J Biochem 253*, 469-475.
33. Holtzapple, M. T., Caram, H. S., and Humphrey, A. E. (1984) The Hch-1 Model of Enzymatic Cellulose Hydrolysis, *Biotechnol Bioeng 26*, 775-780.
34. Kim, D. W., Kim, T. S., Jeong, Y. K., and Lee, J. K. (1992) Adsorption-Kinetics and Behaviors of Cellulase Components on Microcrystalline Cellulose, *J Ferment Bioeng 73*, 461-466.
35. Hong, J., Ye, X. H., and Zhang, Y. H. P. (2007) Quantitative determination of cellulose accessibility to cellulase based on adsorption of a nonhydrolytic fusion protein containing CBM and GFP with its applications, *Langmuir 23*, 12535-12540.
36. Nidetzky, B., Steiner, W., and Claeyssens, M. (1994) Cellulose Hydrolysis by the Cellulases from Trichoderma-Reesei - Adsorptions of 2 Cellobiohydrolases, 2 Endocellulases and Their Core Proteins on Filter-Paper and Their Relation to Hydrolysis, *Biochem J 303*, 817-823.
37. Zaar, K. (1977) Biogenesis of Cellulose by Acetobacter Xylinum, *Cytobiologie 16*, 1-15.
38. Dubois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. A., and Smith, F. (1956) Colorimetric Method for Determination of Sugars and Related Substances, *Analytical Chemistry 28*, 350-356.
39. Gusakov, A. V., Protas, O. V., Chernoglazov, V. M., Sinitsyn, A. P., Kovalysheva, G. V., Shpanchenko, O. V., and Ermolova, O. V. (1991) Transglycosylation activity of cellobiohydrolase I from Trichoderma longibrachiatum on synthetic and natural substrates, *Biochim Biophys Acta 1073*, 481-485.
40. Harjunpaa, V., Helin, J., Koivula, A., Siika-aho, M., and Drakenberg, T. (1999) A comparative study of two retaining enzymes of Trichoderma reesei: transglycosylation of oligosaccharides catalysed by the cellobiohydrolase I, Cel7A, and the beta-mannanase, Man5A, *Febs Lett 443*, 149-153.
41. Praestgaard, E., Elmerdahl, J., Murphy, L., Nymand, S., McFarland, K. C., Borch, K., and Westh, P. (2011) A kinetic model for the burst phase of processive cellulases, *Febs J 278*, 1547-1560.

42. Bezerra, R. M., Dias, A. A., Fraga, I., and Pereira, A. N. (2011) Cellulose Hydrolysis by Cellobiohydrolase Cel7A Shows Mixed Hyperbolic Product Inhibition, *Appl Biochem Biotechnol 165*, 178-189.
43. Teeri, T. T., Koivula, A., Linder, M., Wohlfahrt, G., Divne, C., and Jones, T. A. (1998) Trichoderma reesei cellobiohydrolases: why so efficient on crystalline cellulose?, *Biochem Soc T 26*, 173-178.
44. Michaelis, and Menten. (1913) Kinetics of Invertase Action, *Biochemische Zeitschrift 49*, 333-369.
45. Jalak, J., and Valjamae, P. (2010) Mechanism of initial rapid rate retardation in cellobiohydrolase catalyzed cellulose hydrolysis, *Biotechnol Bioeng 106*, 871-883.
46. Gruno, M., Valjamae, P., Pettersson, G., and Johansson, G. (2004) Inhibition of the Trichoderma reesei cellulases by cellobiose is strongly dependent on the nature of the substrate, *Biotechnol Bioeng 86*, 503-511.
47. Becker, D., Braet, C., Brumer, H., Claeyssens, M., Divne, C., Fagerstrom, B. R., Harris, M., Jones, T. A., Kleywegt, G. J., Koivula, A., Mahdi, S., Piens, K., Sinnott, M. L., Stahlberg, J., Teeri, T. T., Underwood, M., and Wohlfahrt, G. (2001) Engineering of a glycosidase Family 7 cellobiohydrolase to more alkaline pH optimum: the pH behaviour of Trichoderma reesei CeI7A and its E223S/A224H/L225V/T226A/D262G mutant, *Biochem J 356*, 19-30.
48. Voutilainen, S. P., Puranen, T., Siika-Aho, M., Lappalainen, A., Alapuranen, M., Kallio, J., Hooman, S., Viikri, L., Vehmaanpera, J., and Koivula, A. (2008) Cloning, expression, and characterization of novel thermostable family 7 cellobiohydrolases, *Biotechnol Bioeng 101*, 515-528.
49. von Ossowski, I., Stahlberg, J., Koivula, A., Piens, K., Becker, D., Boer, H., Harle, R., Harris, M., Divne, C., Mahdi, S., Zhao, Y. X., Driguez, H., Claeyssens, M., Sinnott, M. L., and Teeri, T. T. (2003) Engineering the exo-loop of Trichoderma reesei cellobiohydrolase, Ce17A. A comparison with Phanerochaete chrysosporium Cel7D, *Journal of Molecular Biology 333*, 817-829.
50. Ooshima, H., Sakata, M., and Harano, Y. (1983) Adsorption of Cellulase from Trichoderma-Viride on Cellulose, *Biotechnol Bioeng 25*, 3103-3114.
51. Ding, H. S., and Xu, F. (2004) Productive Cellulase Adsorption on Cellulose, In *Lignocellulose Biodegradation*, pp 154-169.
52. Hoshino, E., Shiroishi, M., Amano, Y., Nomura, M., and Kanda, T. (1997) Synergistic actions of exo-type cellulases in the hydrolysis of cellulose with different crystallinities, *J Ferment Bioeng 84*, 300-306.
53. Linder, M., and Teeri, T. T. (1996) The cellulose-binding domain of the major cellobiohydrolase of Trichoderma reesei exhibits true reversibility and a high exchange rate on crystalline cellulose, *P Natl Acad Sci USA 93*, 12251-12255.
54. Kipper, K., Valjamae, P., and Johansson, G. (2005) Processive action of cellobiohydrolase Cel7A from Trichoderma reesei is revealed as 'burst' kinetics on fluorescent polymeric model substrates, *Biochem J 385*, 527-535.
55. Eriksson, T., Karlsson, J., and Tjerneld, F. (2002) A model explaining declining rate in hydrolysis of lignocellulose substrates with cellobiohydrolase I (Cel7A) and endoglucanase I (Cel7B) of Trichoderma reesei, *Appl Biochem Biotech 101*, 41-60.
56. Stahlberg, J., Johansson, G., and Pettersson, G. (1991) A New Model for Enzymatic-Hydrolysis of Cellulose Based on the 2-Domain Structure of Cellobiohydrolase-I, *Bio-Technology 9*, 286-290.
57. Xu, F., and Ding, H. S. (2007) A new kinetic model for heterogeneous (or spatially confined) enzymatic catalysis: Contributions from the fractal and jamming (overcrowding) effects, *Appl Catal a-Gen 317*, 70-81.
58. Nishiyama, Y., Sugiyama, J., Chanzy, H., and Langan, P. (2003) Crystal structure and hydrogen bonding system in cellulose 1(alpha), from synchrotron X-ray and neutron fiber diffraction, *Journal of the American Chemical Society 125*, 14300-14306.
59. Liu, Y.-S., Baker, J. O., Zeng, Y., Himmel, M. E., Haas, T., and Ding, H. S. (2011) Cellobiohydrolase hydrolyzes crystalline cellulose on hydrophobic faces, *J Biol Chem*.

60. Meinke, A., Damude, H. G., Tomme, P., Kwan, E., Kilburn, D. G., Miller, R. C., Jr., Warren, R. A., and Gilkes, N. R. (1995) Enhancement of the endo-beta-1,4-glucanase activity of an exocellobiohydrolase by deletion of a surface loop, *J Biol Chem 270*, 4383-4386.
61. Drissen, R. E. T., Maas, R. H. W., Van Der Maarel, M. J. E. C., Kabel, M. A., Schols, H. A., Tramper, J., and Beeftink, H. H. (2007) A generic model for glucose production from various cellulose sources by a commercial cellulase complex, *Biocatal Biotransfor 25*, 419-429.
62. Holtzapple, M., Cognata, M., Shu, Y., and Hendrickson, C. (1990) Inhibition of Trichoderma-Reesei Cellulase by Sugars and Solvents, *Biotechnol Bioeng 36*, 275-287.
63. Bu, L. T., Beckham, G. T., Shirts, M. R., Nimlos, M. R., Adney, W. S., Himmel, M. E., and Crowley, M. F. (2011) Probing Carbohydrate Product Expulsion from a Processive Cellulase with Multiple Absolute Binding Free Energy Methods, *J Biol Chem 286*, 18161-18169.
64. Tian, C. G., Beeson, W. T., Iavarone, A. T., Sun, J. P., Marletta, M. A., Cate, J. H. D., and Glass, N. L. (2009) Systems analysis of plant cell wall degradation by the model filamentous fungus Neurospora crassa, *P Natl Acad Sci USA 106*, 22157-22162.
65. Carrard, G., Koivula, A., Soderlund, H., and Beguin, P. (2000) Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose, *Proc Natl Acad Sci U S A 97*, 10342-10347.
66. Herve, C., Rogowski, A., Blake, A. W., Marcus, S. E., Gilbert, H. J., and Knox, J. P. (2010) Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects, *P Natl Acad Sci USA 107*, 15293-15298.
67. Harris, P. V., Welner, D., McFarland, K. C., Re, E., Navarro Poulsen, J. C., Brown, K., Salbo, R., Ding, H., Vlasenko, E., Merino, S., Xu, F., Cherry, J., Larsen, S., and Lo Leggio, L. (2010) Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family, *Biochemistry 49*, 3305-3316.
68. Vaaje-Kolstad, G., Westereng, B., Horn, S. J., Liu, Z., Zhai, H., Sorlie, M., and Eijsink, V. G. (2010) An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides, *Science 330*, 219-222.
69. (2011) Carbohydrate-Binding Module Family Classification, In *Carbohydrate-Active Enzyme Database Cazy.org*.
70. Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition, *Biochem J 382*, 769-781.
71. Boraston, A. B., Kwan, E., Chiu, P., Warren, R. A. J., and Kilburn, D. G. (2003) Recognition and hydrolysis of noncrystalline cellulose, *J Biol Chem 278*, 6120-6127.
72. Beckham, G. T., Matthews, J. F., Peters, B., Bomble, Y. J., Himmel, M. E., and Crowley, M. F. (2011) Molecular-Level Origins of Biomass Recalcitrance: Decrystallization Free Energies for Four Common Cellulose Polymorphs, *J Phys Chem B 115*, 4118-4127.
73. Zhong, L. H., Matthews, J. F., Hansen, P. I., Crowley, M. F., Cleary, J. M., Walker, R. C., Nimlos, M. R., Brooks, C. L., Adney, W. S., Himmel, M. E., and Brady, J. W. (2009) Computational simulations of the Trichoderma reesei cellobiohydrolase I acting on microcrystalline cellulose I beta: the enzyme-substrate complex, *Carbohyd Res 344*, 1984-1992.
74. Chundawat, S. P., Bellesia, G., Uppugundla, N., da Costa Sousa, L., Gao, D., Cheh, A. M., Agarwal, U. P., Bianchetti, C. M., Phillips, G. N., Jr., Langan, P., Balan, V., Gnanakaran, S., and Dale, B. E. (2011) Restructuring the crystalline cellulose hydrogen bond network enhances its depolymerization rate, *J Am Chem Soc 133*, 11163-11174.
75. Sathitsuksanoh, N., Zhu, Z., Wi, S., and Zhang, Y. H. (2011) Cellulose solvent-based biomass pretreatment breaks highly ordered hydrogen bonds in cellulose fibers of switchgrass, *Biotechnol Bioeng 108*, 521-529.

Chapter 6: Superresolution Images Reveal Binding Targets for the Cooperative Action of Cellulolytic Enzymes

6.1 Abstract

The mechanisms of enzyme activity on solid substrates are not well understood. Cellulolytic enzymes, which catalyze the depolymerization of cellulose, have received particular attention in efforts aimed at reducing cellulose-to-glucose conversion costs, one of the largest expenses of cellulosic biofuels production. These enzymes exhibit binding specificities for a range of cellulose morphologies, but the influence of these specificities on the synergy exhibited by multi-enzyme mixtures has remained difficult to study without methods for determining binding site organization on the surface of heterogeneous cellulosic substrates. Using photoactivated localization microscopy (PALM) to produce pointillistic maps of carbohydrate-binding modules (CBMs) bound to cotton, we developed an order parameter to quantify the different spatial arrangements of adsorbed CBMs and used that order parameter to explain synergy between cellulase enzymes designed to target different surface structures. We show that combinations of cellulolytic enzymes that bind similar but non-identical cellulose targets can exhibit activities significantly higher than those attainable by individual enzymes. These results demonstrate a strategy for improving the activity of cellulolytic mixtures and demonstrate a versatile method for investigating protein organization on heterogeneous surfaces and in complex structures.

6.2 Introduction

Cellulosic biofuels can provide a renewable alternative to fossil fuels, but high costs associated with the breakdown of cellulose into fermentable sugars have hindered their commercialization (*1-3*). Cellulose, a polymer of β-[1,4]-linked cellobiose units, is the main constituent of the plant secondary cell wall. There, cellulose is organized on 1, 10, and 100 nm length scales into a spectrum of morphologies and packing densities (*4*). Cellulolytic enzymes employ distinct substrate-binding structural motifs or carbohydrate-binding modules (CBMs) to interact preferentially with particular cellulose structures (*5-8*). Increased cellulose-to-glucose conversion efficiencies, and thus reduced biofuel production costs, require cellulolytic enzyme cocktails that are optimally matched to the structural organizations of particular biomass substrates and which exploit all potential avenues of enzyme synergy.

Conventional imaging and biophysical techniques have struggled to quantify the differential targeting of cellulolytic enzymes to heterogeneous cellulosic surfaces that resemble those of the native plant cell wall (*9-11*). The binding targets of cellulolytic enzymes are typically characterized as either amorphous or crystalline, but these categories are probably more reflective of the limitations of current measurement methods rather than the underlying organization of the substrate. Indeed, confocal imaging of CBMs bound to tobacco stem sections has revealed binding patterns that differ not only between these categories but also within them (*12*). Atomic force microscopy and electron microscopy have the spatial resolution needed to visualize and quantify the continuum of cellulose architectures (*13, 14*), but these techniques cannot resolve individual proteins on densely labeled heterogeneous surfaces where the morphological binding targets of CBMs are interspersed (Figs. 6.1A-1B). The lack of methods for identifying CBM binding site preferences and for evaluating the influence of those preferences on the kinetics of multi-enzyme mixtures has limited attempts to understand the

structural origins of cell wall recalcitrance and has impeded efforts to optimize the efficacy of cellulolytic mixtures.

Recent advances in light microscopy allow single fluorescent molecules to be localized with resolutions consistent with the length scales of cellulose organization. Photoactivated localization microscopy (PALM), in particular, allows large numbers of photoactivatable fluorescent proteins to be localized with a precision of 10-50 nm via the stochastic activation and excitation of small subsets of molecules over time (*15*). The localization precision afforded by PALM is thus 4-20 fold finer than the resolution achievable with diffraction-limited techniques (~200 nm).

6.3 Results and Discussion

We employed 2D TIRF-PALM to localize six different CBMs bound to dewaxed cotton. Previously, PALM and defocused orientation and position imaging (DOPI) have been used to confirm the preferential targeting of crystalline-specific CBMs to the hydrophobic (110) faces of cellulose nanocrystals (*16, 17*). In the present work, we employ PALM to investigate the differential targeting of six different CBMs on a structurally heterogeneous cellulosic substrate that more closely resembles the structural complexity of the plant cell wall. These CBMs are representative of the two common categories of CBMs found on cellulolytic enzymes (crystalline- and amorphous-specific, Table 1). The photoactivatable fluorescent protein mEos2 was linked to each CBM via a 30-residue linker from the *T. reesei* Cel7A cellobiohydrolase (SI Materials and Methods). The monomeric green-to-red protein mEos2 was selected for its high contrast between activated and unactivated states (*18*); the *T. reesei* Cel7A linker, which is representative of the threonine- and proline-rich linkers found in many cellulase enzymes (*19-21*), was chosen to prevent the mEos2 from interfering with the binding of the CBM. This family of localizable mEos2-CBM fusions provides a unique reagent to map surface features of cellulose substrates.

Name	Origin	Specificity	Mode of binding
CBM1	*Trichoderma reesei* cellobiohydrolase Cel7A	Crystalline	Use a planar surface of tryptophan and/or tyrosine residues to adsorb onto microfibrils, which are crystalline aggregates of 24-36 cellulose chains (*4, 22, 31*).
CBM1	*Trichoderma reesei* cellobiohydrolase Cel6A		
CBM2A	*Acidothermus cellulolyticus* GH5 endoglucanase		
CBM3A	*Clostridium thermocellum* CipA scaffoldin		
CBM17	*Clostridium cellulovorans* Cel5A endoglucanase	Amorphous	Employ binding clefts to engage single chains within disordered networks (*24, 25, 32*).
CBM28	*Bacillus sp.* 1139 Cel5a endoglucanase		

Table 6.1 The CBM domains investigated within this study. Each domain was linked to the photoactivatable fluorescent protein mEos2 for imaging and to the catalytic domain from the *Acidothermus Cellulolyticus* GH5 endoglucanase for activity studies.

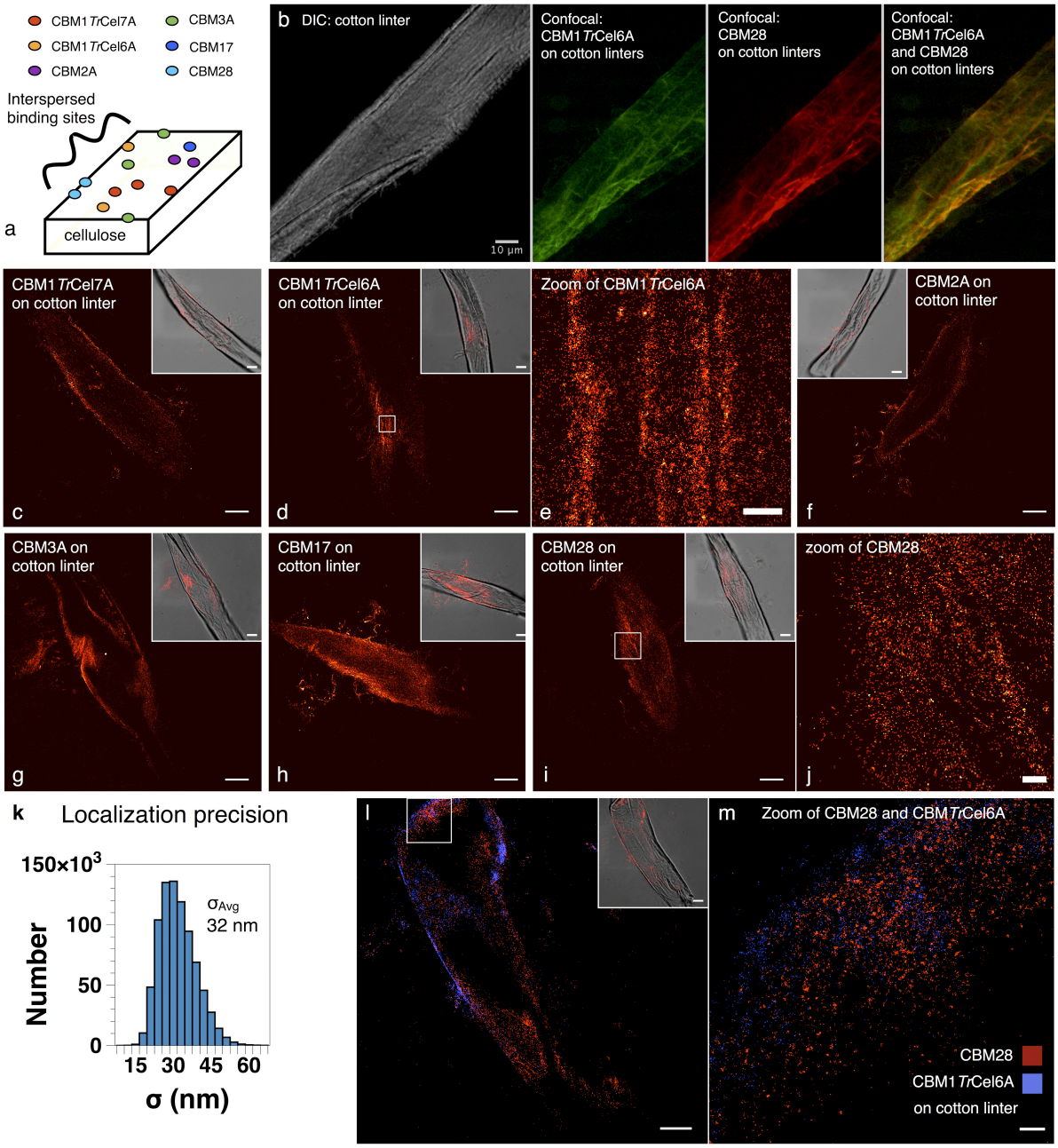


Figure 6.1 Equilibrium binding patterns of CBMs bound to cellulose. (a) Cartoon illustrating the interspersed binding sites of CBMs with different morphological targets. (b) Images of PSCFP2-CBM*Tr*Cel6A and mEos2-CBM28 bound to a cotton linter at 0.24 µmol protein / g cellulose loadings (each): Differential interference contrast (DIC) to show the cotton linter (grayscale), 442 nm confocal to show PSCFP-CBM*Tr*Cel6A (green), 488 nm confocal to show mEos2-CBM28 (red), and a 442/488 nm overlay to show both CBMs (yellow). (c-j) PALM images of mEos2-CBM constructs bound to cotton linters at 0.12 µmol protein / g cellulose: (c) mEos2-CBM1*Tr*Cel7A, (d) mEos2-CBM1*Tr*Cel6A, (e) zoom of boxed region of (d), (f) mEos2-CBM2A, (g) mEos2-CBM3A, (h) mEos2-CBM17, (i) mEos2-CBM28, (j) zoom of boxed region in (i). Insets of (c-d) and (f-i) show overlays of bright field and PALM images. (k) Histogram of localization precisions for mEos2-CBM positions used in our analysis. (l) Dual color image of PSCFP2-CBM1*Tr*Cel6A and mEos2-CBM28 bound to a cotton linter at 0.12 µmol protein/g cellulose (each). The inset contains an overlay of a bright field image with both the CBM1*Tr*Cel6A and CBM28 positions in red. (m) Zoom of boxed region in (l). Scale bars in (c-d), (f-i), (l), and the insets represent 10 µm. Scale bars in all zoomed images represent 1 µm.

After being expressed and purified in the dark, each mEos2-CBM construct was equilibrated with a dilute suspension of cotton linters at loadings of 0.12 and 0.24 μmol mEo2-CBM/g cellulose (pH 7.0, 18°C). At each loading, the mEos2-CBM constructs were localized on the surface of cotton with single-color 2D TIRF PALM (SI Materials and Methods). Equilibration and imaging of each mEos2-CBM construct was carried out 3 times at each loading for a total of 36 separate experiments. The mean localization precision of the labeled CBMs was 32 nm (Fig. 6.1K) and the Nyquist resolution, a measure of resolution that incorporates the spatial density of localized molecules, was 77 nm (SI Appendix 7).

Although each cotton linter had a different size, shape, and position relative to the coverslip, general trends between crystalline- and amorphous-specific CBMs were observable by eye. Crystalline-specific CBMs (CBM1Cel7A, CBM1Cel6A, CBM2A, CBM3A) tended to cluster on ridges (Figs. 6.1C-1G), occasionally highlighting linear regions running parallel to one another (Figs. 6.1E). Amorphous-specific CBMs, bound more homogeneously across the surface (Figs. 6.1H-1J), occasionally localizing to sinuous protrusions from the fiber edges (Fig. 6.1H). Dual color superresolution imaging showed that the sample-to-sample structural variability in cotton linters was not the primary source of the observed differences in binding behavior between domains. In Figs. 6.1L and 1M, a two color PALM image of PACFP2-CBM1*Tr*Cel6A and mEos2-CBM28 bound to the same cotton linter shows contrast between the preferential localization of CBM1*Tr*Cel6A (blue) to edges and the more uniform binding of CBM28 (red).

To investigate subtle organizational differences of CBMs adsorbed on the cellulose surface, we sought a metric to quantify CBM-substrate binding patterns. Because the substrate is differentially ordered on many length scales, we required this metric to be a function of length scale and to incorporate the directional dependence of CBM arrangements. For each PALM image, circular regions with 3.1 μm diameters were selected (Fig. 6.2A). Points within these regions were projected onto eight axes: one corresponding to the direction of the principal component of the points within the image, and seven corresponding to axes located at various angles (multiples of 22.5°) to the principal component (Fig. 6.2B). Pair correlation functions (PCF's) were then constructed along each axis, and the distance between the y-intercepts of the two PCF's that differed most in amplitude was calculated (Fig. 6.2C). This number, which we term the "CBM order parameter" and represent as Ω, indicates the local directional dependence of a CBM's spatial organization on the cellulose surface. The local emphasis is a result of using the PCF y-intercepts, which weight the organizational differences of nearest neighbors. For each CBM, an overall order parameter value Ω was determined by averaging the individually calculated values from at least 15 different circular regions within 3 separate PALM datasets (Table 2). Examples of circular regions from PALM datasets are shown in Figs. 6.2D-2I with their associated order parameter.

CBM	Order Parameter at 0.12 µmol/g	Order Parameter at 0.24 µmol/g
1*Tr*Cel7A	0.85(0.23)	0.85(0.25)
1*Tr*Cel6A	0.74(0.22)	0.52(0.12)
2A	0.58(0.15)	0.52(0.11)
3A	0.47(0.22)	0.41(0.09)
17	0.44(0.09)	0.72(0.16)
28	0.34(0.06)	0.71(0.13)

Table 6.2 Order parameters calculated for spatial point patterns from PALM images of mEos2-CBM constructs bound to cotton linters at loadings of 0.12 µmol protein/g cellulose and 0.24 µmol protein/g cellulose.

Order parameters indicate the binding site preferences of CBMs by revealing the organization of the structures to which they bind. The targets of CBM28 and CBM1*Tr*Cel7A represent two extremes in a morphological spectrum (Fig. 6.3A). At a low loading, CBM28, which isothermal titration calorimetry experiments have shown binds tightly to single glycan chains, had an order parameter of 0.34, and CBM*Tr*Cel7A, which electron microscopy studies have shown binds to the 110 hydrophobic face of crystalline cellulose microfibrils, had an order parameter of 0.85 (*22, 23*). PALM reveals that CBM28 binds broadly across the surface without highlighting rough surface features and that CBM1*Tr*Cel7A clusters along edges and along linear ridges. The low order parameter associated with CBM28 appears to result from broadly distributed single chains; whereas, the high order parameter associated with CBM1*Tr*Cel7A appears to result from a concentration of ordered microfibrils within edges and ridges. This link between order parameters and CBM affinity for such structures was confirmed with an independent spatial adsorption model (SI Appendix 3). The intermediate order parameters describing the arrangements of CBM17, CBM3A, CBM2A, and CBM1*Tr*Cel6A are indicative of intermediary binding site preferences.

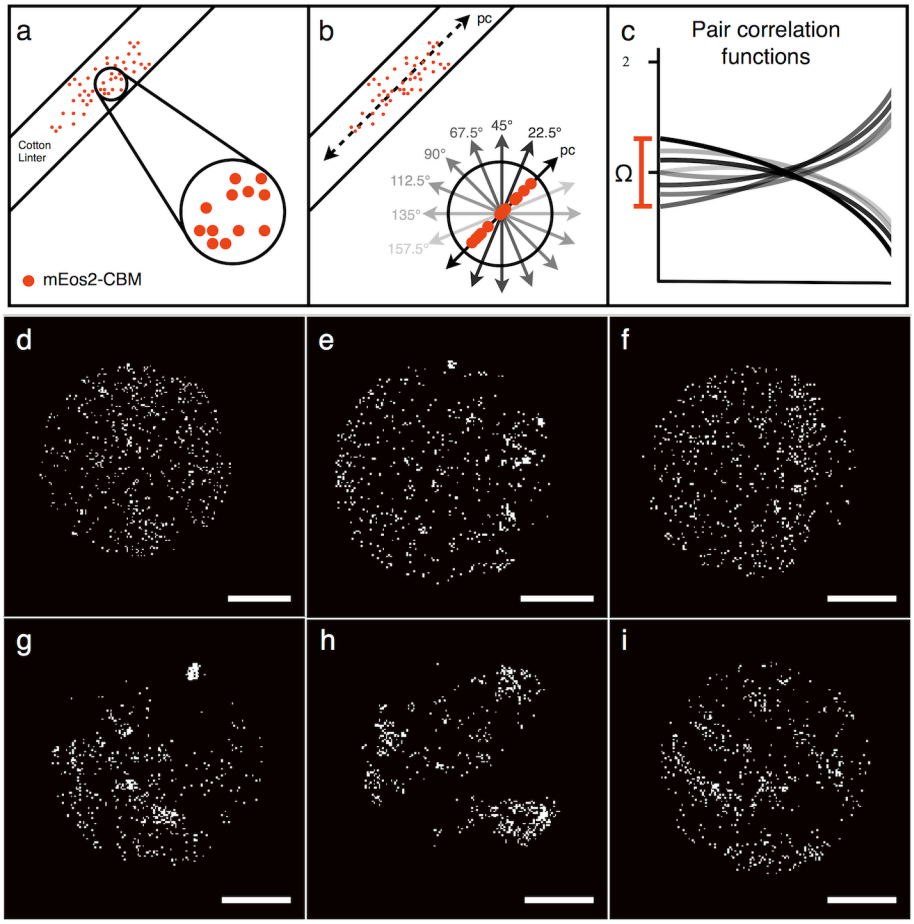


Figure 6.2 The CBM order parameter. (a-c) Steps for analyzing spatial distributions of CBMs on the cotton surface. In (c), the difference between the extrema of the y-intercepts is indicated with an orange bracket. (d-i) Example circular regions from PALM images of each mEos2-CBM construct (1 μm scale bars): (d) mEos2-CBM28, (e) mEos2-CBM17, (f) mEos2-CBM3A, (g) mEo2-CBM2A, (h) mEo2-CBM*Tr*Cel6A, and (i) mEos2-Cm*Tr*Cel7A. Only 700 points from each region have been plotted to facilitate comparison.

The binding behaviors suggested by the PALM order parameter values are consistent with published adsorption studies. The low order parameters of CBM28 and CBM17 (respectively, 0.34 and 0.44) agree with reports that these CBMs bind chains within disordered regions, but that CBM17 has a higher affinity for crystalline cellulose than does CBM28 (*23-25*). Analogously, the high order parameters of CBM1*Tr*Cel7A and CBM1*Tr*Cel6A (respectively, 0.85 and 0.74) are indicative of a high affinity for crystalline ridges, which these homologous CBMs have been shown to have (*22*). The lower Ω value of CBM1*Tr*Cel6A may be explained by this CBM's ability to accommodate greater morphological diversity than CBM1*Tr*Cel7A; indeed, a study by Linder *et al* showed that CBM1*Tr*Cel6A binds chitin, a polymer of N-acetylglucosamine, with a 3-fold greater affinity than CBM1*Tr*Cel7A (*26*). Finally, the order parameters associated with CBM2a and CBM3a (respectively, 0.58 and 0.47) agree with previous binding studies suggesting that CBMs from these families can bind crystalline and non-crystalline substrates, but that CBM2A has a higher affinity for highly crystalline substrates than does CBM3A (*6*). The consistency between the PALM order parameter results (Figs. 6.1 and 6.2) and previous adsorption studies substantiates our interpretation of the order parameter as an indicator of morphological binding specificity.

The sensitivity of the order parameters to CBM loading is indicative of the distribution of high and low affinity binding sites. When the CBM/cellulose loading was increased, the order parameters associated with CBMs 1*Tr*Cel7A, 1*Tr*Cel6A, 2A, and 3A stayed the same or decreased slightly, suggesting that, at the higher loading, these CBMs adsorbed not only to their high affinity sites, but also to some of the more broadly distributed, amorphous-like regions which they have been shown to bind with a lower affinity (*6*). By contrast, the order parameters associated with CBMs 17 and 28 increased, suggesting that these CBMs began to occupy crystalline regions at higher loadings. Such a transition is consistent with previous measurements of the binding site densities for CBM28 and CBM17 on crystalline substrates, which range from 0.08 to 0.26 μmol/g for high affinity sites (likely single chains) and 0.79 to 5.01 for low affinity sites (likely crystalline regions) (*23*).

To relate CBM structural targets and enzymatic activity, we synthesized six endoglucanase chimeras differing only by their CBM domain and assayed their catalytic activity individually and alongside one another for all 15 possible binary combinations. Endoglucanases cleave within cellulose chains by general acid-catalyzed hydrolysis of the β-[1,4]-bonds that link glucose subunits (*27*). Each CBM was connected to the catalytic domain of the *Acidothermus cellulolyticus* GH5 (*Ac*GH5) endoglucanase via the *T. reesei* Cel7A linker (SI Materials and Methods). *Ac*GH5 is a highly active, thermostable endoglucanase (*28*).

The chimeras were assayed against cotton linters at 18°C for 45 hours and at 50°C for 5 hours at GH5-CBM/cotton loadings identical to the mEos2-CBM/cotton loadings of the PALM experiments (SI Materials and Methods). The 18°C temperature was chosen to match the temperature of the room in which the microscopy was carried out; 50°C was chosen to assess the influence of temperature on the performance of each enzyme mixture. Assay durations were based on Arrhenius extrapolations from a 2-hour initial rate regime of *Ac*GH5 on cellulosic substrates at 70°C (*29, 30*) (SI Appendix 4). Soluble sugars released were measured with a glucose oxidase / horseradish peroxidase assay (SI Materials and Methods). The synergy exhibited by each binary combination was calculated as the concentration of soluble sugars released by a combination divided by the concentration of sugars expected from purely additive behavior between the two enzymes (Fig. 6.3B, SI Appendix 6).

The activity of binary GH5-CBM mixtures relative to one another was insensitive to the enzyme loadings and temperatures examined, suggesting a dilute adsorption scenario where the concentrations of total and adsorbed enzyme were approximately equal (SI Appendix 5, Fig. 6.A7). The absence of a shift in the behavior of GH5-CBM28 and GH5-CBM17 between low and high loadings suggests that the activity of the GH5 catalytic domain is similar when it is located near the high and low affinity sites of these CBMs.

A heat map showing the synergy of each binary combination reveals five pairs that exhibit synergy greater than 1.2 (Fig. 6.3B). The synergy of these binary pairs cannot be predicted from their amorphous or crystalline binding classification. Surprisingly, however, when the synergy was plotted against the difference in order parameters ($\Delta\Omega = \Omega_a - \Omega_b$), a pattern emerged (Fig. 6.3C). Of the five highly synergistic combinations, four were composed of GH5-CBM constructs with $\Delta\Omega$s between 0.10 and 0.25. By contrast, the eight combinations with an average synergy of ~ 1 had $\Delta\Omega$s that were either very small ($\Delta\Omega \approx 0$) or very large ($\Delta\Omega >$ 0.25). This pattern is consistent with a dilute adsorption regime where substrate is abundant. When two enzymes target identical sites (*i.e.,* $\Delta\Omega \approx 0$), the mixture behavior is similar to the

behavior of a higher concentration of either enzyme; the combined activity is thus additive (synergy ≈ 1). When two enzymes target very different sites (*i.e.,* large ΔΩ), they hydrolyze cellulose independently, and the combined activity is again additive (synergy ≈ 1). By contrast, when two CBMs bind similar but non-identical targets (0.10 < ΔΩ < 0.25), one enzyme can enhance the susceptibility of the substrate to the action of the other, and the combined activity that results is superadditive, or synergistic.

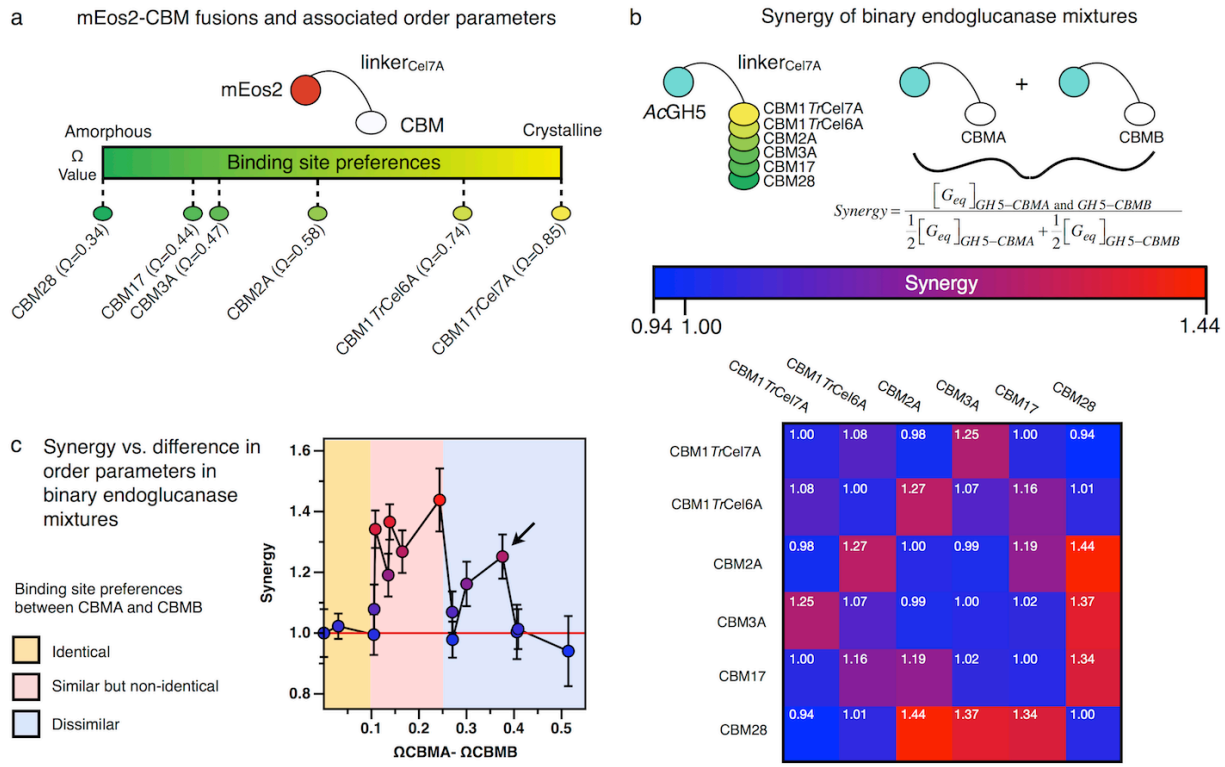


Figure 6.3 Order parameters and synergy values. (a) Diagram of mEos2-CBM architecture above a scale showing the range of CBM order parameters calculated from PALM images of mEos2-CBM constructs bound to cotton linters at 0.12 μmol protein/g cellulose loadings. (b) Diagram of GH5-CBM architecture above a heat map showing the average synergy of each binary GH5-CBM mixture. In the synergy equation, G_{eq} is the glucose equivalent concentration of soluble sugars released from the cotton linters, and CBMA and CBMB represent two non-identical CBMs. Average synergy values are based on the individual synergy values for reaction mixtures assayed on 50.8 mg/L cotton at pH 4.85 under four conditions: 0.12 μmol enzyme/g cellulose for 5 hours at 50°C, 0.24 μmol enzyme/g for 5 hours at 50°C; 0.12 μmol enzyme/g for 45 hours at 18°C; and 0.24 μmol enzyme/g for 45 hours at 18°C. (c) Average synergy values plotted against the difference in order parameters of the two relevant CBMs (Ω_{CBMA}- Ω_{CBMB}) at the 0.12 μmol protein/g cellulose loadings (within a difference range of 0 to 0.51). All points are colored as in the heat map from (b). Zones of identical, similar, and dissimilar binding behavior have been shaded with orange, red, and blue. An arrow has been used to indicate the CBM3A/CBM1*Tr*Cel7A outlier.

Consideration of the major outlier to the ΔΩ-synergy pattern is informative. The CBM1*Tr*Cel7A and CBM3A pair has ΔΩ = 0.38, and yet shows synergistic behavior (Fig. 6.3C, arrow). Previous reports have shown that both CBMs are capable of adsorbing to the hydrophobic 110 face of cellulose microfibrils (*22*) but that CBM3A can bind a multitude of other sites *in addition* to the 110 hydrophobic face. CBM3A's ability to bind a range of structures is consistent with our results (Fig. 6.A3) and those of Carrard *et al* (*11*), which show that CBM3A confers to GH5 enzymes a higher activity on a variety of substrates than do other CBMs. Thus, the relaxed substrate fidelity of CBM3A increases the ΔΩ of the

CBM3A/CBM1*Tr*Cel7A pair, obscuring the ability of these CBMs to bind similar but non-identical sites. When the PALM-based ΔΩ analysis is employed to optimize cellulolytic mixtures, some good synergistic combinations, such as CBM3A/CBM1*Tr*Cel7A, will be missed unless the Ω results are supplemented with adsorption isotherms showing which CBMs are capable of binding a wide range of morphological targets.

6.4 Conclusions

The PALM-based order parameter provides a new technique for assessing the similarity between CBM binding site targets and, in doing so, reveals a strategy for improving the performance of cellulase mixtures. The advantages conferred by multiple CBMs to the *Ac*GH5 endoglucanase have implications for cellobiohydrolases and polysaccharide monooxygenases, the other two main constituents of cellulolytic mixtures. Our results suggest that combinations of cellulolytic enzyme variants that differ only in their binding targets can exhibit higher activities than single enzymes. The inclusion of such variants within cellulolytic enzyme cocktails may improve the overall activity of these mixtures.

6.5 Acknowledgements

We thank Ann L. McEvoy for providing image processing software and J.W. Chu, A. S. Gross, K. Haas, and Ann L. McEvoy for helpful discussions. J. M. F. is the recipient of an NSF pre-doctoral fellowship, and J.L., H.W.B, and D.S.C. acknowledge support from the Energy Biosciences Institute.

6.6 Materials and Methods

6.6.1 Cellulose Preparation

Cotton balls (Walgreens) were de-waxed with a Thermo Scientific ASE 350 Accelerated Solvent Extractor. Cotton was packed into 100 mL steel extraction cells, which were then saturated with chloroform at 100°C and 10 MPa for 5 minutes. This process was repeated a total of three times, and between each 5-minute holding time, 40% of the column volume was emptied and replaced with fresh chloroform. After the third round of extraction, a 60-second nitrogen purge was used to remove residual solvent.

Cotton linters were generated through pulsed blending and subsequent washing of the de-waxed cotton. The chloroform-extracted cotton balls were submerged in 1 L of water, and the resulting suspension was blended for 1.5 minutes on high in a 1 L Waring Blender. Cotton was removed from the blade and cut with sterile scissors. This blending/cutting process was repeated five times. The final suspension was mixed with sodium hypochlorite to create a final aqueous solution of 2% sodium hypochlorite. This cotton suspension was heated for 2 hours at 70°C and cooled to 4°C overnight. The final bleached product was spun down at 6000 rpm in 1 L centrifuge containers for 20 minutes, the supernatant was decanted, and replaced with Nanopure water. This washing process was repeated four times. The final 3L suspension was mixed at 150 rpm on a stir plate, and the top 500 ml was withdrawn. This was enriched in small cotton linters; larger particles and clumps fell to the bottom. The final product was stored in 0.06% sodium azide to prevent contamination by microorganisms. The final cellulose concentration was determined with the phenol / sulfuric acid method as described previously (*33*).

6.6.2 Design of Fusion Constructs

Fusion constructs were generated with overlap extension PCR. For CBM1*Tr*Cel7A, CBM2A, CBM17, and CBM28, the mEos2 and *Ac*GH5 domains were attached to the N-terminus of the CBMs via the following primers:

Forward primer for mEos2:	AAAAACCATGGGGGCGATTAAGCCAGACAT
Reverse primer for mEos2:	ACCCGGGCTGCTACCGGTGGTGGTCGCCGGACGACGGGT-GGTGGTGGTACCACGGGATCCTCGTCTGGCATTGTCAGG
Forward primer for *Ac*GH5:	AAAACCATGTCGGGCGGTGGTTATTGGC
Reverse primer for *Ac*GH5:	ACCCGGGCTGCTACCGGTGGTGGTCGCCGGACGACGG-GTGGTGGTGGTACCACGACCAACAGGATCGAAGATC-GACGATTTA
Forward primer for CBM1*Tr*Cel7A:	CGTGGTACCACCACCACCCGTCGTCCGGCGACCACCACC
Reverse primer for CBM1*Tr*Cel7A:	GAATTCTTAATGATGATGATGATGATGTTTATCA-TCATCATCCAGGCACTGGC
Forward primer for CBM2A:	CCACCCGTCGTCCGGCGACCACCACCGGTAGCAGCCCG-GGTCCGACCCAGAGCACTGCGGCGAGCGGC
Reverse primer for CBM2A:	GAATTCTTAATGATGATGATGATGATGGGTTTTATC-ATCATCATCACTCGCCGCGCACGC
Forward primer for CBM17:	ACCCGTCGTCCGGCGACCACCACCGGTAGCAGCC-CGGGTCCGACCCAGAGCCAACCGACCGCCCCGAAA
Reverse primer for CBM17:	AAAAAAGAATTCTTAATGGTGATGGTGATGGTGG
Forward primer for CBM28:	ACCCGTCGTCCGGCGACCACCACCGGTAGCAGCCCGGGT-CCGACCCAGAGCGGCACCGAAGTGGAAATTCCG
Reverse primer for CBM28:	AAAAAAGAATTCTTAATGATGATGATGATGATGGG TTTTATCATC

For CBM1*Tr*Cel6A and CBM3A, the mEos2 and *Ac*GH5 domains were attached to the C-terminus of the CBMs via the following primers:

Forward primer for mEos2:	ACCACCCGTCGTCCGGCGACCACCACCGGTAGCAGC-CCGGGTCCGACCCAGAGCATGGGGGCGATTAAGCCA-GACAT
Reverse primer for mEos2:	GAATTCTTAGGATCCTCGTCTGGCATTGTC
Forward primer for *Ac*GH5:	CCACCCGTCGTCCGGCGACCACCACCGGTAGCAGCC CGGGTCCGACCCAGAGCTCGGGCGGTGGTTATTGGC
Reverse primer for *Ac*GH5:	GAATTCTTAACCAACAGGATCGAAGATCGACGATTTA
Forward primer for CBM1*Tr*Cel6A:	AAAACCATGGCACATCATCATCATCATCATACCG-ATGATGATGATAAAACCCAGGCGTGCAGCAGCG
Reverse primer for CBM1*Tr*Cel6A:	ACCCGGGCTGCTACCGGTGGTGGTCGCCGGACGACG GGTGGTGGTGGTACCACGCAGGCACTGGCTATAATAA TCGTTGCT
Forward primer for CBM3A:	AAAACCATGGCACATCATCATCATCATCATACC-GATGATGATGATAAAGGCATGGCGAATACCCCGGTGTCT
Reverse primer for CBM3A:	ACCCGGGCTGCTACCGGTGGTGGTCGCCGGACGAC GGGTGGTGGTGGTACCACGCGGTTCTTTACCCC AAACCAGCAC

For CBM1*Tr*Cel6A, the PSCFP2 domain was attached to the C-terminus of the CBM via the following primers:

Forward primer for PSCFP2:	CCACCCGTCGTCCGGCGACCACCACCGGTAGCAGC-CCGGGTCCGACCCAGAGCATGAGCAAGGGCGCCG
Reverse primer for PSCFP2:	AAAAAGAATTCTCACTTGTACAGCTCATCCATGCCGTG

Each half of each construct was subcloned within a separate amplification reaction consisting of the relevant template and the two relevant primers. Fusion constructs were generated in amplification reactions consisting of the two parent templates and the non-overlapping primers corresponding to the N- and C-termini of the complete protein. The product of each final amplification reaction was gel purified, digested with NcoI-HF and EcoRI-HF restriction enzymes (New England Biolabs), and ligated into a pTrcHisC plasmid (Invitrogen).

6.6.3 Expression and Purification of Fusion Constructs

Fusion proteins were expressed in Origami B *E. coli*, which is capable of forming disulfide bonds. Top10 chemically competent *E. coli* cells (Invitrogen) were transformed with the product of each ligation reaction and grown on LB plates containing 100 mg/L of Carbenicillin. A subset of the resulting colonies were checked for error-free plasmids via sequencing, and an error-free plasmid for each construct was used to transform Origami B *E. coli* (Novagen). For each construct, a plasmid-containing colony of Origami B was grown up overnight in LB media containing 50 mg/L of Carbenicillin (37°C shaker). Approximately 20 ml of overnight culture was used to inoculate 2 L of LB media, which was grown in a 37°C shaker in the dark. The 2L cultures were induced at an OD_{600} of approximately 0.65 with 1 mM IPTG, and induction was carried out for 18 hours in the dark. Stopped cultures were spun down in 1 L centrifuge tubes at 6000 rpm for 25 minutes, and pellets were lysed with a homogenizer.

Fluorescent fusion proteins were purified with a fast protein liquid chromatography system (ATKA Explorer, GE Healthcare) in the dark. Cell lysate from each 2 L culture was equilibrated with sodium phosphate buffer (20 mM sodium phosphate, 500 mM NaCl, pH 8.10), and the final mixture was filtered with a 0.45 μm sterile filter. This solution was loaded onto a 5 mL His-FF column (GE Healthcare) attached to a refrigerated FPLC, and the purified protein was eluted with a stepwise gradient of sodium phosphate buffer containing imidazole (20 mM sodium phosphate, 500 mM NaCl, 500 mM Imidazole, pH 8.10). The eluted fraction was concentrated with a 10 kDa centrifuge concentrator and resuspended in Tris-HCl buffer (20 mM Tris-HCl, 200 mM NaCl, 2 mM $CaCl_2$, pH 7.2). To remove the polyhistidine tag from each fusion construct, recombinant light chain enterokinase (New England Biolabs) was incubated with each protein solution in a ratio of 0.001 wt/wt % (enterokinase/protein) for 8 hours at room temperature in the dark. The resulting solution was again loaded onto the 5 ml His-FF column, and the flow-through was collected. In several cases where this flow-through showed contaminants on an SDS-denaturing gel stained with Coomassie Blue, an additional Mono Q 10/100 GL anion exchange column was used to ensure single band purity (gradient elution with 25 mM HEPES, pH 7.35, 1.0 M NaCl). Non-fluorescent fusion proteins were purified in an identical manner, but without the protection from light.

6.6.4 Instrumentation for PALM

PALM imaging was performed on an Olympus IX81 inverted microscope that was customized as in reference (*34*). Imaging was carried out with a 63X 1.49 N.A. oil immersion objective with Type DF nondrying immersion oil (Cat. No 16242 Cargille Laboratories) and glass coverslips (No. 1.5, VWR 48336-227). The arrangements of the 561 nm laser (Coherent compasss-561-100) and 488 nm laser (Coherent Sapphire 488-100) were similar within this setup. Beams from each of these lasers were passed through laser clean up filters (Semrock #LL02-561-12.5 and Semrock #LL01-488-12.5 respectively) and expanded to beam diameters of approximately 3mm. A mixing filter (Semrock #LM01-552-25) was used to combine these beams onto a mutual path that passed through an acousto-optic tunable filter or AOTF (AA optoelectronic **AOTFnC-VIS**). The AOTF allowed for pulsed laser imaging and facilitated adjustments to laser power reaching the sample. The 405nm laser (Coherent cube-405-100) was passed through a clean up filter (Semrock LD01-405/10-12.5), expanded to approximately 3mm and combined with the 561nm and 488nm beams using another laser mixing filter (Semrock LM01-427-25). This combined beam was then expanded and passed through a series of lenses to provide an appropriately–sized TIRF spot on the microscope objective (Olympus APON 60X0TIRF).

6.6.5 Imaging Procedure for PALM

Sample preparation for PALM imaging took approximately 2.5 hours. For each sample, a 500-μL suspension with the following composition was prepared in a 1.5 ml Eppendorf tube: 40 mg/L cotton, 1X PBS buffer (137 mM NaCl, 10 mM Phosphate, 2.7 mM KCl, and 200 μM $CaCl_2$, pH 7.4), and 4.8 or 9.6 nM mEos2-CBM construct, depending on whether or not the loading was 0.12 or 0.24 μmol/g. This suspension was incubated in the dark (covered in foil) at room temperature for 2 hours on a rocking platform. After this time, 400 μL of suspension was mixed with 4 μL of 25 nm gold nanorods (550 LSPR, Nanopartz), and the resulting mixture was spun down in an imaging well containing a 22x22 mm glass coverslip (No. 1.5, VWR 48336-227) at 2500 rpm for 20 minutes at 18°C.

Imaging was performed in TIRF mode at ~18°C. Under illumination from approximately 12mW of 561 nm laser light, a cotton section was identified that contained both (1) a fluorophore-coated region spanning several microns and (2) detectable gold nanorods (for drift correction). This image was photographed under bright field illumination. Under TIRF mode, this section was then subjected to 561 nm laser again for 5 minutes prior to imaging. Through this process, many of the mEos2 fluorophores activated through exposure to the bright field light were bleached out. Imaging was then carried out through simultaneous illumination with 12mW 561 nm and 2mW of 405 nm light; illumination was pulsed at 20 Hz with an AOTF. The dichroic and emission filters were Chroma T495lp and Semrock optics 617/73, respectively. Twenty images were captured every second for 85 minutes. During this period, when the frequency with which activated mEos2 molecules were detected began to dwindle, the power of the 405 nm laser was increased.

For the dual color image, several subsequent steps were taken. After the 85-minute illumination period, the 405 nm laser was turned off for 5 minutes, and recording was continued. During this period, much of the activated mEos2 was bleached out. The 488 nm laser, set to 10mW, was then turned on, and the dichroic and emission filters were changed to Semrock

FF562-Di02 and Chroma HQ525/50M, respectively. The 561 nm laser was turned off, and the 488 nm laser was held on for 5 minutes to bleach out the previously activated PSCFP2 molecules. Imaging began with simultaneous illumination with 488 and 405 nm lasers (at 10 and 4 mW, respectively) for 85 more minutes; illumination was pulsed at 20 Hz with the AOTF as before. During this period, when the frequency with which activated PSCFP2 molecules were detected began to dwindle, the power of the 405 nm laser was increased.

Drift correction was accomplished through tracking of 25 nm gold nanorods (550 LSPR, Nanopartz), which were visible in each channel. Localization and image rendering was carried out as described elsewhere (*15, 34*).

6.6.6 Confocal Imaging for Fig. 6.1B

Sample preparation for confocal imaging took approximately 2 hours. A 500-µL suspension with the following composition was prepared in a 1.5 ml Eppendorf tube: 40 mg/L cotton, 1X PBS buffer as above, and 9.6 nM each of mEos2-CBM28 and PSCFP2-CBM1*Tr*Cel6A construct, for a loading of 0.24 µmol each construct/g cotton. This suspension was incubated in the dark (covered in foil) at room temperature for 2 hours. After this time, 30 µL of suspension was placed onto a glass microscope slide (Fisherbrand 12-550-343) and covered with a glass coverslip (Fisherbrand 24X30-1.5X) prior to imaging. Confocal imaging was carried out with a 63X PL APO N.A. 1.2 W objective attached to a Leica DIM6000B microscope equipped with a CSU-X1 confocal scanner unit (Yokogawa), a Quantum 512 SC camera (Photometrics), and 442 and 488 nm lasers controlled by Metamorph Software. Under DIC, 442 nm confocal, and 488 nm confocal imaging conditions, 30 z-sections of thickness 0.384 µm/section were collected separately; 20 from each illumination condition were reassembled with ImageJ (http://rsbweb.nih.gov/ij/).

6.6.7 Hydrolysis Assays of GH5-CBM Constructs on Cotton Linters

Hydrolysis assays were conducted in round-bottom 96-well plates. Each plate consisted of 100-µL reactions containing 50.8 mg/L cotton, sodium acetate buffer (50 mM sodium acetate, 500 µM $CaCl_2$, pH 4.85), and either 6 nM or 12 nM of GH5-CBM enzyme, depending on whether the enzyme/cellulose loading was 0.12 or 0.24 µmol/g, in the ratios described in Table 3. All reactions of a given loading were conducted in triplicate within a single plate. The remaining 15 reactions consisted of an enzyme-free control conducted in triplicate, and three known concentrations of glucose (0 µM, 5 µM, and 10 µM) conducted in quadruplicate. A thermal sealer was used to seal each plate with foil prior to incubation at 18°C for 45 hours or 50°C for 5 hours. A total of four 96-well plates were used, one at each loading for each temperature.

6.6.8 Glucose Oxidase / Horseradish Peroxidase Assay

The glucose-equivalent concentration of soluble sugars released by the activity of GH5-CBM enzymes on cotton was measured with the glucose oxidase / horseradish peroxidase/ Amplex Red assay. From each 100 µL reaction, 16 µL of solution was withdrawn and reacted with 16 µL of 5 g/L β-glucosidase solution in a separate black, round-bottom 96-well plate at room temperature for 1 hour. During this period, the β-glucosidase hydrolyzed all soluble cello-oligosaccharides into glucose. To the resulting mixture, 64 µL of a solution containing 12.5 U glucose oxidase, 12.5 U horseradish peroxidase, 62.5 µM Amplex UltraRed reagent (Invitrogen),

and 125 mM HEPES buffer (pH 7.3) was added. This mixture was given 20 minutes to react in the dark, and the fluorescence at 595 nm was measured in a Paradigm plate reader (Beckman Coulter). Glucose concentrations were calculated from fluorescence intensity measurements based on known glucose concentrations that were subjected to the same treatment as the reaction mixtures.

For each of the four 96-well plates used in the hydrolysis assays, the soluble sugar assay was conducted four times. The values that are reported in Fig. 6.A7 correspond to interquartile averages of the 12 data points corresponding to the activity of each binary mixture assayed under a given set of conditions. Standard rules for error propagation were used to calculate error for synergy and average synergy calculations.

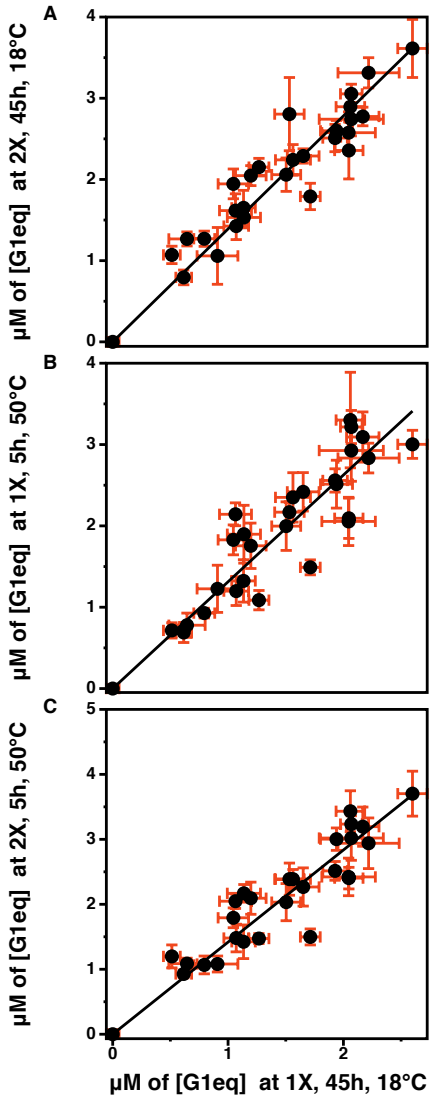


Fig. 6.A7 A comparison of soluble sugars released (measured in glucose equivalent) from 50.8 mg/L of cotton linters at pH 4.85 by the GH5-CBM combinations listed in Table 6.A3 under different reaction conditions. The x-axis corresponds to the soluble sugars released by a total GH5-CBM loading of 0.12 µmol/g cellulose (1X) for 45 hours at 18°C. The y-axes correspond to the soluble sugars released by (A) 0.24 µmol enzyme/g cellulose (2X) for 45 hours at 18°C, (B) 0.12 µmol enzyme/g cellulose (1X) for 5 hours at 50°C, and (C) 0.24 µmol enzyme/g cellulose (2X) for 5 hours at 50°C. The correlation coefficients for the lines are 0.88 (A), 0.82 (B), and 0.85 (C).

6.7 Appendix 1: Filtration of Data

Points that were weakly detected or likely to be erroneous were removed prior to analysis. Weakly detected points were removed through the exclusion of fluorophores that were localized with less than 14 photons or with a Gaussian fit standard deviation of 2 pixels or more. Bright contaminants such as gold nanorods and dirt were removed through the exclusion of fluorophores that were localized with more than 162 photons. Double counting of reactivated fluorophores, which have been shown to occur with mEos2 (*35*), was prevented through the removal of points occurring within 11 nm of any point within a 30-second time period after its initial appearance.

6.8 Appendix 2: Detail of Mathematical Analysis

Points within each PALM image were subjected to a mathematical analysis designed to reveal the directional dependence of CBM organization on the cellulose surface. This analysis consisted of the following steps:

1. Calculation of the principal component of all points within the image using the MATLAB function *princomp*. The vector returned is normalized (a unit vector).
2. Calculation of seven additional unit vectors located at multiples of 22.5° to the principal component:
$$\begin{bmatrix} P_1(\theta)' \\ P_2(\theta)' \end{bmatrix} = \begin{bmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{bmatrix} \times \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} \quad (6.A1)$$
where (P_1,P_2) is the principal component unit vector, and $(P_1(\theta)',P_2'(\theta))$ is a unit vector located at angle θ to the principal component. Values of θ included 67.5°, 45°, 22.5°, 0°, -22.5°, -45°, and -90°. Thus, a total of eight unit vectors were calculated.
3. Selection of the eight densest regions with a diameter of 14 pixels (3.1μm). These have no relation to the eight unit vectors.
4. Projection of the points within each circle onto each of the eight unit vectors:
$$\begin{bmatrix} P_1(\theta)' \cdot P_1(\theta)' & P_1(\theta)' \cdot P_2(\theta)' \\ P_1(\theta)' \cdot P_2(\theta)' & P_2(\theta)' \cdot P_2(\theta)' \end{bmatrix} \times \begin{bmatrix} x_1 & y_1 \\ \vdots & \vdots \\ x_n & y_n \end{bmatrix}^T = \begin{bmatrix} x_1' & \cdots & x_n' \\ y_1' & \cdots & y_n' \end{bmatrix} \quad (6.A2)$$
where each (x_i',y_i') pair corresponds to each (x_i,y_i) pair.
5. Construction of a pair correlation function within each circular region for the points along each axis:
$$g_\theta(x) = \rho_\theta(x)dx \quad (6.A3)$$
where g_θ is a pair correlation corresponding to the axis located at angle θ to the principal component, and ρ_θ is the number density of the points projected along the axis.
6. Calculation of the difference in y-intercepts of the pair correlation functions g_θ and g_θ' which have the greatest difference in y-intercept. This number, which we term the "CBM order parameter" and represent as Ω, indicates the local directional dependence of a CBM's organization on the cellulose surface. The local emphasis is a result of the y-intercept, which weights the organizational differences of nearest neighbors.

Note that steps 4,5, and 6 are carried out separately for each circular region.

After steps 1-6 were carried out for all 36 PALM images, order parameters corresponding to anomalous circular regions (e. g., those located off the main fiber or near blanked gold nanorod locations) were removed through two steps:

1. Removal of data from circular regions on cellulosic material that was not located on the main cotton linter within a PALM image.
2. Removal of order parameter values lower than 0.15 and higher than 1.5, which appeared to correspond to discontinuities in an range of order parameter values that was otherwise continuous between these two limits (Fig. 6.A1).

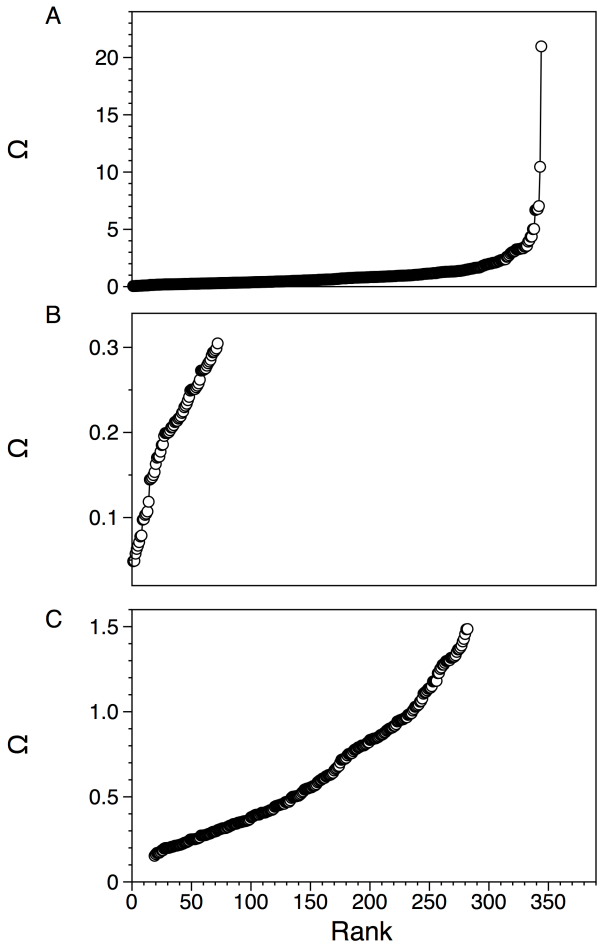


Figure 6.A1 Order parameter values from 0.12 and 0.24 μmol protein/g cellulose loadings are plotted from smallest to largest. Plots include various subsets as follows: (A) all order parameters, (B) order parameters lower than 0.3, and (C) order parameters between 0.15 and 1.5 (the order parameter values used in our analysis).

Each of the final order parameter values listed in Table 6.2 corresponds to an interquartile average of at least 15 order parameters calculated from circular regions within three separate PALM images.

6.9 Appendix 3: Model of CBM Adsorption to cellulose

A model was employed to investigate the connection between binding site locations and order parameter values. Adsorption of CBMs with varying degrees specificity for crystalline and amorphous regions was modeled by representing cellulose as a square where these morphologies were present within ridge and non-ridge sections, respectively (Fig. 6.A2). The CBMs were partitioned onto ridge and non-ridge regions in accordance with the preference for binding each, but their position within each type of region was random. Order parameters for the point distributions produced by the model were calculated in a fashion identical to that used for PALM images.

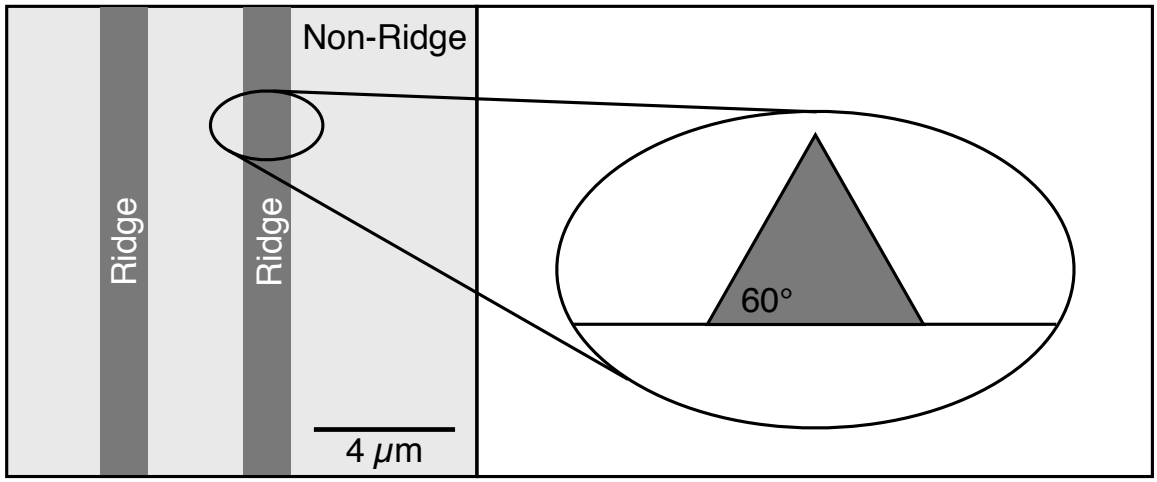


Figure 6.A2 Diagram of the heterogeneous cellulose square upon which CBM adsorption was modeled. A close-up shows the Z-profile of the surface. Ridge regions have cross-sections which resemble equilateral triangles. When viewed from overhead, the ridge area available for a CBM to bind is twice that of the observable area (the area of the triangle base).

The enhanced surface area associated with ridge-like features, which protrude into the z-direction of the 100-nm thick imaging plane, was simulated by modeling ridges as having equilateral triangle cross-sections (Fig. 6.A2). When viewed from overhead, a ridge will appear to have a surface area equal to that of its base area. To account for this within the model, each triangular ridge was given an area equal to twice that of its base (the area observed); this is the area of the two exposed ridge sides when the angle between them is 60°. This accounting is important to the adsorption calculations below.

Adsorption was modeled under dilute conditions. The linear regime of a two-site Langmuir binding isotherm was assumed:

$$[CBM^*] \approx K_1[*_1]_0[CBM] + K_2[*_2]_0[CBM] \qquad (6.A4)$$

Equivalently,

$$[CBM^*] \approx \frac{K_c[*_c]_0[CBM]_0}{1+K_c[*_c]_0+K_a[*_a]_0} + \frac{K_a[*_a]_0[CBM]_0}{1+K_c[*_c]_0+K_a[*_a]_0} \qquad (6.A5)$$

where $*_c$ and $*_a$ represent the total number of binding sites located within the ridge and non-ridge regions, respectively, and K_c and K_a are the associated binding constants for these regions; [CBM*] is the total concentration of adsorbed CBM (ridge and non-ridge regions). Based on the discussion in Appendix 5, we further assume that all of the CBM adsorbs to the cellulose surface:

$$[CBM^*] \approx \frac{K_c[*_c]_0[CBM]_0}{K_c[*_c]_0+K_a[*_a]_0} + \frac{K_a[*_a]_0[CBM]_0}{K_c[*_c]_0+K_a[*]_0} \quad (6.A6)$$

where

$$\frac{K_c[*_c]_0}{K_c[*_c]_0+K_a[*_a]_0} = fraction\ bound\ to\ ridges \quad (6.A7)$$

and

$$\frac{K_a[*_a]_0}{K_c[*_c]_0+K_a[*_a]_0} = fraction\ bound\ to\ non-ridge\ regions \quad (6.A8)$$

Within the model, Eqs. S4 and S5 are used to distribute CBMs onto ridge and non-ridge regions of a 13.33 x 13.33 μm square of cellulose containing two 1.33 x 13.33 μm ridges spaced 2.66 μm apart (Fig. 6.A2). These dimensions are similar to the dimensions of surface features revealed by PALM images. Values of $[*_c]_0$ and $[*_a]_0$ were calculated as follows:

$$[*_c]_0 \propto area\ of\ ridges = 2\times(2\times1.33\times13.33)\ \mu m^2 \quad (6.A9)$$

$$[*_a]_0 \propto area\ off\ ridges = (13.33)^2\ \mu m^2 - (2\times1.33\times13.33)\ \mu m^2 \quad (6.A10)$$

Scenarios examined by the model correspond to different CBM/cellulose loadings and varying levels of preference for binding ridge regions over non-ridge regions (K_c/K_a). Loadings of 5,000, 10,000, and 20,000 CBMs were used for the seven different K_c/K_a values listed in Table 6.A1.

Table 6.A1 Spatial adsorption model results

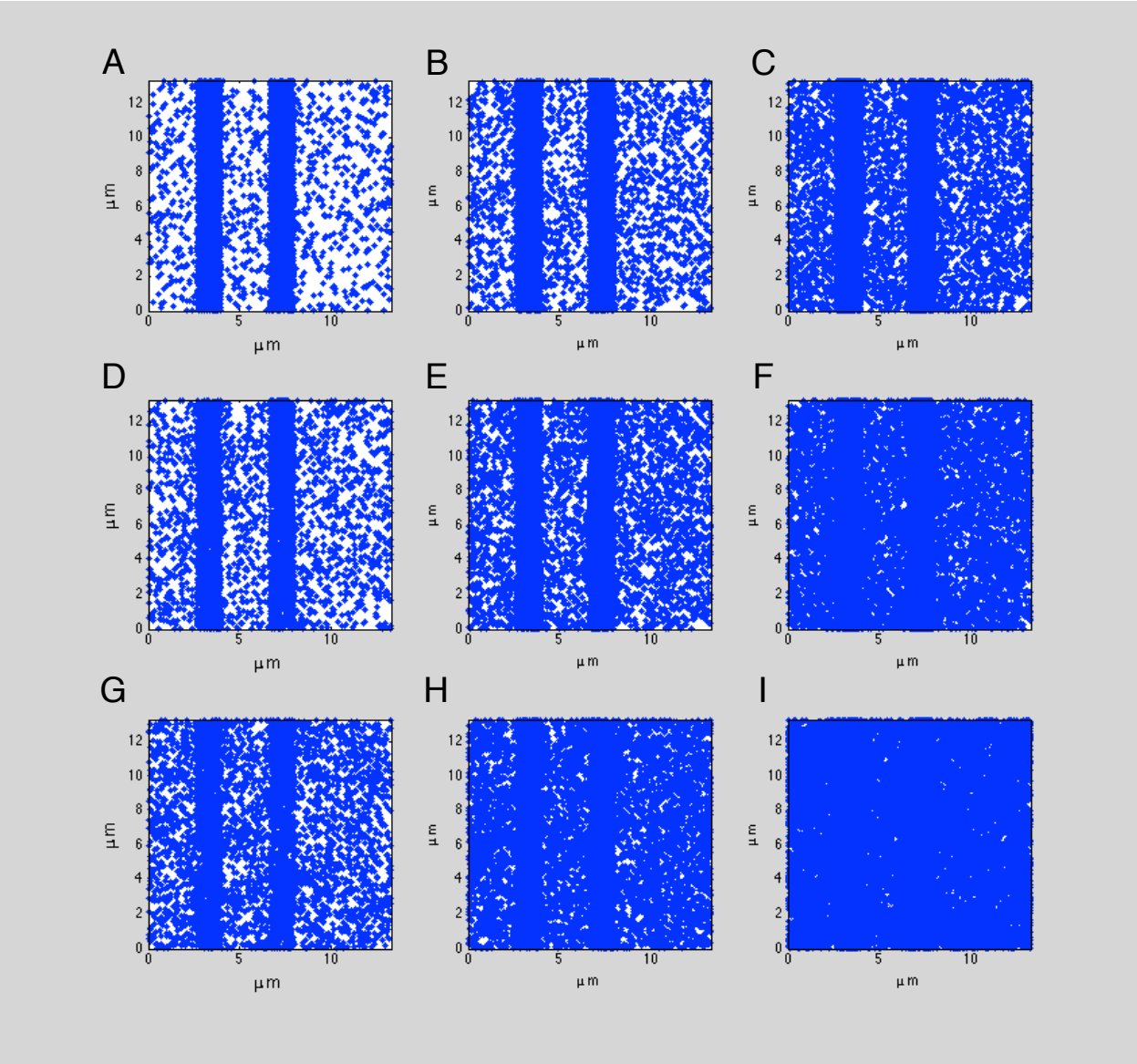
K_c/K_a	Order Parameter with 10,000 proteins bound	Order Parameter with 20,000 proteins bound
10	2.24	2.25
5	1.97	1.92
2	1.17	1.18
1	0.63	0.43
0.5	0.17	0.15
0.2	0.25	0.20
0.05	0.16	0.11

Table 6.A1. Order parameters calculated for spatial point patterns modeled at two different loadings on a 13.33 x 13.33 μm cellulose square with two 1.33 x 13.33 μm crystalline regions. The K_c/K_a value represents the relative preference of the CBM to bind crystalline sites within these regions over amorphous sites outside of them.

The images produced by this model are shown in Figs. 6.A4-6.A6. The results of the 10,000 and 20,000 CBM loadings are given Table 6.A1. Order parameter trends were consistent with our interpretation of the experimental order parameter results: Ω values were highest for CBMs

with the highest preference for binding ridge regions and lowest for those with a preference for binding inter-ridge regions.

Figure 6.A4



Figsures 6.A4-6.A6. Images showing the arrangement of CBMs adsorbed to the model cellulose surface described in Fig. 6.A2. The scenarios that are depicted differ by the total number of adsorbed CBM and by the relative affinity of that CBM for crystalline cellulose over amorphous cellulose. The left, middle, and right columns correspond to scenarios with 5,000, 10,000, and 20,000 adsorbed CBMs. The K_c/K_a ratios for the images are as follows: 10 for 4A-4C; 5 for 4D-4F, 2 for 4G-4I; 1 for 5A-5C, 0.5 for 5D-5F; 0.2 for 5G-5I; and 0.05 for 6A-6C.

Figure 6.A5

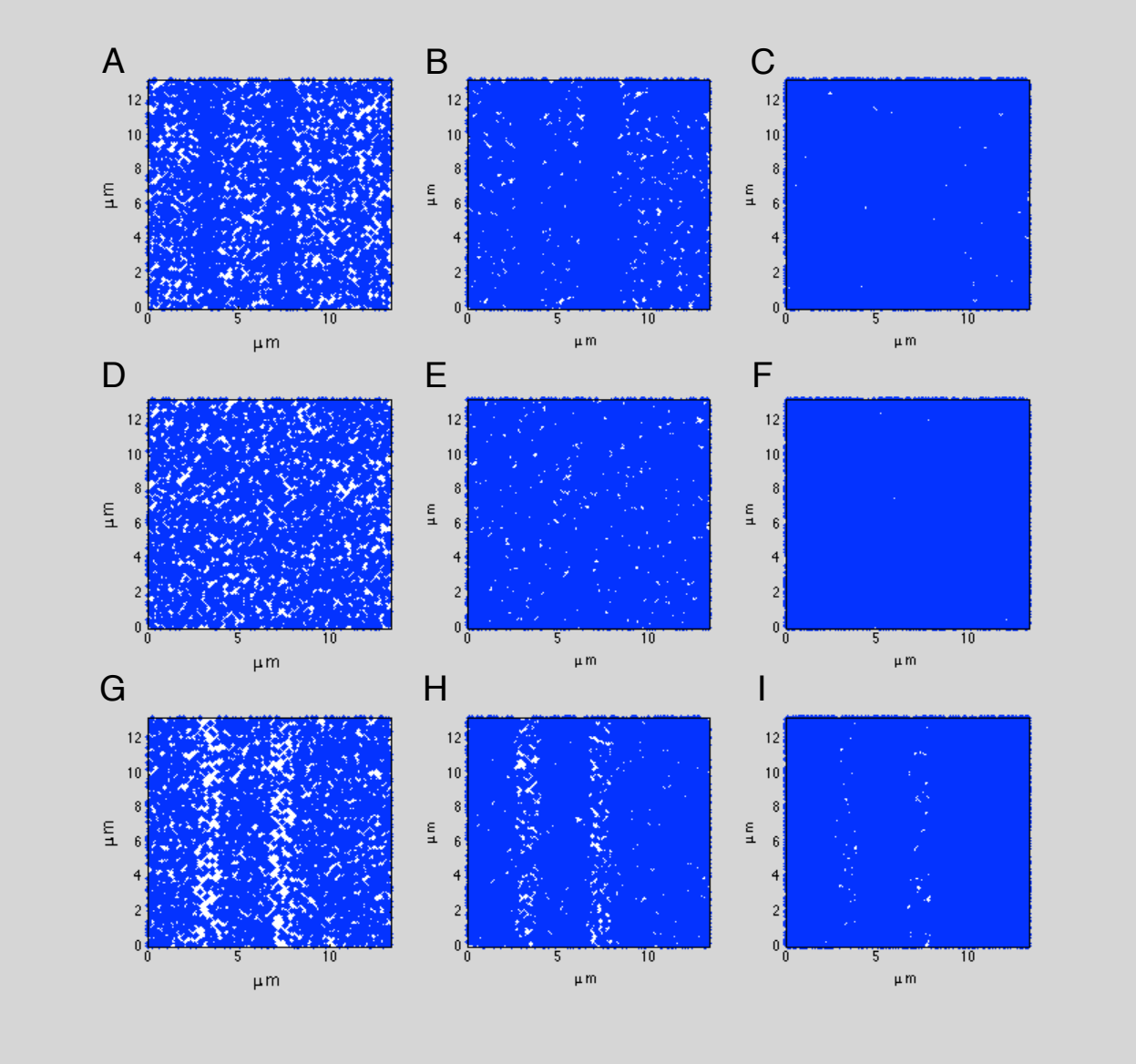
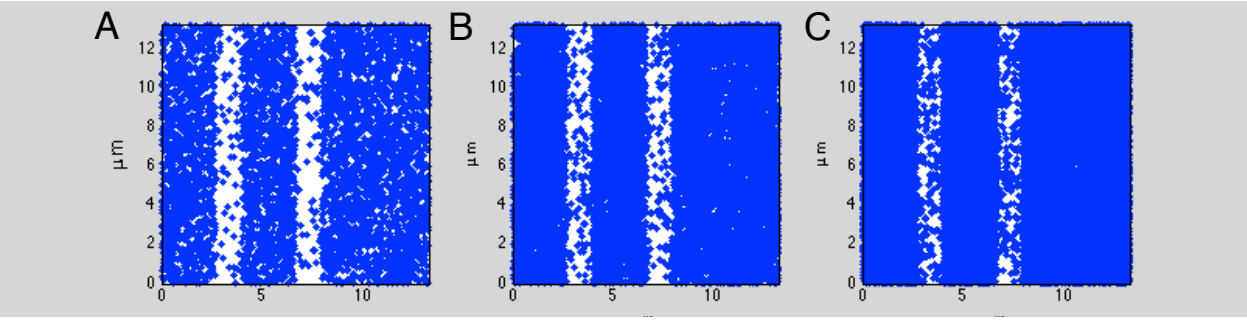


Figure 6.A6



6.10 Appendix 4: Arrhenius Basis of Reaction Times

The activity of *Ac*GH5 on insoluble cellulose has been shown to obey the Arrhenius relationship:

$$k = Ae^{-E_a/RT} \quad (6.A11)$$

where k is the catalytic constant, A is the pre-exponential factor, E_a is the activation energy, R is the ideal gas constant, and T is the temperature (in Kelvin). Using data from three previous investigations showing the influence of temperature on the activity of either *Ac*GH5 endoglucanase REF 30 (*36, 37*) or of the secreted *A. cellulolyticus* cellulolytic mixture (*28*), we used the Arrhenius equation to estimate time periods likely to occur within the initial-rate regime of *Ac*GH5-catalyzed cellulose hydrolysis at 18° and 50°C. A study by Kim et al showed that the activity of *Ac*GH5-CBM chimeras on Avicel was constant for 5-10 hours at 70°C (*29*). We based our calculations on a conservative 2-hour initial-rate regime using the following equation:

$$time(T_2) = time(T_1) \cdot \frac{k(T_1)}{k(T_2)} \quad (6.A12)$$

where T_1= 343.15 K (or 70°C), time(T_1) = 2 hours, and k(T_2) was value of k at 70°C predicted by the Arrhenius fit from each of the aforementioned references. The reaction durations resulting from this calculation are reported in Table 6.A2; the final reaction durations (45h for 18° and 5h for 50°C) are based on averages of the three extrapolated durations from each of the three references.

Initial-rate regime periods at various temperatures

Temperature (°C)	Dai et al (*36*)	Tucker et al (*28*)	Lindenmuth et al (*37*)	Average Time	Time Used
70	2	2	2	2	**2**
50	5.5	4.0	7.1	5.5	**5**
18	37.9	14.4	77	43	**45**

Table 6.A2 Arrhenius-based extrapolations of initial-rate regime periods for *A. cellulolyticus* GH5 endoglucanase at various temperatures.

6.11 Appendix 5: Derivation of Dilute Adsorption Equation

The insensitivity of GH5-CBM cellulase performance to the enzyme loadings and temperatures examined in this work evidences a dilute adsorption regime where the concentrations of total and adsorbed enzyme are approximately equal. This connection is best explained through the derivation of an equation for CBM adsorption under dilute conditions. Several groups have showed that CBMs adsorb to high and low affinity sites on cellulose (*23, 38*). Under low CBM/cellulose loadings, the CBMs adsorb primarily to high affinity sites, and adsorption can be represented by a one-site Langmuir isotherm:

$$[CBM^*] \approx \frac{K[*]_0[CBM]}{1+K[CBM]} \quad (6.A13)$$

where [CBM*], [CBM], and $[*]_0$ are the volumetric concentrations of bound CBM, free CBM, and total adsorption sites, respectively. The binding constant, K, represents a CBM's affinity for a particular site:

$$K = \frac{[CBM^*]}{[*][CBM]} \quad (6.A14)$$

where [*] is the concentration of free adsorption sites. Within the linear regime of a Langmuir isotherm, the concentration of free protein is low enough that K[CBM]<<1. In this scenario, the concentration of bound protein is proportional to the concentration of free protein:

$$[CBM^*] \approx K[*]_0[CBM] \quad (6.A15)$$

Equivalently,

$$[CBM^*] \approx \frac{K[*]_0[CBM]_0}{1+K[*]_0} \quad (S16)$$

where $[CBM]_0$ is the volumetric concentration of total protein. The insensitivity of relative GH5-CBM mixture activities to total enzyme loading suggests that this bound protein/total protein proportionality holds for all six constructs (Fig. 6.A7); for each GH5-CBM variant, a doubling in enzyme concentration leads to the same increase in concentration of GH5-CBM on the cellulose surface. The applicability of a linear adsorption regime was confirmed by a separate experiment where seven different concentrations of each GH5-CBM construct were assayed against 50.8 mg/L cotton linters for 5 hours at 50°C and pH 4.85 (50 mM sodium acetate buffer with 500 µM $CaCl_2$). The relationship between activity and enzyme loading was linear for all constructs (Fig. 6.A3).

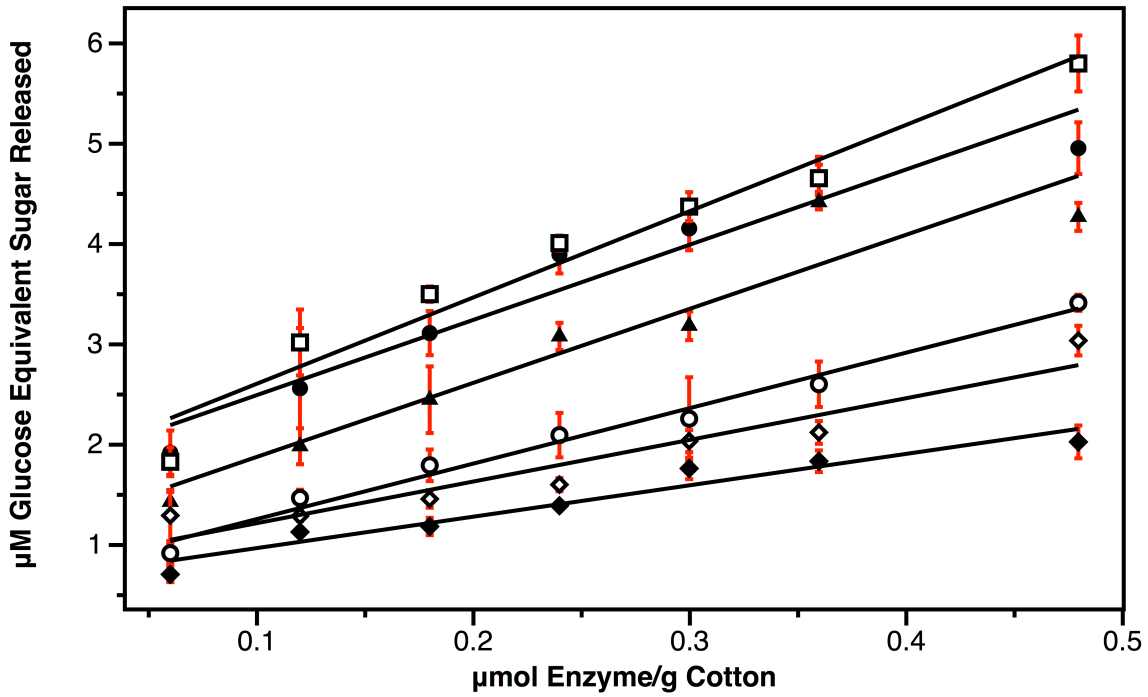


Figure 6.A3 Plot showing the influence of enzyme/cotton loading on the concentration of soluble sugars released by each GH5-CBM construct acting on 50.8 mg/L cotton linters at 50°C for 5 hours (pH 4.85). The glucose-equivalent concentration of soluble sugars is plotted against the enzyme/cotton in loading, which is given in µmol/g cellulose. The symbols correspond to GH5-CBM1*Tr*Cel7A (open circle), GH5-CBM1*Tr*Cel6A (filled circle), GH5-CBM2A (filled triangle), GH5-CBM3A (open square), GH5-CBM17 (open diamond), and GH5-CBM28 (filled diamond).

A special case of Eq. S16 where adsorption is additionally insensitive to temperature may apply to the loading and temperature conditions examined in this work. When $K[N]_0 >> 1$, Eq. S16 can be approximated as follows:

$$[CBM^*] \approx [CBM]_0 \quad (6.A17)$$

In such a scenario, changes in K resulting from changes in temperature do not significantly alter adsorption behavior. Such a regime is one possible explanation for the insensitivity of the relative performance of GH5-CBM mixtures to the temperatures examined in the present study. This insensitivity could also be the result of binding constants that are not heavily influenced by temperature, but previous calorimetry studies of CBM adsorption to cellulose have shown that temperature dependencies can vary greatly between CBMs. The adsorption of amorphous-specific CBMs is heavily influenced by enthalpic effects (*39*), and the adsorption of crystalline-specific CBMs is heavily influenced by entropic contributions (*38*); binding constants of crystalline-specific CBMs have stronger temperature dependencies (*39, 40*). Thus, the insensitivity of relative GH5-CBM mixture performance to temperature is likely the result of a binding scenario where most of the enzyme is bound to the cellulose surface. In such a regime, changes in K occur, but they are not significant enough to change the applicability of the approximation given by Eq. S17, and the relative performance of the different GH5-CBM mixtures will remain similar at 18°C and 50°C.

6.12 Appendix 6: Synergy Calculation for Binary Mixtures

The synergy exhibited by each binary combination was calculated as the concentration of soluble sugars released by the enzyme combination divided by the concentration of sugars expected from purely additive behavior between the two enzymes:

$$Synergy = \frac{[Geq]_{GH5-CBMA\ and\ GH5-CBMB}}{\frac{1}{2}[Geq]_{GH5-CBMA} + \frac{1}{2}[Geq]_{GH5-CBMB}} \quad (6.A18)$$

where G_{eq} is the glucose equivalent concentration of soluble sugars released from the cotton linters, and CBMA and CBMB are two non-identical CBMs.

As an example, we calculate the synergy exhibited by the mixture of GH5-CBM17 and GH5-CBM28 enzymes. At 0.12 μmol enzyme/g cellulose, when exposed to 50.8 mg/L of cotton linters for 5 hours at 50°C, GH5-CBM17 produced 1.16 μM glucose-equivalent soluble sugars, GH5-CBM28 produced 0.95 μM, and an equimolar mixture of both produced 1.51 μM. If the enzymes act independently of each other when present simultaneously, the concentration of soluble sugars expected to be released by a binary mixture containing half the concentration of each would be 1/2*1.16+1/2*0.95 = 1.055 μM. The synergy that the mixture of both enzymes exhibits as a result of them not acting independently is then 1.51/1.055=1.43. Synergy values such as this one were calculated in an identical fashion for each binary mixture under the four sets of conditions investigated within the present work.

6.13 Appendix 7: Calculation of Nyquist Resolution and Localization Precision

The localization precision and Nyquist spatial resolution were calculated with standard procedures. The two-dimensional localization precision of a point source of light is given by the following (*41*):

$$\sigma_{xy}^2 = \frac{s^2+p^2/12}{N} + \frac{8\pi s^4 b^2}{p^2 N^2} \quad (6.A19)$$

Where σ_{xy} is the localization precision, s is the standard deviation of the fitted Gaussian approximation of the PSF, N is the total number of photons collected, p is the effective size of an image pixel within the focal plane, and b is the background noise per pixel. When background noise is negligible relative to the collected signal, the following approximation can be made:

$$\sigma_{xy}^2 = \frac{s^2}{N} \quad (6.A20)$$

Eq. S20 was used to calculate the localization precision for each of densest eight 3.1-μm circles identified within each of the 32 images used for analysis. The mean localization precision was 32 nm +/- 7 nm. Though not as fine as localization precision values reported for other systems studied with PALM (*15, 42*), this value is adequate for our imaging system: 63X N.A. 1.49 objective with a glass coverslip (No. 1.5, VWR 48336-227). A 100X N.A. 1.65 objective and a sapphire coverslip, which were used in reference 11, would have permitted a higher localization precision by enhancing the cone of light entering the microscope from any given fluorophore (i.e., more photons/localized protein); however, this system requires immersion oil to be replaced hourly, and it does not allow for a large enough field of view to isolate multiple densely labeled regions within a single sample. Moreover, there is no evidence to suggest that a twofold increase in localization precision would impact the results of this study.

The 2-dimensional Nyquist spatial resolution was calculated within the densest 0.44-μm diameter circular regions within each of the 3.1-μm diameter circles using a standard equation based on the Nyquist criterion for two-dimensional patterns:

$$r_N = \frac{2}{\sqrt{\rho}} \quad (6.A21)$$

where r_N is the Nyquist resolution (in nm), and ρ is the density of localized fluorophores within the circle (points/nm). The mean Nyquist resolution was 77 +/- 21 nm.

The Nyquist resolution may not be the most appropriate resolution measure for our analysis. The CBM order parameters are calculated from distances between localized fluorophores, and they are weighted towards nearest neighbor distances. Thus, the localization-based resolution, which represents the smallest structure that can be imaged by PALM, may be a more applicable resolution measure. The localization-based resolution is defined as follows (*43*):

$$r_L = \sqrt{\sigma_{xy}^2 + r_{NN}^2} \quad (6.A22)$$

where r_L is the localization-based resolution and r_{NN} is the nearest neighbor distance between fluorophores. When $r_{NN}>>\sigma_{xy}$, as we find in the current study, this equation can be approximated as follows:

$$r_L = \sigma_{xy} \quad (6.A23)$$

Thus, while the Nyquist resolution was only 81 +/- 25 nm, the localization-based resolution, which is more directly related to the nearest neighbor distances used in the order parameter calculation, was 32 nm +/- 7 nm.

6.14 Appendix 8: Compositions of GH5-CBM Mixtures Assayed Against Cotton

GH5-CBM		1*Tr*Cel7A	1*Tr*Cel6A	2A	3A	17	28
1*Tr*Cel7A	1:0	1:1					
1*Tr*Cel6A	1:0	1:1	1:1				
2A	1:0	1:1	**1:1**	1:1			
3A	1:0	**1:1**	1:1	1:1	1:1		
17	1:0	1:1	1:1	**1:1**	1:1	1:1	
28	1:0	1:1	1:1	**1:1**	**1:1**	**1:1**	1:1

Table 6.A3 The compositions of GH5-CBM mixtures that were assayed against 50.8 mg/L of cotton under loadings of 0.12 and 0.24 µmol total enzyme/g cotton at pH 4.85 (50 mM sodium acetate with 500 µM $CaCl_2$). The 1:1 combinations represent binary mixtures where each enzyme is present at a 0.06 or 0.12 µmol/g to achieve total loadings of 0.12 and 0.24 µmol/g, respectively; the 1:0 combinations represent single enzyme reactions at half loadings where each GH5-CBM is present at 0.06 or 0.12 µmol/g cotton. Combinations highlighted in orange yielded greater than 20% synergy.

6.15 Appendix 9: Synergy values and associated order parameter differences

Ω_{CBMA}-Ω_{CBMB}	Synergy	Synergy σ	CBMA	CBMB
0.00	1.00	0.03	CBMi	CBMi
0.03	1.01	0.05	3A	17
0.11	0.99	0.07	2A	3A
0.11	1.05	0.08	1Cel6A	1Cel7A
0.11	1.33	0.06	17	28
0.14	1.23	0.08	2A	17
0.14	1.42	0.11	3A	28
0.16	1.27	0.07	1Cel6A	2A
0.24	1.49	0.11	2A	28
0.27	1.06	0.07	1Cel6A	3A
0.27	0.97	0.06	1Cel7A	2A
0.30	1.13	0.07	1Cel6A	17
0.38	1.27	0.09	1Cel7A	3A
0.41	1.00	0.09	1Cel7A	17
0.41	0.99	0.07	1Cel6A	28
0.51	0.97	0.13	1Cel7A	28

Table 6.A4 Synergy values of various GH5-CBM combinations listed by increasing difference in the order parameters of the associated CBMs (Ω_{CBMA}-Ω_{CBMB}). The CBM termed CBMi corresponds to any individually acting CBM.

6.16 References

1. M. E. Himmel *et al.*, Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* **315**, 804 (Feb 9, 2007).
2. S. P. Chundawat, G. T. Beckham, M. E. Himmel, B. E. Dale, Deconstruction of lignocellulosic biomass to fuels and chemicals. *Annu Rev Chem Biomol Eng* **2**, 121 (2011).
3. K. Igarashi *et al.*, Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface. *Science* **333**, 1279 (Sep 2, 2011).
4. A. N. Fernandes *et al.*, Nanostructure of cellulose microfibrils in spruce wood. *Proc Natl Acad Sci U S A* **108**, E1195 (Nov 22, 2011).
5. X. Li, W. T. t. Beeson, C. M. Phillips, M. A. Marletta, J. H. Cate, Structural basis for substrate targeting and catalysis by fungal polysaccharide monooxygenases. *Structure* **20**, 1051 (Jun 6, 2012).
6. B. W. McLean *et al.*, Carbohydrate-binding modules recognize fine substructures of cellulose. *J Biol Chem* **277**, 50245 (Dec 27, 2002).
7. A. B. Boraston, D. N. Bolam, H. J. Gilbert, G. J. Davies, Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J* **382**, 769 (Sep 15, 2004).
8. S. J. Horn, G. Vaaje-Kolstad, B. Westereng, V. G. Eijsink, Novel enzymes for the degradation of cellulose. *Biotechnol Biofuels* **5**, 45 (Jul 2, 2012).
9. L. R. Lynd *et al.*, How biotech can transform biofuels. *Nat Biotechnol* **26**, 169 (Feb, 2008).
10. R. J. Quinlan *et al.*, Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc Natl Acad Sci U S A* **108**, 15079 (Sep 13, 2011).
11. G. Carrard, A. Koivula, H. Soderlund, P. Beguin, Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose. *Proc Natl Acad Sci U S A* **97**, 10342 (Sep 12, 2000).
12. A. W. Blake *et al.*, Understanding the biological rationale for the diversity of cellulose-directed carbohydrate-binding modules in prokaryotic enzymes. *J Biol Chem* **281**, 29321 (Sep 29, 2006).
13. T. T. Teeri, Crystalline cellulose degradation: New insight into the function of cellobiohydrolases. *Trends Biotechnol* **15**, 160 (May, 1997).
14. D. Harris, V. Bulone, S. Y. Ding, S. DeBolt, Tools for Cellulose Analysis in Plant Cell Walls. *Plant Physiol* **153**, 420 (Jun, 2010).
15. E. Betzig *et al.*, Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642 (Sep 15, 2006).
16. D. J. Dagel *et al.*, In Situ Imaging of Single Carbohydrate-Binding Modules on Cellulose Microfibrils. *Journal of Physical Chemistry B* **115**, 635 (Feb 3, 2011).
17. D. J. Dagel *et al.*, DOPI and PALM Imaging of Single Carbohydrate-binding Modules Bound to Cellulose Nanocrystals. *Proceedings of SPIE* **79050**, 1 (2011).
18. S. A. McKinney, C. S. Murphy, K. L. Hazelwood, M. W. Davidson, L. L. Looger, A bright and photostable photoconvertible fluorescent protein. *Nat Methods* **6**, 131 (Feb, 2009).
19. G. T. Beckham *et al.*, The O-glycosylated linker from the Trichoderma reesei Family 7 cellulase is a flexible, disordered protein. *Biophys J* **99**, 3773 (Dec 1, 2010).
20. V. Receveur, M. Czjzek, M. Schulein, P. Panine, B. Henrissat, Dimension, shape, and conformational flexibility of a two domain fungal cellulase in solution probed by small angle X-ray scattering. *J Biol Chem* **277**, 40887 (Oct 25, 2002).
21. D. K. Y. Poon, S. G. Withers, L. P. McIntosh, Direct demonstration of the flexibility of the glycosylated proline-threonine linker in the Cellulomonas fimi xylanase Cex through NMR spectroscopic analysis. *Journal of Biological Chemistry* **282**, 2091 (Jan 19, 2007).
22. J. Lehtio *et al.*, The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. *Proc Natl Acad Sci U S A* **100**, 484 (Jan 21, 2003).
23. A. B. Boraston, E. Kwan, P. Chiu, R. A. Warren, D. G. Kilburn, Recognition and hydrolysis of noncrystalline cellulose. *J Biol Chem* **278**, 6120 (Feb 21, 2003).

24. S. Jamal, D. Nurizzo, A. B. Boraston, G. J. Davies, X-ray crystal structure of a non-crystalline cellulose-specific carbohydrate-binding module: CBM28. *J Mol Biol* **339**, 253 (May 28, 2004).
25. V. Notenboom *et al.*, Recognition of cello-oligosaccharides by a family 17 carbohydrate-binding module: an X-ray crystallographic, thermodynamic and mutagenic study. *J Mol Biol* **314**, 797 (Dec 7, 2001).
26. M. Linder, I. Salovuori, L. Ruohonen, T. T. Teeri, Characterization of a Double Cellulose-binding Domain. *J Biol Chem* **271**, 21268 (1996).
27. J. Sakon, W. S. Adney, M. E. Himmel, S. R. Thomas, P. A. Karplus, Crystal structure of thermostable family 5 endocellulase E1 from Acidothermus cellulolyticus in complex with cellotetraose. *Biochemistry* **35**, 10648 (Aug 20, 1996).
28. M. P. Tucker, A. Mohagheghi, K. Grohmann, M. E. Himmel, Ultra-Thermostable Cellulases from Acidothermus-Cellulolyticus - Comparison of Temperature Optima with Previously Reported Cellulases. *Bio-Technol* **7**, 817 (Aug, 1989).
29. T. W. Kim, H. A. Chokhawala, D. C. Nadler, H. W. Blanch, D. S. Clark, Binding modules alter the activity of chimeric cellulases: Effects of biomass pretreatment and enzyme source. *Biotechnol Bioeng* **107**, 601 (Nov 1, 2010).
30. Z. Dai, B. S. Hooker, D. B. Anderson, S. R. Thomas, Expression of Acidothermus cellulolyticus endoglucanase E1 in transgenic tobacco: biochemical characteristics and physiological effects. *Transgenic Res* **9**, 43 (Feb, 2000).
31. B. W. McLean *et al.*, Analysis of binding of the family 2a carbohydrate-binding module from Cellulomonas fimi xylanase 10A to cellulose: specificity and identification of functionally important amino acid residues. *Protein Eng* **13**, 801 (Nov, 2000).
32. A. B. Boraston, P. Chiu, R. A. Warren, D. G. Kilburn, Specificity and affinity of substrate binding by a family 17 carbohydrate-binding module from Clostridium cellulovorans cellulase 5A. *Biochemistry* **39**, 11129 (Sep 12, 2000).
33. M. Dubois, K. Gilles, J. K. Hamilton, P. A. Rebers, F. Smith, A colorimetric method for the determination of sugars. *Nature* **168**, 167 (Jul 28, 1951).
34. H. Shroff *et al.*, Dual-color superresolution imaging of genetically expressed probes within individual adhesion complexes. *Proc Natl Acad Sci U S A* **104**, 20308 (Dec 18, 2007).
35. P. Annibale, M. Scarselli, A. Kodiyan, A. Radenovic, Photoactivatable Fluorescent Protein mEos2 Displays Repeated Photoactivation after a Long-Lived Dark State in the Red Photoconverted Form. *J Phys Chem Lett* **1**, 1506 (May 6, 2010).
36. Z. Y. Dai, B. S. Hooker, D. B. Anderson, S. R. Thomas, Expression of Acidothermus cellulolyticus endoglucanase E1 in transgenic tobacco: biochemical characteristics and physiological effects. *Transgenic Res* **9**, 43 (Feb, 2000).
37. B. E. Lindenmuth, K. A. McDonald, Production and characterization of Acidothermus cellulolyticus endoglucanase in Pichia pastoris. *Protein Expr Purif* **77**, 153 (Jun, 2011).
38. A. L. Creagh, E. Ong, E. Jervis, D. G. Kilburn, C. A. Haynes, Binding of the cellulose-binding domain of exoglucanase Cex from Cellulomonas fimi to insoluble microcrystalline cellulose is entropically driven. *Proc Natl Acad Sci U S A* **93**, 12229 (Oct 29, 1996).
39. A. B. Boraston, The interaction of carbohydrate-binding modules with insoluble non-crystalline cellulose is enthalpically driven. *Biochem J* **385**, 479 (Jan 15, 2005).
40. G. Carrard, M. Linder, Widely different off rates of two closely related cellulose-binding domains from Trichoderma reesei. *Eur J Biochem* **262**, 637 (Jun, 1999).
41. R. E. Thompson, D. R. Larson, W. W. Webb, Precise nanometer localization analysis for individual fluorescent probes. *Biophys J* **82**, 2775 (May, 2002).
42. D. Greenfield *et al.*, Self-organization of the Escherichia coli chemotaxis network imaged with super-resolution light microscopy. *PLoS Biol* **7**, e1000137 (Jun 16, 2009).
43. S. T. Hess, T. P. Girirajan, M. D. Mason, Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys J* **91**, 4258 (Dec 1, 2006).

Chapter 7: Conclusions and Future Directions

In this section, we discuss additional experiments and improvements to the hydrolysis model that are suggested by our results and by the results of other recently published investigations of cellulolytic enzymes.

7.1 The Cellobiohydrolase

Most, if not all, enzymes that enhance the activity of cellulolytic mixtures do so by allowing cellobiohydrolases to work more effectively. Cellobiohydrolases comprise the largest component of cellulolytic mixtures likely to be implemented in industrial biofuels processes (*1*); in the Ascomycete fungus *Trichoderma reesei*, the industrial standard for cellulase production, cellobiohydrolase I (Cel7A) and cellobiohydrolase II (Cel6A) make up over 80% of the secretome (*2*). In recent years, evidence has mounted that these enzymes, which are unique in their ability to degrade crystalline cellulose at substantial rates, experience rate limitations imposed by physical phenomena, not intrinsic kinetics (*3-5*). By functioning to alleviate these kinetic impediments, companion enzymes can enhance cellobiohydrolase-catalyzed hydrolysis rates and thus the activity of cellulolytic mixtures.

Our work with *Talaromyces emersonii* endoglucanase II (*Tem*GH5) evidences two main mechanisms by which enzymes can enhance cellobiohydrolase activity (*4*): (1) free end generation and (2) obstacle removal. By estimating initial- and processive-cut products of the *Trichoderma longibrachiatum* cellobiohydrolase Cel7A (*Tl*Cel7A) during catalysis, we showed that the rate of *Tl*Cel7A-catalyzed hydrolysis of crystalline cellulose is limited by the rate of enzyme complexation with glycan chains, a rate that can be enhanced by endoglucanase-generated free ends. We also gave evidence that *Tl*Cel7A processive length is impeded by surface morphological obstacles, which can be removed by endoglucanase enzymes or potentially overcome by the concerted action of multiple cellobiohydrolases engaging parallel chains. Recent kinetic and atomic force microscopy studies have also shown that cellobiohydrolase processivity can be arrested by cellulose- and protein-based obstructions (*3, 5*).

Focused enzyme activities may be advantageous for cellulolytic enzyme mixture optimization strategies. We worked with an endoglucanase enzyme (*Tem*GH5) that was capable of accomplishing both free end generation and obstacle removal, but we hypothesize that two enzymes optimized to carry out only one of these tasks could be more effective than one enzyme that accomplishes both. The ratio of these enzymes could be tuned to overcome substrate-specific challenges: absence of free ends or excess of pseudo-amorphous obstacles.

The polysaccharide monooxygenases (PMO) is a recently discovered example of an enzyme that has been optimized to generate free ends (*6, 7*). PMOs cleave internal glycosidic bonds located within crystalline regions; their free-end-generating activity is tightly focused on the cellulose surfaces that are most susceptible to cellobiohydrolase action. Though PMO crystal structures showing different cellulose surface targets suggest that synergistic behavior is likely to vary between different PMO-cellobiohydrolase combinations, PMOs demonstrate the extent to which focused behavior can be accomplished within cellulolytic mixtures (*8*). Future

mechanistic investigations that reveal precisely how cellobiohydrolase activity is enhanced by specific companion enzymes are likely to yield new strategies for enzyme design and mixture optimization.

7.2 An Optimal Cellulase Cocktail

Optimal cellulolytic mixture compositions are dependent on the nature of the cellulosic substrate and the conditions under which it is hydrolyzed. We have shown how surface characteristics such as degree of polymerization, particle shape, and surface area can dramatically influence the kinetics of enzymatic cellulose hydrolysis (*9, 10*); we have demonstrated that morphological targeting plays a critical role in enzyme-enzyme synergy (*4, 11*); and we have shown that slow catalysis at 37°C eclipses potential gains from *S. cerevisiae* consumption of the cellobiose inhibitor during SSF (*12*). These results suggest strategies for optimizing enzymatic cellulose hydrolysis.

A cellulolytic mixture can be tuned to match cellulose surface characteristics with (1) modification of cellulase architecture and (2) changes in mixture composition. We found that endoglucanase enzymes can enhance the behavior of cellobiohydrolases both by generating free ends and by removing surface obstructions, and we demonstrated that combinations of enzymes differing only in their CBM domains can exhibit activities significantly higher than individual enzymes (*4, 11*). Enzymes designed to function by (1) primarily clearing away surface obstructions, (2) primarily generating free ends, or (3) focusing their activity within specific morphologies may be advantageous within enzyme cocktails optimized to efficiently degrade particular cellulose structures. On lignocellulosic substrates, enzymes less prone to inhibition by hemicellulose or lignin, which have been shown to lower hydrolysis rates (*13*), would also be advantageous. Future investigations that probe the influence of specific enzyme-surface interactions on the overall kinetics of multi-enzyme mixtures acting on heterogeneous substrates are likely to reveal new strategies for optimizing cocktail composition.

Mixture compositions and enzyme activities can also be adjusted to meet particular process goals. Our results suggest that the combined residence time for hydrolysis and fermentation might be lowered by cellobiohydrolases that have been engineered to be more resistant to cellobiose inhibition; these enzymes could lower the β-glucosidase requirement during hydrolysis, increase the cellobiose/glucose ratio in the fermentation feed stream, and lower glucose inhibition of xylose uptake in an *S. cerevisiae* strain engineered to consume glucose, cellobiose, and xylose. The lower β-glucosidase requirement and faster xylose fermentation might result in a cost savings. Alternatively, to reduce the risk of microbial contamination during hydrolysis, cellulase mixtures can be designed for optimal performance above 65° (*14*). As binding behavior is temperature sensitive, and as half-lives are enzyme-specific, such mixture compositions are likely to be temperature-dependent. In general, in addition to substrate characteristics, overall reactor engineering and materials cost considerations can also be used to guide the design of optimal cellulase cocktails with particular capabilities.

7.3 Particle Shape

Particle shape influences the kinetics of enzymatic cellulose hydrolysis by affecting the initial substrate surface area and the manner in which that surface area evolves throughout depolymerization. In our model, cellulose particles are represented as spheres. We have given evidence that polydisperse distributions of spheres can be used to capture the initial volume, initial surface area, and the time-dependent surface evolution of any particle shape (*9*); however, we do not supply a mathematical algorithm to generate such distributions. In future versions of the kinetic model, we must (1) evaluate the influence of particle shape on the relative performance of different enzyme mixtures and (2) generate methods for mapping specific shapes to polydisperse distributions of spheres.

We must explore the first objective before assigning priority to the second. The shape-dependent evolution of surface area affects the amount of substrate that is exposed to enzymatic attack at any point in the reaction. As this surface is exposed to all enzymes, the relative performance of different enzymes will not necessarily be shape-dependent. If, however, cellulose is represented with different morphologies, which are differentially targeted by the various components of cellulolytic mixtures, the evolution of particular cellulose architectures and thus the activities of different enzymes will be shape-dependent. Future versions of the kinetic model that include multiple morphologies will be necessary to assess the influence of particle shape on reaction kinetics and thereby justify attempts to capture shape behavior with mathematical rigor.

There are two main strategies for mapping the time-dependent evolution of a particular shape to the time-dependent evolution of a polydisperse distribution of spheres. A well-defined equation could be developed: f(number of particles, particle shape, particle volume) = (number of sphere of radius r_1, r_2,…). Alternatively, an iterative computational approach could be used; surface-area-dependent shrinking of a given shape can be fit to surface-area-dependent shrinking of several spheres with the same total initial surface area and volume. Though unnecessary for studying the general influence of particle shape on the kinetics of enzymatic hydrolysis, such methods are needed to examine the hydrolysis of a particular particle shape such as one based on the image of a real particle. Future applications of the model to experimental scenarios may require shape matching.

7.4 Surface Heterogeneities

Our results indicate that surface heterogeneities must be incorporated into future versions of the kinetic model. Using photoactivation localization microscopy to localize CBMs bound to the surface of cotton, we mapped CBM binding targets onto a morphological spectrum (*11*). By fusing these CBMs to an identical catalytic domain, we also showed that these morphological targets influence enzyme activity and enzyme-enzyme synergy. These results, in addition to the aforementioned findings that endoglucanases and polysaccharide monooxygenases enhance cellobiohydrolase activity in a morphology-dependent manner, reveal the critical role of surface structure in guiding interactions between different cellulolytic mixture components. Estimates of morphology distributions throughout cellulosic substrates and of morphology-specific enzyme-

cellulose binding constants will be necessary to generate a hydrolysis model that captures the influence of surface heterogeneities.

Lignin and hemicellulose must be incorporated into the kinetic model if it is to be used with lignocellulosic substrates. These polymers are hypothesized to influence cellulase behavior by (1) serving as non-productive surfaces for enzyme adsorption, (2) lowering the effective cellulose surface area, and (3) impeding the processive action of cellobiohydrolases (*5, 13*). In analogy to the multiple morphology case, lignin/hemicellulose distributions within lignocellulosic substrates and lignin/hemicellulose binding constants will be necessary to generate a hydrolysis model that is compatible with lignocellulosic substrates.

7.4.1 Multiple Types of Sites

Multiple morphologies can be incorporated into the existing hydrolysis model through the introduction of multiple sites. For example, for a two-site model (crystalline and amorphous), the cellobiohydrolase-substrate material balances can be modified as follows:

One site model:

$$\begin{aligned}\frac{d[E_{CBH-ads}C_i^{'}]}{dt} &= +k_{comp-CBH-cel}\theta_{CBH}[E_{CBH-ads}^{'}][C_i^{'}] \\ &\quad -(k_{decomp-CBH-cel}+k_{cat-CBH})[E_{CBH-ads}C_i^{'}] \\ &\quad +k_{cat-CBH}[E_{CBH-ads}C_{i+2}^{'}]\end{aligned} \tag{7.1}$$

Two site model:

$$\begin{aligned}\frac{d[E_{CBH-ads}C_i^{'}]}{dt} &= k^1_{comp-CBH-cel}\theta^1_{CBH}[E^{'1}_{CBH-ads}][C_i^{'1}] \\ &\quad -(k^1_{decomp-CBH-cel}+k^1_{cat-CBH})[E^{'1}_{CBH-ads}C_i^{'1}] \\ &\quad +k^1_{cat-CBH}[E^{'1}_{CBH-ads}C_{i+2}^{'1}] \\ &\quad +k^2_{comp-CBH-cel}\theta^2_{CBH}[E^{'2}_{CBH-ads}][C_i^{'2}] \\ &\quad -(k^2_{decomp-CBH-cel}+k^2_{cat-CBH})[E^{'2}_{CBH-ads}C_i^{'2}] \\ &\quad +k^2_{cat-CBH}[E^{'2}_{CBH-ads}C_{i+2}^{'2}]\end{aligned} \tag{7.2}$$

Superscripts in Eq. 7.2 represent amorphous (7.1) and crystalline (7.2) sites. In all model equations, the surface concentration terms for cellulose chains, adsorbed enzymes, and adsorbed-complexed enzymes become morphology-specific. In a two-site model, there will be approximately twice as many surface species terms and equations as in a one-site alternative (depending on the difference in the average degree of polymerization for both morphologies).

Parameter estimation presents the largest challenge for an expanded model. For each morphology, there are 4 parameters relating to degree of polymerization and surface concentration, 8 parameters relating to the surface activity of each cellobiohydrolase, and 6 parameters relating to the surface activity of each endoglucanase (Table 7.1).

The 18 parameters that must be estimated for each type of site in a two-enzyme system can be divided into two categories: (1) those that must be estimated from a morphologically homogeneous substrate (green in Table 7.1) and (2) those that can be estimated from any cellulosic substrate (blue in Table 7.1). The first category consists of degree of polymerization estimates and of all parameters describing complexation and catalysis events on the cellulose surface. The average degree of polymerization of a given cellulose sample is estimated by measuring the total concentration of reducing ends or dissolved cellulose chains (*2*), and complexation/catalytic constants are estimated by fitting the model to solid substrate hydrolysis results (*9, 10*). All of these parameters are measured through system averages; if they were measured on a substrate of mixed morphology, their calculated values would reflect an average of all morphological influences. The second category of parameters, by contrast, can be estimated from morphology-specific observables. The initial concentration and surface area associated with a particular binding site morphology can be estimate from binding isotherms of morphology-specific CBMs on any cellulosic sample (*15, 16*); similar constructs can be used to calculate the on rates, off rates, and/or equilibrium binding constants of these CBMs and, by analogy, the enzymes to which they are attached.

Multi-site models that extend beyond two kinds of sites (crystalline and amorphous) would be difficult to support with experimentally determined parameters. Crystalline cellulose is typically prepared with sodium hydroxide extraction or dilute acid pretreatment of bacterial cellulose to produce bacterial microcrystalline cellulose (*2, 17*); and amorphous cellulose is typically prepared with concentrated phosphoric acid treatment of Avicel to produce phosphoric acid pretreated cellulose (PASC) (*2*). While both BMCC and PASC are enriched in ordered and disordered sites, respectively, they are far from homogeneous. Competitive binding isotherms have revealed several types of binding sites on each type of preparation (*16, 18*); additionally, while there are CBMs with higher binding affinities for one type over the other, there are none that exclusively bind only to one. The existence of structural overlap between these two drastically different cellulose surfaces suggests that attempts to parameterize intermediary morphologies with intermediary preparations may be difficult.

If representation of more than two morphologies within the model is desired, extension of parameters from only category 2 above (blue in Table 7.1) may be sufficient. Many of the parameters for enzyme complexation and catalysis on the cellulose surface (category 1) are governed by amino acids involved in substrate binding near the active site; thus the influence of surface structure on their values may be small. Category 1 parameters could either be held constant or extended to two sets of values, depending on whether a particular site was more amorphous-like or crystalline-like. Clearly, parameter estimation for a multi-site model presents a time-consuming challenge; the model's sensitivity to particular parameters should be examined to priority to such representations.

Table 7.1: Parameters relating to surface reactions

Parameter	Description	Estimation Strategy
DP_0^p	Initial degree of polymerization for morphology type p	Reducing end assay
DP^p_{max}	Maximum degree of polymerization for morphology type p	Based on the Poisson distribution
$*'^p_{T0}$	Initial surface concentration of free sites of morphology type p (mol dm^{-1})	Binding isotherms with eGFP-CBM constructs
S^p_0	Initial surface area of morphology type p (dm^2 g^{-1})	Binding isotherms with eGFP-CBM constructs
$k^p_{CBH\text{-}ads}$ $k^p_{EG\text{-}ads}$	Adsorption of a cellobiohydrolase or endoglucanase onto a cellulose surface of morphology type p (L $mmol^{-1}$ s^{-1})	Short period binding experiments on a specific cellulosic sample.
$k^p_{CBH\text{-}des}$ $k^p_{EG\text{-}des}$	Desorption of a cellobiohydrolase or endoglucanase onto a cellulose surface of morphology type p (L $mmol^{-1}$ s^{-1})	Short period binding experiments on a specific cellulosic sample.
$k^p_{cat\text{-}CBH\text{-}cel}$, $k^p_{cat\text{-}EG\text{-}cel}$	Catalytic constant for a cellobiohydrolase or endoglucanase acting on a solid cellulose chain of morphology type p (s^{-1})	Fit to experimental data collected on a morphologically homogeneous substrate
$k^p_{cat\text{-}EG\text{-}term}$	Catalytic constant for an endoglucanase acting the terminal glycosidic linkage of a cellulose chain of morphology type p (s^{-1})	Fit to experimental data collected on a morphologically homogeneous substrate
$k^p_{comp\text{-}CBH}$	Kinetic rate constant for complexation of a cellobiohydrolase with the end of a solid cellulose chain of morphology type p (m^2 $mmol^{-1}$ s^{-1})	Fit to experimental data collected on a morphologically homogeneous substrate
$k^p_{decomp\text{-}CBH}$	Kinetic rate constant for decomplexation of a cellobiohydrolase with the end of a solid cellulose chain of morphology type p (s^{-1})	Fit to experimental data collected on a morphologically homogeneous substrate
$K^p_{comp\text{-}EG}$	Complexation equilibrium constant for endoglucanase with a cellulose chain of morphology type p (mmol m^{-2})	Fit to experimental data collected on a morphologically homogeneous substrate
$K^p_{comp\text{-}EG\text{-}term}$	Complexation equilibrium constant for endoglucanase with a terminal glycosidic linkage of a cellulose chain of morphology type p (s^{-1})	Fit to experimental data collected on a morphologically homogeneous substrate
p^p_i where i=1,2,3	Probability of initial catalytic event of a cellobiohydrolase after Complexation will form an i length cello-oligosaccharide	Fit to experimental data collected on a morphologically homogeneous substrate

7.4.2 The Spatial Arrangement of Sites

In chapter 6, we measured the spatial distributions of cellulose morphologies on the surface of cotton and showed that they were non-identical (*11*). Unfortunately, the existing kinetic model contains no spatial dependencies. An improved model could make use of estimates of the spatial distribution of various morphologies supplied by PALM results, but it would also require directionality to be incorporated into all mechanistic steps on the surface. In essence, the kinetic model would have to be recreated. Before such an effort is undertaken, a toy model should be developed to assess the sensitivity of hydrolysis kinetics to particular site arrangements in a simplified cellulase-cellulose system.

7.5 Cellulase Glycosylation

Recently, evidence has emerged that cellulase glycosylation may influence binding behavior. Several modeling studies suggest that N-linked glycosylation on CBM domains and O-linked glycosylation on linkers can enhance cellulase binding affinities for crystalline cellulose (*19, 20*). As many CBM studies rely on bacterial expression systems that do not glycosylate, these findings bring the validity of some experimental results (including our own) into question.

Binding isotherms of eGFP-CBM fusions expressed in *E. coli* and *N. crassa* could be used to investigate the influence of glycosylation on binding behavior. Using the family 1 CBM domains from *T. reesei* Cel7A and Cel6A cellobiohydrolases, we have constructed two eGFP-CBM fusions containing *T. reesei* Cel7A linkers. When expressed in *N. crassa*, these constructs should be glycosylated on their CBM (N-linked) and linkers (O-linked); when expressed in *E. coli*, they will have no glycosylation. Binding isotherms of constructs expressed in each organism will be non-identical if glycosylation imparts an enhanced or reduced cellulose affinity to these CBMs. The glycosylation sites most responsible for either effect can then be pinpointed through selective site removal via mutation by site-directed mutagenesis. The results of such studies are likely to be regime specific; many CBMs bind to both high and low affinity sites, occupying the latter only at large CBM/cellulose loadings. As the low affinity sites tend to overlap between different CBMs, binding to these regions may involve fewer specific interactions and may thus be more easily influenced by glycosylation. Regardless of the interactions most affected, the impact of glycosylation-influenced binding on the kinetics of hydrolysis could be explored in a multi-site model. In addition to revealing the importance of using fungal expression systems to express constructs for PALM and adsorption studies, investigations of the role of glycosylation on CBM and cellulase binding to low and high affinity cellulose sites may inform new strategies for enzyme design, which has previously been limited to amino acid sequence adjustments.

7.6 Oxidative Enzymes

Oxidative enzymes can be easily incorporated into the hydrolysis model. What follows is a plan for incorporating kinetic steps for cellobiose dehydrogenase (CDH) and polysaccharide monooxgenase (PMO) enzymes; this plan is based on current theories of PMO and CDH mechanism (*6, 7*). Before engaging in catalysis, a CDH undergoes the same kinetic steps as a

cellobiohydrolase: while in solution, it complexes with the reducing end of a soluble cellooligosaccharide; while adsorbed on cellulose, it complexes with the reducing end of a soluble cellooligosaccharide or a surface cellulose chain. Then, the CDH generates a lactone and decomplexes from its substrate. A PMO engages in similar catalytic steps to an endoglucanase, but its adsorption and complexation steps are not separate: it (1) adsorbs and complexes, (2) cleaves a glycosidic bond, and (3) decomplexes and desorbs. The remaining details relate to (1) binding site preference and (2) electron transfer. The crystalline-specific activities of PMOs can be accommodated with the multi-site model discussed above, and the single electron transfer steps from CDHs to PMOs can be represented as bimolecular reactions between the two enzymes (on the surface or in solution). As mechanistic details pertaining to the action and interaction of these two enzymes continue to evolve, their associated kinetic steps may need to be modified within the model; however, as these changes are unlikely to involve behaviors that are significantly different from those of cellulase enzymes pre-catalysis, modifications to model equations are likely to be minor.

7.7 Conclusion

We conclude by noting that the hydrolysis model is intended to be a useful tool to understand and engineer cellulolytic enzymes; it is not intended to recreate reality. The motivation for introducing non-spherical particles, multiple sites, or oxidative enzymes does not stem from a general desire to recreate *in silico* the complexity of natural lignocellulosic substrates *in vivo*, it arises from our desire for a more complete understanding of the manner in which different binding behaviors or morphology-specific enzyme activities influence the overall hydrolysis kinetics of cellulolytic enzyme mixtures acting on heterogeneous substrates.

The investigations discussed in this work have been part of a general effort to investigate the kinetic impediments encountered by cellulase enzymes as they catalyze the hydrolysis of cellulose. Our results suggest that cellobiohydrolase activity on crystalline cellulose is limited by its rate of complexation with cellulose chains and by its tendency to encounter physical impediments to processivity; we further posit that the kinetics of enzymatic cellulose hydrolysis are limited by cellobiohydrolase activity and that enzymes capable of enhancing the activity of cellulolytic mixtures likely do so by allowing cellobiohydrolases to work more effectively. With regard to such enzyme capabilities, we find that substrate particle shape, surface area, degree of polymerization, and heterogeneity play a critical role in influencing enzyme activity and enzyme-enzyme synergy. Our results reveal strategies for using morphological targeting to enhance the activity of cellulolytic mixtures, and they motivate experimental investigations and modeling studies that probe the molecular level details of specific enzyme-surface interactions and the influence of those interactions on the performance of multi-enzyme systems.

7.8 References

1. L. R. Lynd, P. J. Weimer, W. H. van Zyl, I. S. Pretorius, Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev* **66**, 506 (Sep, 2002).
2. Y. H. Zhang, L. R. Lynd, Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems. *Biotechnol Bioeng* **88**, 797 (Dec 30, 2004).

3. M. Kurasin, P. Valjamae, Processivity of cellobiohydrolases is limited by the substrate. *J Biol Chem* **286**, 169 (Jan 7, 2011).
4. J. M. Fox, S. E. Levine, D. S. Clark, H. W. Blanch, Initial- and processive-cut products reveal cellobiohydrolase rate limitations and the role of companion enzymes. *Biochemistry* **51**, 442 (Jan 10, 2012).
5. K. Igarashi *et al.*, Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface. *Science* **333**, 1279 (Sep 2, 2011).
6. W. T. Beeson, C. M. Phillips, J. H. Cate, M. A. Marletta, Oxidative cleavage of cellulose by fungal copper-dependent polysaccharide monooxygenases. *Journal of the American Chemical Society* **134**, 890 (Jan 18, 2012).
7. C. M. Phillips, W. T. Beeson, J. H. Cate, M. A. Marletta, Cellobiose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by Neurospora crassa. *ACS Chem Biol* **6**, 1399 (Dec 16, 2011).
8. X. Li, W. T. t. Beeson, C. M. Phillips, M. A. Marletta, J. H. Cate, Structural basis for substrate targeting and catalysis by fungal polysaccharide monooxygenases. *Structure* **20**, 1051 (Jun 6, 2012).
9. S. E. Levine, J. M. Fox, H. W. Blanch, D. S. Clark, A mechanistic model of the enzymatic hydrolysis of cellulose. *Biotechnol Bioeng* **107**, 37 (Sep 1, 2010).
10. S. E. Levine, J. M. Fox, D. S. Clark, H. W. Blanch, A mechanistic model for rational design of optimal cellulase mixtures. *Biotechnol Bioeng* **108**, 2561 (Nov, 2011).
11. J. M. Fox *et al.*, Superresolution Images Reveal Binding Targets for the Cooperative Action of Cellulolytic Enzyme Mixtures. *Submitted*, (2012).
12. J. M. Fox, S. E. Levine, H. W. Blanch, D. S. Clark, An evaluation of cellulose saccharification and fermentation with an engineered Saccharomyces cerevisiae capable of cellobiose and xylose utilization. *Biotechnol J*, (Jan 9, 2012).
13. S. P. Chundawat, G. T. Beckham, M. E. Himmel, B. E. Dale, Deconstruction of lignocellulosic biomass to fuels and chemicals. *Annu Rev Chem Biomol Eng* **2**, 121 (2011).
14. C. M. Dana *et al.*, Biased clique shuffling reveals stabilizing mutations in cellulase Cel7A. *Biotechnol Bioeng* **109**, 2710 (Nov, 2012).
15. J. Hong, X. Ye, Y. H. Zhang, Quantitative determination of cellulose accessibility to cellulase based on adsorption of a nonhydrolytic fusion protein containing CBM and GFP with its applications. *Langmuir* **23**, 12535 (Dec 4, 2007).
16. B. W. McLean *et al.*, Carbohydrate-binding modules recognize fine substructures of cellulose. *J Biol Chem* **277**, 50245 (Dec 27, 2002).
17. N. R. Gilkes *et al.*, The adsorption of a bacterial cellulase and its two isolated domains to crystalline cellulose. *J Biol Chem* **267**, 6743 (Apr 5, 1992).
18. A. B. Boraston, E. Kwan, P. Chiu, R. A. Warren, D. G. Kilburn, Recognition and hydrolysis of noncrystalline cellulose. *J Biol Chem* **278**, 6120 (Feb 21, 2003).
19. G. T. Beckham *et al.*, Harnessing glycosylation to improve cellulase activity. *Curr Opin Biotechnol* **23**, 338 (Jun, 2012).
20. C. B. Taylor *et al.*, Computational investigation of glycosylation effects on a family 1 carbohydrate-binding module. *J Biol Chem* **287**, 3147 (Jan 27, 2012).