Title
Engineering Cellulase Enzymes for Bioenergy

Author
Atreya, Meera Elizabeth

Publication Date
2015

Peer reviewed|Thesis/dissertation

Engineering Cellulase Enzymes for Bioenergy

By

Meera Elizabeth Atreya

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Douglas S. Clark, Co-Chair
Professor David F. Savage, Co-Chair
Professor Matthew B. Francis
Professor Chris R. Somerville

Summer 2015

Abstract

Engineering Cellulase Enzymes for Bioenergy

by

Meera Elizabeth Atreya

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Douglas S. Clark, Co-Chair

Professor David F. Savage, Co-Chair

Sustainable energy sources, such as biofuels, offer increasingly important alternatives to fossil fuels that contribute less to global climate change. The energy contained within cellulosic biofuels derives from sunlight energy stored in the form of carbon-carbon bonds comprising sugars such as glucose. Second-generation biofuels are produced from lignocellulosic biomass feedstocks, including agricultural waste products and non-food crops like *Miscanthus*, that contain lignin and the polysaccharides hemicellulose and cellulose. Cellulose is the most abundant biological material on Earth; it is a polymer of glucose and a structural component of plant cell walls. Accessing the sugar is challenging, as the crystalline structure of cellulose resists degradation; biochemical and thermochemical means can be used to depolymerize cellulose.

Cellulase enzymes catalyze the biochemical depolymerization of cellulose into glucose. Glucose can be used as a carbon source for growth of a biofuel-producing microorganism. When it converts glucose to a hydrocarbon fuel, this microbe completes the biofuels process of transforming sunlight energy into accessible, chemical energy capable of replacing non-renewable transportation fuels.

Due to strong intermolecular interactions between polymer chains, cellulose is significantly more challenging to depolymerize than starch, a more accessible polymer of glucose utilized in first-generation biofuels processes (often derived from corn). While most mammals cannot digest cellulose (dietary fiber), certain fungi and bacteria produce cellulase enzymes capable of hydrolyzing it. These organisms secrete a wide variety of glycoside hydrolase and other classes of enzymes that work in concert.

Because cellulase enzymes are slow-acting and expensive to produce, my aim has been to improve the properties of these enzymes as a means to make a cellulosic biofuels process possible that is more efficient and, consequently, more economical than current methods. Protein engineering targets to improve cellulases include reducing enzyme inhibition, improving inter-enzyme synergy, and increasing enzyme thermotolerance. Ameliorating enzyme inhibition could improve catalytic activity and thus the speed of conversion from biomass to fermentable sugars. Improved enzyme synergy could reduce the enzyme loading required to achieve equivalent biomass conversion. Finally, thermostable enzymes could enable more biomass to be processed at a time, due to high temperatures decreasing the viscosity of biomass slurries. A high-temperature enzyme saccharification reaction could also decrease the risk of contamination in the resulting concentrated sugar solution. Throughout my PhD, I have explored research projects broadly across all of these topics, with the most success in addressing the issue of enzyme inhibition.

Cellulase enzyme Cel7A is the most abundant cellulase employed by natural systems for cellulose hydrolysis. Cellobiohydrolase enzymes like Cel7A break down cellulose into cellobiose (two glucose molecules). Unfortunately, upon cleavage, this product molecule interferes with continued hydrolytic activity of Cel7A; the strong binding of cellobiose in the active site can obstruct the enzyme from processing down the cellulase chain. This phenomenon, known as product inhibition, is a bottleneck to efficient biomass breakdown.

Using insights from computational protein modeling studies, I experimentally generated and tested mutant Cel7A enzymes for improved tolerance to cellobiose. Indeed, this strategy yielded Cel7A enzymes exhibiting reduced product inhibition, including some mutants completely impervious to cellobiose. The improvements in tolerance to cellobiose, however, resulted in an overall reduction of enzymatic activity for the mutants tested. Nevertheless, my findings substantiated computational reports with experimental evidence and pinpointed an amino acid residue in the Cel7A product binding site that is of interest for follow-up mutational studies.

My goal was to improve the effective catalytic activity of cellulase enzymes in industrially-relevant conditions (such as in the presence of high concentrations of cellobiose or at elevated temperatures). The insights gained from my work on enzyme inhibition may inform future efforts to address this important issue. More efficient enzymes should reduce the amount of these proteins needed to break down cellulose to glucose. This, in turn, should decrease the price of the resulting biofuel making it more cost-competitive with fossil fuels and thus encouraging adoption of renewable transportation fuels that reduce our greenhouse gas emissions.

Acknowledgements

I am thankful to my charismatic advisor, Professor Doug Clark, for granting me boundless intellectual freedom to investigate problems of my choosing and for always believing in me (no matter how ambitious my plan!). Thank you so much for your enduring vote of confidence. This arduous journey would not have been possible without the additional support of my brilliant lab-mates, playful friends, and devoted family. Generous financial support was provided by a National Science Foundation Graduate Research Fellowship and by the Energy Biosciences Institute—which I feel lucky to have been a part of throughout.

During my six years in graduate school at UC Berkeley, I was fortunate to have worked in a research area that I am deeply passionate about—sustainable energy. I learned many valuable lessons and grew immensely, both scientifically and personally. Most importantly, I have been humbled by my PhD experience, and for that I am grateful.

All I know is that I know nothing.

(Socrates)

Table of Contents

Chapter 1 – Introduction to Bioenergy.....1

Chapter 2 – Alleviating Product Inhibition in Cellulase Enzyme Cel7A.....21

Chapter 3 – Progress Towards Engineering a Lytic Polysaccharide Monooxygenase Enzyme.....45

Chapter 4 – Exploring Enzyme Synergy and Methods for Purification of *Trichoderma reesei* Cellulases.....59

Chapter 5 – Efforts Towards Enabling High-Throughput Directed Evolution of Cellulase Enzymes via *In Vitro* Compartmentalization.....75

Concluding Remarks.....93

References.....96

Chapter 1

Introduction to Bioenergy

1.1 Climate Change

Evidence that we are changing the world's climate is unequivocal. Anthropogenic emissions of greenhouse gases have resulted in concentrations of carbon dioxide, methane, and nitrous oxide that are unprecedented in nearly the last million years. These emissions, predominantly attributed to industrialization and population and economic growth, have triggered a warming of the climate system. Average temperatures of the atmosphere and ocean have increased, resulting in decreased snow and ice cover and thus rising sea levels (Figure 1). Meanwhile, 30% of anthropogenic carbon dioxide has been absorbed by the oceans, increasing ocean acidity and negatively affecting marine ecosystems. More alarming still is the fact that roughly half of anthropogenic carbon dioxide emissions between 1750 and 2011 occurred in just the last 40 years (Figure 1d)! Greater than three-fourths of these recent emissions are attributed to fossil fuel combustion and industrial processes.[1]

The impacts of global climate change are widespread and ongoing. Human activity is almost certainly responsible for more than half of the global surface temperature increase observed over the past 60 years. Changes in global temperatures and precipitation patterns are projected to get even more severe over the next century (Figure 2).[1] Likewise, we can expect extreme weather and climate events to occur with greater frequency and intensity as greenhouse gas emissions continue. For example, human influence is believed to have already doubled the probability of heat waves in some places. Changes like these will likely drive extinctions of both plant and animal species that are unable to adapt quickly enough; biodiversity will suffer as ecosystems are lost. Additionally, the risks of irreversible alterations to the climate increase with the magnitude of the warming.[1] Most chilling is fact that these significant effects on the Earth's climate are long-lasting—even if anthropogenic emissions are stopped, many impacts of climate change are irreversible and will continue for centuries.

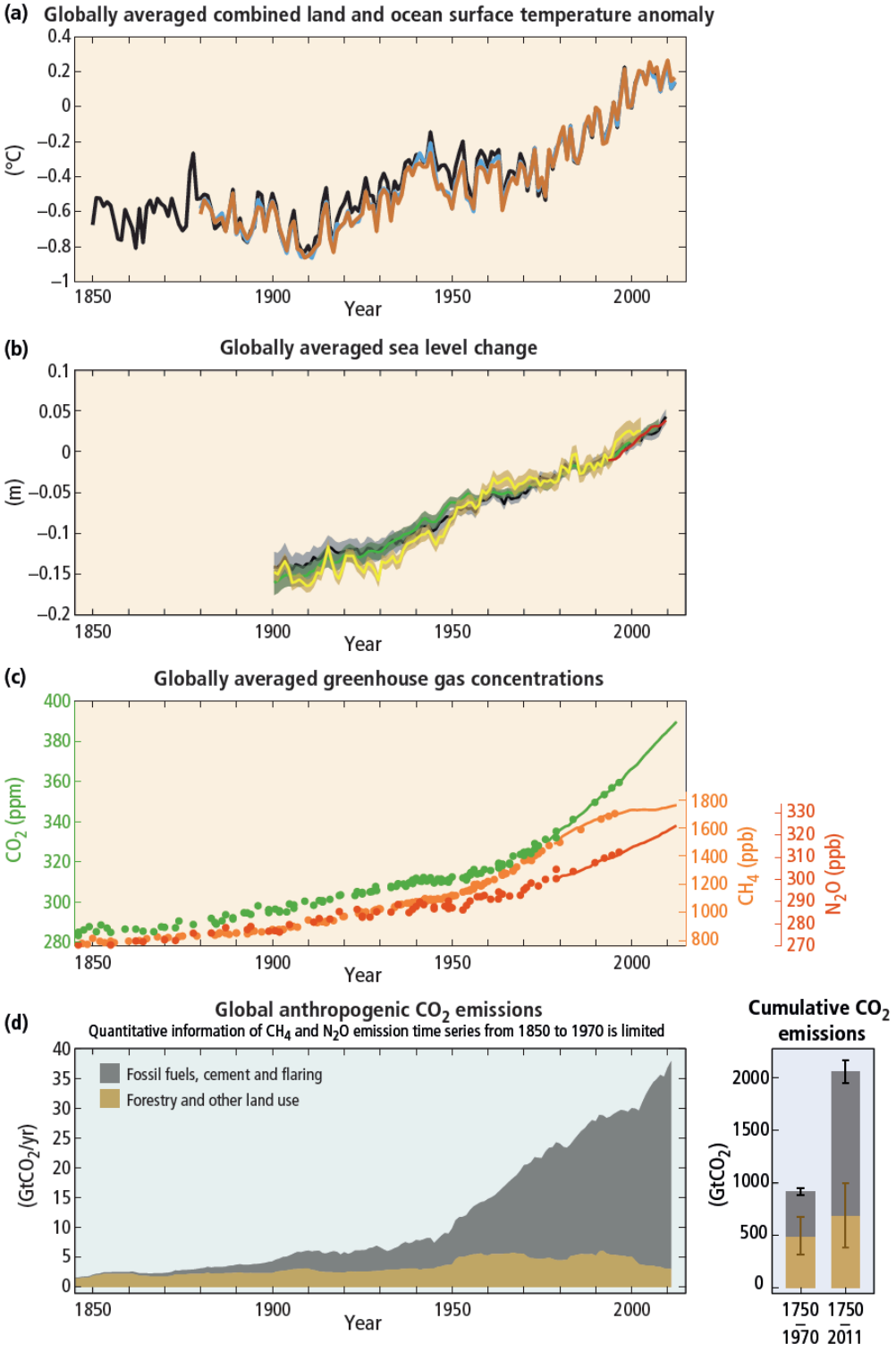


Figure 1. Global temperature (a), sea level (b), and greenhouse gas (c) levels measured mirror the striking increase in recent years of anthropogenic carbon dioxide emissions (d).[1]

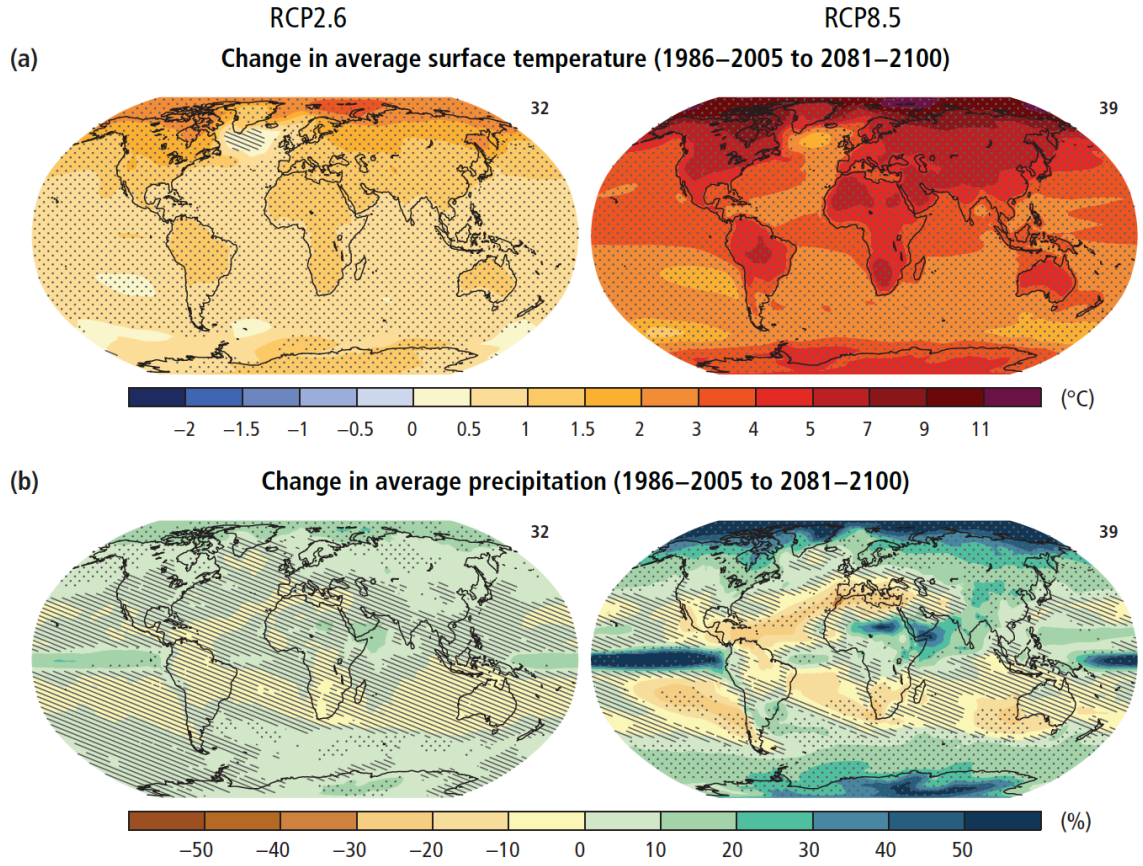


Figure 2. Projected changes in average global surface temperatures (a) and precipitation (b) under two scenarios – stringent greenhouse gas mitigation (left) and unrestrained greenhouse gas emissions (right).[1]

While anthropogenic contributions to climate change are disproportionately caused by the wealthy (enjoying energy-extravagant lifestyles), it is the poor that will unfairly suffer the most from its impacts.[1] Climate effects are not limited to weather, but rather touch the most fundamental elements sustaining human life, including food availability and health (Figure 3). Climate change poses large risks to global food security (as fishery productivity decreases and crop yields waver) and only exacerbates the health problems we face today – particularly in developing regions. People living in rural locations and in circumstances of poor resource and infrastructure availability will be especially vulnerable to the risks increased by climate change, including variations to food and water availability, extreme weather events, flooding, and air pollution. Economic productivity losses will also accelerate as global temperatures increase. Moreover, potential population displacement and resource scarcity will likely increase the risk of violent conflicts.[1]

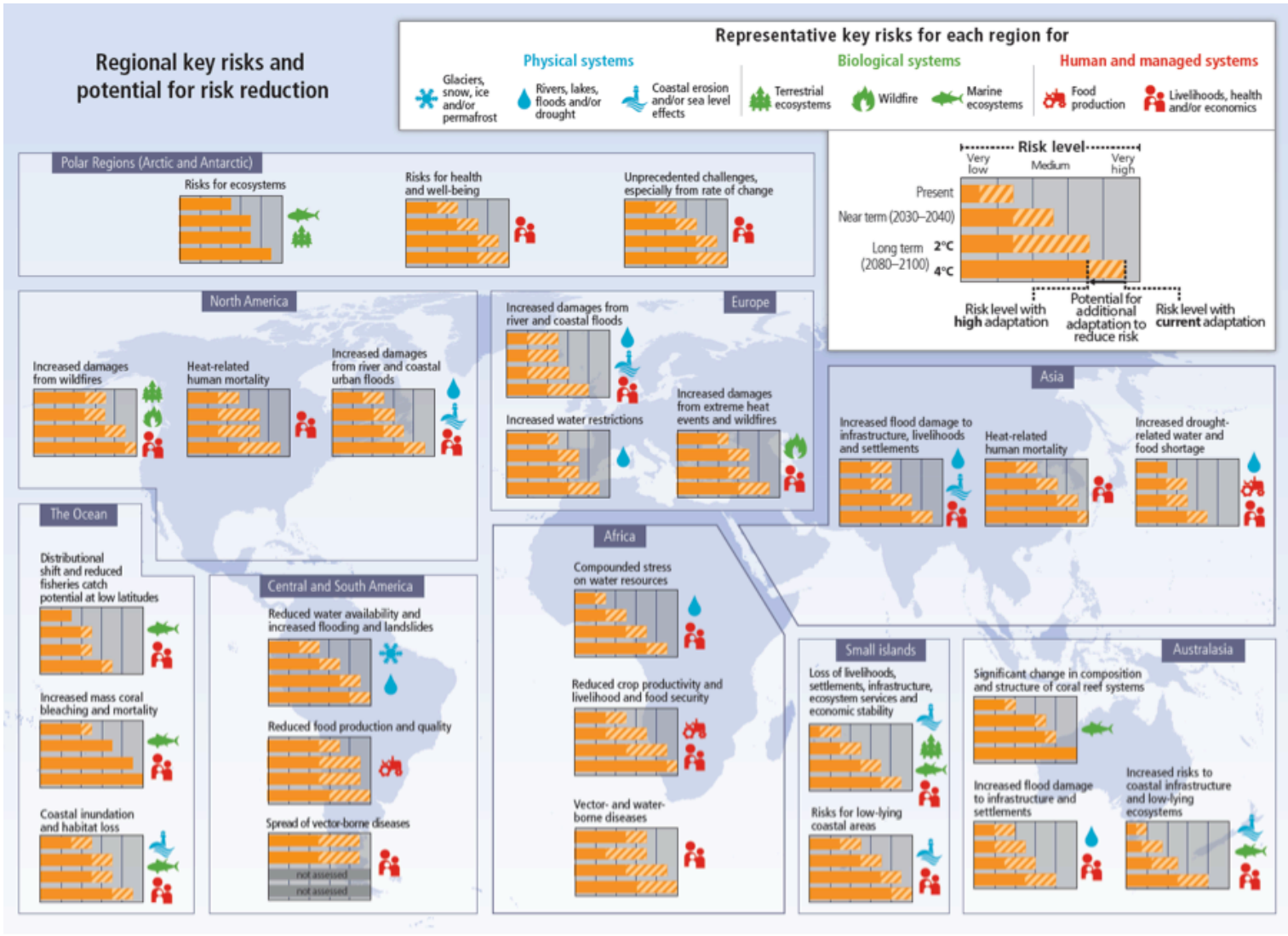


Figure 3. Global climate change risks by region.[1]

The anthropogenic greenhouse gas emissions contributing to the harrowing effects detailed above are detailed by sector in Figure 4. The transportation sector, in particular, contributes to 14% of global greenhouse gas emissions.

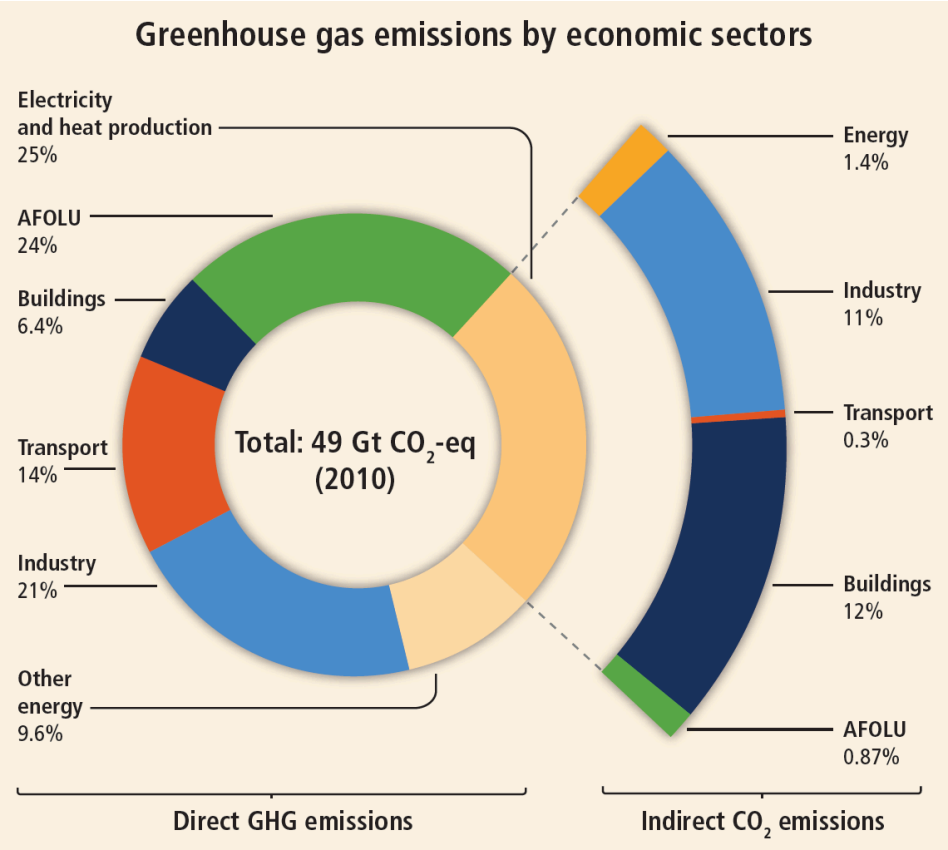


Figure 4. Greenhouse gas emissions by economic sectors. (AFOLU stands for agriculture, forestry, and other land use.)[1]

1.2 Our Challenge

We live in a time where developing world populations, and their corresponding demands for energy, are booming while the resources of the Earth remain finite and our climate system delicate. Ours is a unique challenge—how do we satisfy the irrefutable desire of people to enjoy an energy-rich, modern lifestyle without spoiling that opportunity for generations to come by off-balancing the global climate?

The solution is sustainable energy—energy derived from non-exhaustible resources such that its use can satisfy present demand without compromising the ability of future generations to meet their needs.[2] Sustainable energy encompasses both renewable energy

and energy efficiency; both are critically important for addressing this important problem.

Minimizing climate change necessitates reductions in greenhouse gas emissions that are both significant and sustained.[1] Our transition from fossil fuels to energy sources that are renewable (on timescales far shorter than millions of years) encompasses a range of technologies harnessing energy from (direct) solar, wind, hydro, tidal, geothermal, and biomass sources.

1.3 Renewable Bioenergy

Bioenergy is derived from sunlight energy collected and stored chemically by plants as carbon-carbon bonds (predominantly). Bioenergy is considered a renewable resource because it is naturally replenished on a human timescale. Unlike fossil fuels that release carbon stored underground for millions of years into the atmosphere, biomass energy operates under a theoretically carbon-neutral cycle (Figure 5).

Figure 5. Simplified biomass energy cycle demonstrating carbon recycling.[3]

As illustrated in Figure 5, plants fix carbon dioxide from the atmosphere into organic compounds via photosynthesis. Subsequent combustion of the biomass (or initial conversion of carbohydrates to more energy-dense hydrocarbon fuels, followed by combustion) releases the carbon back into the atmosphere where it can be recycled and reincorporated into carbohydrates by photosynthetic plants. This simplified description of energy from biomass (omitting carbon emissions attributed to production, harvesting, and distribution processes, for example) demonstrates the near carbon-neutral potential of biomass energy.

Biomass is the only renewable carbon source for the production of convenient fuels.[4] Bioenergy is particularly well suited to provide sustainable alternatives for transportation, a sector almost entirely powered by fossil fuel combustion, due to the energy density of chemical energy. We can biologically convert sunlight energy stored by plants as carbohydrates into energy-dense hydrocarbons that can directly replace the nonrenewable transportation fuels used today (such as gasoline and diesel). The transportation sector contributes to a significant amount (14%) of greenhouse gas emissions (Figure 4). Until energy storage (battery) technologies mature and their use in this sector becomes widespread, biofuels could provide a more carbon-neutral alternative that decreases the net greenhouse gas emissions contributing to global climate change. It is estimated that biofuels can offset 30% of fossil fuel usage in the United States; as such, biomass will play a leading role in meeting near- to mid-term sustainable energy targets.[5]

Any biochemical biofuels process consists of two major processing steps: (1) conversion of biomass to accessible sugars and (2) conversion of sugars to fuel molecules (Figure 6).

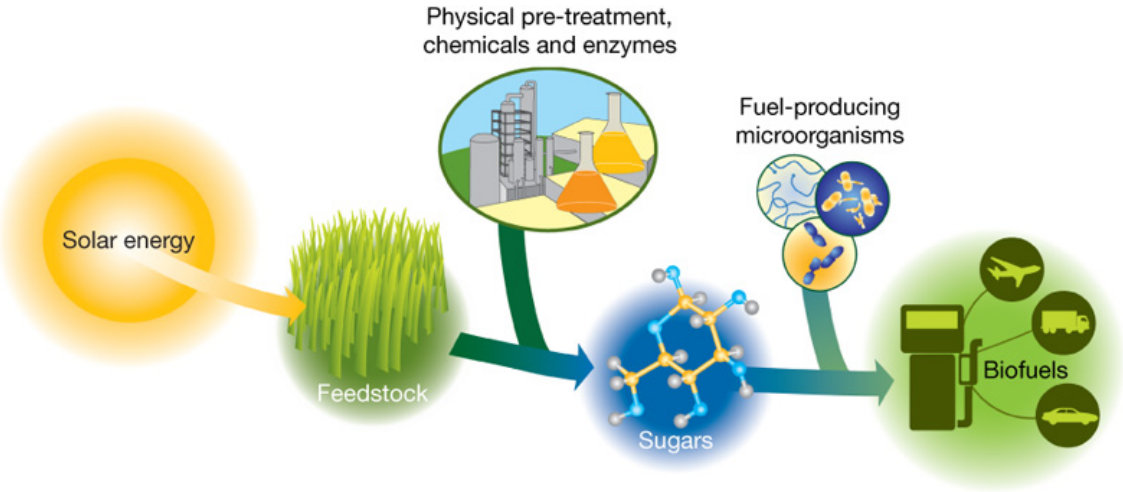


Figure 6. Macroscopic overview of a biochemical biofuels process.[6]

1.3.1 First-Generation Biofuels

To date, the feedstock (biomass energy source) for first-generation biofuels has been predominantly corn (or sugar cane, in enabling climates). About 40% of the corn supply in the United States (US) is used to produce ethanol. Most gasoline sold in the US contains 10% ethanol – this represents a "blend wall" above which only newer, specially-designed engines can handle without corrosion.

Unfortunately, replacing gasoline with corn ethanol has not been shown to substantially reduce emissions. In fact, some research suggests that it may actually increase them! However, the US Environmental Protection Agency (EPA) estimates that, with predicted technology advancements, emissions for corn ethanol will be about 20% lower than those for gasoline by 2022. Factors that affect biofuel emissions include crop yields, fertilizer use, changes in land use and soil carbon, and the efficiency of the conversion process from biomass feedstocks to fuel.[7]

1.3.2 Second-Generation Biofuels

Unlike corn ethanol, second-generation (cellulosic) biofuels are generally considered to produce fewer emissions than gasoline. Relatively few energy inputs are required to produce lignocellulosic biomass feedstocks, and residual lignin can be burned to generate electricity (replacing a fraction of fossil fuel-sourced electricity). Corn stover, in particular, is predicted by the EPA to have the lowest emissions of any cellulosic biofuels feedstock due to land use change not being a factor.[7] Corn stover is the waste biomass remaining after corn production, such as the leaves, stalk, and cob, together making up roughly half of the crop yield.

While 1st generation biofuels are derived from starch, 2nd generation biofuels rely on recalcitrant polysaccharides, like cellulose, as the renewable carbon source. Designed by nature to be robust structural components of plant cell walls, cellulose and hemicellulose are unsurprisingly difficult to degrade into the individual sugars that comprise the polymers. It is this challenge that fundamentally differentiates lignocellulosic biofuels from its relatively successful counterpart, corn ethanol.

Cellulose is the most abundant biological material in the world. It contributes to the structure of plant cell walls (Figure 9) and, together with other polymers hemicellulose and lignin, creates a structural material for plants that is resistant to pathogenic attacks. Cellulose is the main component of biomass, comprising up to half of plant cell walls by dry weight (Figure 8).[8] Its natural abundance and availability as an agricultural waste product make cellulose an attractive feedstock for biofuels.

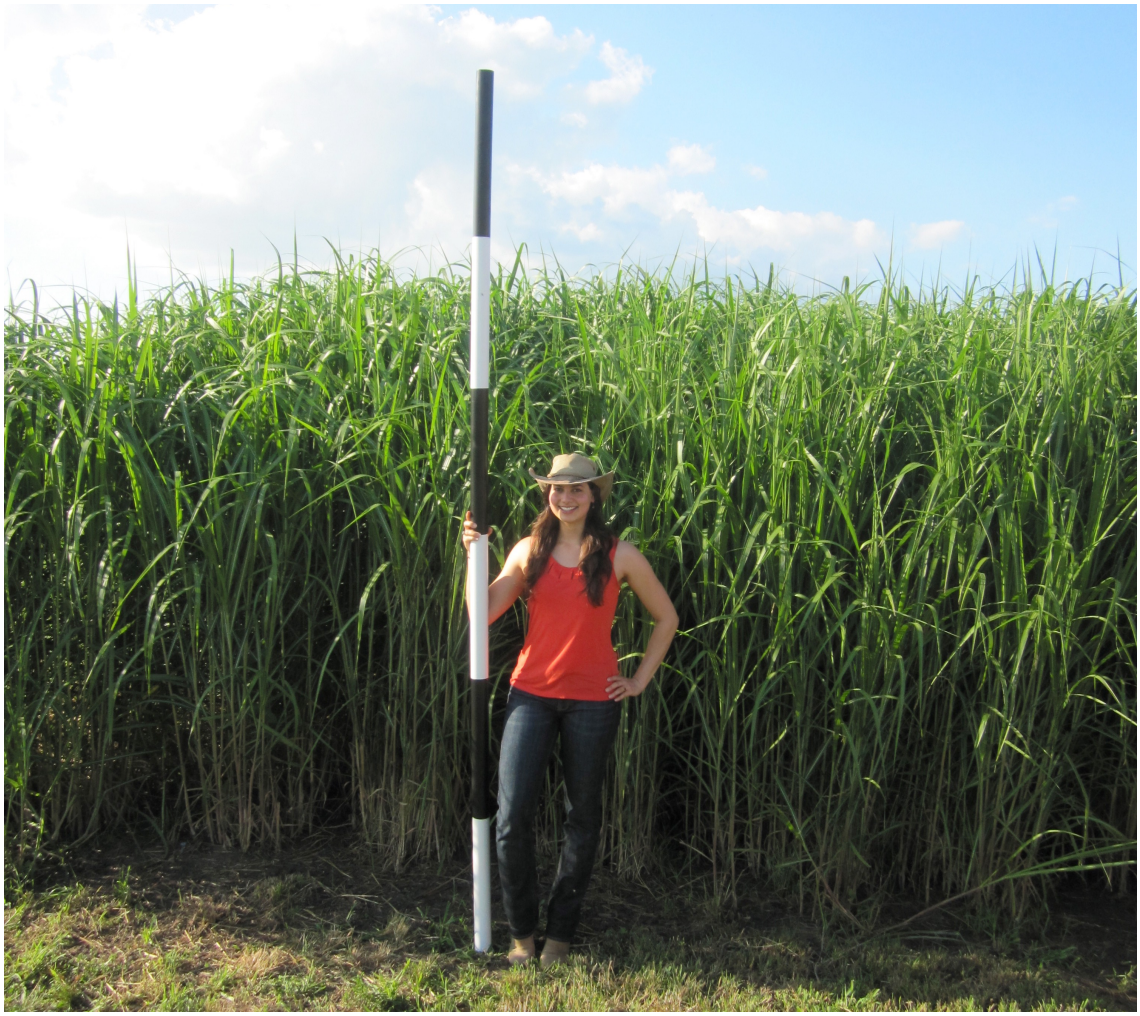


Figure 7. *Miscanthus* is a promising lignocellulosic biofuels feedstock.

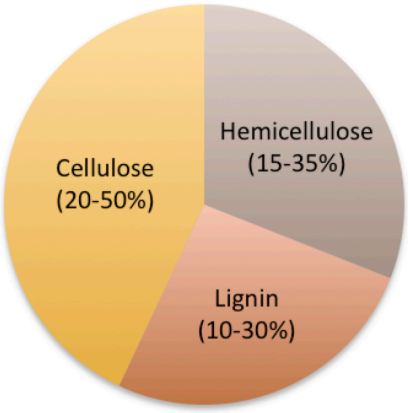


Figure 8. Composition (by dry mass) of plant cell walls from lignocellulosic biomass.[5]

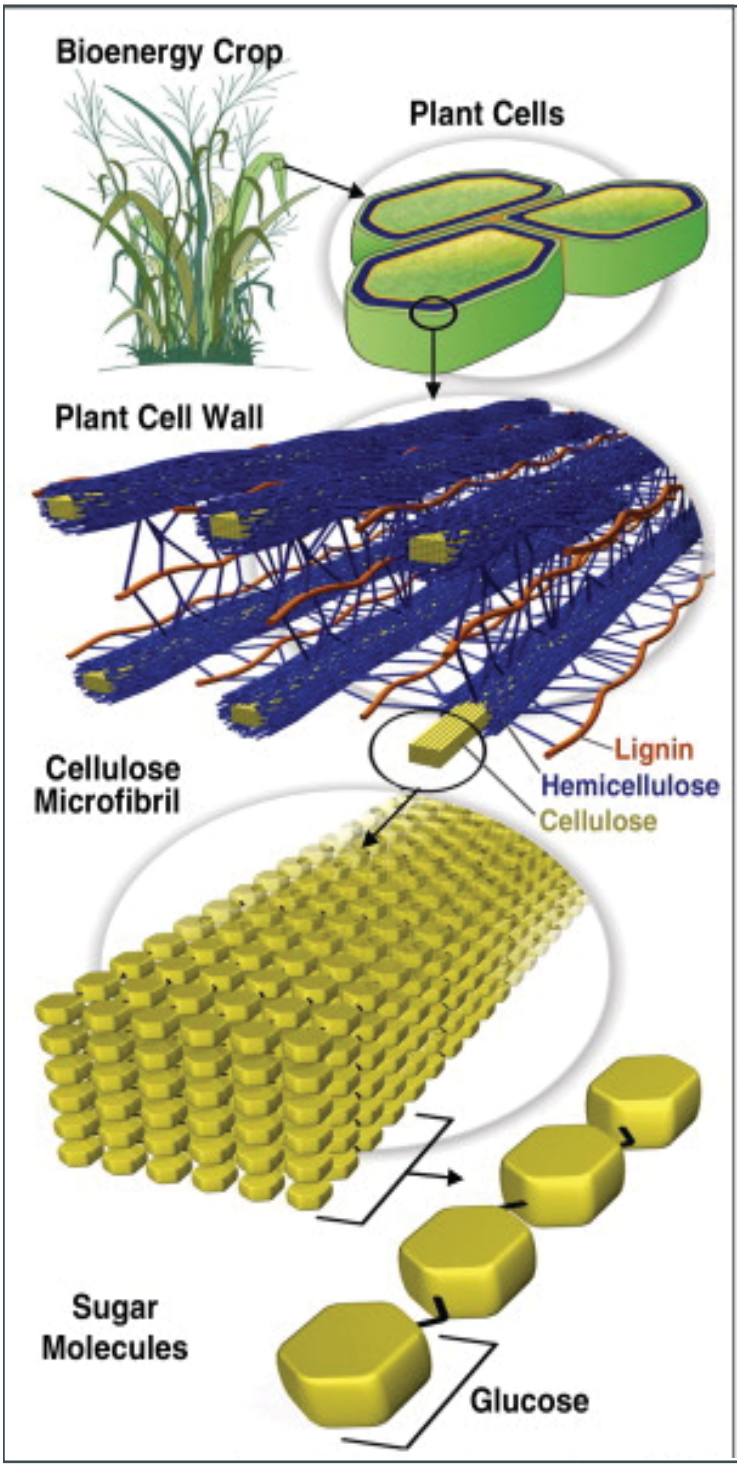


Figure 9. Cellulose, a polymer of glucose, is the primary component of plant cell walls.[9]

Cellulose is a polymer of β-1,4 linked glucose. Starch, another polymer of glucose, differs only by the nature of the chemical bond between sugar molecules occurring in the alpha, not beta, position (Figure 10). While starch is easy to depolymerize into free glucose molecules, cellulose is extremely resistant to degradation. Not only does this polysaccharide have extremely strong internal covalent bonds (twice as stable as linkages in DNA[5]), but many energetically-favorable interactions (hydrogen bonds and van der Waals contacts) form between individual cellulose chains (Figure 10, right). Cellulose microfibrils thus pack together to form crystalline lattices that are resistant to degradation, as only a fraction of the cellulose is surface-exposed (Figure 9). Avicel is a commercially available substrate representing crystalline cellulose that is widely used experimentally. The repeating unit of cellulose polymers is cellobiose, composed of two linked glucose molecules (Figure 11).

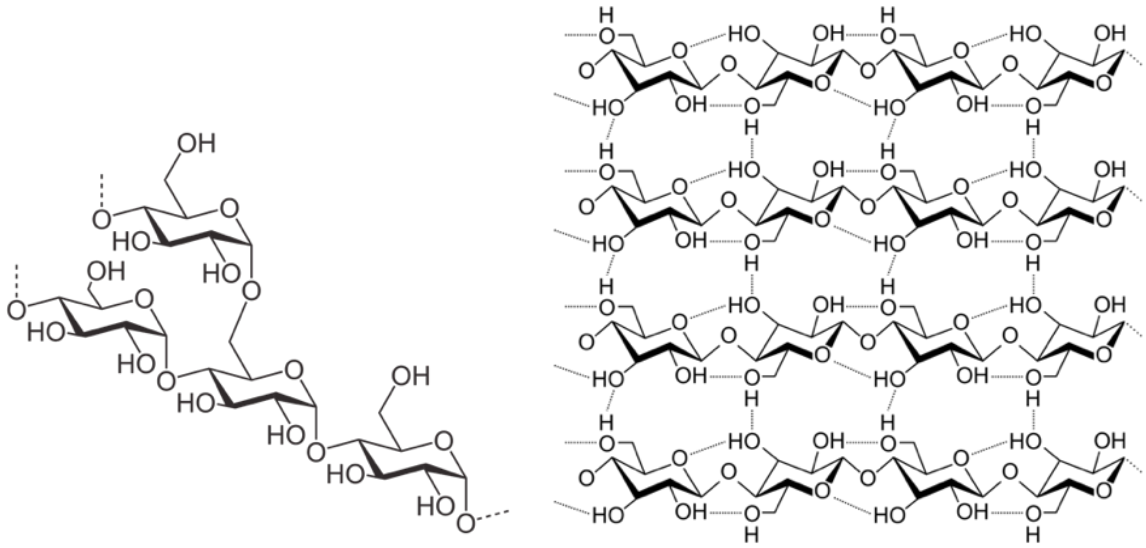


Figure 10. The power of chemistry actuates facile depolymerization of starch (amylopectin, left) into accessible sugars, while cellulose (right) resists degradation by most biological systems. Both polymers are composed of glucose, differing merely by the orientation of a bond.[10]

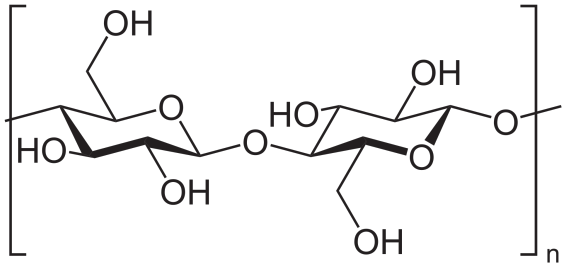


Figure 11. Cellobiose is the repeating unit of the polymer cellulose, consisting of two glucose molecules.

Prior to biochemical conversion of cellulosic feedstocks into accessible sugars, biomass must be milled after harvesting, to reduce its size, and pretreated (with dilute hot acid, for example) to render the plant cell wall accessible to degradation.[5]

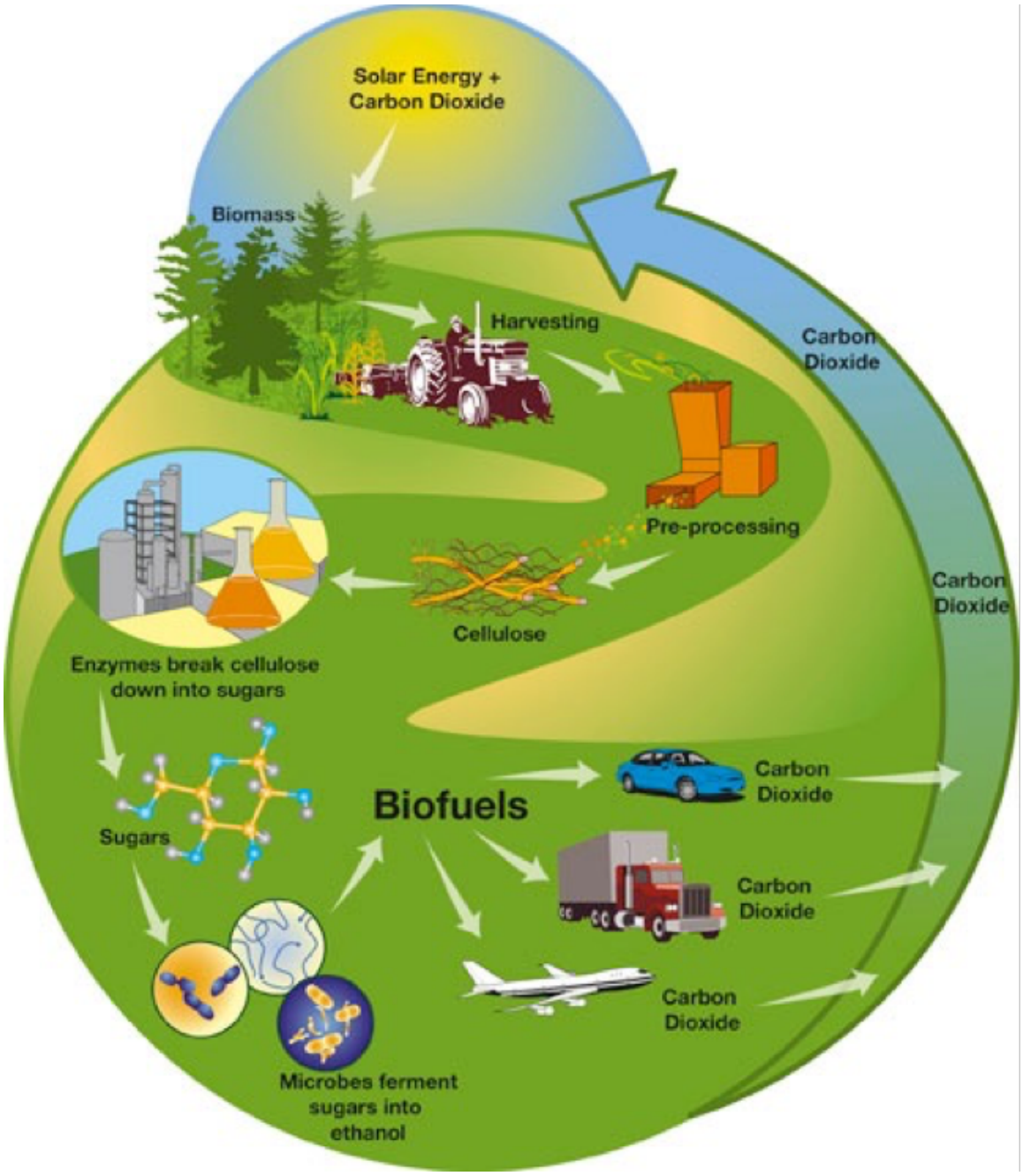


Figure 12. Cellulosic biofuels overview. Physical and thermochemical biomass pretreatments are followed by enzymatic saccharification and subsequent microbial fermentation into biofuels.[11]

1.4 Enzymatic Degradation of Cellulose

In order to effectively degrade crystalline polymers like cellulose, organisms with this ability employ a number of enzymes that work synergistically to break down plant polysaccharides. Unlike most enzymes, these proteins act on an insoluble substrate (which can complicate experimental studies). At neutral pH and without catalysis, the half-life of cellulose is a staggering 5 million years; intact cellulose has even been isolated from fossilized plants. Chemical catalysis by enzymes is remarkably powerful; cellulase enzymes such as glycoside hydrolases can improve the rate of cellulose hydrolysis by as much as 100,000,000,000,000,000-fold![5]

In nature, it is predominantly fungi that degrade biomass. During World War II, a fungus known as *Trichoderma reesei* (*Hypocrea jecorina*) was discovered in the South Pacific for its role degrading US army tents and other cotton fabrics (cotton is nearly pure cellulose).[12,13] This soft-rot, filamentous ascomycete fungi produces high titers of celluloytic enzymes. *T. reesei* is the main mesophilic fungus used industrially for this purpose. When feeding off cellulose as its main carbon source, the organism secretes many types of enzymes that break down biomass together, including cellobiohydrolases Cel7A (50% of total cellulase content) and Cel6A (20%), endoglucanases Cel7B (15%), Cel5A (10%), Cel12A (1%), Cel45A (<1%) and several oxidative enzymes.[14] Cellulase enzymes are classified by families, some of the most studied of which include glycoside hydrolase (GH) enzyme family 7 cellobiohydrolase I (CBH1), known predominantly as GH7 Cel7A.

Cellulose hydrolysis by fungal enzymes is diagrammed in Figure 13. Cellobiohydrolases Cel7A (CBH1) and Cel6A (CBH2) degrade cellulose processively from reducing or non-reducing chain ends, respectively, hydrolyzing the polymer into the disaccharide cellobiose. These enzymes often have multiple domains including a carbohydrate binding module (CBM), linker, and catalytic domain (Figure 14). Beta-glucosidases break up soluble cellobiose units into glucose. Endoglucanases (EGs) hydrolyze the β-1,4 linkages within the polysaccharide; they cleave amorphous regions of cellulose and, due to their open binding cleft, are not processive. More recently discovered oxidative enzymes such as lytic polysaccharide monooxygenases (LPMOs, formerly GH61s) oxidatively cleave crystalline cellulose, generating breaks in the strand and therefore new chain ends (available for further degradation by other cellulases).

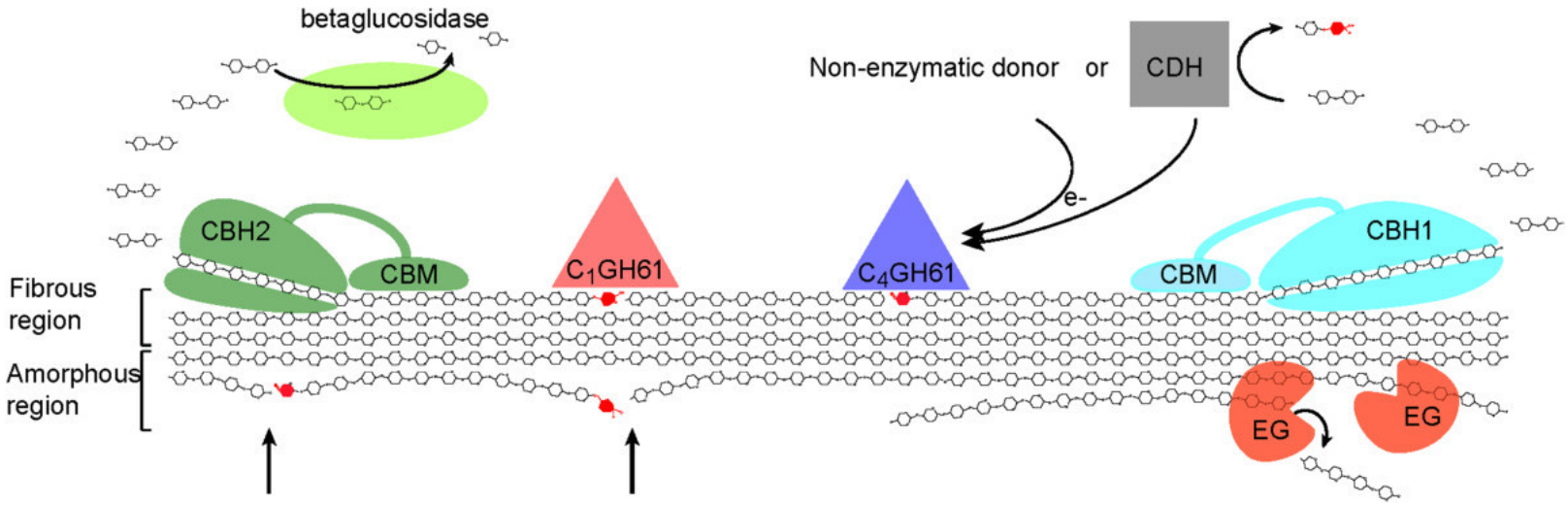


Figure 13. Cellulase enzymes depolymerize cellulose from lignocellulosic biomass into fermentable sugars (for subsequent microbial conversion to biofuels). Endoglucanases (EGs) hydrolyze amorphous cellulose while cellobiohydrolases (Cel7A or CBH1 and Cel6A or CBH2) are localized to more crystalline regions by their carbohydrate binding modules (CBMs). The product of processive cellulose hydrolysis by cellobiohydrolases is cellobiose, which is cleaved by β-glucosidase into glucose. Finally, oxidative enzymes such as LPMOs (formerly GH61s) create cellulose strand breaks in crystalline regions of the substrate following electron transfer from cellobiose dehydrogenase (CDH) or a small molecule reducing agent.[15]

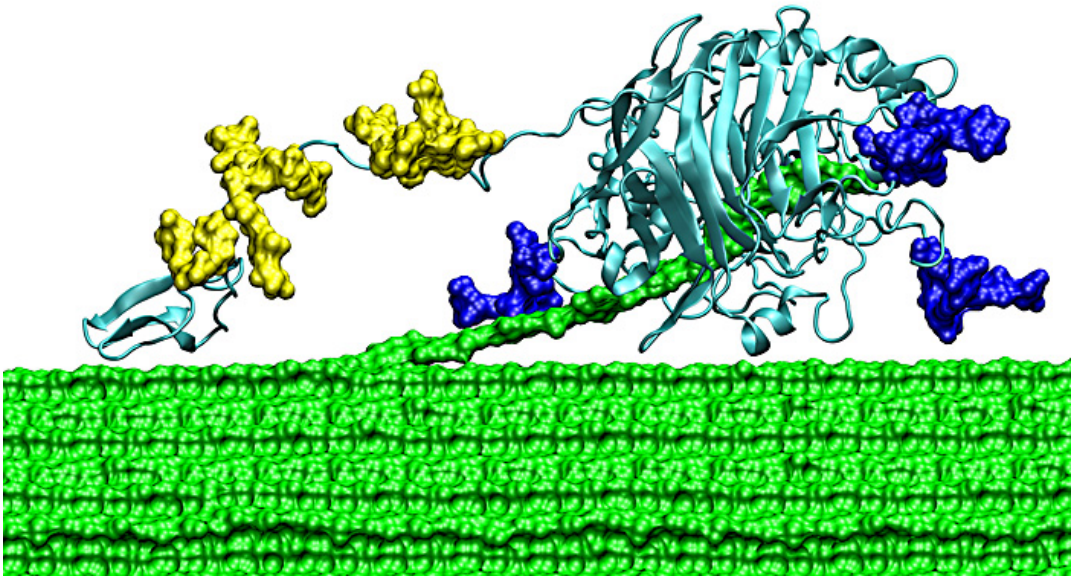


Figure 14. Cellobiohydrolase enzyme Cel7A from *T. reesei* has a carbohydrate binding module (CBM, left) connected with a flexible, glycosylated linker (yellow) to a catalytic domain (right). The enzyme processes along a cellulose chain (green). [15]

Glycoside hydrolase enzymes (GHs) cleave cellulose employing mechanisms that result in either inversion or net retention of stereochemistry at the anomeric carbon (Figure 15). Cel7A, for example, acts from the reducing end of cellulose and utilizes a retaining mechanism (Figure 15B). In the glycosylation step, a nucleophile in the enzyme's active site (e.g. glutamate) attacks the anomeric carbon of the substrate sugar, breaking the cellulose chain and forming a covalent glycosyl-enzyme intermediate. Deglycosylation of the enzyme follows, where an activated water molecule frees the enzyme for continued catalysis and resets the stereochemistry in the polysaccharide.

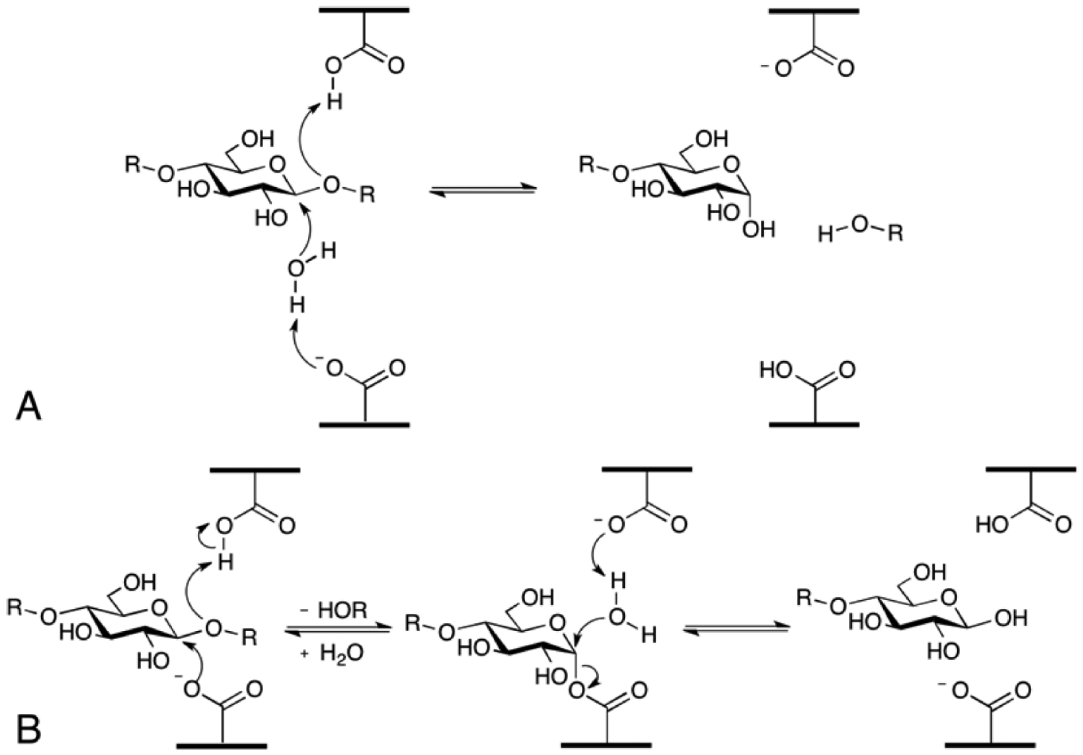


Figure 15. Glycoside hydrolase-catalyzed cellulose hydrolysis. Inverting enzymes invert the stereochemistry at the anomeric carbon (A) while retaining enzymes maintain the stereochemistry, employing a double-displacement catalytic mechanism (B).[5]

1.5 The Renewable Fuel Standard (RFS)

Renewable fuel alternatives, such as biofuels, must have sufficient policy support and/or be economically competitive with fossil fuels for widespread adoption to occur. In the United States, policy supporting biofuels was put in place with the Energy Independence and Security Act of 2007. As a means to reduce greenhouse gas emissions, the Renewable Fuel Standard (RFS) set minimum volumes for renewable fuels to be used in the US transportation sector.[7] While the fuel volumes were meant to increase annually through 2022, suppliers have not been able to produce even the lowest quantities set for cellulosic biofuels due to the complexity and capital intensity of the process. Commercial production began in 2013 with just two plants. This limited production capacity has led to the enormous gap between the policy mandates and actual production (Figure 16). There is accordingly a substantial and immediate need for technological improvements that would enable an efficient, affordable cellulosic biofuels process.

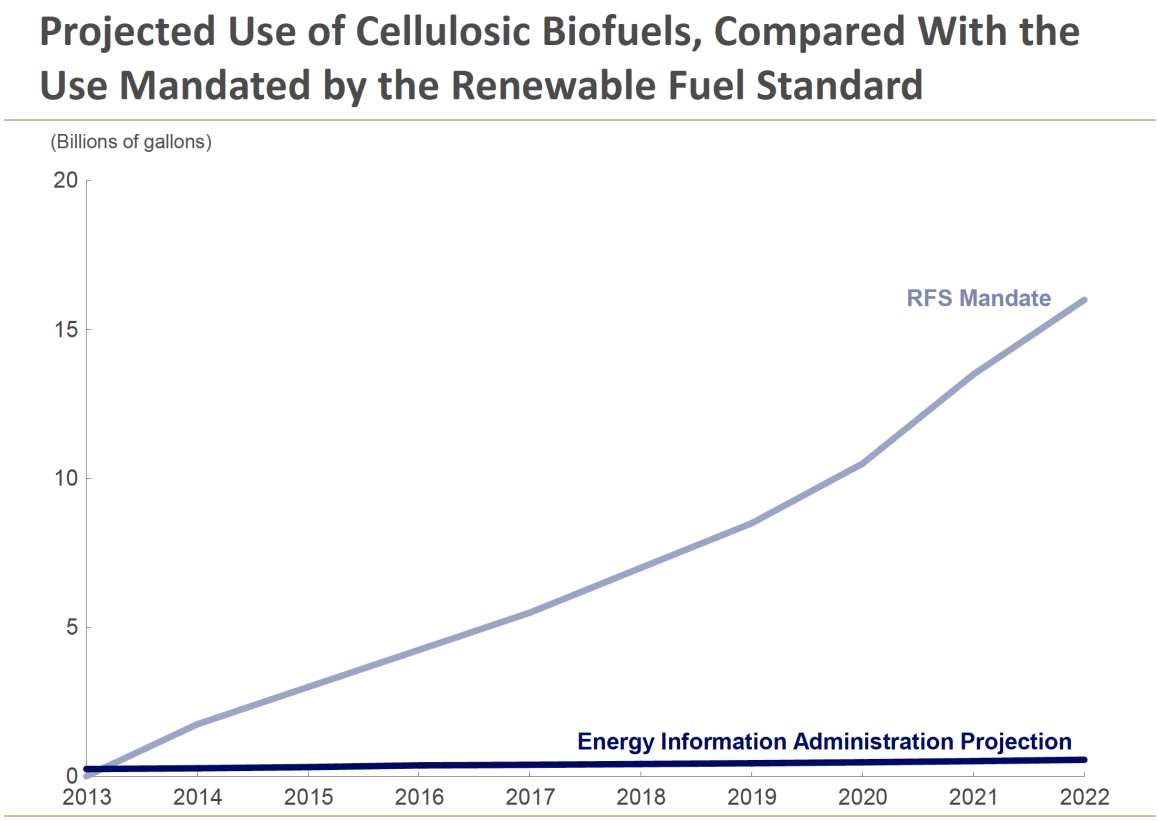


Figure 16. Renewable Fuel Standard (RFS) mandated fuel volumes are orders of magnitude higher than the actual production capacity for cellulosic biofuels.[7]

1.6 Opportunities for Improvement of Cellulosic Biofuels

To economically generate fuels and chemicals from lignocellulosic biomass, we must find a way to access cheap sugar from plant polysaccharides. Of the many processes required to produce cellulosic biofuels, the depolymerization steps, including pretreatment and enzymatic hydrolysis, are considered to be the most expensive components of the conversion.[5] While the degradation of cellulose could be performed thermochemically at high temperatures, biological methods (enzymes) are highly selective for glucose production and result in fewer side-product chemicals that act as inhibitors of downstream processes.[5]

Cellulose depolymerization represents a significant cost in the biofuel production process; enzymes are expensive to produce and may contribute $0.68 or more per gallon of ethanol produced from corn stover.[16] These costs are higher still when considering that ethanol fuel (E100) has only 67% of the chemical energy of gasoline (1.5x the volume of

ethanol is required to reach the gasoline equivalent in energy). This conversion brings the enzyme cost today to roughly $1.00 per gallon of gasoline equivalent.[17]

The US Department of Energy (DOE) has set a goal of reaching a 2017 target of $5 total per gallon of gasoline equivalent from cellulosic feedstocks with a final target of $3 by 2022.[7] To meet these goals, the efficiency of the biofuels process will need to be improved. Potential ways to alleviate cost bottlenecks include understanding and optimizing the process, creating value-added products in addition to commodity chemical fuels, and improving the enzymes used to depolymerize cellulose.

As there is no robust method to recycle enzymes, their high cost and one-time-use nature pose a problem. Generation of cellulase enzymes is challenging, expensive, and limits commercialization, seeding the desire for highly efficient cellulases capable of degrading cellulose quickly, while requiring minimal enzyme concentrations. To reduce the cost of accessing sugar from biomass, the National Renewable Energy Laboratory (NREL) has set a goal of halving the enzyme protein loading required to break down lignocellulosic biomass from today's 20mg enzyme/g cellulose to just 10 mg/g cellulose by 2017, while maintaining a 90% conversion target from cellulose to sugar. Utilizing this enzyme loading and with anticipated technological developments, the enzyme cost is predicted to be $0.37 per gallon of gasoline equivalent (of a total cost of $5.10 per gallon of gasoline equivalent), representing 7% of the production cost of a renewable diesel fuel blendstock.[17] With these goals in mind, there are several opportunities related to enzymes for improvements to a cellulosic biofuels process.

One prospect is to ameliorate enzyme product inhibition. Many cellulases are rendered less efficient by the chemical product of their activity, a trait that poses a substantial challenge for achieving near-complete enzymatic degradation of cellulosic biomass. It is not uncommon for a saccharification reaction to achieve 50% biomass hydrolysis after just 8 hours, yet require up to 96 hours to reach over 85% glucose. Product inhibition is believed to play a critical role in this slowdown, particularly as the industry trends towards high solid loadings (leading to high product concentrations). Development of product-tolerant enzymes could greatly improve process efficiency.

Recently, enzyme synergy has been hailed as a potential avenue for enhancing the enzymatic depolymerization biomass. Many enzymes work in concert to degrade cellulose, often at efficiencies together that are substantially greater than the sum of the enzymes' individual activities. This area of research has been of particular interest following the discovery of synergistic oxidative enzymes such as lytic polysaccharide monooxygenases (LPMOs). Developing synergistic enzyme cocktails tailored to most efficiently convert lignocellulosic biomass to sugars could reduce the enzyme protein loading required to achieve equivalent biomass conversion.

Finally, a cellulosic biofuels process could benefit from improved thermostability of cellulase enzymes. Once cellulose is depolymerized into glucose, there is a risk of contamination of this concentrated sugar stream. Performing the process at an elevated temperature would reduce the risk of contamination while simultaneously enabling a higher biomass solids loading in the reactor due to decreased viscosity, boosting process efficiency. For these reasons, and the fact that reaction rates increase with temperature, thermotolerant cellulase enzymes are of interest.

While cellulosic biofuels may be technologically more challenging than originally anticipated (as Figure 16 reveals), there are many exciting opportunities to improve the efficiency of a cellulosic biofuels process. Engineering cellulase enzymes for reduced product inhibition, improved inter-enzyme synergy, and enhanced thermostability are just a few promising routes towards developing more efficient and cost-competitive biofuels. The following chapters describe my endeavors towards achieving these goals in an attempt to drive adoption of sustainable energy technologies that can reduce our contribution to climate change.

Chapter 2

Alleviating Product Inhibition in Cellulase Enzyme Cel7A

2.1 Abstract

Biofuels produced from non-food crops rich in cellulose suffer from a difficult and expensive production process. Enzymes that degrade cellulose into glucose (for subsequent conversion to fuels and chemicals) are one of the most expensive components of the process. Improving the efficiency of these enzymes has the potential to make biofuels more cost-competitive with fossil fuels, thus encouraging adoption of renewable transportation fuels.

Cellulase enzyme Cel7A is the most abundant enzyme naturally employed by fungi to depolymerize cellulose.[5] Cel7A is inhibited by its product, cellobiose, which contributes to diminishing enzymatic activity over the course of depolymerization.[18] There is therefore great interest in minimizing the detrimental effects of product inhibition on Cel7A. I experimentally generated 10 previously proposed,[19–21] site-directed mutant Cel7A enzymes expected to have reduced cellobiose binding energies[19,21] and tested their resilience to cellobiose as well as their hydrolytic activities on microcrystalline cellulose. Although every mutation tested conferred reduced product inhibition (and even abolished it for some), my results confirm a trade-off between Cel7A tolerance to cellobiose and enzymatic activity: reduced product inhibition was accompanied by lower overall enzymatic activity on solid substrates for the mutants tested. The tempering effect of mutations on inhibition was nearly constant despite relatively large differences in activities of the mutants. My work identifies an amino acid in the Cel7A product binding site of interest for further mutational studies, and highlights both the challenge and the opportunity of enzyme engineering toward improving product tolerance in Cel7A.

2.2 Introduction

Biofuels represent one of many important renewable energy alternatives to fossil fuels with the potential to decrease anthropogenic effects on climate change. Cellulosic biofuels derive energy from chemical bonds stored by plants in the form of cellulose, a polymer of glucose and a primary structural component of plant cell walls.[22] Cellulose-rich biomass can be produced with fewer inputs than first-generation biofuel crops, such as starch-rich corn; however, cellulose is difficult to break down.[7] Once cellulose is depolymerized into glucose, the sugar can be microbially or chemically transformed into fuels and chemicals such as ethanol, butanol or other gasoline, jet fuel, and diesel alternatives.

While cellulose is abundant, accessing the sugar within is challenging. To degrade biomass, several enzymes work in concert, the most abundant of which is cellulase enzyme Cel7A.[5] Cel7A cellobiohydrolase enzymes depolymerize cellulose into its fundamental repeating unit of two glucose molecules—cellobiose. These enzymes suffer from inhibition by this product, which accumulates over the course of a reaction unless

removed by an enzyme such as β-glucosidase, which cleaves cellobiose yielding two glucose molecules.[5] Cel7A experiences mixed inhibition by cellobiose. The molecule can both competitively compete with a cellulose chain for binding in the substrate-binding sites as well as noncompetitively inhibit the enzyme by retarding processive motion as a result of persisting in the product-binding site.[5,23] Measurements on crystalline cellulose show that Cel7A loses half of its activity in the presence of cellobiose concentrations on the order of 2.6mM[24] to 19mM.[18] Enzymatic conversion of lignocellulosic biomass into glucose is especially impeded by product inhibition under the high substrate loadings required for commercial manufacture of biofuels.[5] Unfortunately, addressing this issue with the product inhibition-relieving enzyme β-glucosidase alone is not a comprehensive solution. Beta-glucosidase activity is limited by its own product inhibition from glucose, as well as by gluconic acid (generated by lytic polysaccharide monooxygenase (LPMO) activity).[5]

Enzymes are one of the most expensive components of a biochemical cellulosic biofuels process.[16] Therefore, improving the efficiency of Cel7A by ameliorating product inhibition may result in a lower enzyme requirement for the process and a cheaper renewable fuel that is more cost-competitive with fossil fuels. To this end, several research groups have investigated ways to make Cel7A enzymes less prone to cellobiose inhibition. The prevailing strategy to mitigate product inhibition has been to perturb the binding of cellobiose in the enzyme active site via site-directed mutagenesis of the residues most responsible for this interaction.[5,20,21] Mutations in *Trichoderma reesei* (*Hypocrea jecorina*) Cel7A (*Tr*Cel7A) residues R251 and R394 reportedly resulted in reduced product inhibition.[20] A quintuple *Tr*Cel7A mutant (including mutations T226A and D262G) designed in 2001 to alter the pH optimum was similarly found to both relieve cellobiose inhibition and diminish cellulase activity.[25] More recently, computational point mutation studies in the same enzyme found that mutating residue R251, D259, D262, W376, or Y381 to alanine significantly weakened the calculated binding of cellobiose in the enzyme.[5,19] However, for many of these residues no experimental evidence verifying this has been demonstrated.

Recent molecular dynamics (MD) simulations performed by Silveira and Skaf computationally investigated the effects of various Cel7A mutations on cellobiose binding, as well as any induced structural perturbations to the enzyme.[21] These simulations built upon previous calculations[19] and together point to a handful of mutations predicted to disrupt cellobiose binding affinity (Figures 18 and 19).[21] The aim of this study was to produce a Cel7A variant with reduced cellobiose inhibition. I experimentally generated mutants identified by MD simulations[21] and evaluated their ability to hydrolyze microcrystalline cellulose in the presence of cellobiose. This work experimentally validates that the mutations predicted using computational modeling indeed conferred resistance to product inhibition; however, overall enzymatic activity suffered.

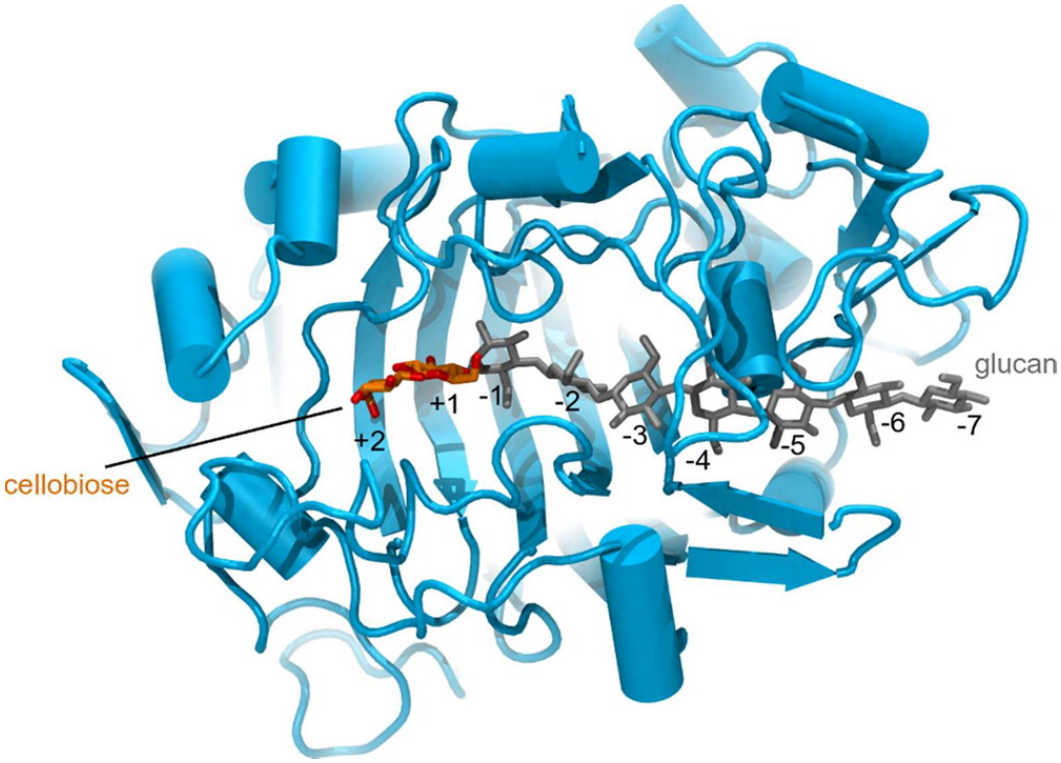


Figure 17. Substrate cellulose chain in the *Tr*Cel7A enzyme active site. The product of enzymatic hydrolysis, cellobiose, is shown in red. Glucose binding sites labeled (+) are towards the reducing end of the sugar polymer and (-) towards the non-reducing end. Hydrolysis occurs between subsites +1 and -1. PDB crystallographic structure: 8CEL.[21]

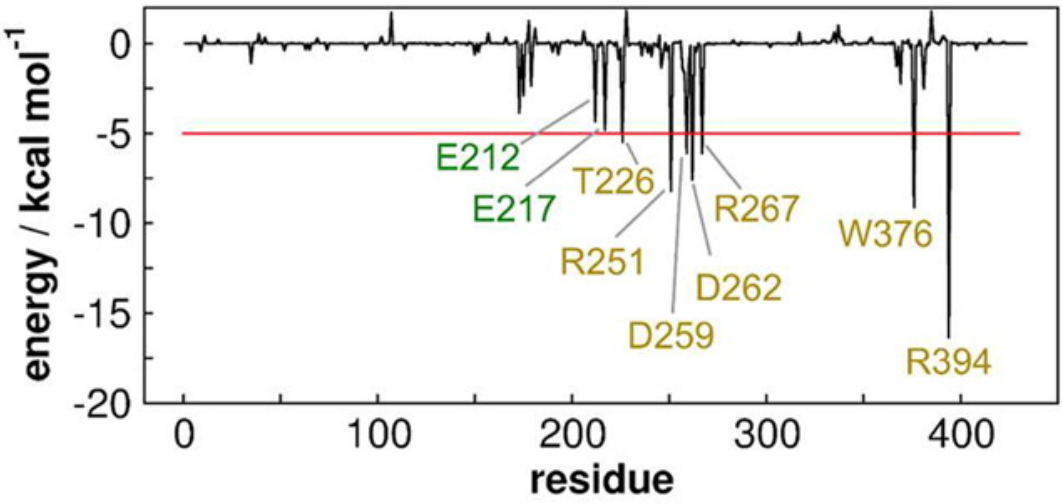


Figure 18. Residues computationally measured to interact with cellobiose at energies below -5 kcal/mol are indicated in yellow. Green labels designate catalytic residues.[21]

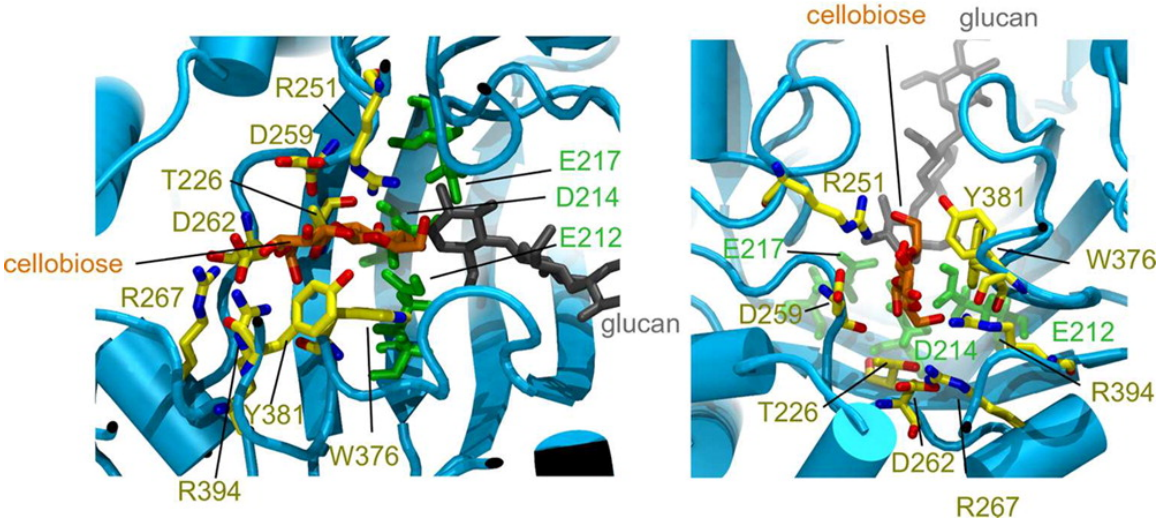


Figure 19. Amino acid residues shown in yellow interact strongly with cellobiose in the active site of *Tr*Cel7A and are candidates for mutation to reduce product inhibition in the enzyme.[21]

2.3 Methods

2.3.1 Selection of Cel7A Mutations

Cel7A mutations for experimental analyses were selected based on computational studies[19,21] (Y381A, D262A, W376A, T226A, R394A, R251A, D259A, R267A) and an industrial patent[20] (R251K and R251K + R394A). These *Trichoderma reesei* Cel7A (*Tr*Cel7A) cellobiose binding site mutations were mapped onto the *Talaromyces emersonii* Cel7A (*Te*Cel7A) enzyme. Due to the structural similarity and highly conserved active sites of these proteins (Figure 20), I expected mutations calculated to relieve product inhibition in *Tr*Cel7A to effect similar outcomes in *Te*Cel7A. Although *T. reesei* is currently the industrial standard for cellulase production, I chose to work with cellulase enzyme *Te*Cel7A because it is more thermotolerant than *Tr*Cel7A[5] and expresses at higher titers in the laboratory host production organism *Saccharomyces cerevisiae* (data not shown). Thermostable cellulases are of interest industrially as high-temperature cellulose hydrolysis decreases the risk of contamination and enables a greater solid substrate loading due to reduced slurry viscosity.[26] Thus *Te*Cel7A is of substantial relevance to the field and has the additional benefit of facile heterologous expression in *S. cerevisiae*.[27]

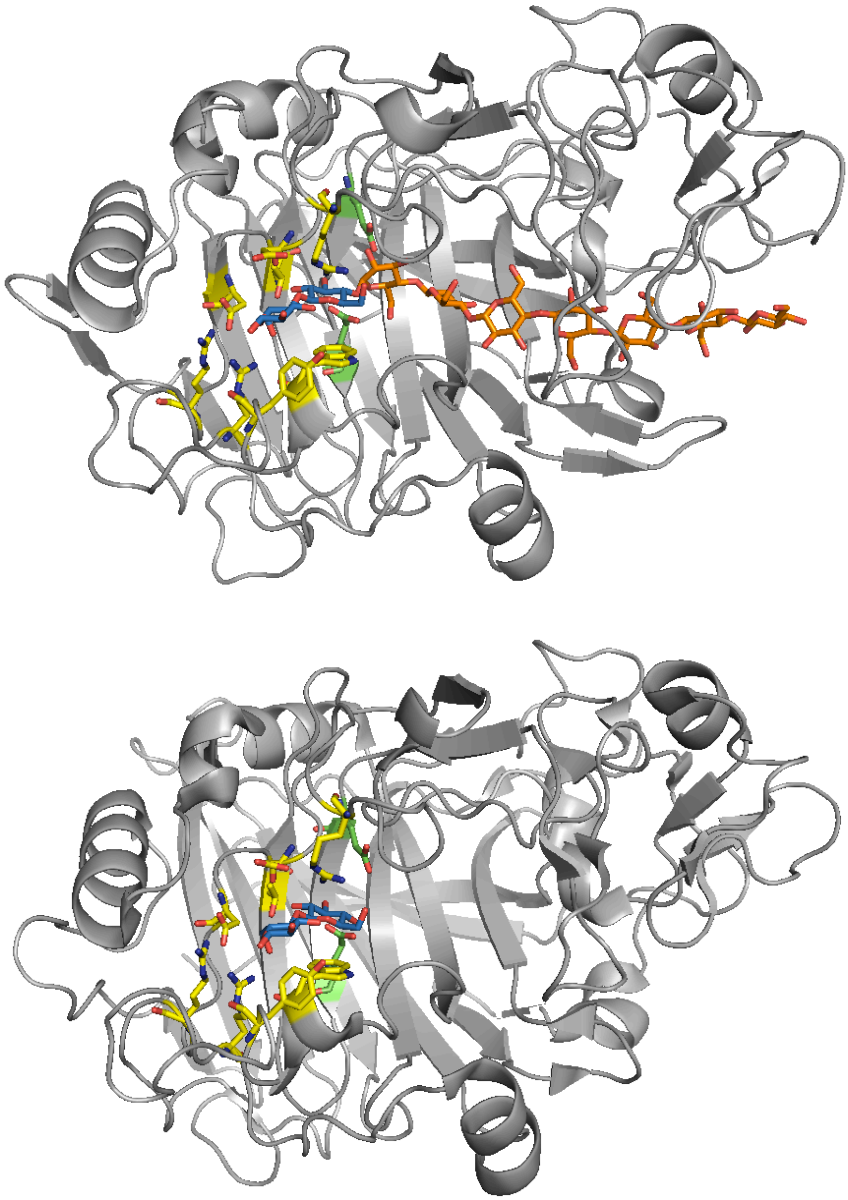


Figure 20. *T. reesei* Cel7A, PDB: 8CEL (top) and *T. emersonii* Cel7A, PDB: 3PFX (bottom). Cellulose chain (orange), cellobiose (blue), and amino acid residues selected for site-directed mutagenesis (yellow) in the enzymes' binding tunnels are highlighted.

Protein sequences of *Tr*Cel7A (Uniprot accession number: P62694) and *Te*Cel7A (Uniprot accession number: Q8TFL9) were aligned using the ExPASy Bioinformatics Resource Portal local similarity program alignment tool (Figure 21). Ten mutations were mapped as shown in Table 1.

Figure 21. Protein alignment of *Te*Cel7A (Uniprot accession number: Q8TFL9) and *Tr*Cel7A (Uniprot accession number: P62694).

Enzyme	Mutation									
***Tr*Cel7A**	R251K/ R394A	Y381A	D262A	W376A	T226A	R394A	R251A	D259A	R267A	R251K
***Te*Cel7A**	R248K/ R398A	Y385A	D259A	W380A	T223A	R398A	R248A	D256A	R264A	R248K

Table 1. Amino acid residues mapped from *T. reesei* Cel7A to *T. emersonii* Cel7A.

2.3.2 Construction of *Te*Cel7A Enzymes

*Te*Cel7A variants were generated in yeast expression vector pCu424.[28] Because *Te*Cel7A does not naturally contain a carbohydrate binding module (CBM), the CBM from *Agaricus bisporus* (Uniprot accession number: Q92400) was appended to the catalytic domain using a short, flexible linker from *Acremonium thermophilum* (Uniprot accession number: A7WNT9) as described previously.[27] Inclusion of the native signal sequence

allowed for secreted enzyme expression and enabled the mature form of the protein to carry the proper N-terminal pyroglutamate following signal sequence cleavage.[29] DNA and protein sequences encoding the wildtype *Te*Cel7A enzyme (with added linker and CBM) are detailed in Figure 22.

DNA Sequence:

5′-ATGCTACGTCGAGCCCTACTTCTGTCTTCATCCGCTATTCTAGCTGTTAAAGCACAACAGGCAGGAAC
AGCTACAGCCGAAAATCATCCCCCGTTAACATGGCAAGAATGTACGGCACCCGGTTCCTGTACAACACAGA
ATGGTGCAGTAGTATTAGATGCCAATTGGCGTTGGGTGCATGATGTTAACGGGTATACTAACTGTTATACC
GGAAATACCTGGGACCCCACATACTGCCCAGACGATGAAACGTGCGCTCAAAACTGTGCATTGGATGGTGC
AGATTACGAAGGAACTTACGGAGTGACATCCTCCGGTAGCAGTTTAAAATTGAATTTTGTGACTGGGTCCA
ACGTTGGTTCAAGACTGTATCTATTGCAGGATGATAGCACCTATCAGATTTTCAAACTTTTGAATCGCGAG
TTCAGTTTCGACGTTGATGTTTCTAACTTGCCTTGCGGTTTAAATGGTGCTTTATACTTTGTTGCTATGGA
CGCCGACGGTGGTGTATCCAAGTACCCCAATAACAAAGCGGGTGCGAAGTATGGGACCGGATACTGTGACA
GCCAATGTCCAAGAGATTTGAAGTTTATTGATGGCGAAGCCAACGTAGAAGGCTGGCAGCCTAGCTCCAAC
AACGCAAATACCGGCATAGGAGATCATGGCTCATGTTGTGCAGAAATGGATGTTTGGGAGGCTAACTCAAT
CAGTAATGCTGTTACCCCCCATCCATGCGATACTCCAGGACAAACGATGTGTTCCGGCGACGATTGCGGAG
GTACATATTCGAATGATAGATATGCCGGAACCTGTGATCCTGATGGTTGTGATTTCAACCCATATAGAATG
GGTAACACGTCTTTTTATGGTCCGGGTAAAATTATAGATACAACAAAGCCATTCACTGTTGTTACCCAGTT
TCTTACCGATGACGGTACCGACACAGGGACACTTAGCGAGATCAAAAGATTTTATATTCAGAACTCAAACG
TTATTCCTCAACCAAATAGTGACATAAGCGGTGTTACTGGCAACTCTATTACGACTGAATTTTGTACGGCT
CAAAAACAAGCCTTTGGAGATACCGATGATTTTAGTCAGCATGGGGGACTGGCTAAAATGGGGGCAGCTAT
GCAACAGGGTATGGTTTTAGTTATGTCATTATGGGATGATTACGCTGCACAAATGCTTTGGTTAGATTCCG
ATTACCCGACTGATGCCGATCCAACAACTCCTGGTATCGCGCGTGGAACATGTCCGACTGACTCTGGCGTT
CCTAGCGACGTTGAATCTCAGAGTCCTAATAGCTATGTCACATACTCCAATATAAAATTTGGTCCTATCAA
TTCAACATTCACCGCCAGC*AATCCCCCAGGTGGTGGTACTACAACAACAACCACCACAACTACCAGTAAGC*
*CGTCAGGTCCAACGACAACTACGAACCCATCCGGACCACAG*CAGACGATGTGGGGACAATGCGGGGGTCAA
GGTTGGACCGGTCCTACAGCCTGTCAGAGTCCTTCGACCTGTCACGTAATCAACGACTTTTACTCTCAATG
TTTCTAA-3′

Mature Protein Sequence:

QQAGTATAENHPPLTWQECTAPGSCTTQNGAVVLDANWRWVHDVNGYTNCYTGNTWDPTYCPDDETCAQNC
ALDGADYEGTYGVTSSGSSLKLNFVTGSNVGSRLYLLQDDSTYQIFKLLNREFSFDVDVSNLPCGLNGALY
FVAMDADGGVSKYPNNKAGAKYGTGYCDSQCPRDLKFIDGEANVEGWQPSSNNANTGIGDHGSCCAEMDVW
EANSISNAVTPHPCDTPGQTMCSGDDCGGTYSNDRYAGTCDPDGCDFNPYRMGNTSFYGPGKIIDTTKPFT
VVTQFLTDDGTDTGTLSEIKRFYIQNSNVIPQPNSDISGVTGNSITTEFCTAQKQAFGDTDDFSQHGGLAK
MGAAMQQGMVLVMSLWDDYAAQMLWLDSDYPTDADPTTPGIARGTCPTDSGVPSDVESQSPNSYVTYSNIK
FGPINSTFTAS*NPPGGGTTTTTTTTTSKPSGPTTTTNPSGPQ*QTMWGQCGGQGWTGPTACQSPSTCHVIND
FYSQCF

Figure 22. DNA and mature protein sequences comprised of the *T. emersonii* native signal sequence (underlined), *T. emersonii* Cel7A (Uniprot accession number: Q8TFL9), *Acremonium thermophilum* linker (Uniprot accession number: A7WNT9, italicized) and *Agaricus bisporus* carbohydrate binding module (CBM) (Uniprot accession number: Q92400).

Site-directed mutagenesis was performed by PCR amplification of the pCu424 *Te*Cel7A DNA construct using overlapping primers to replace wildtype codons with those encoding the desired amino acid substitutions (Table 2). Replacement codons were

selected based on their natural abundance in *S. cerevisiae*. A negative control sample lacking any cellulase gene was also generated by removing the DNA encoding the catalytic domain, linker, and CBM domain from the expression vector. Amplification reactions were verified using agarose gel electrophoresis. Methylated template DNA was digested overnight at 37˚C by restriction endonuclease DpnI (NEB, Ipswich, MA). All resulting plasmids were independently cloned into XL1-Blue *E. coli* cells (Agilent Technologies, Santa Clara, CA) followed by overnight culture growth at 37˚C in Lysogeny Broth (LB) media containing 65mg/mL carbenicillin antibiotic. The amplified vector DNA from the resulting cultures was purified using Quiagen Miniprep kits (Quiagen, Limburg, Netherlands) and thereafter sequenced to verify successful mutagenesis.

***Te*Cel7A Mutation**	**Primer Sequence**	**Annealing Temp.**	**Primer Direction**
R248K and R248K/R398A	5′-GTACATATTCGAATGAT**AAA**TATGCCGGAACCTGTGATCCTG-3′	63˚C	Forward
R248K and R248K/R398A	5′-GTTCCGGCATA**TTT**ATCATTCGAATATGTACCTCC-3′	60.1˚C	Reverse
R398A and R248K/R398A	5′-CTGGTATCGCG**GCT**GGAACATGTCCGACTGACTCTG-3′	66.4˚C	Forward
R398A and R248K/R398A	5′-GTCGGACATGTTCC**AGC**CGCGATACCAGGAGTTGTTG-3′	66.7˚C	Reverse
Y385A	5′-GGTTAGATTCCGAT**GCT**CCGACTGATGCCGATCCAAC-3′	65.1˚C	Forward
Y385A	5′-GGCATCAGTCGG**AGC**ATCGGAATCTAACCAAAGC-3′	63.5˚C	Reverse
D259A	5′-GATCCTGATGGTTGT**GCT**TTCAACCCATATAGAATGG-3′	60.1˚C	Forward
D259A	5′-CTATATGGGTTGAA**AGC**ACAACCATCAGGATCACAGG-3′	61.3˚C	Reverse
W380A	5′-CTGCACAAATGCTT**GCT**TTAGATTCCGATTACCCG-3′	61.1˚C	Forward
W380A	5′-GTAATCGGAATCTAA**AGC**AAGCATTTGTGCAGCGTAATC-3′	60.8˚C	Reverse
T223A	5′-CAGTAATGCTGTT**GCT**CCCCATCCATGCGATACTCC-3′	64.2˚C	Forward
T223A	5′-CGCATGGATGGGG**AGC**AACAGCATTACTGATTGAG-3′	65.4˚C	Reverse
R248A	5′-GTACATATTCGAATGAT**GCT**TATGCCGGAACCTGTGATCCTG-3′	63˚C	Forward
R248A	5′-GTTCCGGCATA**AGC**ATCATTCGAATATGTACCTCC-3′	60.1˚C	Reverse
D256A	5′-GAACCTGTGATCCT**GCT**GGTTGTGATTTCAACCCATATAG-3′	62.3˚C	Forward
D256A	5′-GTTGAAATCACAACC**AGC**AGGATCACAGGTTCCGGCATATC-3′	64.9˚C	Reverse
R264A	5′-GATTTCAACCCATAT**GCT**ATGGGTAACACGTCTTTTTATGG-3′	60.1˚C	Forward
R264A	5′-CGTGTTACCCAT**AGC**ATATGGGTTGAAATCACAACCATCAG-3′	62.4˚C	Reverse

Table 2. Primers used to generate point mutants in *Te*Cel7A; modified codons in bold.

[Note: I also generated plasmids encoding a handful of other *Te*Cel7A mutations as well as those for the expression of all *Tr*Cel7A mutants in *Trichoderma reesei*; however, results of those efforts are not detailed here. Other *Te*Cel7A mutants (suggested by a member of David Baker's computational group at the University of Washington) were not of great interest. However, a few *Tr*Cel7A enzyme mutants may be selected for *T. reesei* expression as a follow-up to this work.]

2.3.3 Expression of *Te*Cel7A Enzymes

Control sample and point-mutant pCU424 *Te*Cel7A DNA were individually transformed using the LiAc method[30] into the enzyme production host organism, *S. cerevisiae* strain YVH10Δ*PMR1* (a strain which limits protein hyperglycosylation).[27] Cells were spread onto selective plates containing 1.5% agar and synthetic complete medium lacking tryptophan (SC-Trp) and incubated for 3 days at 30˚C. Liquid cultures of 100mL SC-Trp for each variant were inoculated with plate colonies and grown overnight at 30˚C with shaking at 220rpm before being used, in turn, to inoculate 2L cultures grown for three days under the same conditions. *Te*Cel7A protein expression was then induced by pelleting the cells via centrifugation at 4,000xg for 15 minutes and resuspending them in yeast peptone dextrose (YPD) medium supplemented with 500μM copper sulfate. The induced cultures were grown for an additional three days at 25˚C with shaking at 220rpm.

2.3.4 Purification of *Te*Cel7A Enzymes

Following protein expression, cultures were centrifuged at 4,000xg for 15 minutes to clarify the supernatants containing the *Te*Cel7A enzymes. Two liters of yeast culture supernatant for each variant were subsequently filtered (to remove residual cells) before being concentrated and buffer exchanged via tangential flow filtration (TFF) into 20mM Tris-HCl pH 7.

Proteins were purified using fast protein liquid chromatography (FPLC) over 5mL HiTrap Q HP columns (GE Healthcare, Little Chalfont, UK) using running buffer 20mM Tris-HCl pH 7 and elution buffer of the same with the addition of 1M sodium chloride. Gradients of 0-25% elution buffer (0-0.25M sodium chloride) over 85mL followed by 25-50% elution buffer (0.25-0.5M sodium chloride) over 35mL were used to separate supernatant proteins.

FPLC fractions were analyzed for cellulase activity using 4-methylumbelliferyl-β-D-cellobioside (MUCell), a fluorescent substrate mimic. Twenty microliters of each FPLC fraction sample was mixed with 80μL of 1.25mM MUCell in 50mM sodium acetate pH 5 and heated for 10 minutes at 50˚C. Reactions were stopped by boiling at 98˚C for two minutes and prepared for analysis by the addition of 10μL of 1M sodium hydroxide. Fluorescence was measured using a multiwell plate reader with an excitation wavelength of 365nm and an emission wavelength of 445nm. Active fractions were analyzed for purity using SDS-PAGE, and those containing uncontaminated *Te*Cel7A enzymes (running at ~75kDa) were combined. Samples were concentrated and buffer exchanged into 50mM sodium acetate pH 5 using 30k MWCO Vivaspin 15 Turbo centrifugal concentrators (Sartorius, Concord, CA). Enzyme concentrations were normalized via absorbance to A_{280}=1, or roughly 13.35μM, and are of single-band purity (Figure 23).

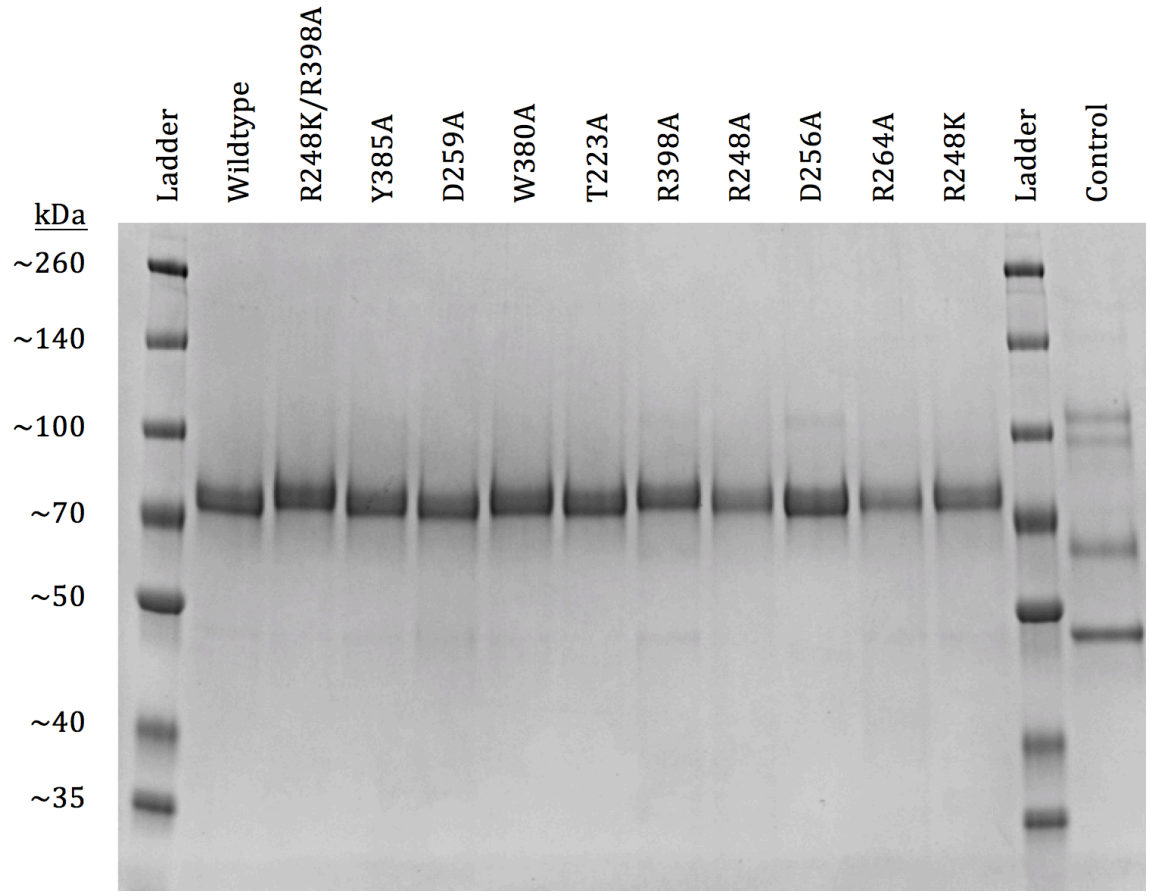


Figure 23. SDS-PAGE demonstrating single-band purity of *Te*Cel7A enzyme samples normalized in concentration by absorbance to A_{280}=1, or approximately 13.35μM. Control sample lacking the *Te*Cel7A gene indeed contains no *Te*Cel7A protein, as indicated by the absence of a band at ~75kDa.

2.3.5 Activity assays

Assays to measure activity and inhibition of purified enzymes on cellulose were performed in 96-well PCR plates with 10 mg/mL Avicel PH-101 (Sigma, St. Louis, MO) substrate and 1.33μM purified *Te*Cel7A enzyme in 50mM sodium acetate pH 5. Three sets of experiments were performed: (1) without β-glucosidase (150μL reaction volumes, performed in triplicate), (2) with β-glucosidase (150μL reaction volumes, performed in triplicate), and (3) with thiocellobiose and β-glucosidase (75μL reaction volumes, performed in duplicate). Reactions including β-glucosidase contained 0.016 g/L β-glucosidase from Novo188 (Novozymes, Bagsvaerd, Denmark). Reactions including thiocellobiose contained 4.385 g/L thiocellobiose (Sigma, St. Louis, MO). All experiments were incubated for 60 hours at 60˚C with constant rotational mixing followed by boiling for five minutes at 95˚C to stop the reactions.

Reactions without β-glucosidase were inhibited by the cellobiose produced during cellulose hydrolysis; more active enzymes generated higher concentrations of cellobiose. The final cellobiose concentrations in reactions without β-glucosidase ranged from 0.20 g/L (0.58 mM, for the least active mutant *Te*Cel7A W380A) to 0.61 g/L (1.8 mM, for the wildtype *Te*Cel7A enzyme). Reactions with thiocellobiose, on the other hand, included uniform concentrations (4.385 g/L, 12.2 mM) of inhibitor. Thiocellobiose, while chemically very similar to cellobiose, is not cleaved by β-glucosidase; its concentration therefore remained constant throughout the hydrolysis reactions.

Although *Te*Cel7A is thermotolerant, reactions were performed at 60˚C in order to preserve the effectiveness of the β-glucosidase enzyme used. At 70˚C, all *Te*Cel7A enzymes were indeed active, but little difference was observed between samples with and without β-glucosidase (Figure 24). This is consistent with a β-glucosidase enzyme that is inactive at this temperature.

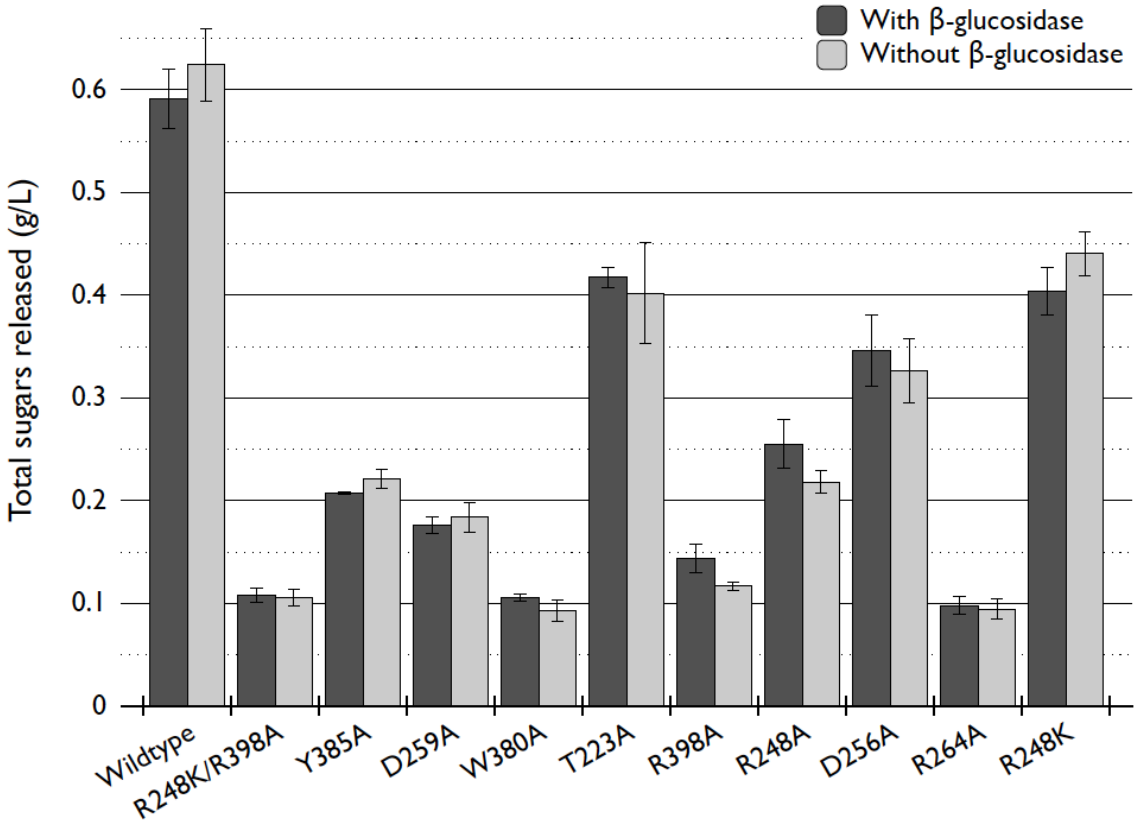


Figure 24. Activity assay at 70˚C is consistent with a β-glucosidase enzyme that is inactive at this temperature. Error bars represent standard error (n=3).

2.3.6 Activity Assay Analysis

To quantify the cellobiose and glucose concentrations in the reactions, samples were filtered through 96-well filter plates with 0.45 μm polypropylene membranes (Seahorse Bioscience, N. Bellerica, MA) and analyzed in 96-conical well plates sealed by aluminum tape using a 1200 series high-pressure liquid chromatography (HPLC) system (Agilent Technologies, Santa Clara, CA, USA) consisting of an autosampler with tray cooling, binary pump, degasser, thermostated column compartment, diode array detector (DAD) and refractive index detector (RI) connected in series. The supernatant (20 μL) was injected onto a 100 mm x 7.8 mm (length x inner diameter) Rezex™ RFQ-Fast Acid H^+ guard column (Phenomenex, Torrance, CA) with 8 μm particle size, 8% cross-linkage equipped with a SecurityGuard™ Standard Carbo H^+ (Phenomenex) column cartridge. Compounds were eluted at 55°C at a flow rate of 1.0 mL using a mobile phase of 5 mM sulfuric acid. Quantification was performed by external calibration with a set of cellobiose and glucose solutions in the ranges of 0.08-10 mg/mL and 0.15-20 mg/mL, respectively.

Data presented represents average values of experiments (controls subtracted) with standard error (n=3 for experiments with and without β-glucosidase, n=2 for experiments with thiocellobiose). Ratio values are quotients of averages with propagated standard error from the two measurements.

2.4 Results and Discussion

Molecular dynamics (MD) simulations by Silveira and Skaf[21] identified several *Tr*Cel7A mutations (many of which were revealed previously[19,20,25]) predicted to reduce product inhibition. Seven of the eight residues selected for mutation to alanine interact with cellobiose at an energy below -5 kcal/mol (Figure 18).[21] I experimentally generated ten mutants of interest (predicted for *Tr*Cel7A, mapped onto *Te*Cel7A, and heterologously expressed in *S. cerevisiae*) and examined their activities under various conditions. Enzyme product inhibition was estimated by comparing enzymatic hydrolysis of cellulose under inhibiting conditions to hydrolysis when the enzyme was minimally inhibited.

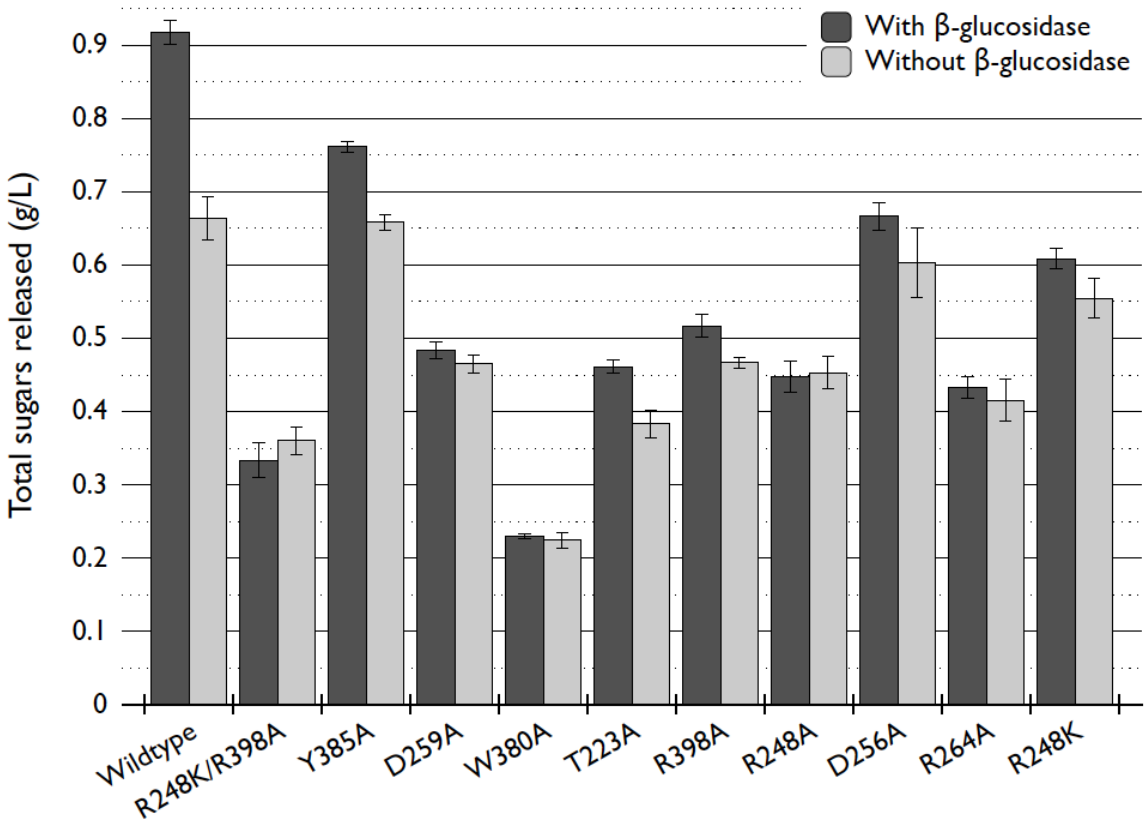


Figure 25. Total sugars (cellobiose + glucose, g/L) released from Avicel hydrolysis by *Te*Cel7A enzymes under minimally inhibiting ("with β-glucosidase") or inhibiting ("without β-glucosidase") conditions after 60 hours at 60˚C. Error bars represent standard error (n=3).

Enzymatic hydrolysis of Avicel with and without β-glucosidase is compared in Figure 25. In reactions containing β-glucosidase, the Cel7A enzymes were negligibly inhibited by cellobiose, as any cellobiose generated was quickly cleaved to glucose by β-glucosidase (no more than 0.02 g/L of cellobiose was measured in these experiments). Reactions lacking β-glucosidase, on the other hand, experienced varying inhibitor concentrations (0.20-0.61 g/L of cellobiose, depending on the enzymes' individual activities) as a result of cellobiose accumulating over the course of hydrolysis.

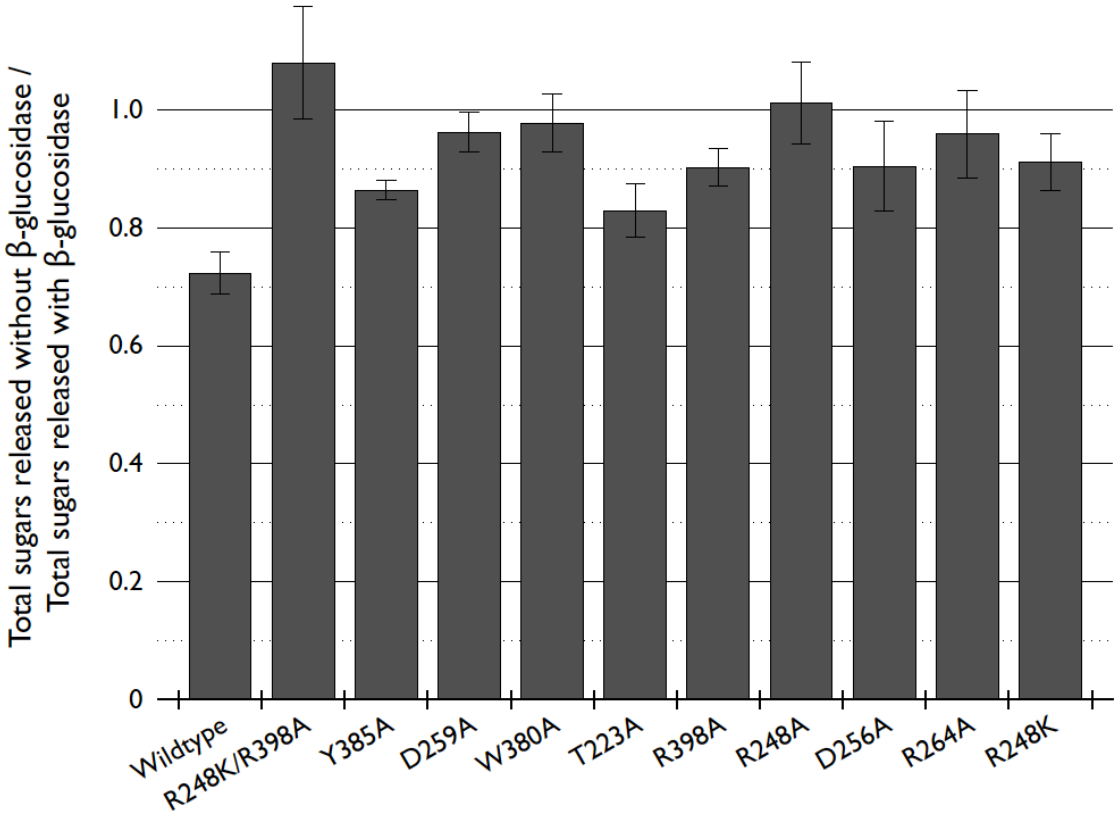


Figure 26. Enzyme tolerance to cellobiose, based on the ratio of activities of *Te*Cel7A enzymes under inhibiting conditions ("without β-glucosidase") to activities under minimally inhibiting conditions ("with β-glucosidase"), as measured by total sugars released from Avicel after 60 hours at 60˚C. Error bars represent propagated standard error.

Enzyme tolerance to the inhibitor cellobiose can be estimated by comparing the extent of hydrolysis under inhibiting conditions to that when inhibition is alleviated.[20] Accordingly, a ratio of one represents the best-case scenario, whereby an enzyme retains 100% of its uninhibited activity in the presence of cellobiose. As predicted,[19–21] every mutation tested improved the enzyme's tolerance to cellobiose (Figure 26); however, in all conditions tested, the wildtype *Te*Cel7A was more active than any mutant toward depolymerizing crystalline cellulose. While the wildtype *Te*Cel7A enzyme retained only 72% of its activity under inhibiting conditions, mutants Y385A, D259A, and W380A, for example, retained 86%, 96%, and 98%, respectively, of their activities when inhibited. Double mutant R248K/R398A and single mutant R248A, in particular, exhibited no measurable loss of activity under these inhibiting conditions. The corresponding two mutants in *Tr*Cel7A were found by BP Biofuels to behave similarly favorably with respect to cellobiose tolerance.[20]

Enzyme inhibition was also studied under conditions providing equivalent inhibition across all samples using the cellobiose mimic, thiocellobiose. Enzyme activities under inhibition by thiocellobiose (Figure 27) corroborate results from hydrolysis under inhibition by auto-generated cellobiose (Figure 25, "without β-glucosidase"). The wildtype *Te*Cel7A was again the most productive enzyme assayed. Thus, the mutations suggested by MD simulations[21] did alleviate product inhibition (Figure 26), but always at the expense of activity.

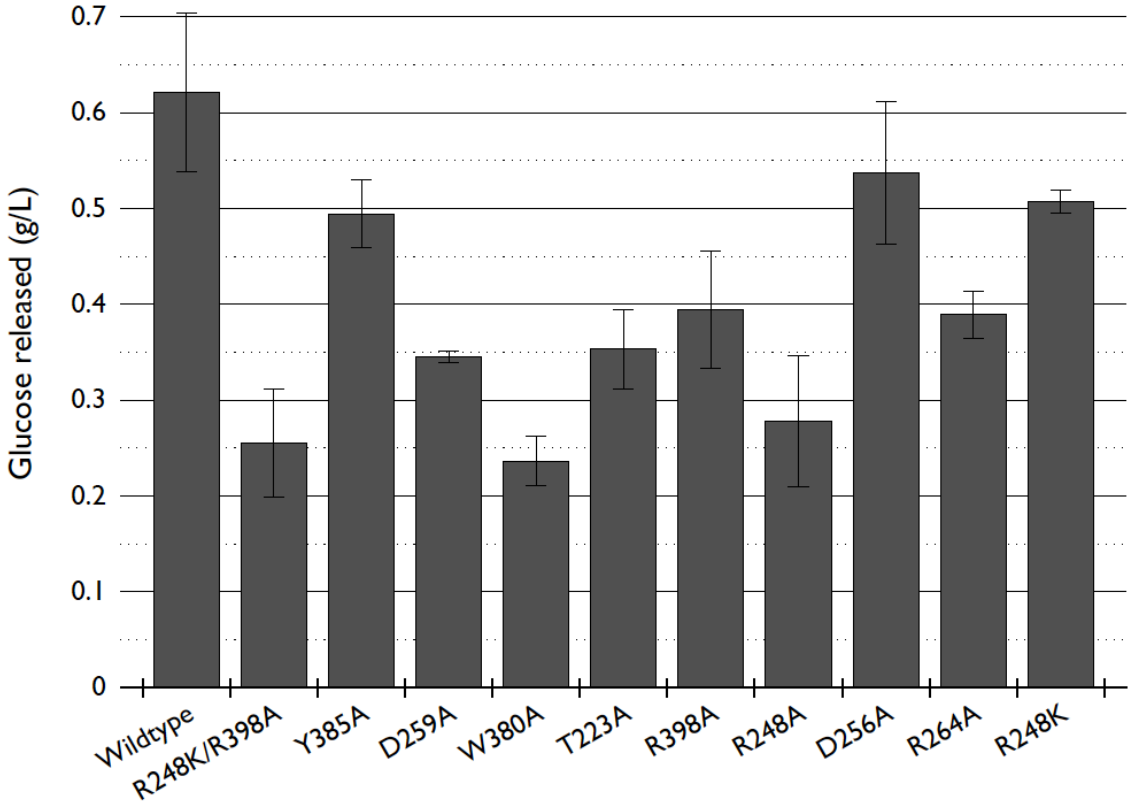


Figure 27. Glucose released from Avicel hydrolysis by *Te*Cel7A enzymes under uniform inhibition by thiocellobiose after 60 hours at 60˚C. Error bars represent standard error (n=2).

Of the *Te*Cel7A variants tested, a few mutants were of particular interest, including Y385A and R248K. Relative to that of the wildtype enzyme, the Y385A mutant demonstrated an improved tolerance to cellobiose (+19%, Figure 26) with the least loss in activity (-17% uninhibited, -1% inhibited, Figure 25). The R248K mutant also showed improved cellobiose tolerance (+26%), but a greater loss in activity relative to the wildtype enzyme (-34% uninhibited, -16% inhibited). Interestingly, MD simulations predicted that residues corresponding to *Te*Cel7A R248 and Y385 can interact to form a closed, tunnel-like conformation (Figure 28B). This conformation, inaccessible until after

hydrolysis of the carbohydrate substrate chain, may obstruct product release.[21] The reduction I observed in cellobiose inhibition may thus be a consequence of removing the protein's ability to adopt this occluding conformation; indeed, MD calculations ranked TrCel7A Y381A (*Te*Cel7A Y385A) as having a substantial impact on cellobiose binding affinity (ΔΔG).[21]

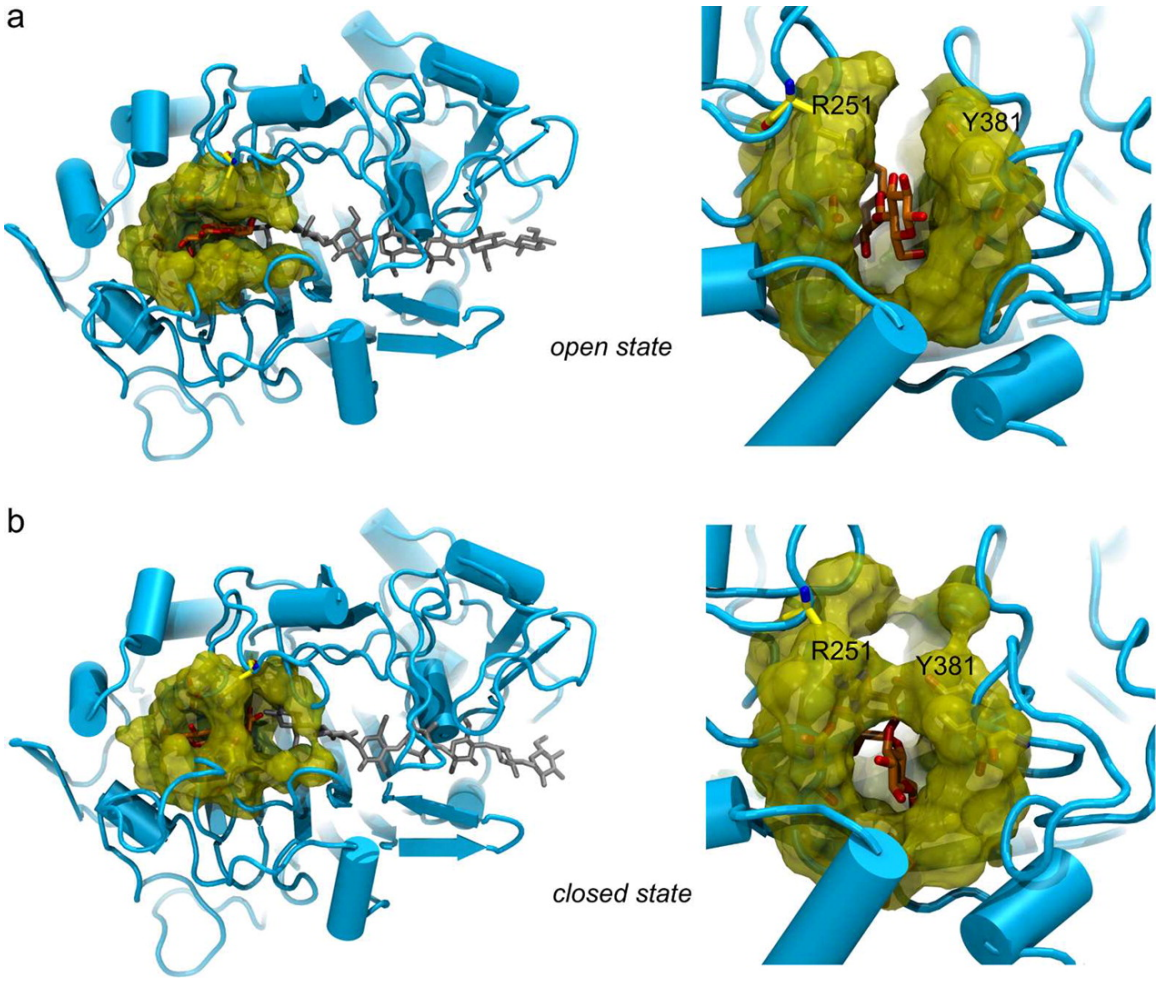


Figure 28. A structural change in the product binding site of *Tr*Cel7A is accessible following enzymatic hydrolysis of the substrate chain. Residues *Tr*Cel7A R251 and Y381 (*Te*Cel7A R248 and Y385, respectively) can interact to form a closed, tunnel-like conformation (b) that obstructs cellobiose product release.[21]

Additionally, as the guanido group of *Tr*Cel7A R251 (*Te*Cel7A R248) is known to make two hydrogen bonds with the sugar at the +1 position,[5] it was unsurprising that eliminating those interactions with mutant R248A reduced cellobiose sensitivity. Despite its favorable inhibited:uninhibited enzymatic activity ratio (Figure 26); however, the mutant's activity relative to the wildtype suffered more than that of R248K (-51% vs -34% uninhibited and -32% vs -16% inhibited, respectively, Figure 25). Indeed, MD simulations

concluded that the product binding site (of the *Tr*Cel7A equivalent of mutant R248A) loses its structural integrity due to the disruption of a conserved salt bridge with D256.[21] Perhaps the less drastic substitution of lysine for arginine (instead of alanine) in *Te*Cel7A residue 248 allowed the enzyme to maintain some electrostatic interactions with the substrate as well as the salt bridge with D256, while disfavoring the protein structural change occluding cellobiose release (resulting from the interaction of amino acids 248 and 385).

The mutant exhibiting the largest reduction in activity relative to the wildtype *Te*Cel7A was W380A (-75% uninhibited, -66% inhibited, Figure 25). Tryptophan 376 of *Tr*Cel7A (W380 of *Te*Cel7A) and three other aromatic residues directly interacting with the cellulose chain remain in place during substrate translocation, stacking with the +1, -2, -4, and -7 glycosyl moieties of the substrate polymer.[5,31,32] These interactions between aromatic amino acids and the carbohydrate are considered important for enzyme processivity.[31] In fact, mutations of these residues are known to handicap processivity and thus hydrolytic activity on crystalline cellulose substrates (while increasing activity on amorphous and soluble substrates).[31] My results are consistent with these findings. In addition, MD simulations predicted this mutation to have a significant impact on cellobiose binding affinity.[21] Indeed, the low affinity of cellobiose for the mutant resulted in the molecule having little inhibitory effect on the enzyme; under inhibiting conditions, the W380A mutant retained 98% of its uninhibited activity (Figure 26).

Mutant Y385A was arguably the best of those tested in that it was the most active relative to the wildtype (Figure 25) while still ameliorating product inhibition (Figure 26). Isothermal titration calorimetry (ITC) experiments (performed by Kathryn Strobel) confirmed that this mutant decreased, but did not eliminate, cellobiose binding affinity to the enzyme. On the other hand, product binding was more substantially weakened in mutant W380A, for example, to the point of being immeasurable by ITC. While W380A showed virtually no sensitivity to cellobiose, the hydrolytic activity of this mutant was the lowest measured. Likewise, combining mutations R248K and R398A resulted in a double mutant with an increased tolerance to cellobiose but reduced hydrolytic activity compared to the single mutants.

A loss of enzyme effectiveness in hydrolyzing cellulose, relative to the wildtype enzyme, was also expected for mutants D256A and R398A. Due almost exclusively to these two amino acids, the foremost glucose unit of the cellulose chain (occupying the +2 position of the enzyme's active site) forms more hydrogen bonds with Cel7A than does any other sugar of the carbohydrate. These strong hydrogen bonds between the leading glycosyl ring and *Tr*Cel7A D259 and R394 (*Te*Cel7A D256 and R398, respectively) stabilize the end point of processive motion.[31] Naturally, disrupting these interactions would lead to decreased hydrolysis activity on solid cellulosic substrates, as I observed experimentally. Additionally, recent MD calculations found that the *Tr*Cel7A D259A (*Te*Cel7A D256A)

mutation disrupts a salt bridge with *Tr*Cel7A R251 (*Te*Cel7A R248), as mentioned above for the *Te*Cel7A R248A mutant.[21] Finally, unintentional structural perturbations outside of the active site may contribute to decreased activity in the *Te*Cel7A mutants. The substrate entrance region of the protein, in particular, was computationally observed to be the most sensitive part of the enzyme to mutations in the product binding site (Figure 29).[21]

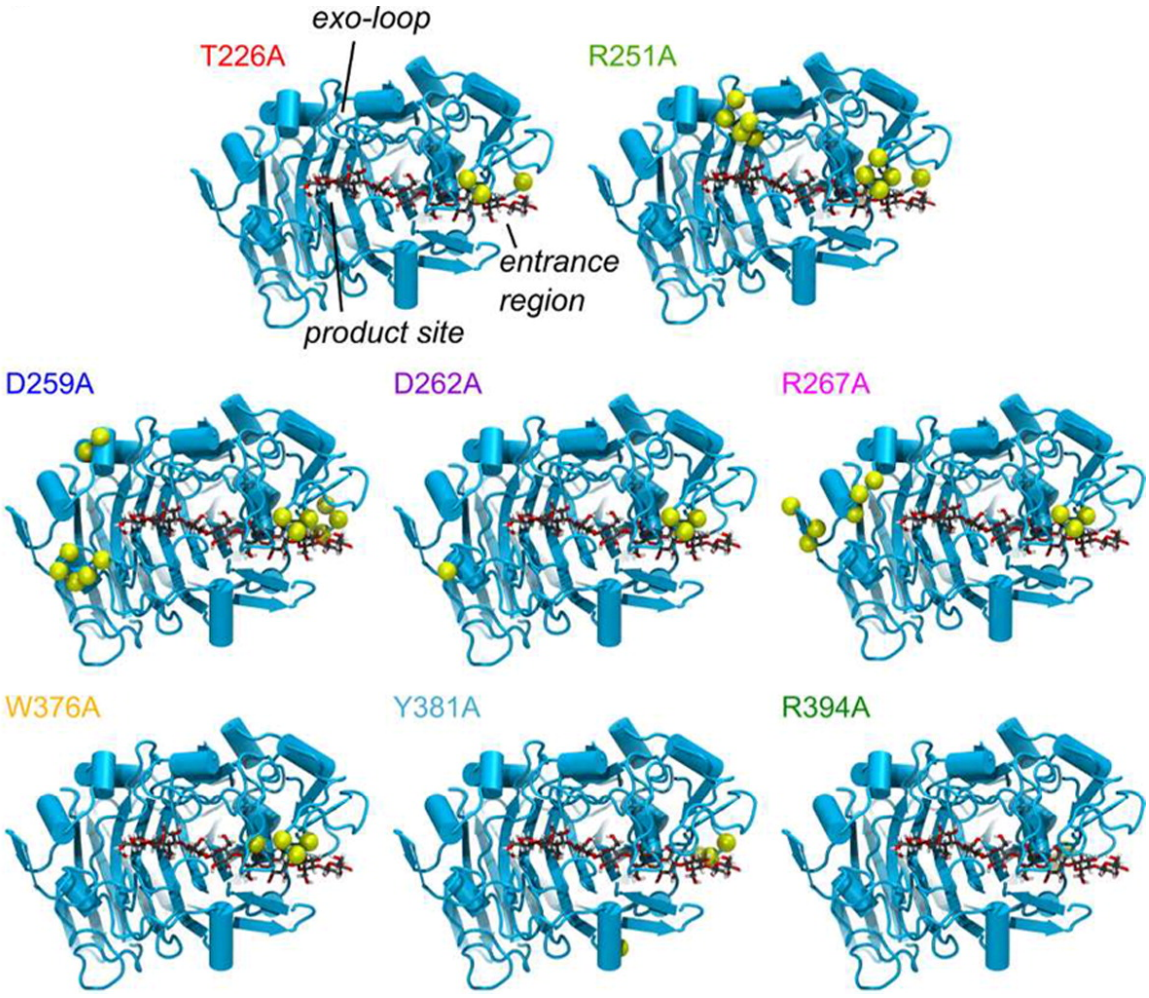


Figure 29. Mutations (shown for *Tr*Cel7A) in the enzyme's active site are computationally predicted to affect the substrate entrance region of the protein, in particular. Structural perturbations are indicated with yellow spheres.[21]

Energy calculations point to a direct link between the binding free energy of cello-oligosaccharides and enzyme processivity.[33] The notable affinity of Cel7A's product binding site for carbohydrate chains is believed to provide the driving force for processivity of the enzyme along the cellulose strand.[31] In fact, cellobiose has been calculated to be 11.2-14.4 kcal/mol more stable in the product binding site of the enzyme

than in free solution.[19] Within the active site, computational studies have quantitatively demonstrated that a cellulose chain binds more tightly to product binding sites than it does to reactant sites of *Tr*Cel7A.[19,33] Recent calorimetry experiments corroborated this conclusion that substrate affinity is highest in the +1 and +2 (product) positions of the enzyme's active tunnel.[34] This strong, preferential binding of the product (resulting in an 11.1 kcal/mol stabilization) likely contributes significantly to both the processivity of cellobiohydrolases and to cellobiose product inhibition.[33] The residues primarily responsible for the strong binding of the leading glycosyl residue of the cellulose chain (providing the "driving force" for processive motion) D259 and R394 in *Tr*Cel7A (D256 and R398, respectively, in *Te*Cel7A) are conserved.[5] As a result, it is not surprising that I observed a trade-off between reduced sensitivity to inhibition by cellobiose, and overall enzyme hydrolysis activity, where tight binding of the product may in fact drive processive motion along a cellulose chain.

In a hydrolysis reaction, inhibition by cellobiose is predominantly a concern for cellobiohydrolases like Cel7A (with a closed substrate binding tunnel) as opposed to endoglucanases (having an open binding cleft facilitating dissociation of this product from the active site).[5,18,24,35] It is hypothesized that, in a synergistic enzyme mixture, the rate-limiting step for glycoside hydrolase family 7 enzymes is processive velocity (encompassing hydrolysis, product expulsion, and processive motion along the substrate chain). Among these, the glycosylation step in the enzyme's retaining mechanism (Figure 15B), specifically, is proposed to be the rate-limiting step for *Tr*Cel7A.[5] Product expulsion (and thus cellobiose inhibition), on the other hand, has been experimentally rejected as the cause of the observed rate retardation in enzymatic hydrolysis.[23,36,37] Enzyme complexation with glycan chains has also been identified as a rate-limiting process for Cel7A-catalyzed hydrolysis of crystalline cellulose.[36] Regardless of the rate-limiting step, alleviating product inhibition of Cel7A should improve cellulose hydrolysis rates by increasing the concentration of catalytically viable Cel7A,[36] provided the beneficial effect is not outweighed by an accompanying loss of catalytic activity.

I found that mutations identified using computational methods could be mapped between enzymes with highly-conserved active sites (and nearly identical three-dimensional structures) to effect the desired decreases in product inhibition in the homolog enzymes. In addition, data published by BP Biofuels on product-tolerant *Tr*Cel7A mutants (Figure 30)[20] were similar to data from my experiments with three corresponding mutants in *Te*Cel7A (R248K/R398A, R248K, and R398A). While *Tr*Cel7A is the industrial standard, the enzyme can be problematic for mutational studies due to the relatively complicated *T. reesei* expression system.[5] For studies on product tolerance, my work suggests that the *S. cerevisiae*-produced *Te*Cel7A serves as a suitable replacement for *Tr*Cel7A that is both facile to work with and more thermostable. In principle, mutations of interest generated in *Te*Cel7A could be mapped back to *Tr*Cel7A for industrial production.

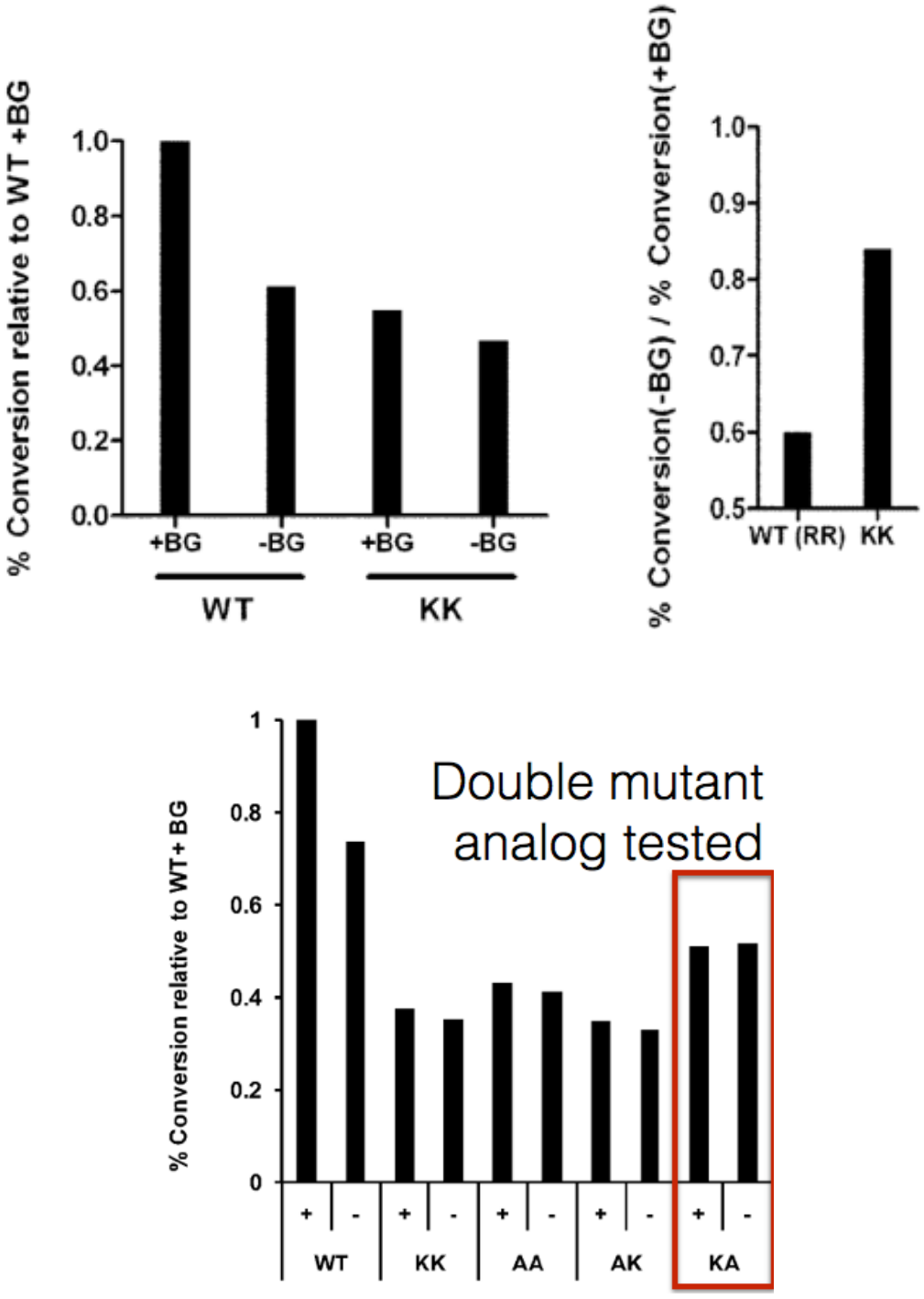


Figure 30. Data from patents by BP Biofuels demonstrating double mutants of *Tr*Cel7A with improved tolerance to cellobiose (but decreased overall activity compared to the wildtype). The double mutant analog tested in my work (*Te*Cel7A R248K/R398A) is boxed in red.[20]

Considering all the mutants together provides insights into how mutations in the product binding site of Cel7A affect both the enzyme's sensitivity to inhibition by cellobiose as well as its overall hydrolytic activity on solid cellulose substrates. Figures 31 and 32 plot the extent of inhibition determined in the two sets of experiments versus the corresponding uninhibited release of product. Both plots demonstrate that the hydrolytic activity of Cel7A suffers as a result of mutations to the product binding site that alleviate product inhibition (the rightmost point in each case corresponds to the wildtype enzyme). Furthermore, the activity of Cel7A is more sensitive to such mutations than is the extent of inhibition (uninhibited activity varies 4-fold compared to just 1.5-fold for the extent of inhibition). It is also noteworthy that for some mutants the effect of cellobiose on enzymatic activity is negligible.

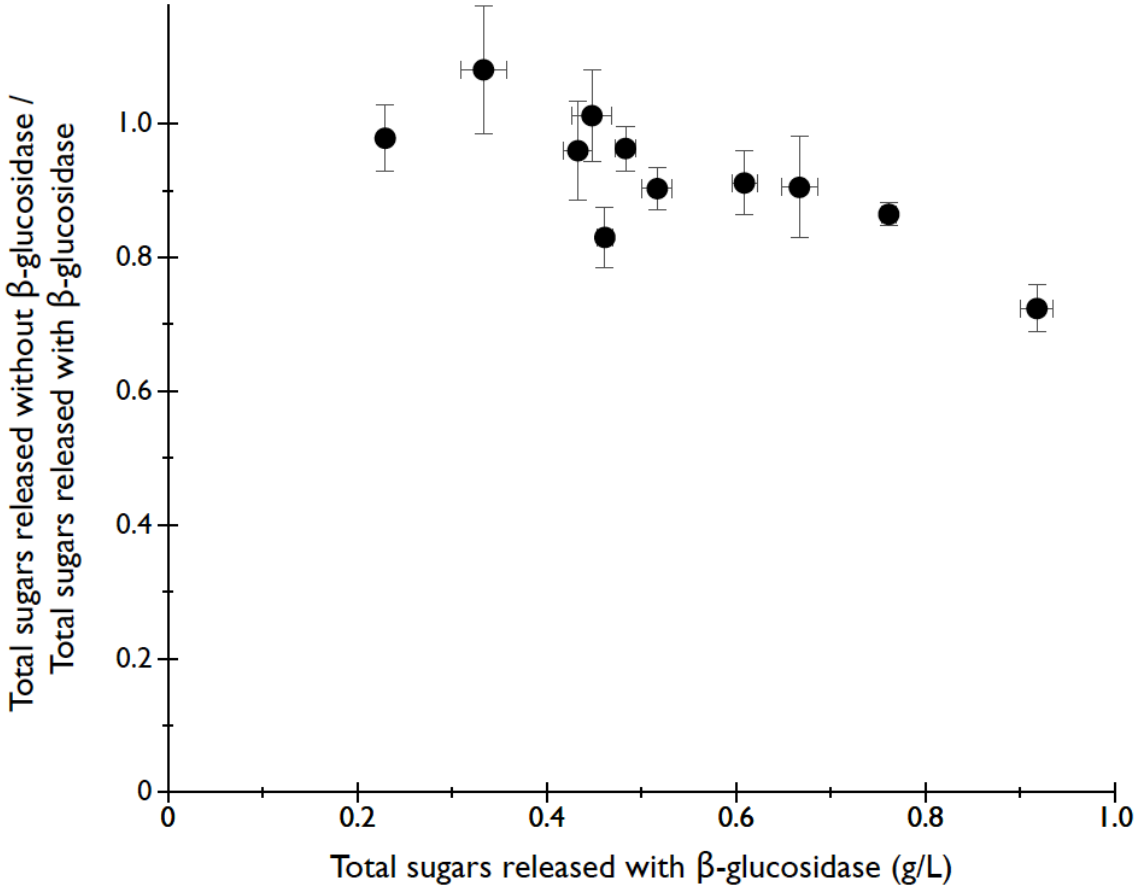


Figure 31. Scatterplot illustrating *Te*Cel7A variants' tolerances to inhibition by cellobiose compared to their uninhibited hydrolytic activities. Error bars represent propagated standard error (y) and standard error (x).

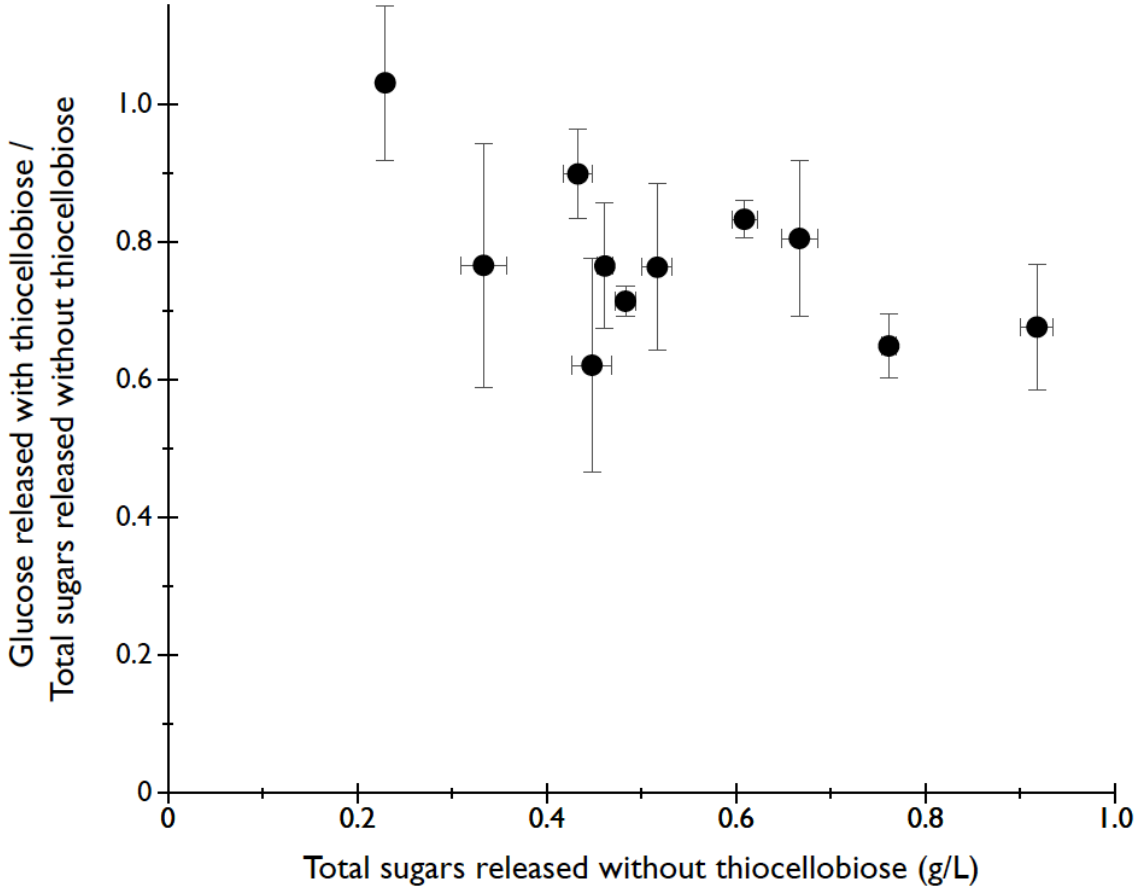


Figure 32. Scatterplot illustrating *Te*Cel7A variants' tolerances to inhibition by thiocellobiose compared to their uninhibited hydrolytic activities. (Note that "total sugars released without thiocellobiose" is equivalent to "total sugars released with β-glucosidase" and reactions with thiocellobiose included β-glucosidase.) Error bars represent propagated standard error (y) and standard error (x).

2.5 Conclusions

Inhibition of Cel7A is of particular relevance to the cellulosic biofuels industry, where high-solids loadings (leading to high cellobiose concentrations upon hydrolysis) are important for generating the concentrated glucose solutions necessary for downstream conversion to fuels and chemicals.[38] Under these conditions, product removal must be swift and efficient due to the adverse contribution of product inhibition to cellulose hydrolysis. As my experiments have demonstrated, alleviating product inhibition in Cel7A requires a delicate balance between maintaining affinity for cellobiose in the active site of the enzyme and allowing for its escape.

My results experimentally validate computational predictions[19,21] for alleviating product inhibition in Cel7A; however, as hypothesized, they reveal a trade-off between catalytic efficiency and product tolerance. All ten *Te*Cel7A mutants examined displayed improved tolerances to cellobiose, with some exhibiting no inhibition; yet, large variations in activity were observed. Mutations of residue Y385, in particular, are of interest due to the demonstrated favorable affect on inhibition without a substantial loss in activity (achieved with an alanine substitution).

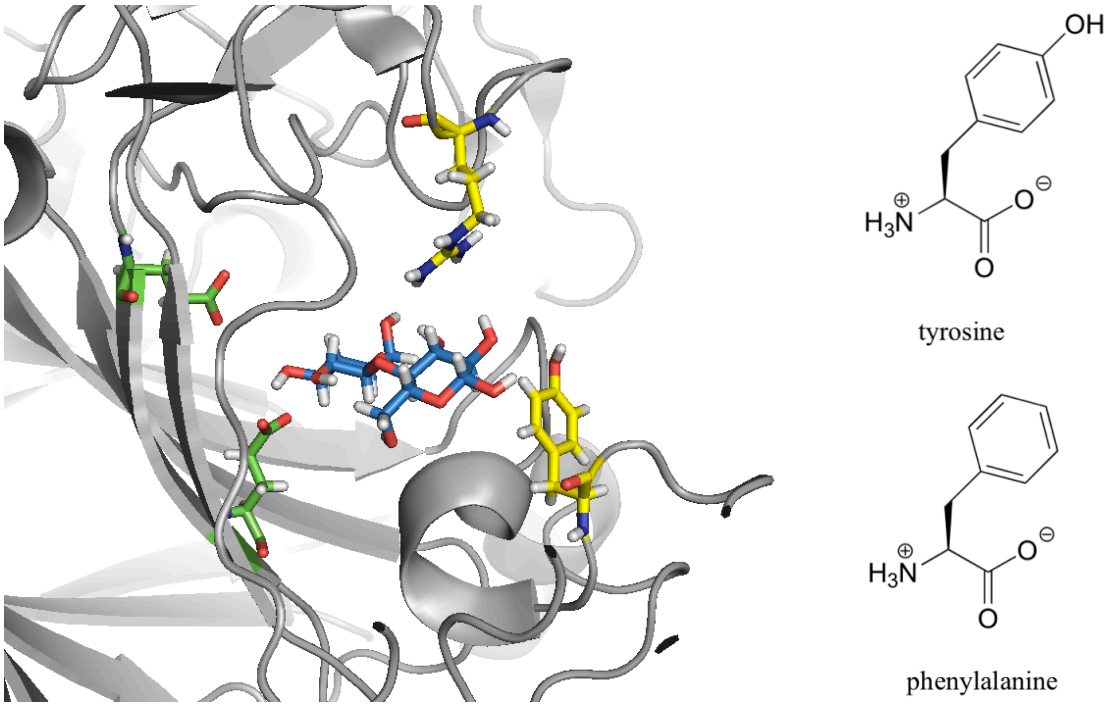


Figure 33. Cellobiose (blue) occupying the product binding sites +1 and +2 of *Te*Cel7A. In yellow are residues R248 (top) and Y385 (bottom), which may interact, perhaps electrostatically, to form a closed, tunnel-like conformation obstructing product release. Catalytic residues are shown in green. Unlike alanine, phenylalanine is structurally very similar to tyrosine, making Y358F a promising Cel7A mutant for future experimental analysis.

Less drastic mutations of this amino acid, perhaps from tyrosine to phenylalanine, may yield similar product tolerance improvements (by conceivably preventing an electrostatic interaction with R248 obstructing cellobiose release, Figures 28 and 33) at a minimal cost to catalytic activity. Future experiments exploring such options may result in a product-tolerant Cel7A mutant with catalytic activity comparable to that of the wildtype enzyme. Such an enzyme would improve the efficiency of cellulose hydrolysis and thereby lower the cost of the resulting biofuel, improving the likelihood of renewable energy adoption.

Chapter 3

Progress Towards Engineering a Lytic Polysaccharide Monooxygenase Enzyme

3.1 Introduction

To effectively break down cellulosic biomass, several enzymes, such as cellobiohydrolases, endoglucanases, and beta-glucosidase, need to work in concert (Figure 13). Recently, additional enzymes, such as lytic polysaccharide monooxygenases, have been found to contribute synergistically to the effectiveness of enzymatic cellulose hydrolysis.[5] Together, the enzyme cocktail is more efficient than the sum of each enzyme's individual contribution. Because these molecular machines make up one of the most expensive inputs to the cellulosic biofuels process, efficiency improvements in enzymatic hydrolysis that result in a reduced enzyme requirement could effect a cost savings, making cellulosic biofuels more cost-competitive with fossil fuels.

Lytic polysaccharide monooxygenases (LPMOs) have been shown to be capable of significantly enhancing the enzymatic depolymerization of cellulosic biomass. Cellulase-active LPMOs are metalloenzymes that oxidize crystalline cellulose, making new cellulose chain ends available for hydrolysis by cellulase enzymes such as cellobiohydrolases and endoglucanases. They create chain breaks in crystalline parts of the substrate where endoglucanases cannot.[5] Thus, LPMOs can be highly synergistic with canonical cellulases, able to improve hydrolytic activity (and consequently reduce enzyme loading) by as much as 2-fold.[39] This ability to synergize with canonical cellulases makes LPMOs of interest as we seek better and more cost-efficient enzyme cocktails for cellulosic biomass hydrolysis.

LPMOs have been identified in both bacteria and fungi (predominantly active on chitin or cellulose, respectively) and each has an N-terminal histidine residue that functions as a bi-dentate ligand for copper (Figure 34). Activity requires an electron donor, either partner enzyme cellobiose dehydrogenase (CDH) or a small molecule reducing agent such as ascorbate or glutathione. Finally, molecular oxygen is necessary for LPMO action.[5]

The enzymes oxidize either the reducing end or non-reducing end of glucose; oxidation at the C1 carbon produces a lactone (and, upon hydrolysis, an aldonic acid) while oxidation at the C4 carbon yields a 4-keto sugar (and subsequently a geminal diol), as shown in Figure 35.[40] LPMOs with these specificities are designated as either type-1 or type-2 enzymes, respectively. The flat, wide binding surface of an LPMO (Figure 36) interacts with several substrate chains of crystalline polysaccharides.[41]

The known diversity of this class of enzymes is constantly expanding; recently, hemicellulose was shown to be an LPMO substrate in addition to chitin and cellulose.[42] While most LPMOs appear to function only on crystalline substrates, there is some evidence for enzymes that can oxidize soluble substrates.[43] The number of LPMOs expressed by cellulolytic organisms varies dramatically. *Trichoderma reesei*, for example, only produces two LPMOs. DNA from *Neurospora crassa*, on the other hand, encodes 14,

and other fungi have genes for over 30 of these proteins.[39] LPMOs have been categorized into three auxiliary activity (AA) families: AA9 enzymes are from fungi, AA10 from bacteria, and AA11 enzymes (the most recently discovered) are believed to be an intermediate of the other two families.[5]

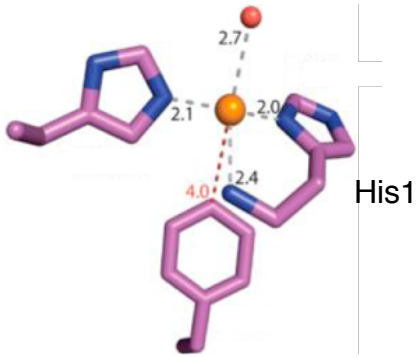


Figure 34. Copper coordination in LPMOs requires an N-terminal histidine residue.[44]

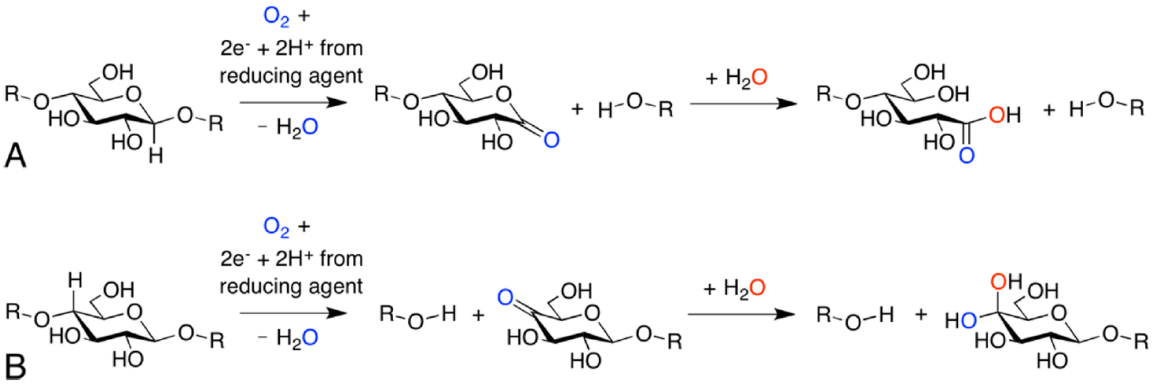


Figure 35. Type-1 LPMO oxidation of cellulose at the C1 carbon of the sugar substrate yields a lactone and subsequent aldonic acid (A). Type-2 LPMO oxidation at the C4 carbon produces a 4-keto sugar that can be hydrolyzed into a gemdiol (B).[5]

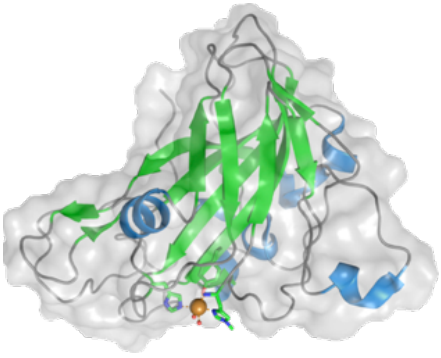


Figure 36. Structure of an LPMO (binding surface is at the bottom).[45]

3.2 A Cellulase-Active, Bacterial LPMO: CelS2

3.2.1 Expression of CelS2 in *E. coli*

Bacterial proteins with structural similarity to fungal LPMOs are best known for their analogous crystallinity-disrupting activity on chitin. However, in 2011 one AA10 protein, CelS2 from soil bacterium *Streptomyces coelicolor* A3(2), was recombinantly expressed in *Escherichia coli* (*E. coli*) and shown to have activity on crystalline cellulose (Avicel).[46]

Demonstrated cellulose-active LPMO expression in a bacterial system made CelS2 a promising candidate for LPMO enzyme engineering, as filamentous fungal expression systems are not amenable to high-throughput enzyme expression. Thus, it was my goal to express CelS2 in bacteria, verify its activity on cellulose, study its native thermostability and, if applicable, use directed evolution to evolve a more thermostable CelS2 mutant. An engineered CelS2 mutant could complement other "improved" enzymes that have been developed, including thermostable Cel7A (CBH1) and Cel7B (EG1) mutants, providing a more complete and efficient enzyme cocktail suitable for high-temperature cellulose hydrolysis. I also hoped to characterize the CelS2 enzyme to understand its temperature optimum and stability range, pH optimum, activity on substrates subjected to a variety of pretreatments, and optimal ratio with hydrolases for maximum synergy.

Unfortunately, expressing CelS2 in *E. coli* was unexpectedly challenging. I had the gene encoding CelS2 (codon-optimized for either *E. coli* or *Bacillus subtilis*) synthesized. I cloned these genes into vectors for *E. coli* expression, engineering in a method for cleaving off a tag (Figure 37) in order to generate the N-terminal histidine residue required for copper coordination (Figure 34) and oxidative activity.

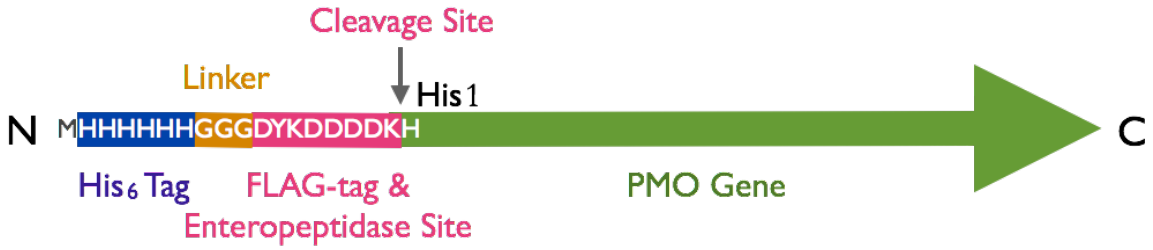


Figure 37. Protein design for expression and processing of CelS2 in *E. coli*.

After extensive troubleshooting, this strategy appeared to be working in *E. coli* (only for the *B. subtilis* codon-optimized gene, surprisingly). I was able to express a modest amount of CelS2 in *E. coli* and follow tag removal over time by protease enzyme enteropeptidase using SDS-PAGE (Figure 38) and western blots with anti-6x-histidine tag antibodies (Figure 39). Despite this, however, the enzyme did not appear to be active.

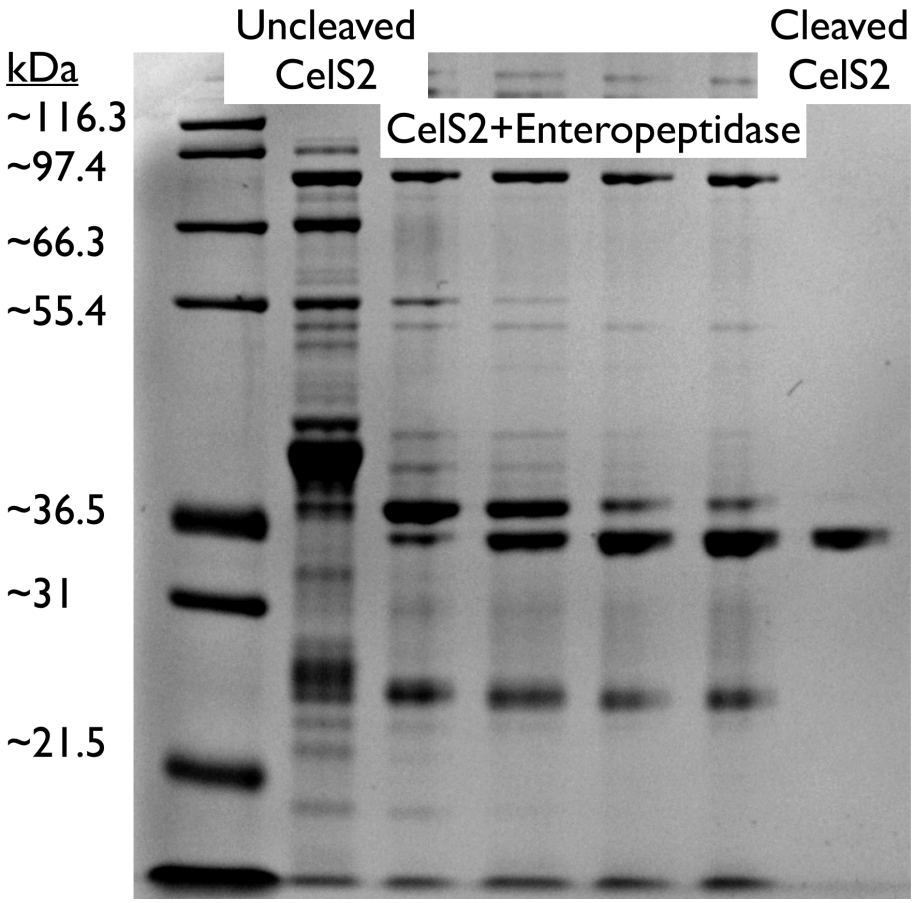


Figure 38. SDS-PAGE illustrating cleavage of CelS2 by enteropeptidase over time. Uncleaved CelS2 (with a His-tag and FLAG tag) had a calculated molar mass of 36.7 kDa. Cleaved CelS2 (with an N-terminal histidine) had a calculated molar mass of 34.6 kDa.

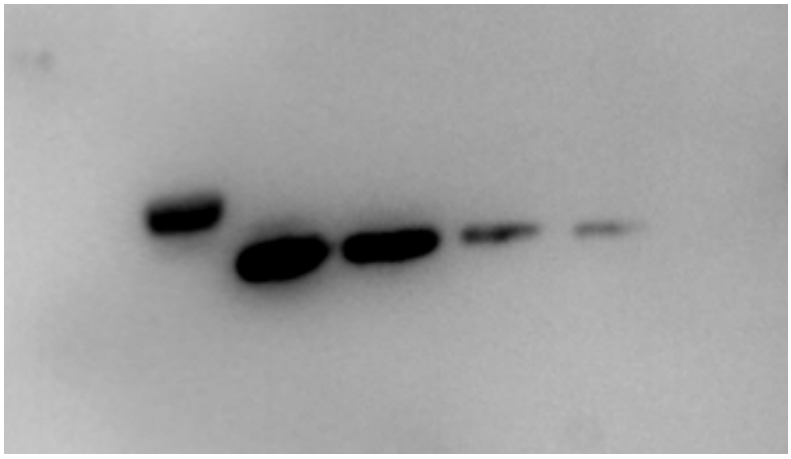


Figure 39. Successful removal of the N-terminal 6x-histidine tag over time could be followed by a western blot.

Eventually, N-terminal protein sequencing revealed that the protease tag on the protein was being improperly cleaved at the wrong lysine residue (Figure 40). The LPMO could not be active without the N-terminal histidine, as the N terminus itself assists in copper coordination (Figure 34).

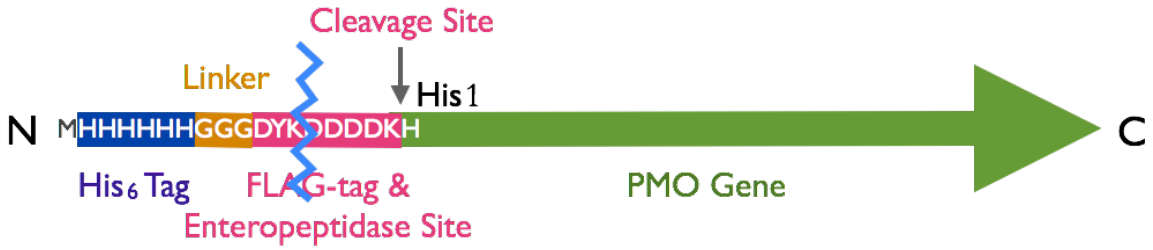


Figure 40. Enteropeptidase processed this protein incorrectly by cleaving after the first lysine residue (indicated with a blue squiggle) instead of directly before the desired N-terminal histidine (gray arrow) of CelS2.

To circumvent this problem, I attempted to express CelS2 in the *E. coli* periplasm; but, unfortunately this approach failed using the *ompA* system and worked only marginally better under the *pelB* system.

3.2.2 Expression of CelS2 in *B. subtilis*

After limited success expressing CelS2 in *E. coli*, I changed strategies and instead sought an organism that could, on its own, both properly express the bacterial CelS2 gene and generate the desired N-terminus without requiring external processing (such as tag removal by proteases).

Postdoctoral scholar Dr. Lars Giger and I developed an alternative expression method in *Bacillus subtilis*. I cloned and expressed the gene for bacterial LPMO CelS2 (with an N-terminal secretion tag and a C-terminal 6x-histidine tag) in the bacterium, which can naturally remove secretion tags from proteins. Indeed, *B. subtilis* was able to both secrete CelS2 and cleave off the signal peptide to yield the proper N-terminus. I verified that the secretion tag on the N-terminus was properly removed using N-terminal sequencing, SDS-PAGE (Figure 41), and LC-MS mass spectrometry (Figure 42). I also confirmed copper presence using electron paramagnetic resonance (EPR) spectroscopy (results not shown).

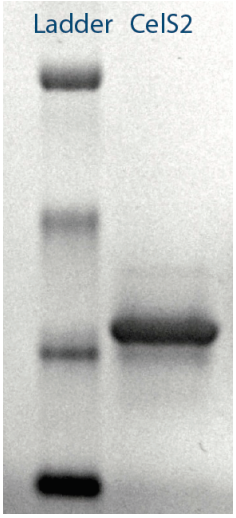


Figure 41. SDS-PAGE showing pure, correctly-processed CelS2 protein.

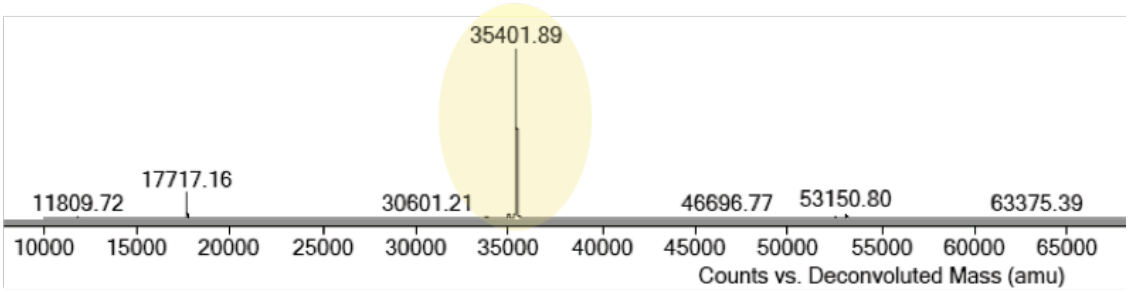


Figure 42. LC-MS (liquid chromatography mass spectrometry) results confirm CelS2 (with an N-terminal histidine residue and a C-terminal 6x-histidine tag) size and purity. The calculated molecular mass is 35390.4 Da.

As other researchers had successfully expressed four fungal LPMOs in yeast *Pichia pastoris*,[47] I cloned four fungal LPMOs from *N. crassa* cDNA (NCU01050 NCU02240, NCU07898, and NCU08760) into *B. subtilis* vectors to test expression in that system. Unsurprisingly, I could not get any of these fungal proteins to successfully express in the bacterium. Because fungal LPMOs may be more industrially relevant (but more difficult to study and engineer) than bacterial ones, I also attempted expression of these *N. crassa* enzymes in *Saccharomyces cerevisiae*, but was unsuccessful. My aim was to use a facile expression host such as *S. cerevisiae* to engineer fungal LPMOs using directed evolution. *B. subtilis* and *S. cerevisiae* expression systems are more suitable for high-throughput enzyme engineering than filamentous fungal systems such as *N. crassa* or *T. reesei*. An LPMO enzyme with activity at 65˚C or 70˚C and a long operational half-life would be suitable for enzyme recycle (should it become a possibility) and could greatly improve the efficiency of enzyme-catalyzed cellulose hydrolysis at high temperatures.

3.2.3 Engineering Targets for CelS2

To my knowledge, there has been no published, applied LPMO research—I aimed to demonstrate the first LPMO enzyme engineering effort. A thermostable LPMO mutant may lead to a more complete and efficient enzyme cocktail suitable for high-temperature cellulose hydrolysis.

Expression in a bacterial host makes CelS2 amenable to high-throughput protein engineering. CelS2 (with a C-terminal 6x-histidine tag) has a transition midpoint temperature of 56˚C, as measured by differential scanning calorimetry (Figure 43). A more thermostable LPMO (e.g. with a T_m of ~70˚C) could be engineered to help create a thermostable, synergistic enzyme cocktail. If an enzyme is active and amenable to high-throughput engineering, a library of mutants can be generated using error-prone PCR (or B-factor analysis of the crystal structure) followed by a screen for synergistic activity with thermostable Cel7A at elevated temperatures. An alternative strategy to generate a more thermostable LPMO is to mine genomes for naturally thermostable LPMOs and shuffle the genes of these homologs.

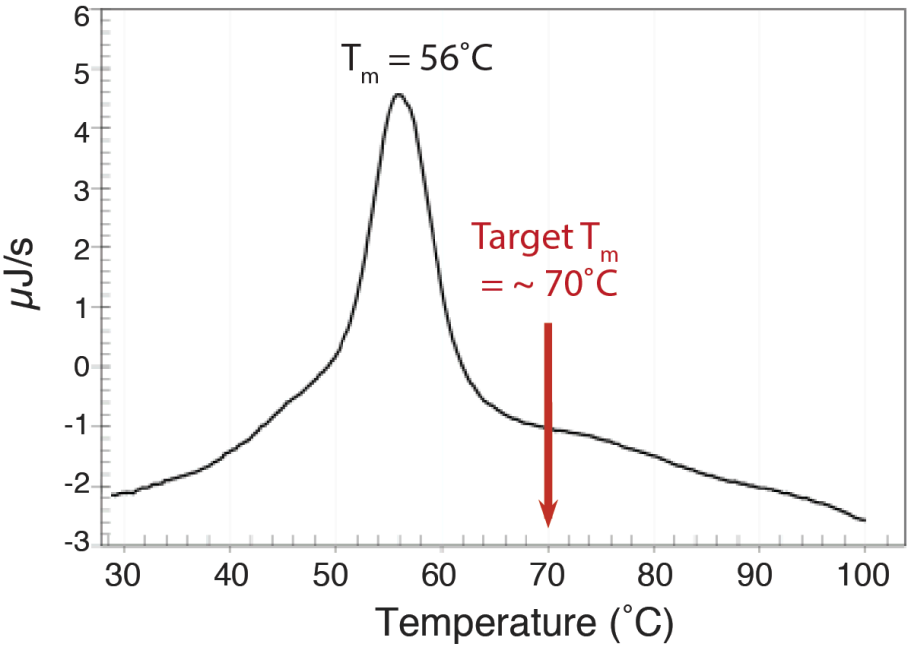


Figure 43. A differential scanning calorimetry (DSC) curve reveals that CelS2 (with a C-terminal 6x-histidine tag) has a transition midpoint (T_m) temperature of 56˚C; a target T_m might be 70˚C.

3.2.4 Activity Analysis of CelS2

I verified the *B. subtilis*-produced CelS2 as an active, Type-1 LPMO capable of producing oxidized sugar products from Avicel, a model crystalline cellulose substrate (Figure 44,

red arrows). Type-1 LPMOs hydroxylate the C1 position of pyranose sugars, producing aldonolactones (Figure 35).

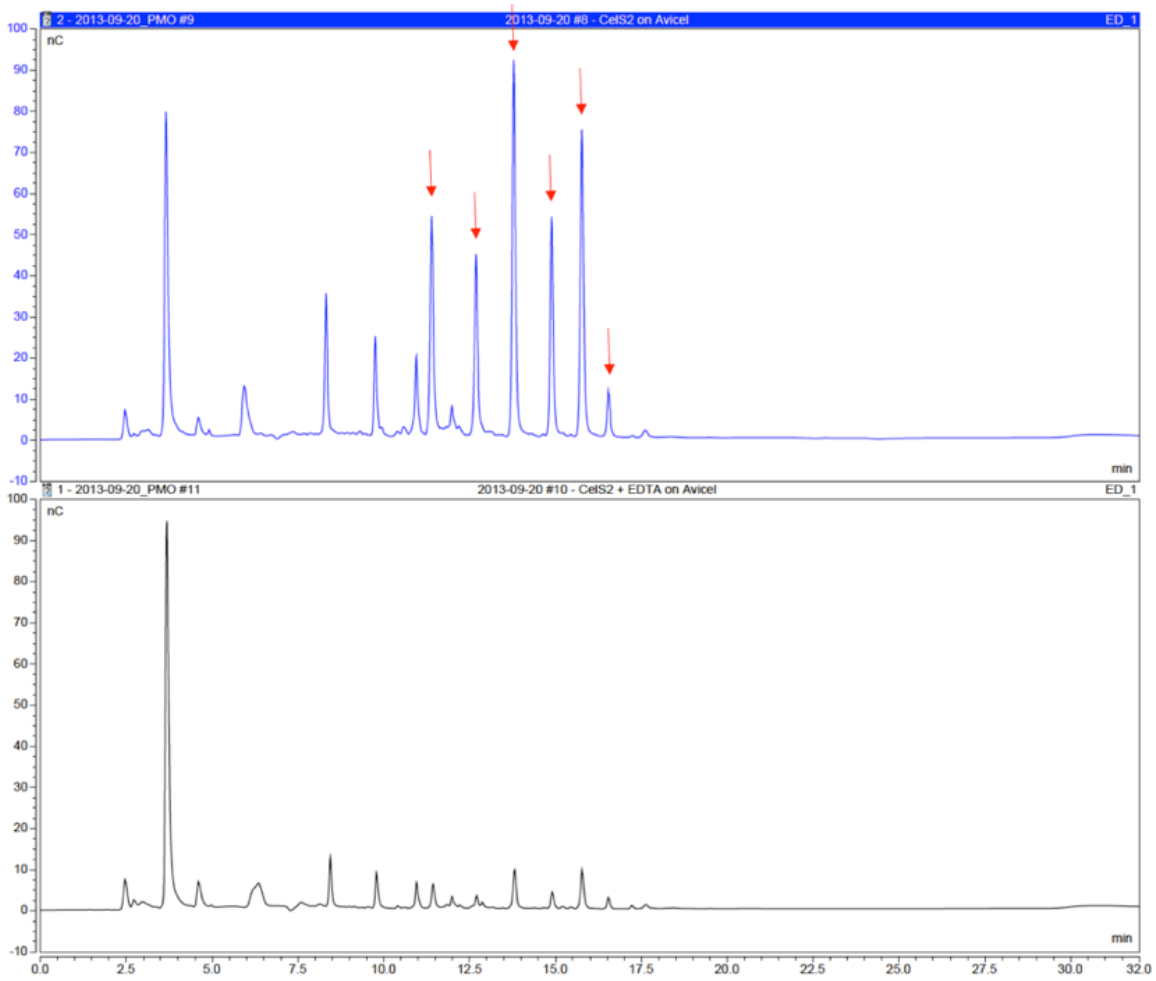


Figure 44. Dionex HPLC data confirms oxidative type-1 LPMO activity of CelS2 on Avicel (above); red arrows indicate aldonic acids. As expected, metalloenzyme CelS2 loses activity without Cu^{2+} and in the presence of metal chelator EDTA (below).

To facilitate potential high-throughput activity analysis, I also tried using a simple fluorescent LPMO activity assay that utilizes a soluble substrate (Figure 45). Once copper in the LPMO has been reduced by either CDH or ascorbate, the enzyme activates oxygen. Peroxide can be detected by an enzyme-coupled reaction with horseradish peroxidase and Amplex Red, a molecule that is converted to fluorescent resorufin.[47] Unfortunately, I was unable to successfully monitor enzymatic activity using this method.

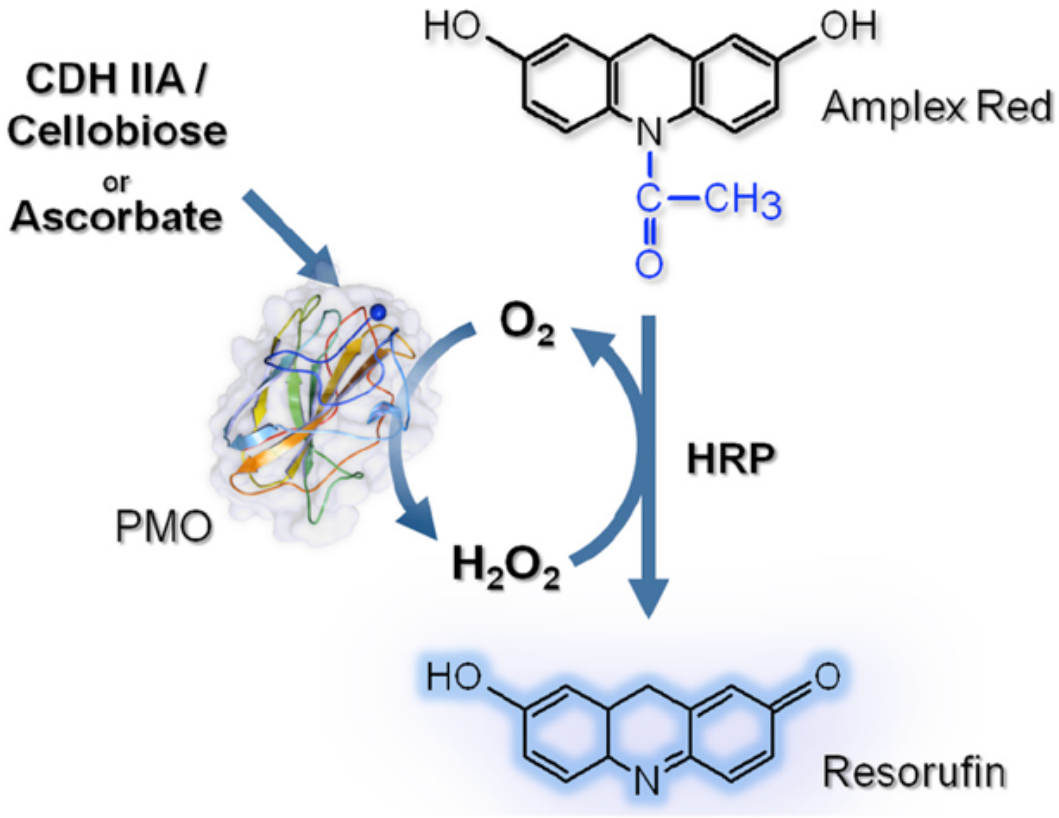


Figure 45. A fluorescent LPMO activity assay.[47]

3.2.5 Investigating Synergy with CelS2

Once I had verified CelS2 expression and oxidative activity using high performance liquid chromatography (HPLC, Figure 44), I sought to confirm this LPMO's reported synergy with glycosyl hydrolases. As shown in Figure 46, synergistic activity of CelS2 with canonical cellulases was observed, but less than expected.

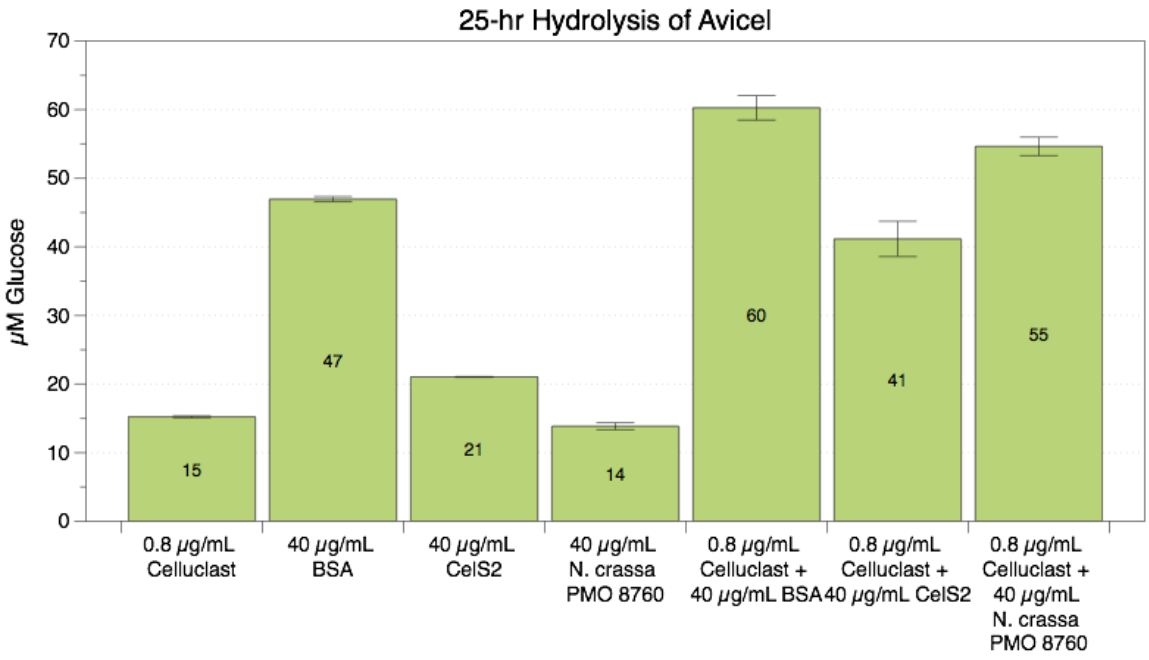


Figure 46. Glucose released by proteins alone (left) or in combination (right).

As anticipated, there appeared to be an enhancement in cellulose hydrolysis by Celluclast in the presence of CelS2 that was (at least mildly) synergistic (41μM glucose > 15μM + 21μM). Fungal type-1 LPMO *N. crassa* NCU08760 exhibited significant synergy in this experiment. However, notable enhancements also resulted from control samples containing bovine serum albumin (BSA) protein. Similar observations were made with crude *E. coli* cell lysate (data not shown). Unfortunately, CelS2 neither stood apart in these assays nor was this activity diminished by attempted removal of Cu^{2+} and addition of metal chelator EDTA, which should inhibit the metalloenzyme (Figure 44). Together, these results suggested that the observed boosting effects in enzymatic hydrolysis were due to (any) protein being present and not the result of oxidative LPMO activity.

It has actually been reported in the literature that BSA can increase hydrolytic activity of cellulases by a factor of two.[48,49] However, in the original 2006 study, this effect was evident on lignocellulosic substrates only and not on Avicel (pure, crystalline cellulose). Because BSA did not adsorb significantly on Avicel, the authors concluded that BSA irreversibly bound to lignin and prevented the enzymes from non-productively doing so, thus improving their ability to hydrolyze the cellulose.[48] In my experiments, however, I observed a hydrolysis boost with BSA on Avicel, which conflicts with this conclusion. A 2015 study, on the other hand, is consistent with my work in that it does report a BSA boosting effect on filter paper (cellulose) hydrolysis.[49]

Even more surprising than seeing "synergy" with control samples was the fact that data in the literature about CelS2 activity with cellulase mixtures (such as Celluclast) clearly show a boosting effect above that provided by BSA (Figure 47).

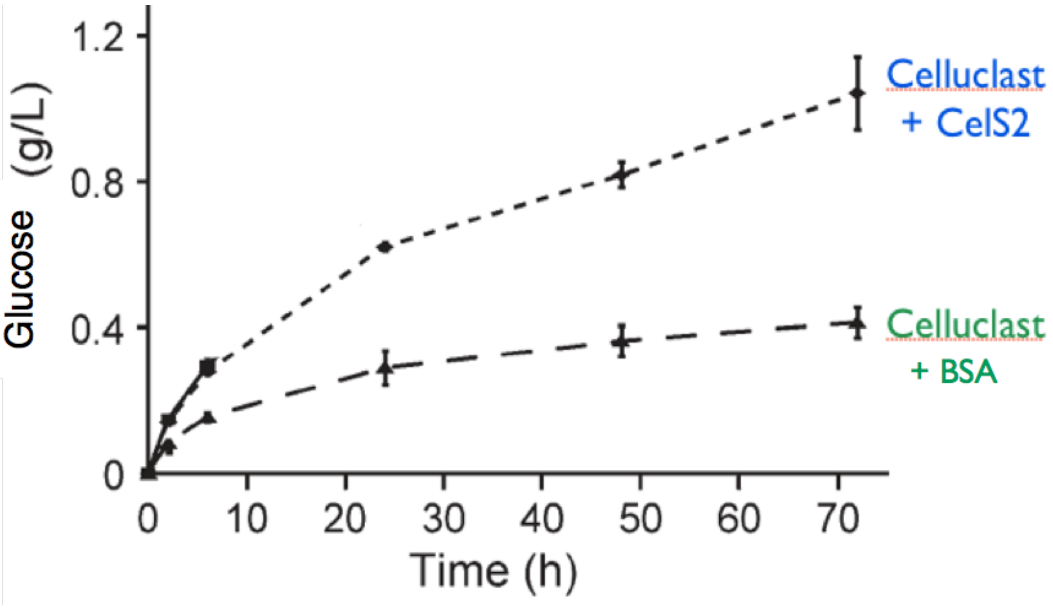


Figure 47. Others reported a 2-fold improvement in hydrolytic activity (compared to that of control samples containing BSA) when CelS2 was added to cellulase enzyme mixture Celluclast.[46]

Unfortunately, I was never able to repeat the data shown in Figure 47. Using an updated cellulase mixture (CTec2 vs Celluclast, both from Novozymes), I did not observe LPMO synergy with CelS2 (Figure 48).

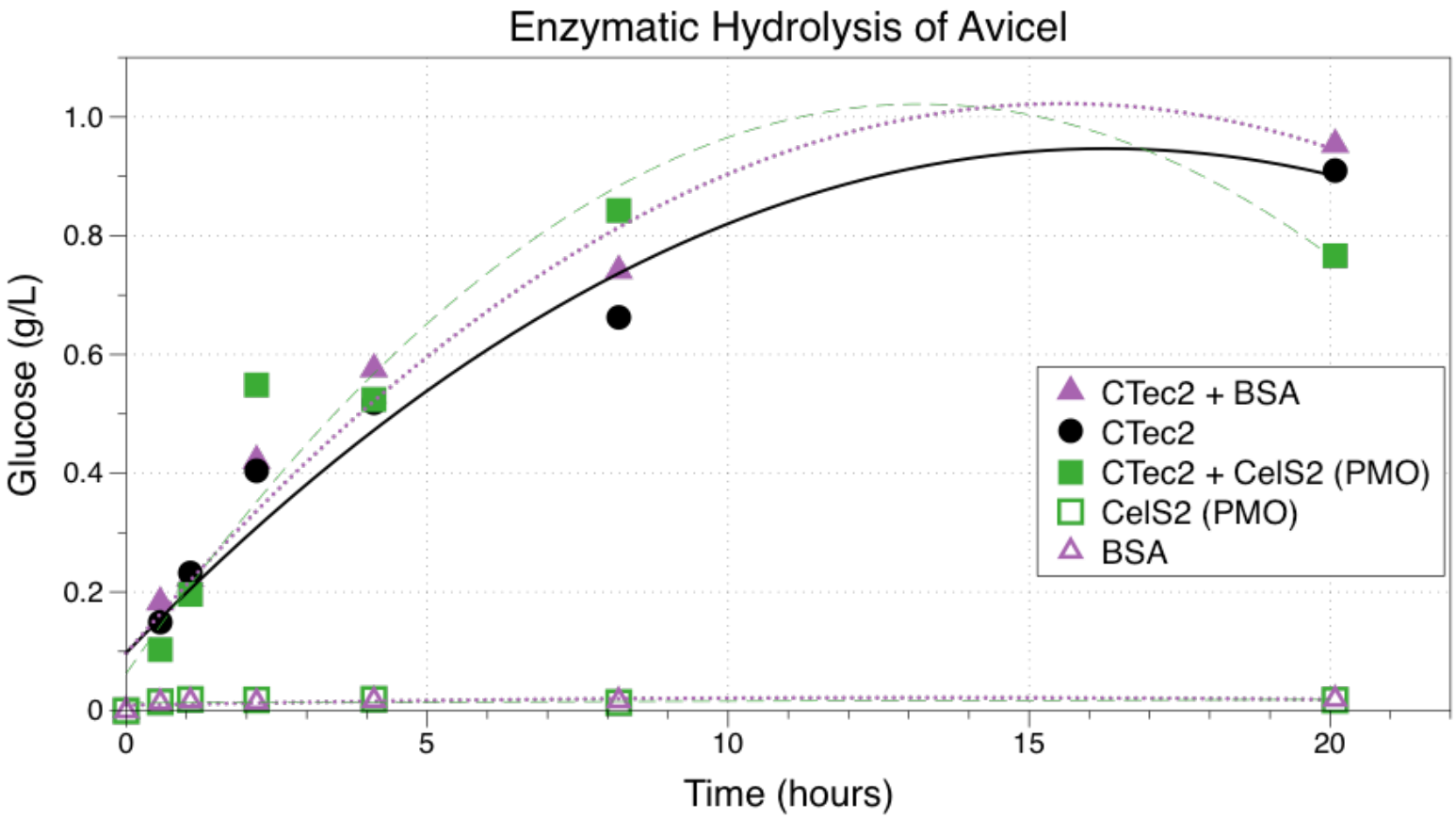


Figure 48. Enzymatic hydrolysis of cellulose over time. Assay conditions: 10 mg/mL Avicel, 20mM sodium acetate buffer pH 5.5, 2mM ascorbic acid, 0.177 mg/mL CTec2 fungal cellulase enzyme cocktail (source: Novozymes), 0.177 mg/mL LPMO (5μM CelS2), 0.177 mg/mL BSA (source: NEB). One-hour pre-incubation of Avicel with BSA or LPMO (CelS2) as indicated prior to CTec2 addition. 1mL reaction volumes. 50˚C, shaking at 800rpm. Glucose measured via Dionex HPLC.

3.3 Conclusions

Despite bacterial LPMO CelS2 being active and capable of oxidizing crystalline cellulose (at least in some preparations), it did not appear to exhibit notable synergy with commercial cellulase enzyme cocktails under the conditions tested. Perhaps CelS2 is synergistic with bacterial cellulases but less so with fungal ones. More work is necessary to determine why some LPMOs enhance cellulase activity more than others. For example, *T. reesei* contains several LPMO genes, yet the addition of LPMOs from a separate fungus increases enzymatic activity.[50] A better understanding of these fascinating enzymes, and under what conditions they synergize with hydrolases, is needed.

This project did not work as envisioned for a number of reasons. Expression of CelS2, even in *B. subtilis*, was inconsistent and enzyme synergy could not be reliably observed (if at all). CelS2 was recently shown to work in synergy with a type-2 LPMO,[51] without which CelS2 has less cellulose-oxidizing activity, a fact which may have contributed to the observed lack of synergy in my experiments. Fungal LPMOs are also known to have better synergistic properties, but I was unable to express any AA9 enzymes using *E. coli*, *B. subtilis*, and *S. cerevisiae* expression systems.

A few additional challenges persist as well. Transition midpoint temperatures of fungal LPMOs are already fairly high, around 66.9°C, 63.0°C, 68.9°C and 67.9°C for LPMO-01867, LPMO-02916, LPMO-03328 and LPMO-08760, respectively.[47] Perhaps there is no need to engineer a thermostable LPMO. Additionally, the ratio of LPMO CelS2 to Cellulast is 40:1 in Figure 47.[46] A realistic ratio for boosting activity with a synergistic enzyme would be closer to the opposite, or 1:40 for LPMOs to other cellulases. Finally, LPMO activity can create gluconic acid, a molecule that inhibits β-glucosidase and may complicate development of a cocktail with high cellulase activity. For these many reasons, I was unable to meet my goals for this research.

Chapter 4

Exploring Enzyme Synergy and

Methods for Purification of *Trichoderma reesei* Cellulases

4.1 Enzyme Synergy

Cellulolytic organisms produce a multitude of cellulase enzymes that work in concert to depolymerize cellulose. Breaking down cellulose to glucose might, in theory, only require a few enzymes—an endoglucanase to cleave internal bonds, cellobiohydrolases I and II to hydrolyze cellulose from the reducing and non-reducing ends, respectively, and β-glucosidase to hydrolyze cellobiose to glucose. Indeed, cellulolytic fungi secrete at least three extracellular enzymes: an endoglucanase, cellobiohydrolase, and β-glucosidase;[52] however, the genomes of filamentous fungi contain many more than three cellulases. Ten cellulases have been predicted in *Trichoderma reesei* (*Hypocrea jecorina*)[53] and as many as 60 in *Podospora anserina*.[54]

It is interesting to consider why these fungal cellulase complexes are so large; perhaps interactions between various enzymes play an important role. Endoglucanases and cellobiohydrolases are known to interact synergistically.[55] We can define enzyme synergy as meaning cellulose degradation activity together that is greater than the sum of the enzyme's activities alone. I was interested in engineering cellulase enzymes for improved properties (thermostability, pH tolerance, specific activity, etc.) to enable large-scale commercialization of lignocellulosic biofuel technologies. Given the multitude of cellulase enzymes in fungal broths, an improved enzyme would need to function, and show improved performance, in the presence of these others.

Directed evolution experiments have generally improved cellulase enzymes in isolation; however, this technique is most powerful when the screening conditions accurately mimic the application conditions. Therefore, I proposed to engineer cellulases for desired attributes in the presence of complementary enzymes on industrially-relevant solid biomass substrates (such as acid-pretreated *Miscanthus*). Using directed evolution, I aimed to improve the hydrolytic activity of fungal cellulase suites as a whole by screening enzyme libraries of one or two cellulases in the company of other enzymes. By avoiding library screens in isolation, I expected to preserve, and perhaps evolve for, inter-enzyme synergy. I anticipated that these directed evolution experiments would generate cellulase cocktails with improved activities on solid biomass substrates.

Specifically, I wished to screen a Cel7A cellobiohydrolase I library for thermostability and improved specific activity either alone or in the presence of other cellulase enzymes—perhaps with *N. crassa* ΔCBH1 supernatant or a thermostable endoglucanase I (Cel7B, EG1) mutant—and evaluate if the most interesting enzymes identified differ when screened alone or in a mixture. Cellulases would not be used in isolation for industrial biomass hydrolysis; thus, it is important to determine if current directed evolution experiments screening enzymes in isolation are relevant, or if future screens should instead utilize enzyme cocktails as a more realistic representation of enzyme use for biofuels.

First, to examine if I could measure enzyme synergy, I studied the hydrolysis activities of enzymes alone and together (Figure 49 and 50) and, indeed, observed a synergistic effect.

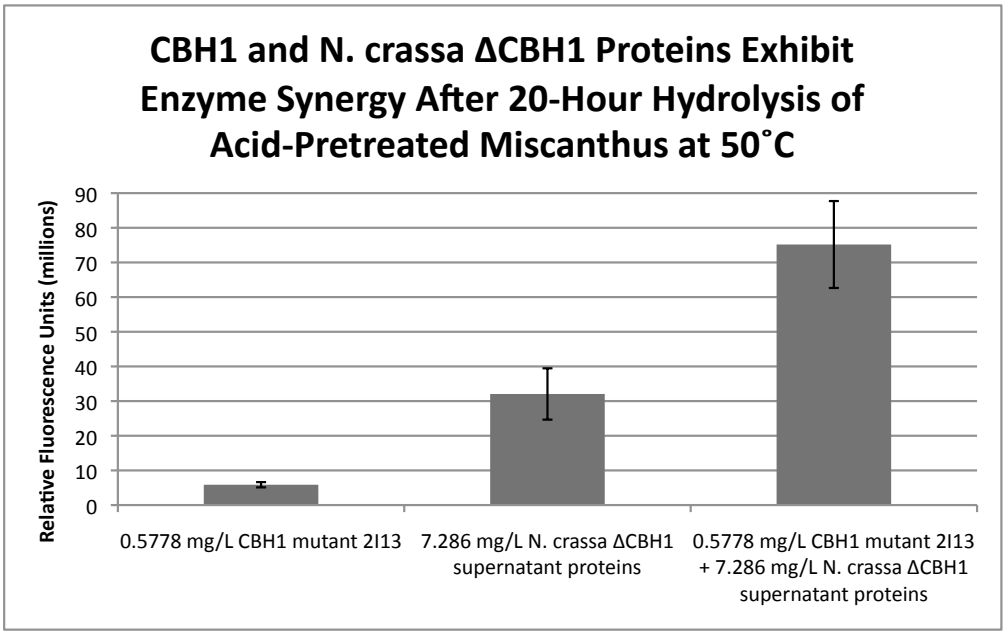


Figure 49. Cel7A (CBH1) and *N. crassa* supernatant proteins lacking Cel7A exhibit synergy when combined. (Approximately 1 million Relative Fluorescence Units represents 1.25μM glucose released.)

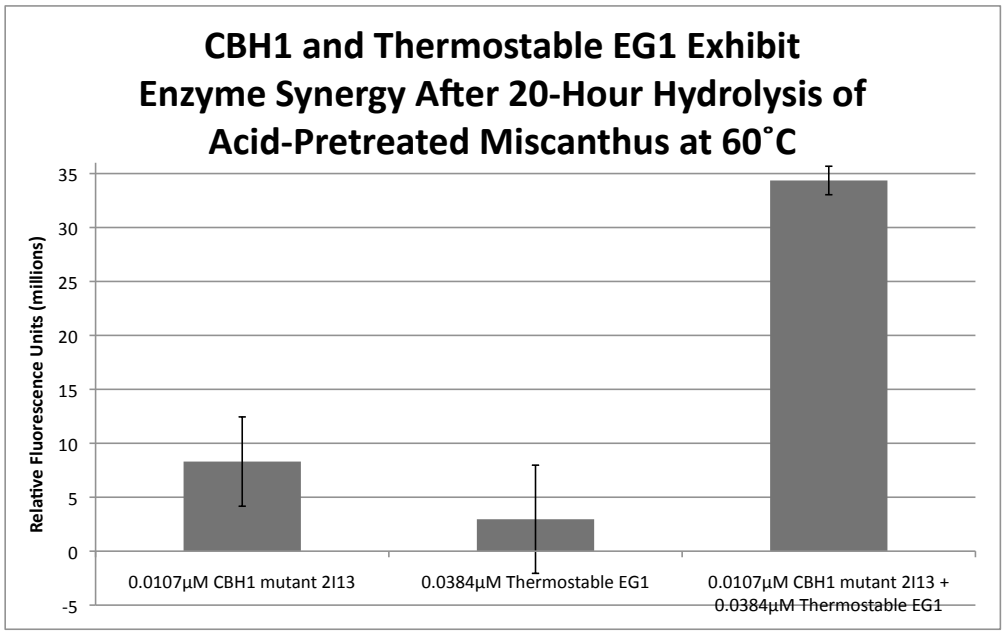


Figure 50. Cel7A (CBH1) and a thermostable endoglucanase I (Cel7B, EG1) enzyme exhibit synergy at 60˚C. (Approximately 1 million Relative Fluorescence Units represents 1.25μM glucose released.)

I subsequently performed a time-course hydrolysis experiment comparing wildtype (WT) *N. crassa* supernatant proteins, *N. crassa* ΔCBH1 supernatant proteins (lacking the Cel7A enzyme), and *N. crassa* ΔCBH1 supernatant proteins + the following concentrations of *T. longibrachiatum* CBH1 (Cel7A):

- 0.5x the amount of CBH1 found in WT *N. crassa* -> 19.75% CBH1 by mass
- 1x the amount of CBH1 found in WT *N. crassa* -> 39.5% CBH1 by mass
- 2x the amount of CBH1 found in WT *N. crassa* -> 79% CBH1 by mass
- 0.5x the amount of CBH1 found in WT *N. crassa* (while keeping *N. crassa* ΔCBH1 supernatant constant)
- 2x the amount of CBH1 found in WT *N. crassa* (while keeping *N. crassa* ΔCBH1 supernatant constant)

I performed these experiments on lignocellulosic substrate *Miscanthus* (Figure 51) or pur cellulose substrate Avicel (Figure 52) with reaction volumes of 70µL and glucan concentrations of 1.59%.

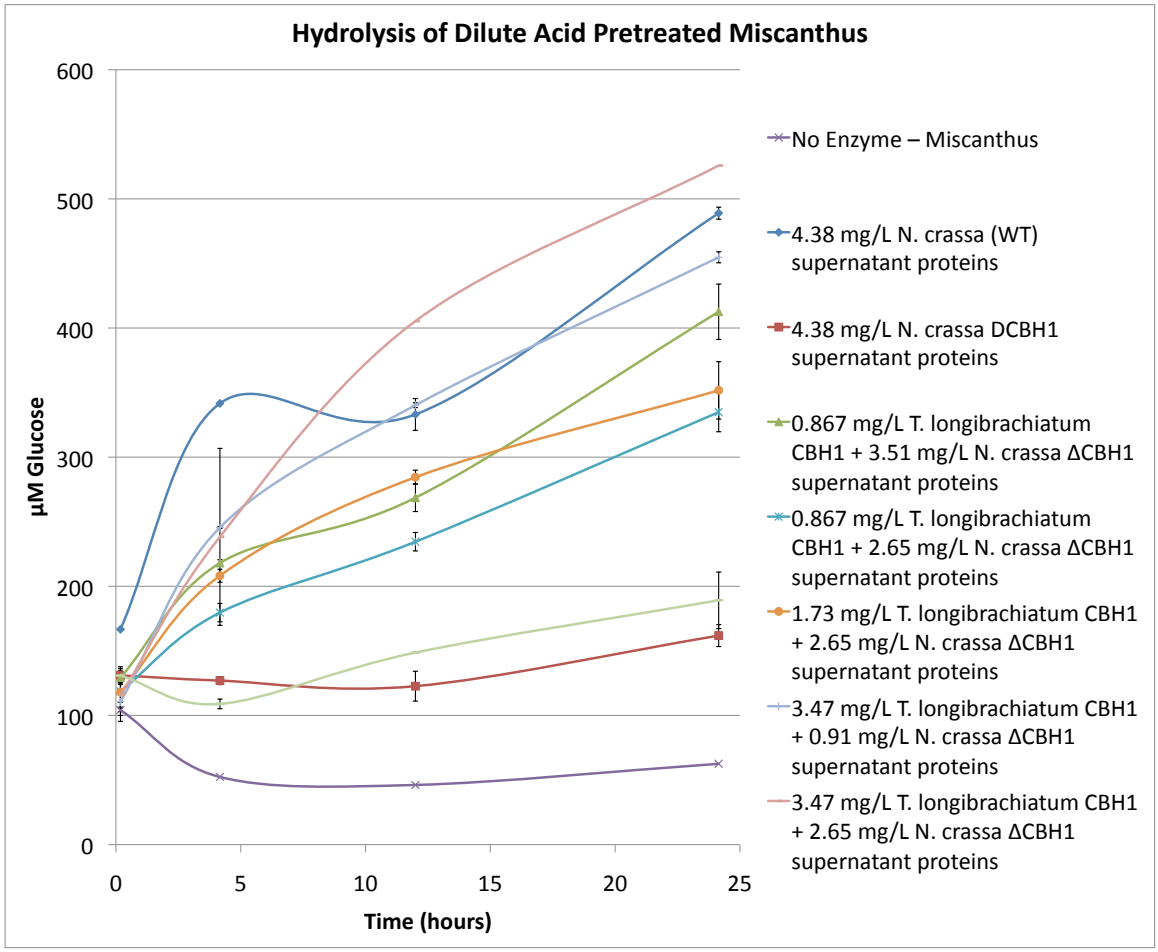


Figure 51. Enzymatic hydrolysis of *Miscanthus*.

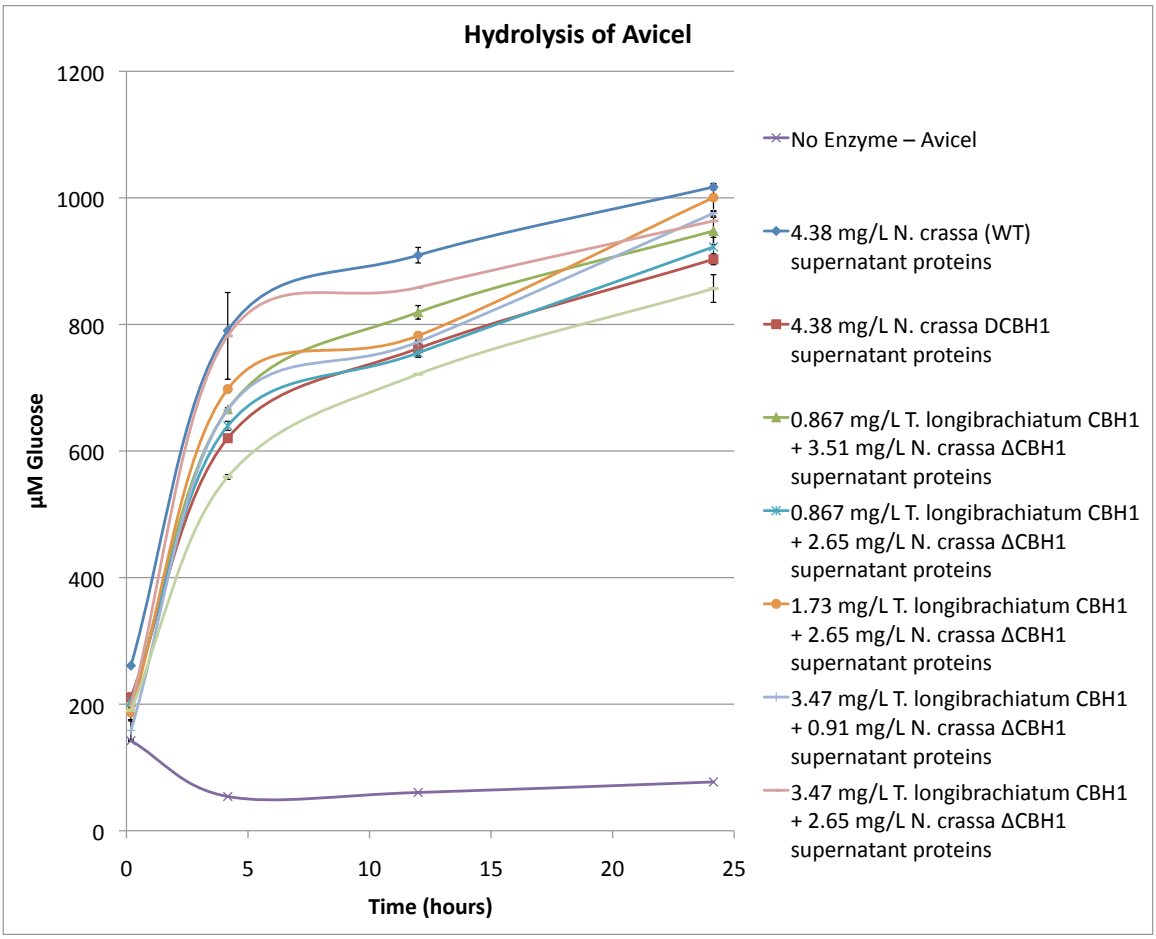


Figure 52. Enzymatic hydrolysis of Avicel. (Only about 1% of the substrate was successfully converted to glucose, thus the enzyme/substrate ratio was low and these results are likely due to hydrolysis of free ends.)

Interestingly, in both experiments the wildtype (WT) *N. crassa* supernatant proteins were nearly the best at hydrolyzing cellulose. This brought into question whether I could actually evolve for cellulase enzyme synergy (if little difference was observed when complementing reactions with *T. longibrachiatum* Cel7A) and if there are, in fact, gains to be made in the efficiency of a cellulase complex as a whole by evolving individual enzymes in the presence of others.

In light of the complications around using an undefined enzyme mixture (such as *N. crassa* supernatant proteins or Novozymes' Celluclast, for example), I opted to refine future experiments by purifying individual *T. reesei* enzymes to use as defined enzyme "background" mixtures for such studies. The hope was that by screening for improved synergy with a discrete set of complementary enzymes I could reduce variability in the data.

4.2 Methods for Purification of *Trichoderma reesei* Cellulases

The motivation for purifying *T. reesei* cellulase enzymes was to isolate individual cellulases for synergy studies, develop reliable reference measurements, create standardized protocols and assays, and make enzyme comparisons with engineered variants. Comparisons of interest included small-scale vs. large-scale experiments and enzymatic activity on various solid substrates following different biomass pretreatments.

Our team's approach was to purify Cel7A (CBH1), Cel6A (CBH2), Cel7B (EG1), and Cel5A (EG2) from a commercial crude cellulase mixture (Novozymes' Celluclast) using Fast Protein Liquid Chromatography (FPLC). After confirming enzyme identity and purity using liquid chromatography and mass spectrometry (LC-MS), we would characterize the enzyme, measuring enzyme specific activities, pH and temperature optima, and kinetic parameters using standardized assays. We would also create a resource of purified enzyme stocks (for use in individual projects) and the foundation for making enzyme comparisons in a consistent manner using standardized protocols and assay conditions.

This project was a joint effort and required substantial protocol development, which I describe below.

4.2.1 Separating Celluclast Proteins

To purify individual enzymes from Novozymes' *T. reesei* cellulase enzyme mixture, Celluclast, I first desalted a 7.5 g/L enzyme solution using a HiPrep 26/10 Desalting Column and 25mM HEPES buffer pH 7.35. Next, I separated the proteins via anion exchange over a HiLoad 16/10 Q Sepharose High Performance Column using 25mM HEPES buffer pH 7.35 with or without 1M sodium chloride (Figures 53 and 54).

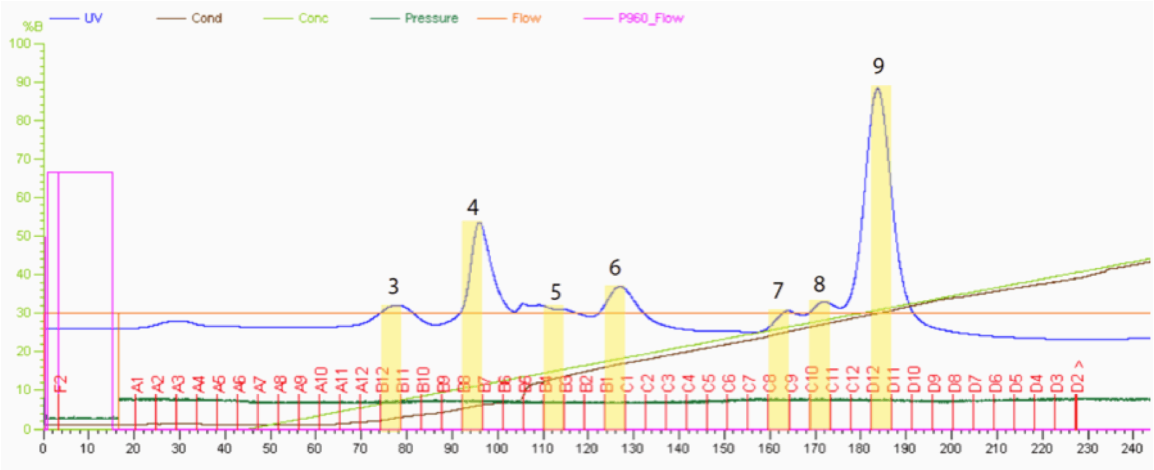


Figure 53. FPLC chromatogram showing the separation of proteins from the cellulase mixture Celluclast using anion exchange.

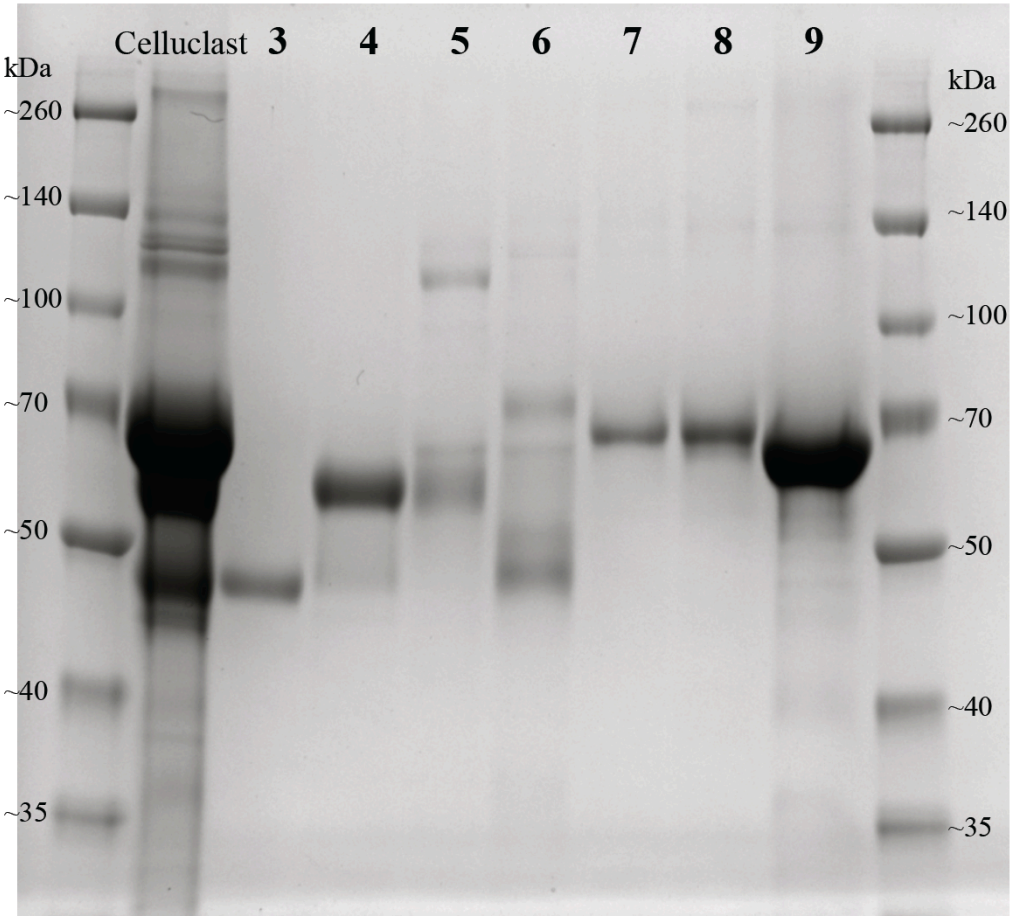


Figure 54. SDS-PAGE corresponding to the anion exchange purification of Celluclast (Figure 53).

4.2.2 Cel5A (Endoglucanase II)

Endoglucanase activity was detected using (3,5-dinitrosalicylate) DNS (Figure 55) to determine which FPLC fractions after anion exchange were of interest (Figure 56).

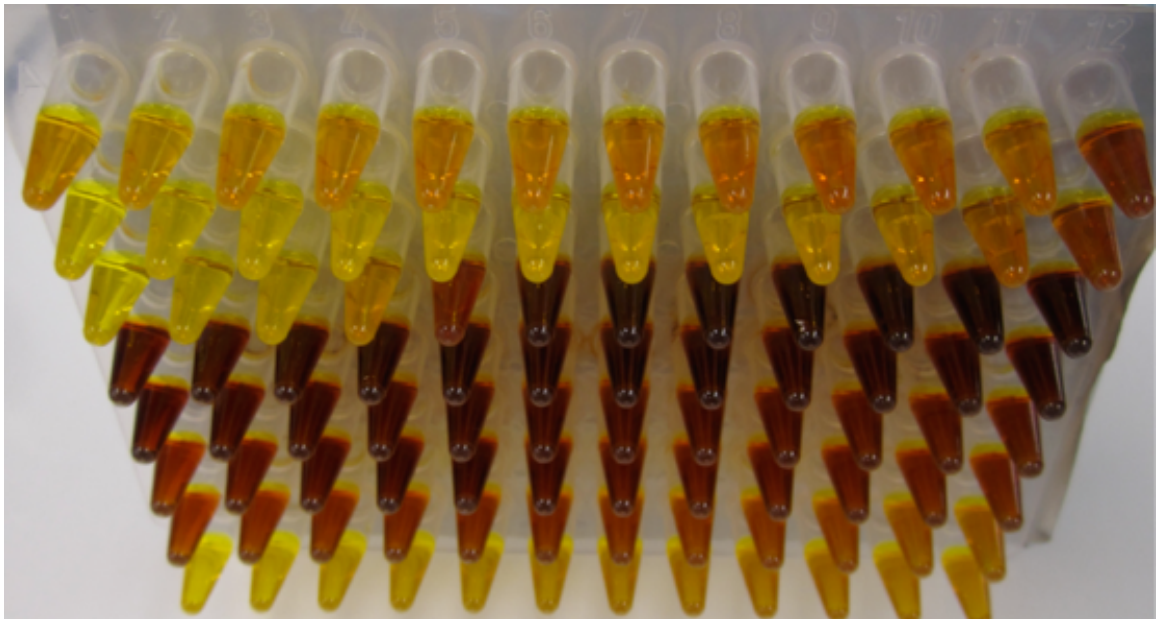


Figure 55. 3,5-dinitrosalicylate (DNS) was used to detect endoglucanases activity of FPLC fractions. Dark wells correspond to activity.

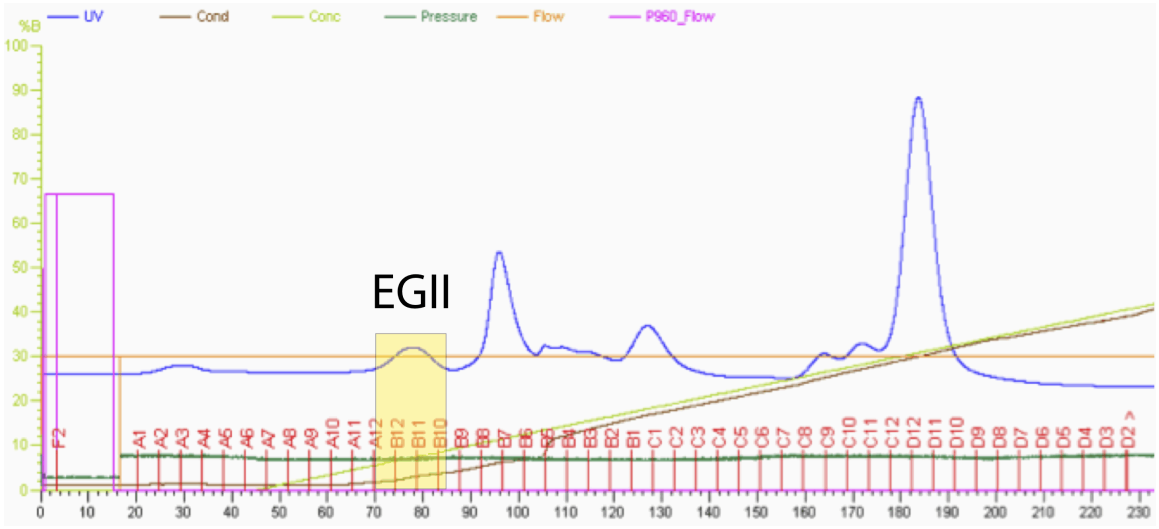


Figure 56. FPLC fractions from Celluclast separation corresponding to Cel5A (EG2) are highlighted in yellow.

The identified fractions were analyzed for purity using SDS-PAGE (Figure 57).

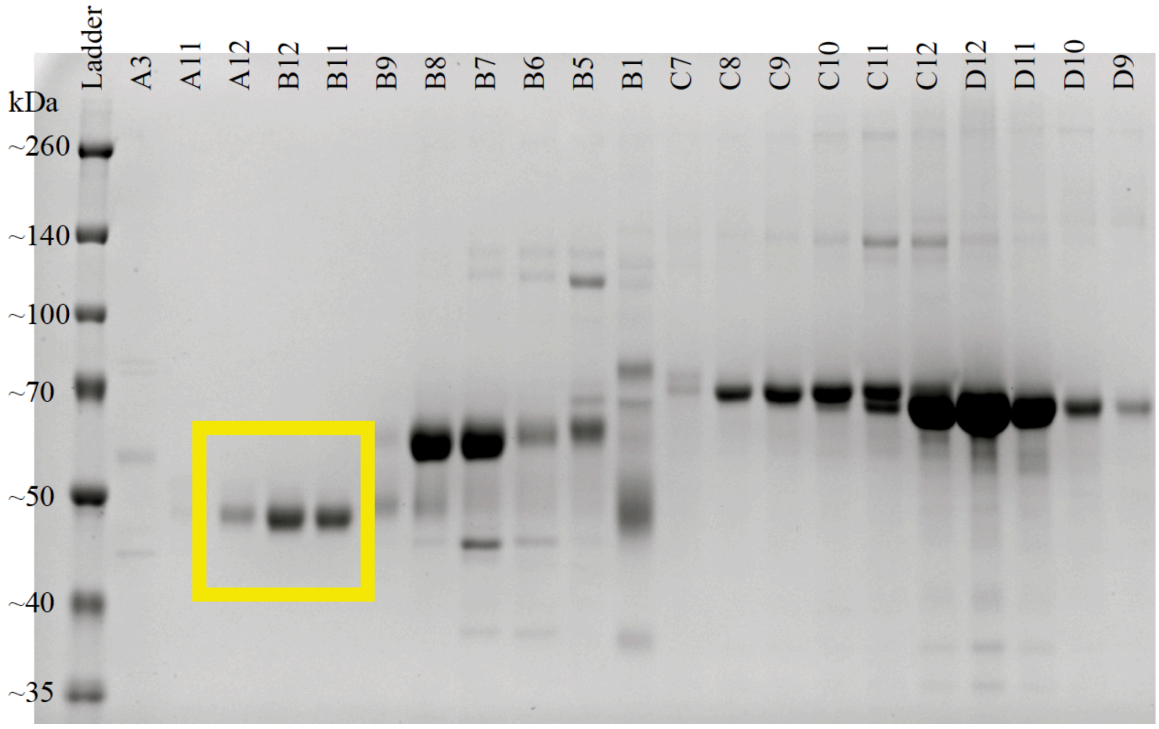
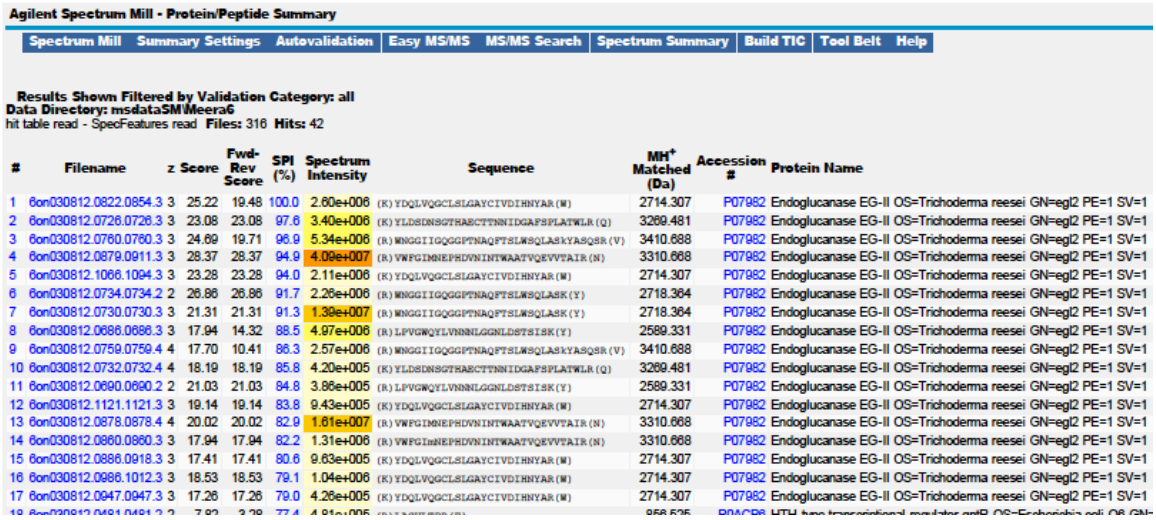


Figure 57. SDS-PAGE analysis highlighting fractions containing Cel5A (EG2).

Pure samples were combined and buffer exchanged into 50mM sodium acetate pH 4.85. Protein concentration was measured via absorbance at A_{280}. Finally, following a trypsin protease digest, proteomic analysis of the sample using LC-MS/MS confirmed the protein identity as Cel5A (Figure 58).



Agilent Spectrum Mill - Protein/Peptide Summary

Spectrum Mill | Summary Settings | Autovalidation | Easy MS/MS | MS/MS Search | Spectrum Summary | Build TIC | Tool Belt | Help

Results Shown Filtered by Validation Category: all
Data Directory: msdataSM\Meera6
hit table read - SpecFeatures read **Files:** 316 **Hits:** 42

#	Filename	z	Score	Fwd-Rev Score	SPI (%)	Spectrum Intensity	Sequence	MH+ Matched (Da)	Accession #	Protein Name
1	6on030812.0822.0854.3	3	25.22	19.48	100.0	2.60e+006	(K)YDQLVQGCLSLGAYCIVDIHNYAR(W)	2714.307	P07982	Endoglucanase EG-II OS=Trichoderma reesei GN=egl2 PE=1 SV=1
2	6on030812.0726.0726.3	3	23.08	23.08	97.6	3.40e+006	(K)YLDSDNSGTHAECTTNNIDGAFSPLATWLR(Q)	3269.481	P07982	Endoglucanase EG-II OS=Trichoderma reesei GN=egl2 PE=1 SV=1
3	6on030812.0760.0760.3	3	24.69	19.71	96.9	5.34e+006	(R)WNGGIIGQGGPTNAQFTSLWSQLASKYASQSR(V)	3410.688	P07982	Endoglucanase EG-II OS=Trichoderma reesei GN=egl2 PE=1 SV=1
4	6on030812.0879.0911.3	3	28.37	28.37	94.9	4.09e+007	(R)VWFGIMNEPHDVNINTWAATVQEVVTAIR(N)	3310.668	P07982	Endoglucanase EG-II OS=Trichoderma reesei GN=egl2 PE=1 SV=1
5	6on030812.1066.1094.3	3	23.28	23.28	94.0	2.11e+006	(K)YDQLVQGCLSLGAYCIVDIHNYAR(W)	2714.307	P07982	Endoglucanase EG-II OS=Trichoderma reesei GN=egl2 PE=1 SV=1
6	6on030812.0734.0734.2	2	26.86	26.86	91.7	2.26e+006	(R)WNGGIIGQGGPTNAQFTSLWSQLASK(Y)	2718.364	P07982	Endoglucanase EG-II OS=Trichoderma reesei GN=egl2 PE=1 SV=1
7	6on030812.0730.0730.3	3	21.31	21.31	91.3	1.39e+007	(R)WNGGIIGQGGPTNAQFTSLWSQLASK(Y)	2718.364	P07982	Endoglucanase EG-II OS=Trichoderma reesei GN=egl2 PE=1 SV=1
8	6on030812.0686.0686.3	3	17.94	14.32	88.5	4.97e+006	(R)LPVGWQYLVNNNLGGNLDSTSISK(Y)	2589.331	P07982	Endoglucanase EG-II OS=Trichoderma reesei GN=egl2 PE=1 SV=1
9	6on030812.0759.0759.4	4	17.70	10.41	86.3	2.57e+006	(R)WNGGIIGQGGPTNAQFTSLWSQLASKYASQSR(V)	3410.688	P07982	Endoglucanase EG-II OS=Trichoderma reesei GN=egl2 PE=1 SV=1
10	6on030812.0732.0732.4	4	18.19	18.19	85.8	4.20e+005	(K)YLDSDNSGTHAECTTNNIDGAFSPLATWLR(Q)	3269.481	P07982	Endoglucanase EG-II OS=Trichoderma reesei GN=egl2 PE=1 SV=1
11	6on030812.0690.0690.2	2	21.03	21.03	84.8	3.86e+005	(R)LPVGWQYLVNNNLGGNLDSTSISK(Y)	2589.331	P07982	Endoglucanase EG-II OS=Trichoderma reesei GN=egl2 PE=1 SV=1
12	6on030812.1121.1121.3	3	19.14	19.14	83.8	9.43e+005	(K)YDQLVQGCLSLGAYCIVDIHNYAR(W)	2714.307	P07982	Endoglucanase EG-II OS=Trichoderma reesei GN=egl2 PE=1 SV=1
13	6on030812.0878.0878.4	4	20.02	20.02	82.9	1.61e+007	(R)VWFGIMNEPHDVNINTWAATVQEVVTAIR(N)	3310.668	P07982	Endoglucanase EG-II OS=Trichoderma reesei GN=egl2 PE=1 SV=1
14	6on030812.0860.0860.3	3	17.94	17.94	82.2	1.31e+006	(R)VWFGIMNEPHDVNINTWAATVQEVVTAIR(N)	3310.668	P07982	Endoglucanase EG-II OS=Trichoderma reesei GN=egl2 PE=1 SV=1
15	6on030812.0886.0918.3	3	17.41	17.41	80.6	9.63e+005	(K)YDQLVQGCLSLGAYCIVDIHNYAR(W)	2714.307	P07982	Endoglucanase EG-II OS=Trichoderma reesei GN=egl2 PE=1 SV=1
16	6on030812.0986.1012.3	3	18.53	18.53	79.1	1.04e+006	(K)YDQLVQGCLSLGAYCIVDIHNYAR(W)	2714.307	P07982	Endoglucanase EG-II OS=Trichoderma reesei GN=egl2 PE=1 SV=1
17	6on030812.0947.0947.3	3	17.26	17.26	79.0	4.26e+005	(K)YDQLVQGCLSLGAYCIVDIHNYAR(W)	2714.307	P07982	Endoglucanase EG-II OS=Trichoderma reesei GN=egl2 PE=1 SV=1

Figure 58. LC-MS/MS peptide sequencing confirmed protein identity as Cel5A (endoglucanase II).

To further confirm purity and identification, I performed full-length protein analysis using LC-MS of the intact protein (Figure 59). This cellulase was pure to the detection limit and exhibited a distribution of glycosylation (Figure 60).

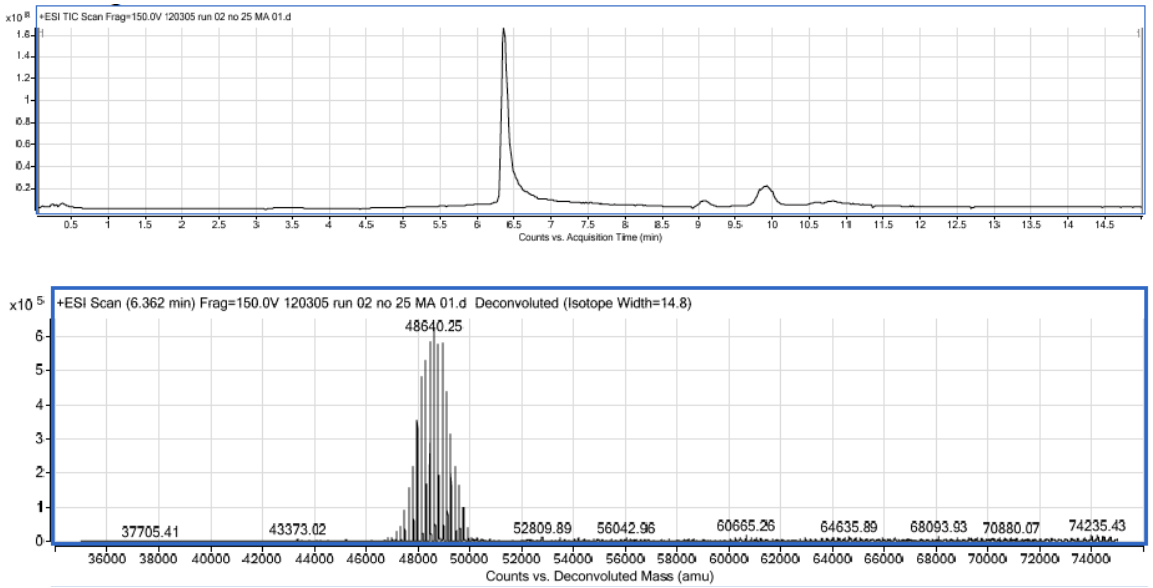


Figure 59. LC-MS intact protein analysis of Cel5A sample indicates pure protein with a mass of 48.64 kDa.

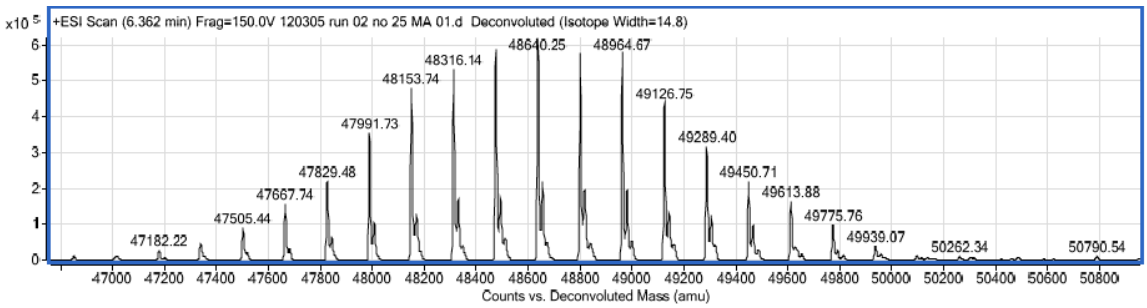


Figure 60. Zoom-in of intact protein LC-MS peak with equal spacing of masses showing the distribution of glycosylation.

After generating approximately 3mL of pure, 30µM *T. reesei* Cel5A (EG2) protein, I began characterization experiments and measured the enzyme's pH optimum (Figure 61).

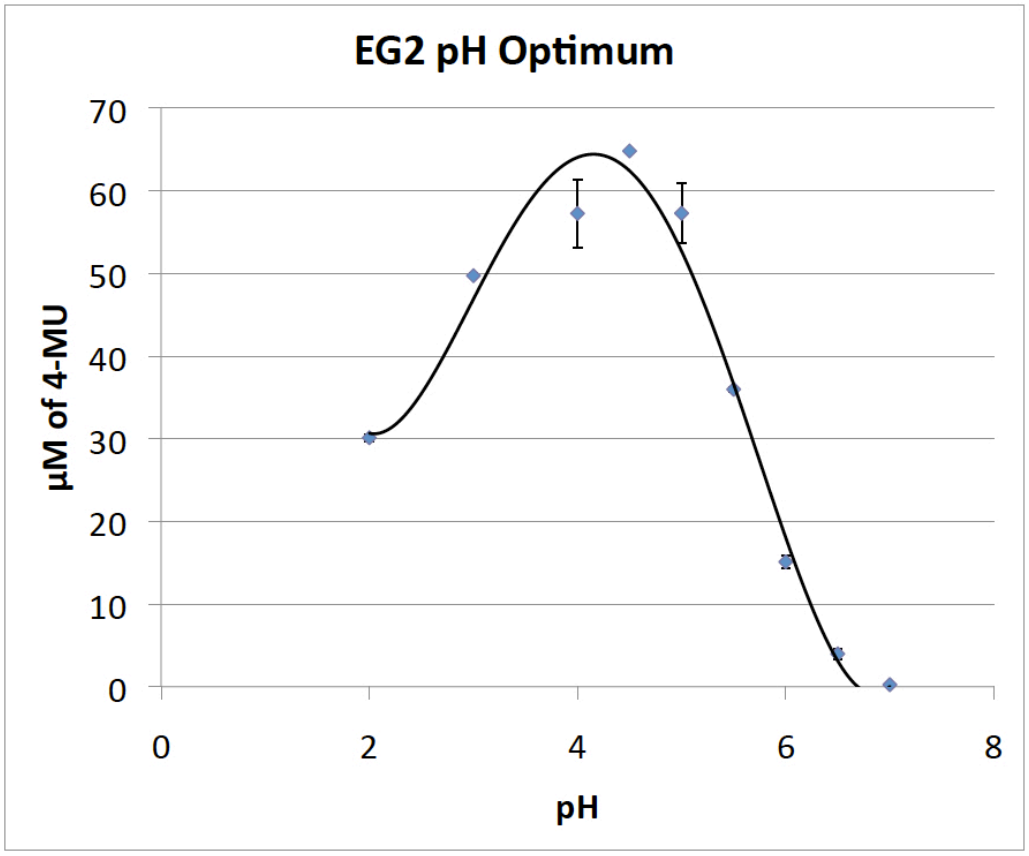


Figure 61. Measured pH optimum of Cel5A (EG2). Reactions were performed in duplicate at 40µL volumes containing 0.15µM Cel5A enzyme, 1mM 4-MU-Cellobioside substrate, and 100mM Britton-Robinson Buffer (33.33mM H_3BO_3, 33.33mM H_3PO_4, 33.33mM Acetic Acid), incubated for 10 minutes at 50°C.

4.2.3 Cel6A (Cellobiohydrolase II)

I used a similar protocol to purify Cel6A (cellobiohydrolase II).

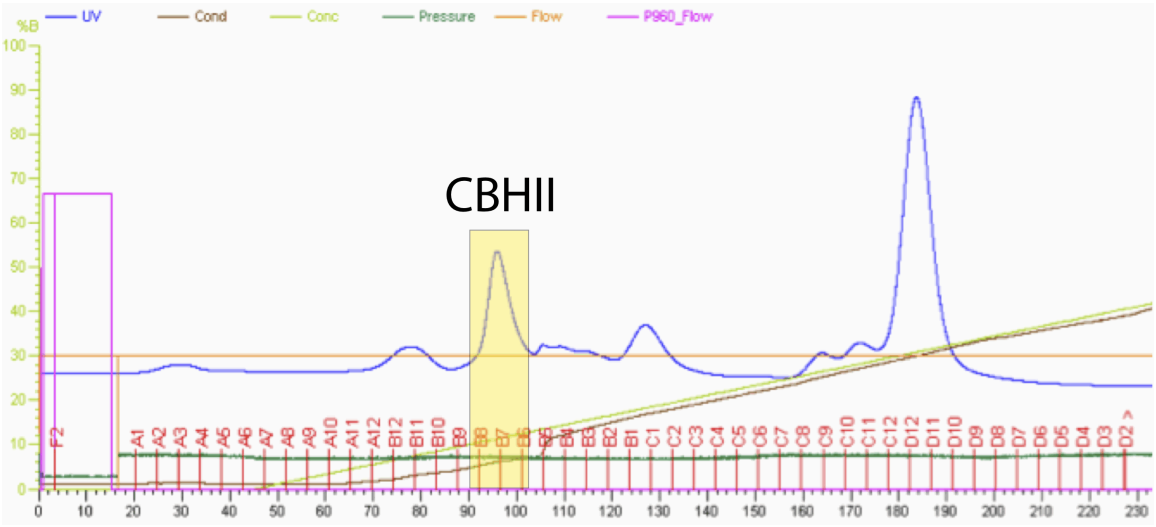


Figure 62. FPLC fractions from Celluclast separation corresponding to Cel6A (CBH2) are highlighted in yellow.

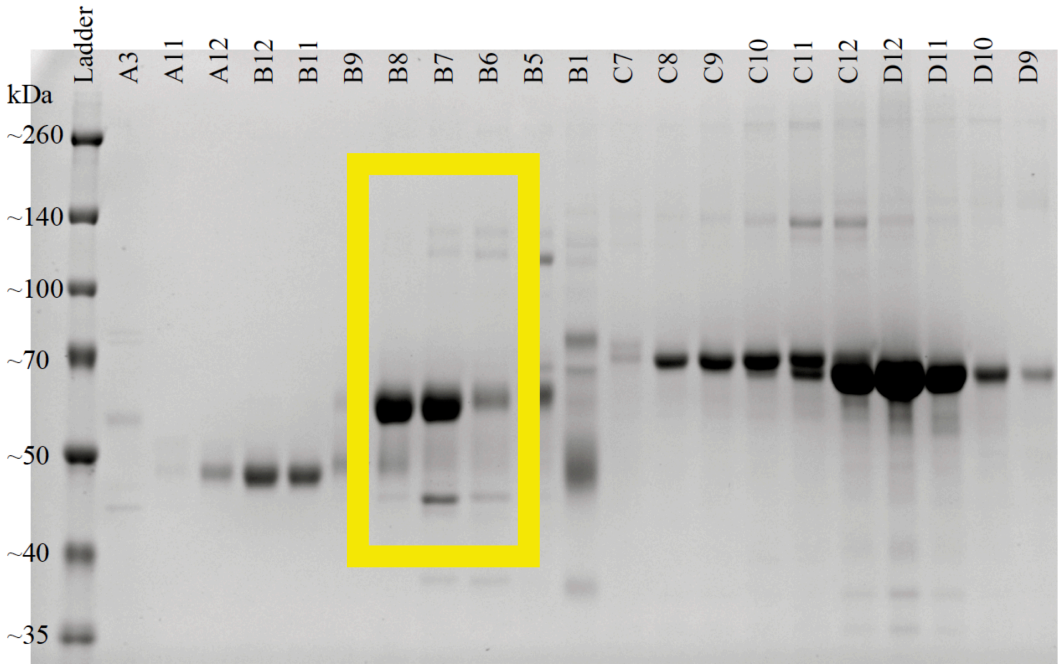


Figure 63. SDS-PAGE analysis highlighting fractions containing Cel6A (CBH2).

Additional purification was necessary to separate Cel6A from Cel5A, for which I used another anion exchange step over a MonoQ column (Figure 64).

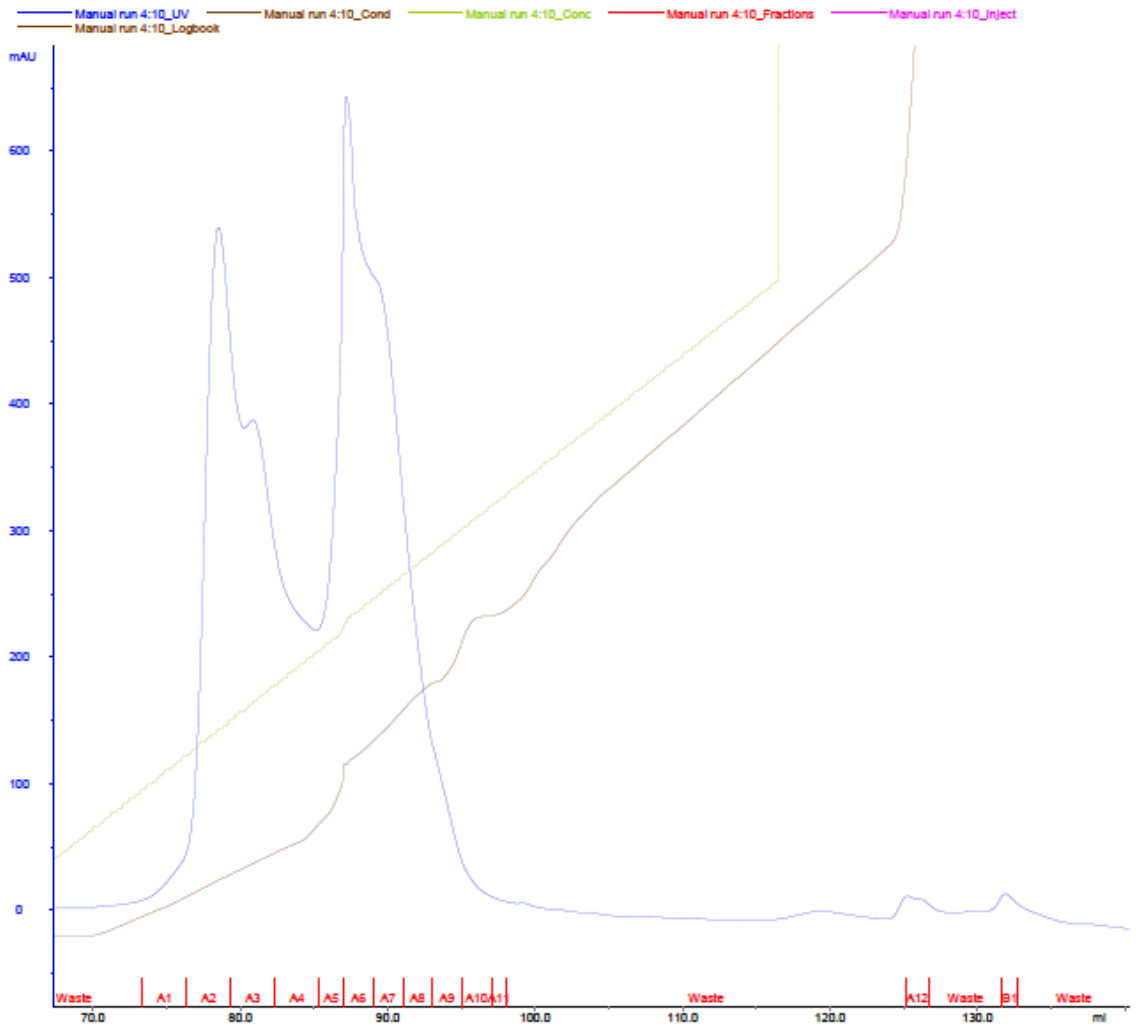


Figure 64. MonoQ purification separated Cel6A (left peak) from contamination with Cel5A (right peak).

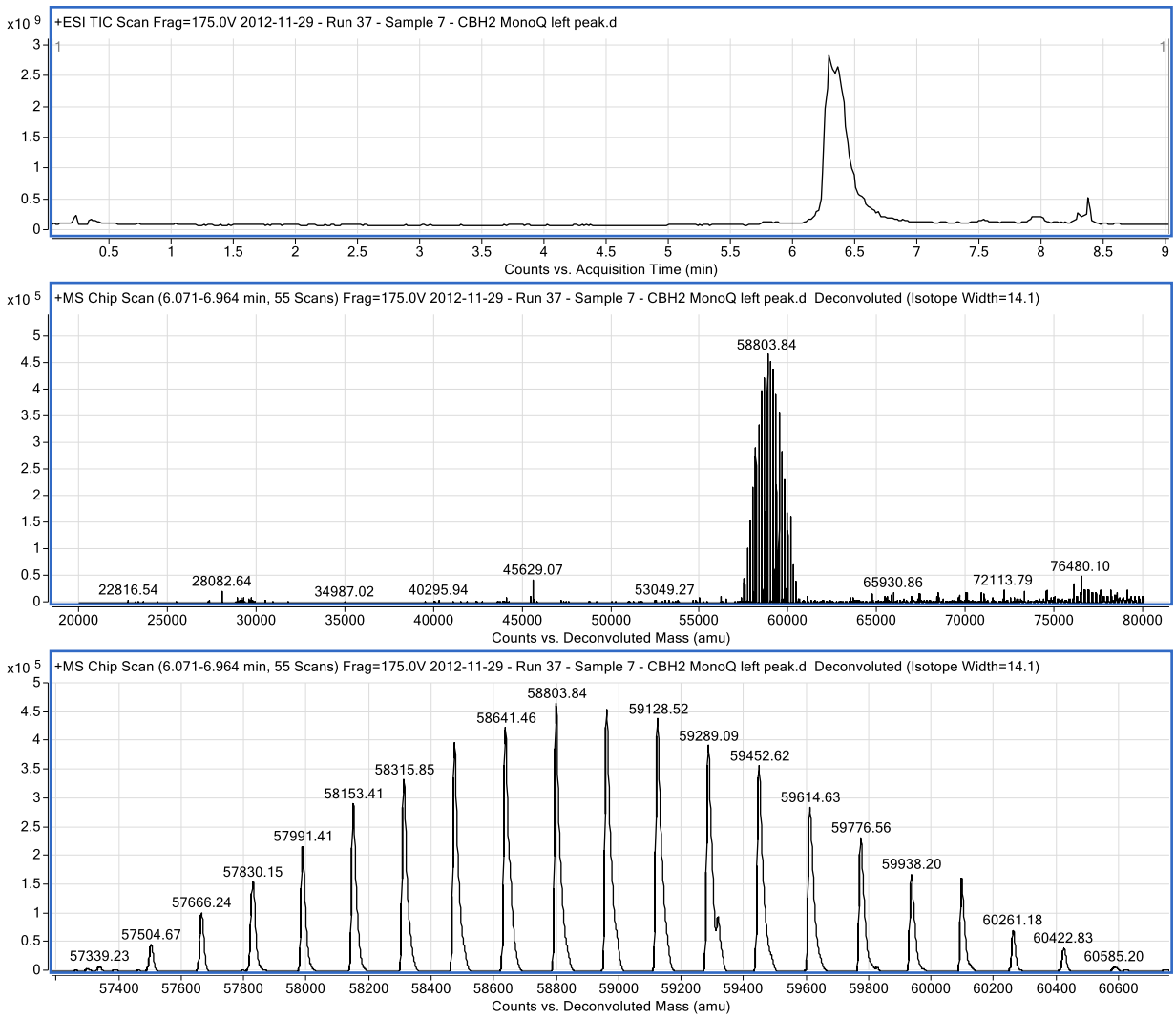


Figure 65. Intact protein LC-MS analysis of Cel6A (CBH2) shows enzyme purity and glycosylation patterns.

4.2.4 Cel7A (Cellobiohydrolase I)

Cel7A is, by far, the most abundant cellulase in Celluclast. This protein binds strongly to the anion exchange column and elutes last (Figure 66).

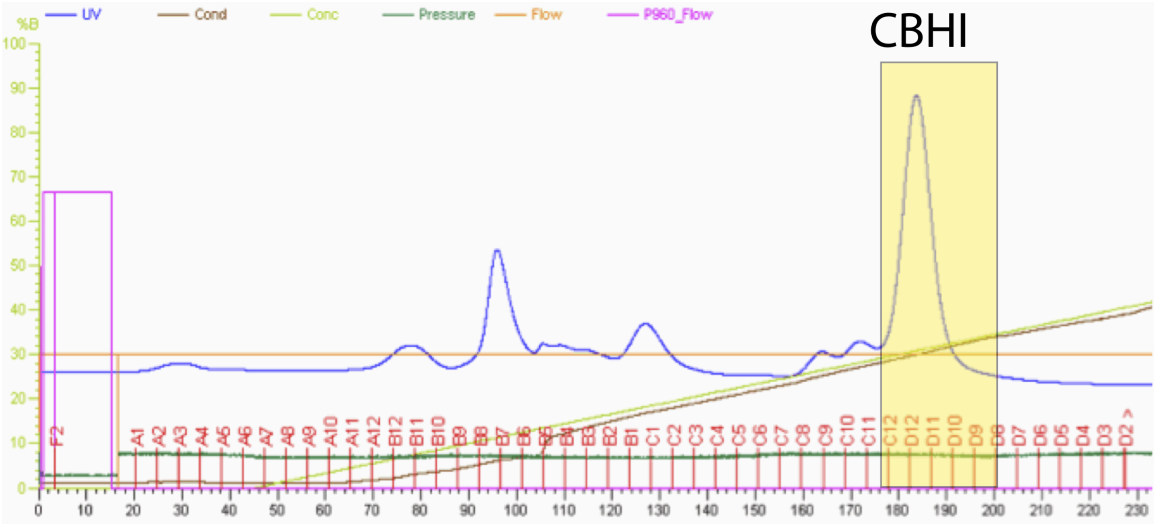


Figure 66. FPLC fractions from Celluclast separation corresponding to Cel7A (CBH1) are highlighted in yellow.

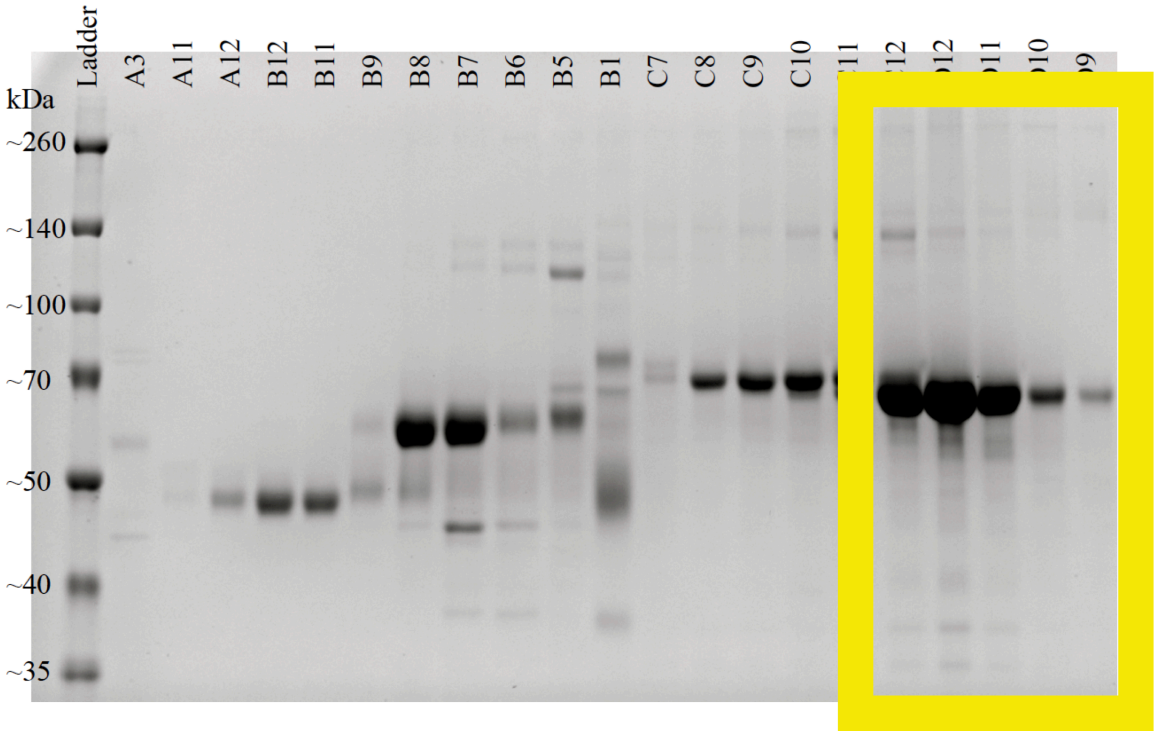


Figure 67. SDS-PAGE analysis highlighting fractions containing Cel7A (CBH1).

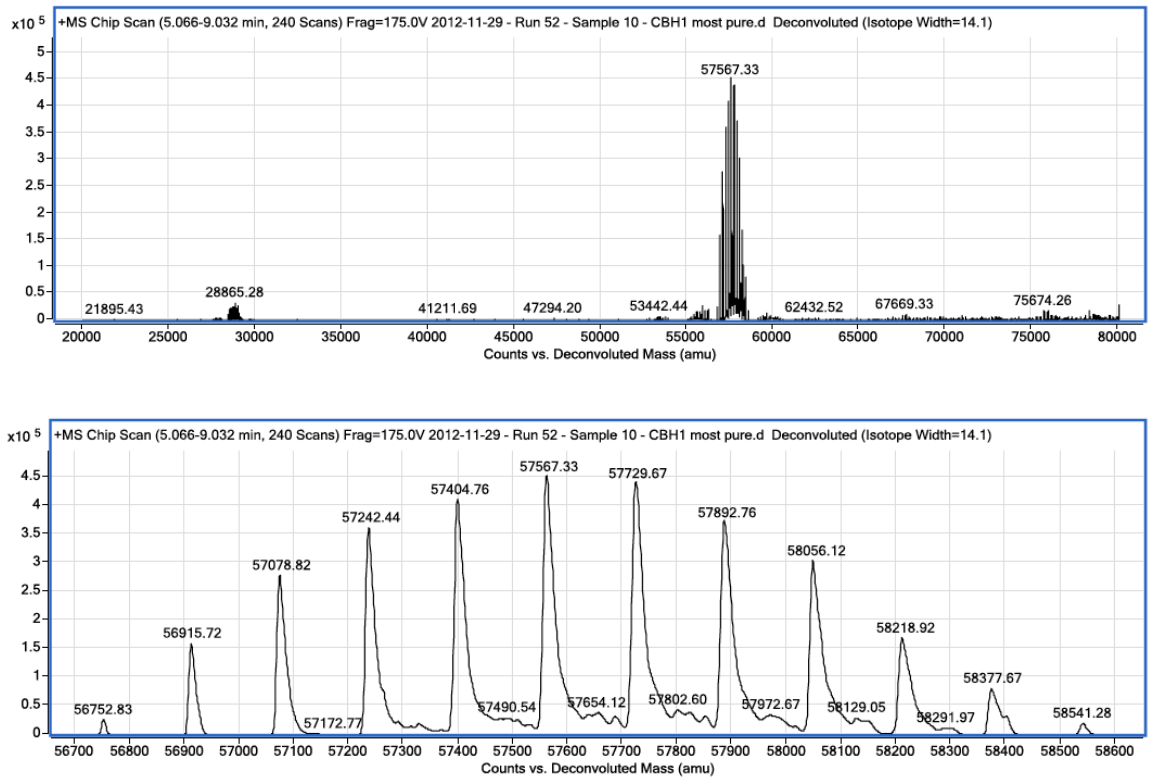


Figure 68. Intact protein LC-MS analysis of Cel7A (CBH1) shows enzyme purity and glycosylation patterns.

4.3 Conclusions

My work was successful in developing and disseminating protocols for purifying and identifying individual cellulase enzyme components of Celluclast. Despite significant effort, I was never able to purify Cel7B (EG1). This is not surprising, however, as most studies in the literature reported similar luck. I did not have a chance to perform the synergy experiments I had originally planned, but other researchers found these protocols and purified enzymes useful for their experiments.

Chapter 5

Efforts Towards Enabling High-Throughput Directed Evolution of Cellulase Enzymes via *In Vitro* Compartmentalization

5.1 Abstract

Global climate change from greenhouse gas emissions is driving the world towards sustainable energy. The transportation sector accounts for 14% of global energy-related emissions[1] and is a potential target for large-scale emission reductions. Biofuels are liquid transportation fuels that are renewable, compatible with existing fuel infrastructure, and can operate with minimal net emissions. Cellulosic biofuels are one of the most promising types due to the high abundance of cellulose, a structural component of plant cell walls.

Cellulosic biofuels face several barriers to commercialization. To produce them, one must use high volumes of expensive cellulase enzymes[56] to catalyze the depolymerization of cellulose into glucose, which is subsequently converted by microorganisms to fuel molecules of interest. By developing cellulase enzymes that have greater thermostability and specific activity on biomass substrates, while maintaining inter-enzyme synergy, biomass can be processed using fewer enzymes. This reduction of enzyme loading would result in biofuels that are more cost-competitive with fossil fuels.

Cellulosic biomass is solid and insoluble; however, most cellulase evolution methods substitute lignocellulosic biomass with soluble substrate analogs to simplify the screening process. Unfortunately, enzymatic activity on soluble substrates has little correlation to activity on natural cellulosic substrates.[57] Despite this, there exist few methods to evolve cellulases using solid substrates, all of which are limited in throughput to multi-well plates at best.[58,59]

Directed evolution can be used to improve enzyme properties, such as activity and stability. I aimed to develop a method for high-throughput directed evolution of cellulase enzymes that was amenable to solid substrates. Genotype-phenotype linkage, crucial for any directed evolution experiment, would be achieved using surfactant-stabilized water-in-oil emulsion droplets that serve as artificial cell-like compartments.[60] One milliliter of emulsion contains ~10^{10} individual droplets that would each act as a reaction vessel. This method, *in vitro* compartmentalization (IVC), has been successfully used for enzyme evolution and allows for screening of multiple enzyme turnovers.[61] When enzymatic activity is coupled to fluorescence, fluorescence-activated cell sorting (FACS) can separate compartments containing active enzymes from those with inactive enzymes. FACS can collect droplets (with fluorescence greater than a set threshold) at a rate of over 10^7 droplets per hour.[62] Together, these methods could create a powerful and efficient method for evolving cellulase enzymes.

5.2 Directed Evolution

Directed evolution is a powerful tool for generating enzymes with desirable properties. Genes encoding enzymes are modified (using a variety of methods from error-prone PCR to gene shuffling) and "sequence space" explored as genetic diversity is introduced. While most mutations are deleterious, as in nature, a small fraction will confer traits of interest. With each subsequent round of evolution, mutants with desirable properties are selected and chosen for subsequent mutagenesis. Over many rounds, directed evolution enables the creation of an enzyme with increased fitness (activity), as shown in Figure 69.

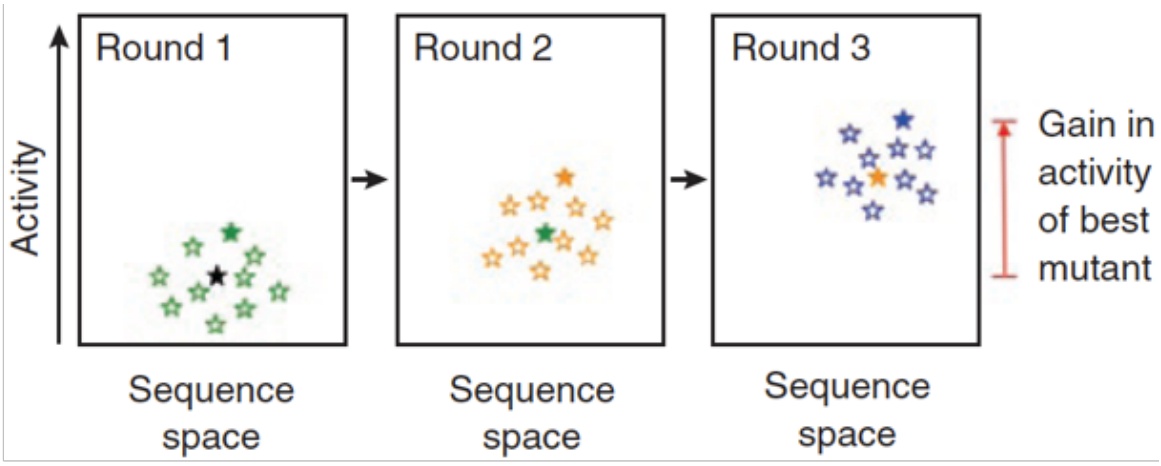


Figure 69. Directed evolution enables the generation of mutant enzymes with increased activity.[63]

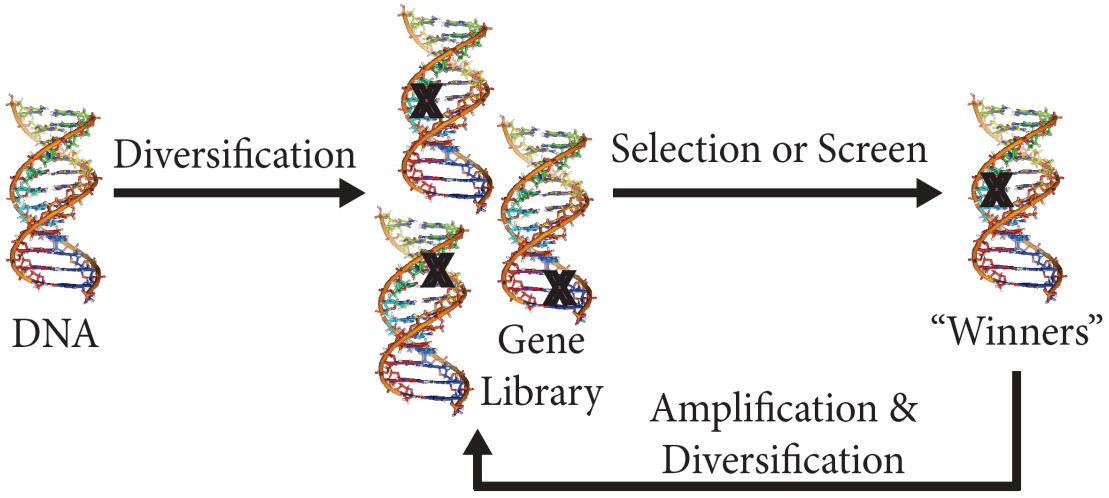


Figure 70. Directed evolution scheme.

A key component of directed evolution experiments is that they must somehow link genotype to phenotype. This linkage could be achieved by producing and analyzing enzymatic activity *in vivo* within a cellular compartment, or *in vitro* by assessing catalytic activity in a test tube following enzyme production via cells or cell-free methods.

5.3 *In Vitro* Compartmentalization

In vitro Compartmentalization (IVC) describes surfactant-stabilized emulsions, such as water-in-oil droplets, that mimic cellular compartments (Figure 71). These compartments, like cells, can physically link genotype to phenotype by containing both a gene and its encoded protein within a droplet. The main advantage of IVC is the ability to produce up to 10^{10} individual compartments in just one milliliter of solution—a significant departure from 96- or 384-well high-throughput assay plates.

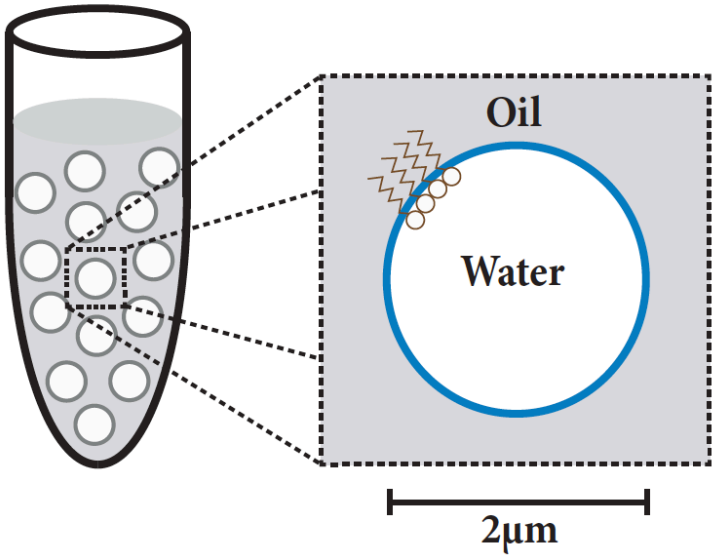


Figure 71. *In vitro* compartmentalization involves oil-in-water droplets that act as miniature reaction compartments.

5.4 Directed Evolution of Cellulases

Standard methods for cellulase enzyme evolution (using relevant substrates) involve either producing cellulases *in vivo* or *in vitro* (via cell-free protein synthesis) and screening for activity on solid cellulose substrates in reaction tubes or multi-well plates (Figure 72). The current scheme enables approximately 10^3-10^4 variants to be screened per day. My goal was to develop a method for high-throughput screening of cellulase

activity on solid cellulose substrates. I envisioned using directed evolution to evolve cellulases for improved thermostability (at temperatures around 70˚C) and specific activity on solid substrates (both individually and in the presence of other cellulases).

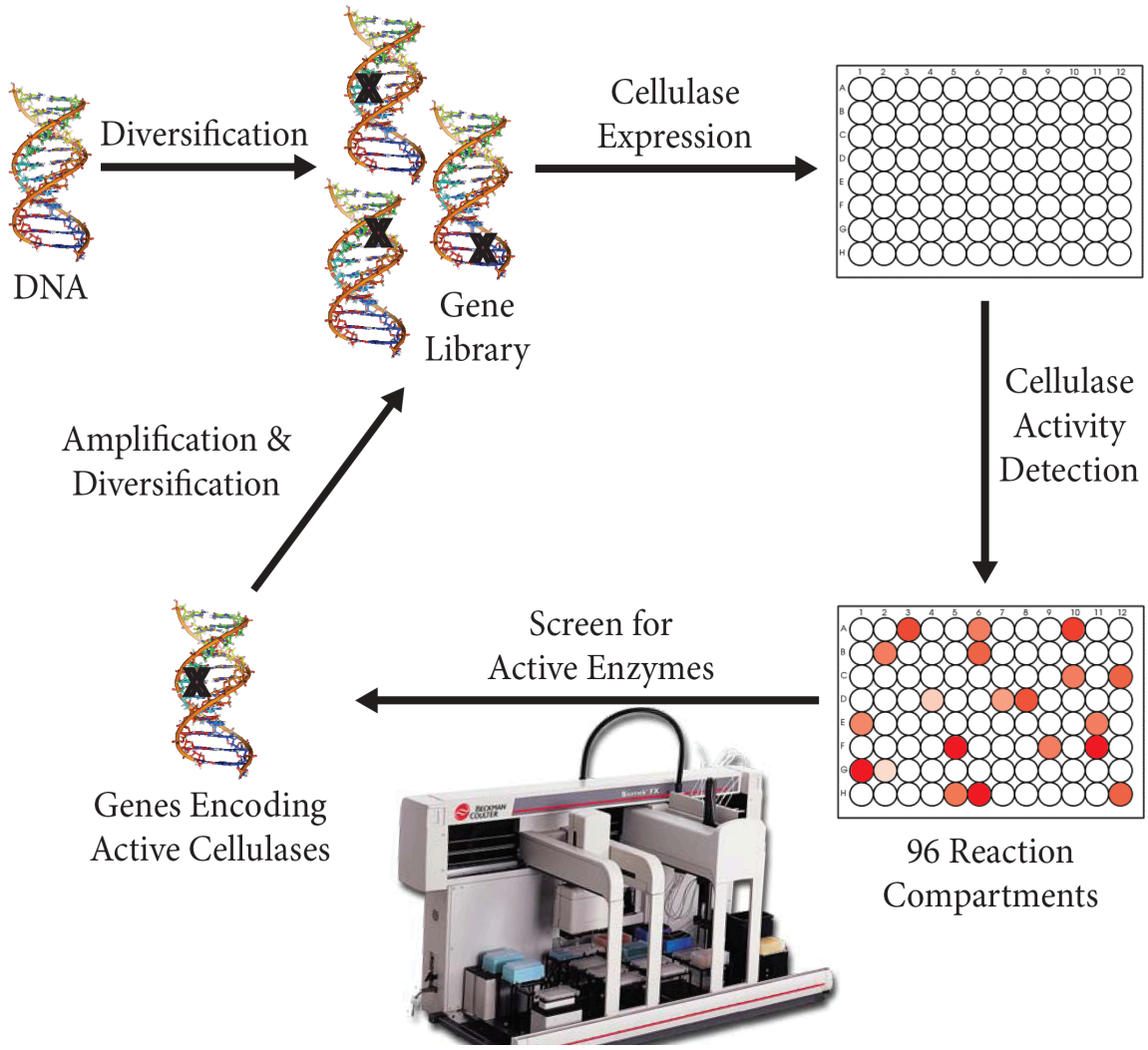


Figure 72. Current scheme for directed evolution of cellulases on solid substrates using multi-well plates.

Due to the nature of working with solid substrates, current directed evolution techniques rely on brute force methods of low to moderate throughput, requiring significant researcher (or robotic) time to screen library members. As current technology limits throughput to 10^3-10^4 variants screened per day and directed evolution benefits from multiple rounds of selection, a high-throughput method for screening cellulases is needed. I aimed to link genotype to phenotype by encapsulating the enzyme system in emulsion droplets using *in vitro* compartmentalization (IVC). My proposed scheme

(Figure 73) could enable screening of approximately 10^8 variants per day—a 10,000-100,000-fold increase in screening efficiency.

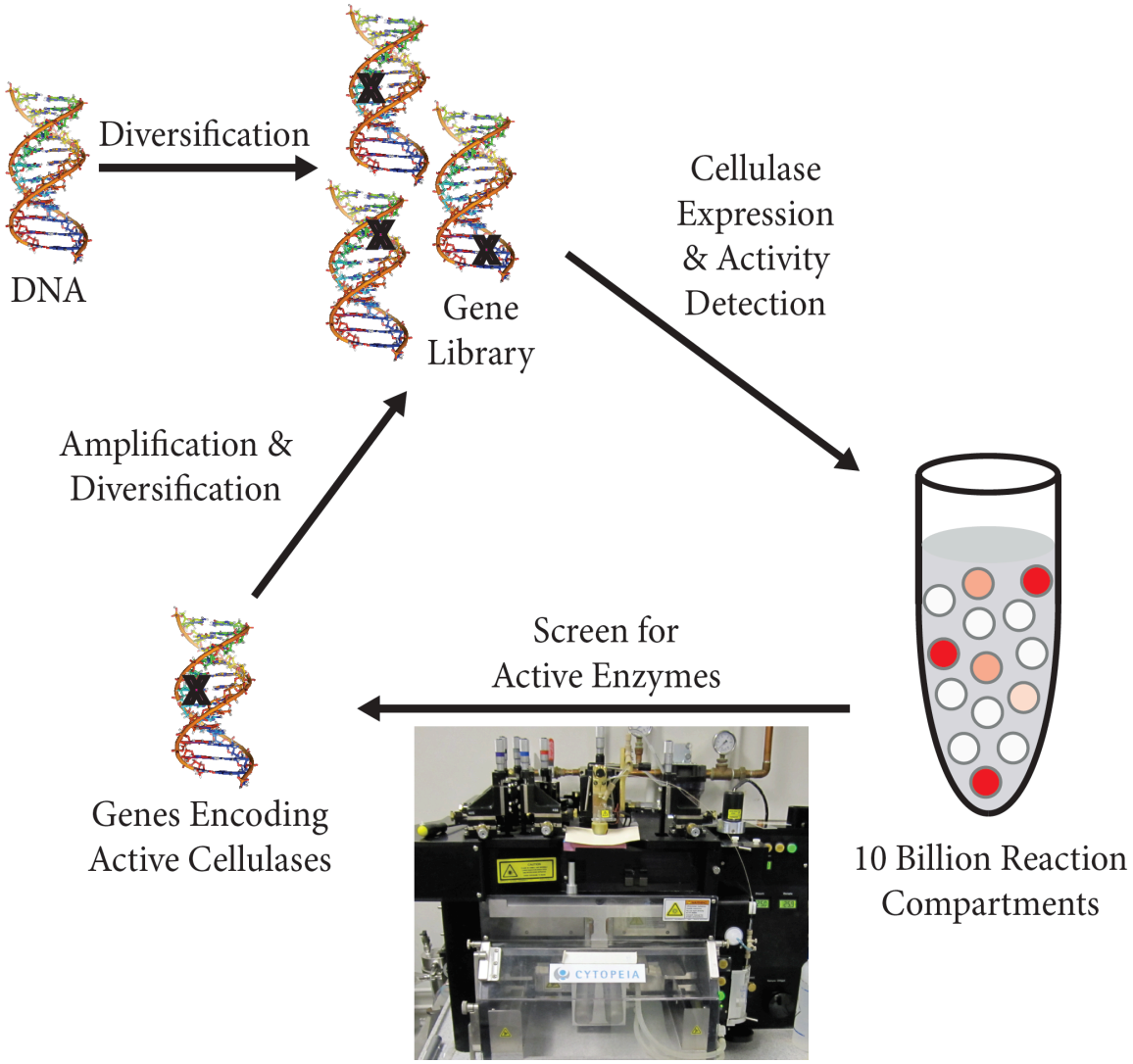


Figure 73. Proposed scheme for directed evolution of cellulases on solid substrates using *in vitro* compartmentalization (IVC).

5.5 Detection of Cellulase Enzyme Activity

Amplex Red is a colorless reagent that can be used to detect glucose (Figure 74), the product of cellulase hydrolysis of cellulose. The molecule reacts with hydrogen peroxide via a peroxidase to produce resorufin, which is highly fluorescent. This system is very

sensitive (it can detect as little as 10 picomoles of hydrogen peroxide) and can produce a fluorescent signal corresponding to the activity of glucose-producing cellulase enzymes.

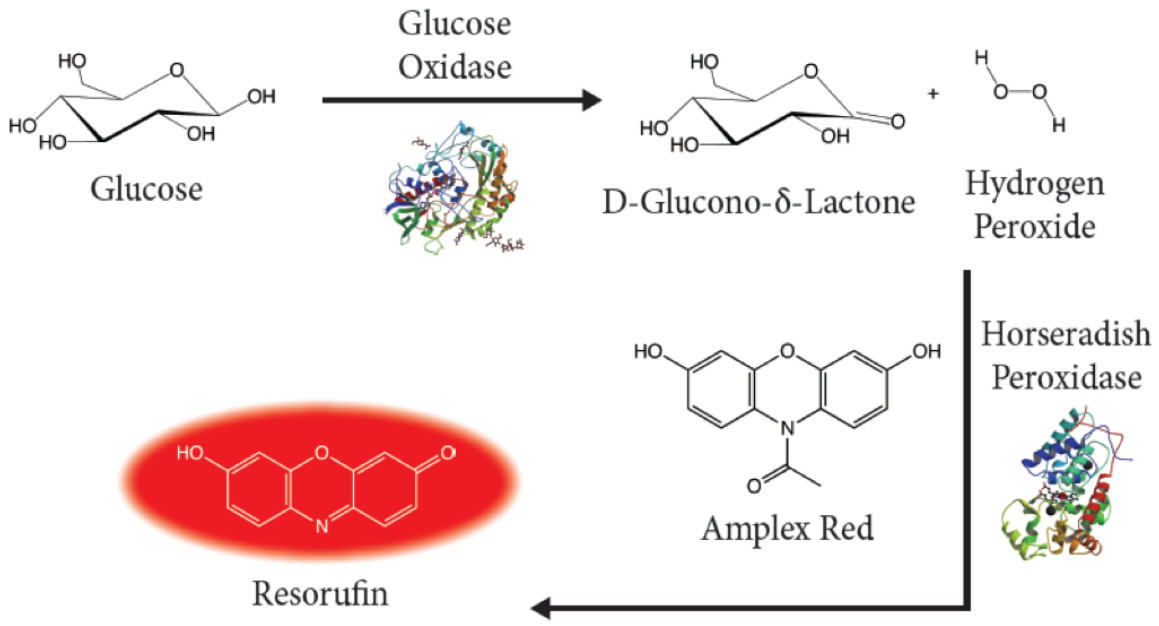


Figure 74. Glucose production by cellulases can be monitored by fluorescence using an Amplex Red assay.

5.6 Fluorescence-Activated Cell Sorting (FACS)

Fluorescence-activated cell sorting (FACS) is a powerful method that can be used when enzymatic activity is coupled to fluorescence. FACS is a high-throughput screening method using flow cytometry. A laser beam of a single wavelength is focused on a stream of fluid and the instrument measures fluorescence quickly and quantitatively. Detectors physically separate the cells or droplets from a heterogeneous mixture depending on a set fluorescence threshold (Figure 75).

Because active cellulase enzymes can be detected by fluorescence (via the Amplex Red assay, Figure 74), I proposed to screen enzyme library members for cellulose hydrolysis activity via FACS. This method could, in theory, sort a cellulase library of up to 10^8 variants in one day! IVC emulsion droplets have been shown to be amenable to high-throughput screening via FACS after single water-in-oil emulsions are re-emulsified to form water-in-oil-in-water double emulsions. Emulsion compartments can be made stable enough to withstand FACS, as well as temperatures up to 95˚C.[64,65] Various methods of droplet delivery also make it possible to modify droplet contents of formed emulsions.[66] Such a system would provide a completely *in vitro* high-throughput method for cellulase enzyme evolution without requiring modified substrates or physical DNA-

protein attachments. IVC methods have not yet been used to evolve cellulases or enzymes with activity towards solid substrates.

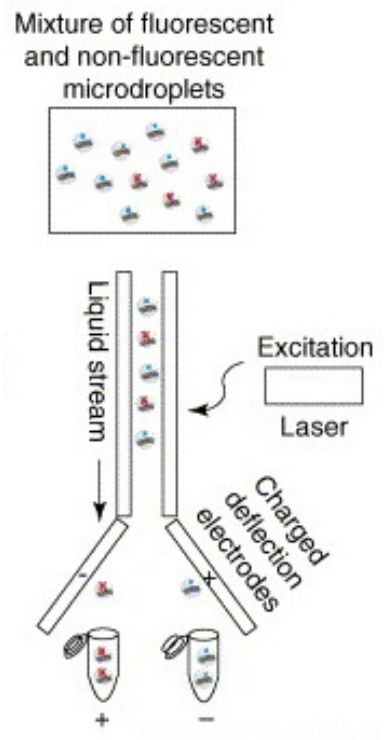


Figure 75. Fluorescence-activated cell sorting (FACS).[67]

5.7 Progress Towards Directed Evolution of Cellulases via IVC

In my proposed directed evolution method, each IVC droplet compartment would contain, on average, a single cellulase gene, materials for cell-free protein expression, a solid cellulose particle, and reagents for fluorogenic detection of enzymatic activity (via glucose production). Figure 76 illustrates my methodology.

The DNA encoding a cellulase enzyme would be mutated, creating a gene library. These genes would be distributed into emulsion compartments such that each compartment contains a single gene. This DNA would be translated within the droplet, using cell-free protein synthesis, producing the cellulase enzyme variant. After reagent delivery and double emulsion formation, active cellulases (and their encoding DNA) within fluorescent droplets would be screened and separated using FACS. The recovered genes, corresponding to cellulases with properties of interest (e.g. activity at elevated temperatures), would be used as templates for subsequent rounds of evolution until the target goal has been met (Figure 69).

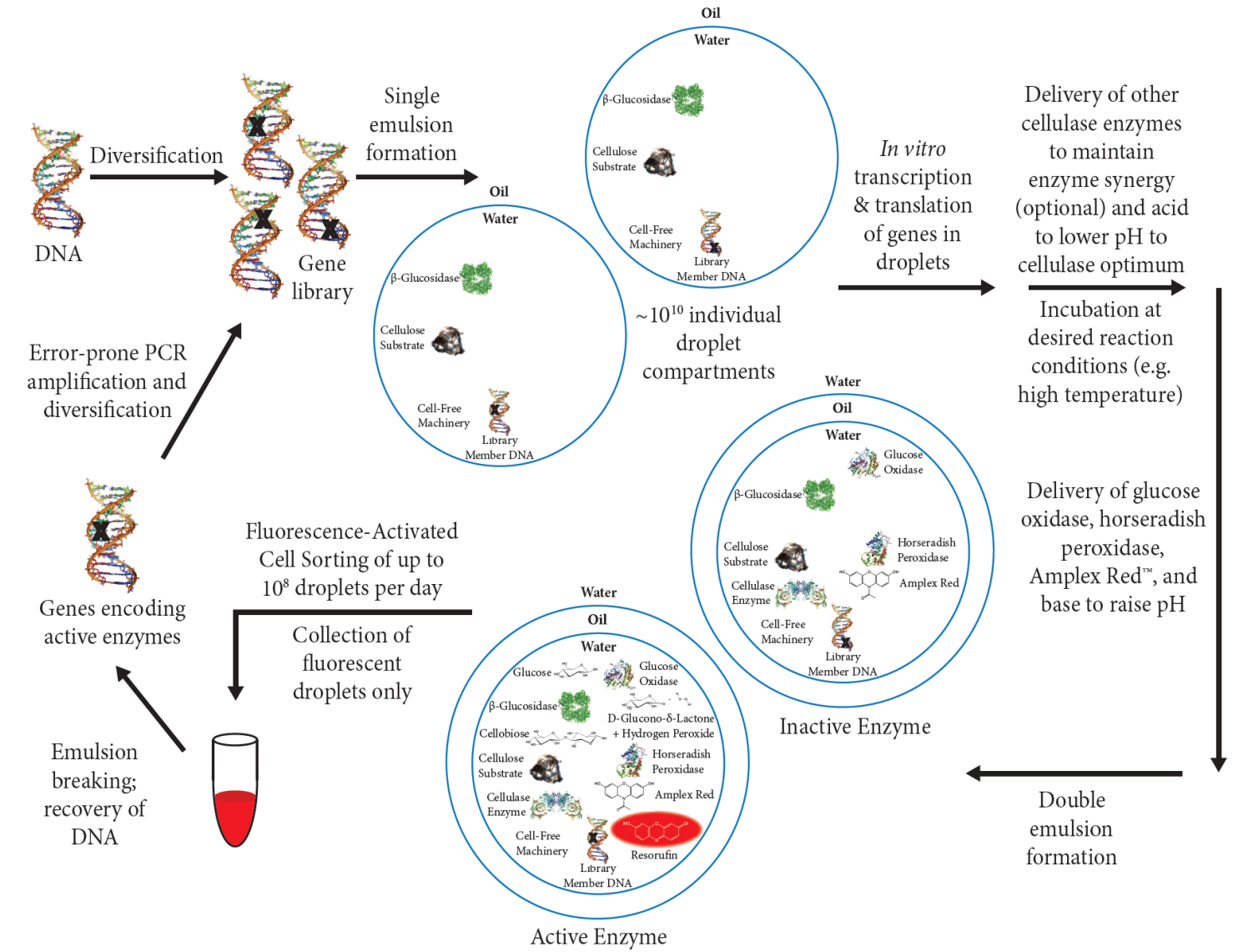


Figure 76. Cellulase evolution strategy using *in vitro* compartmentalization (IVC).

Using IVC and FACS, I hoped to evolve cellulases on natural substrates with unprecedented speed. As demonstrated in the figures below, I successfully formed single and double emulsions, encapsulated solid cellulose particles, demonstrated cell-free cellulase synthesis in emulsions, delivered reagents, and detected cellulase activity within double emulsions via FACS. To achieve uniform double emulsion droplet size, I also explored microfluidics as a means of homogenous emulsion formation.

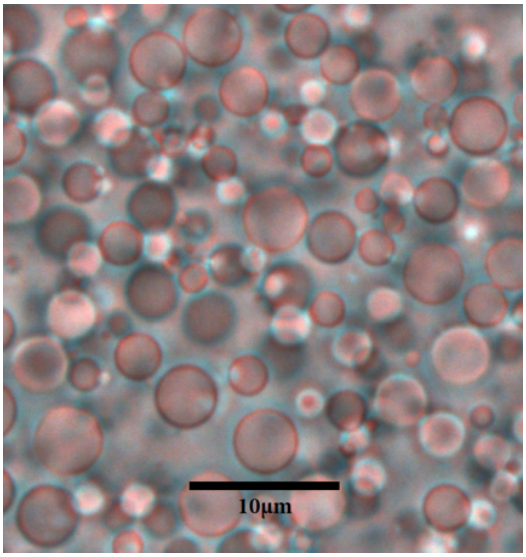


Figure 77. A single emulsion was created by homogenization of 50µL of an aqueous phase with 950µL of light mineral oil containing 5% ABIL EM90 surfactant at 8000rpm for 5 minutes.

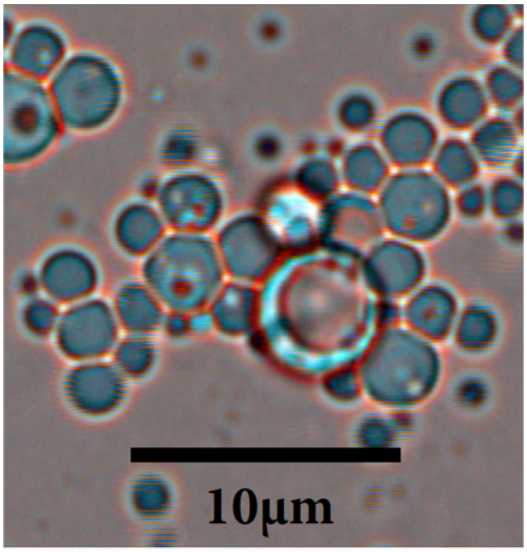


Figure 78. A double emulsion (necessary for FACS sorting) was formed by homogenizing a single emulsion with a surfactant and an equal volume of a second aqueous phase.

Figure 79. Successful cellulose particle encapsulation within an emulsion droplet. Solid particles had not been encapsulated with IVC emulsions previously.

Cellulase enzyme production in emulsions was performed by expressing *Trichoderma reesei* endoglucanase 1 (EG1, Cel7B) using cell-free protein synthesis, after which the emulsion was broken and the aqueous fraction (containing the cellulase) collected. Enzymatic activity was compared to activity of EG1 produced outside of an emulsion with the same cell-free techniques; the cellulase produced within the emulsion was active (Figure 80).

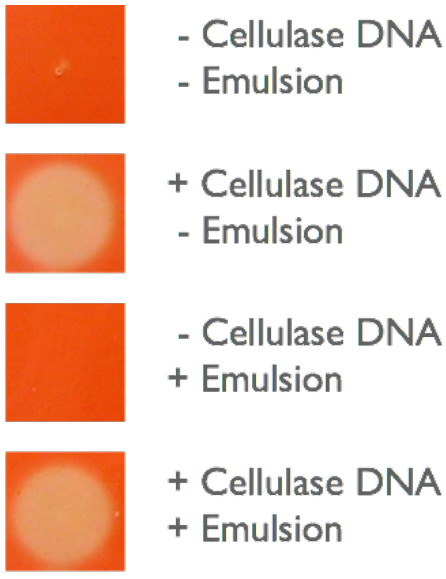


Figure 80. Successful cell-free cellulase synthesis in emulsions. A halo indicates an active cellulase enzyme.

I was also able to demonstrate successful reagent delivery to emulsion compartments (Figure 81) and detection of droplets containing cellulase enzymes using a fluorogenic substrate and FACS (Figure 82).

Figure 81. Reagent delivery to droplets appeared successful. Amplex Red in DMSO was added to pre-formed single emulsions and stirred for 30 sec at 1100rpm. Amplex Red remained in the aqueous phase, as desired.

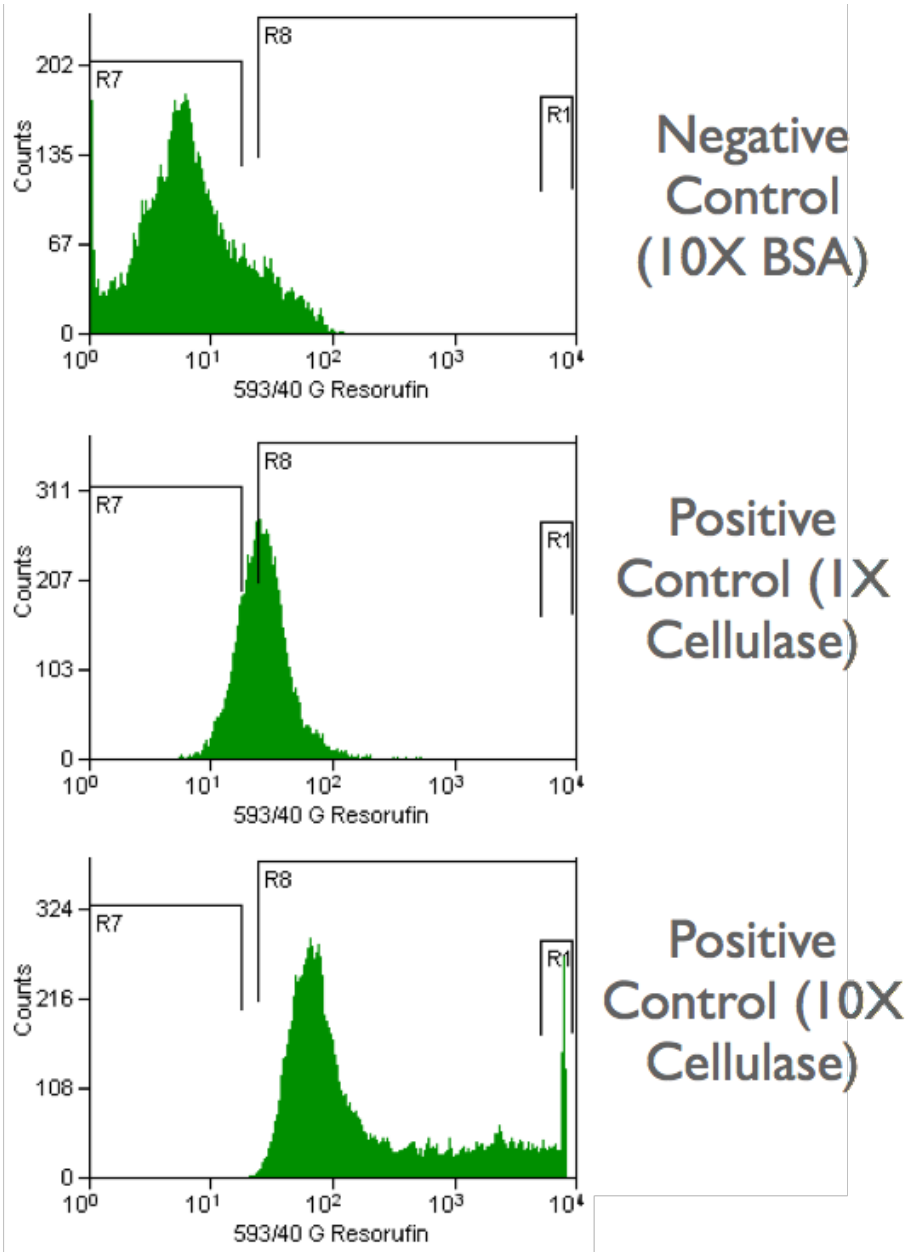


Figure 82. Detection of active enzymes in droplets via FACS using a soluble, fluorogenic substrate. Droplets with and without cellulases were distinguishable via FACS (individually, not sorted from a mixture). A higher concentration of cellulase enzyme simulated improved activity and indeed resulted in higher measured fluorescence (x-axis).

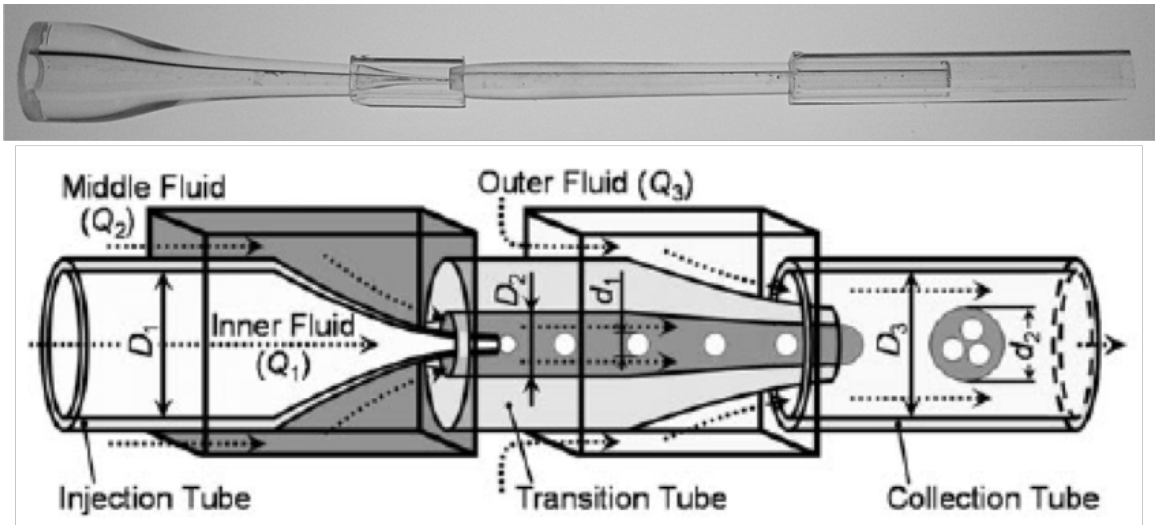


Figure 83. An attempt at generating a glass microfluidic device for producing uniform double emulsion droplets (top) was based on a published scheme (below).[68]

I also examined emulsion thermostability and found little change after 70˚C heat treatment (Figures 84 and 85). The emulsions appeared intact, indicating suitability for high-temperature enzyme evolution studies.

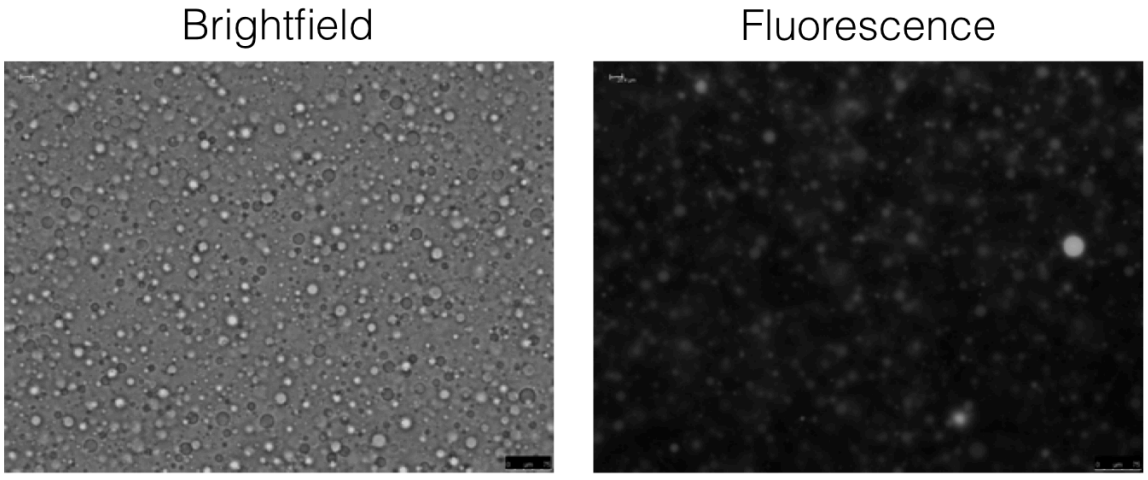


Figure 84. Single emulsions containing GFP (at room temperature).

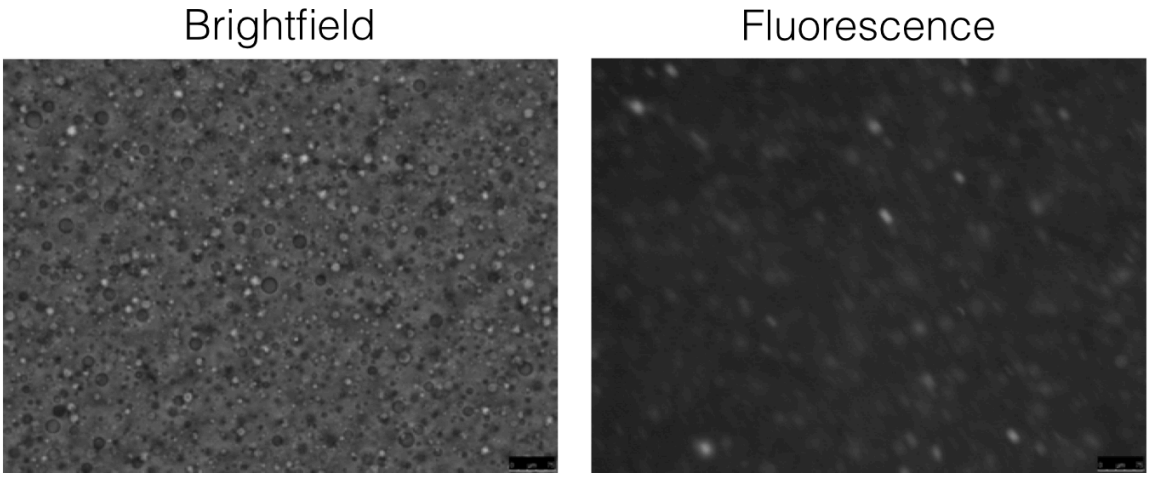


Figure 85. Single emulsions containing GFP after 70˚C heat treatment.

I wanted to first test a simplified system to see if I could distinguish between droplets carrying negative or positive (cellulase gene-containing) DNA. By using a water-soluble fluorescent substrate, I would eliminate the needs to encapsulate a solid cellulose particle and deliver glucose-detecting reagents, and I could create double and single emulsions consecutively. Active enzymes would result in fluorescent droplets, which would be sorted by FACS. PCR amplification of the recovered DNA would determine whether I successfully enriched for the cellulase gene or not.

Towards this end, I added two different DNA markers to separate double emulsions, mixed the emulsions, and attempted to selectively recover droplets containing the "positive" DNA marker (which was associated with a fluorophore). When I analyzed the DNA within the sorted droplets using PCR amplification, I found that I could not perfectly separate DNA from "negative" samples apart from "positive" samples (Figure 86), but the enrichment in positive samples was encouraging.

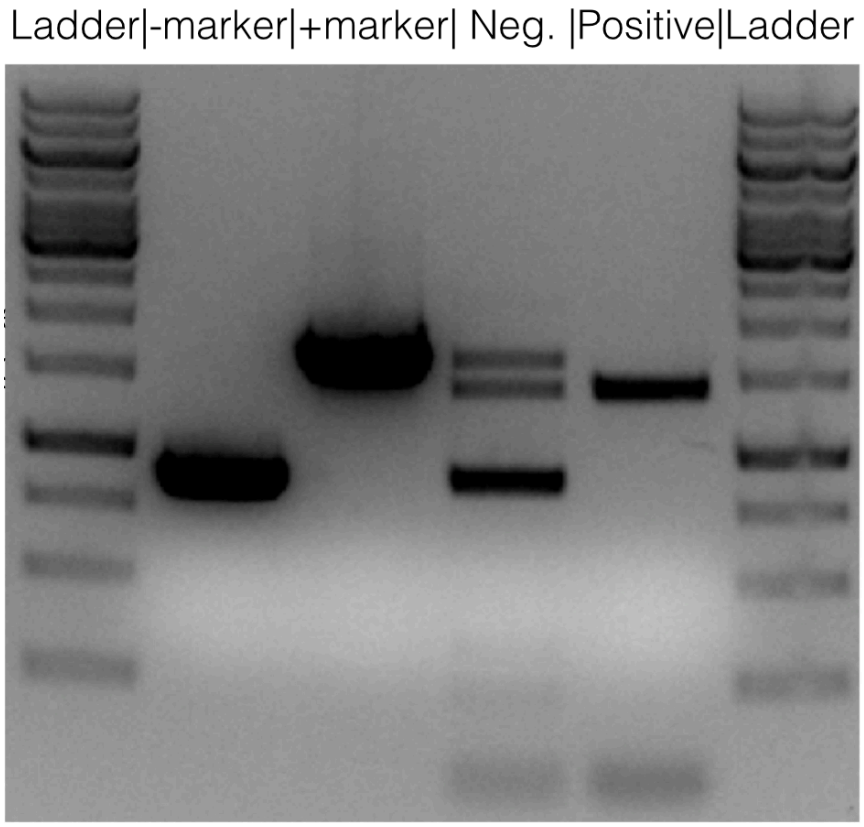


Figure 86. An attempt at separating two separate genes in different emulsions using FACS.

I also prepared double emulsions containing either green fluorescent protein (GFP) or resorufin, and mixed the emulsions in an equal ratio. Instead of seeing two distinct populations, however, FACS identified droplets that appeared to contain both GFP and resorufin after emulsion mixing, indicating compromised droplet integrity (Figure 87). I observed a similar effect when mixing emulsions with and without resorufin. Concerned that this droplet cross-talk was due to the fluorophore diffusing out of the droplets, I replaced resorufin with a charged fluorophore (which should not cross a hydrophobic oil layer) but was unable to correct the problem.

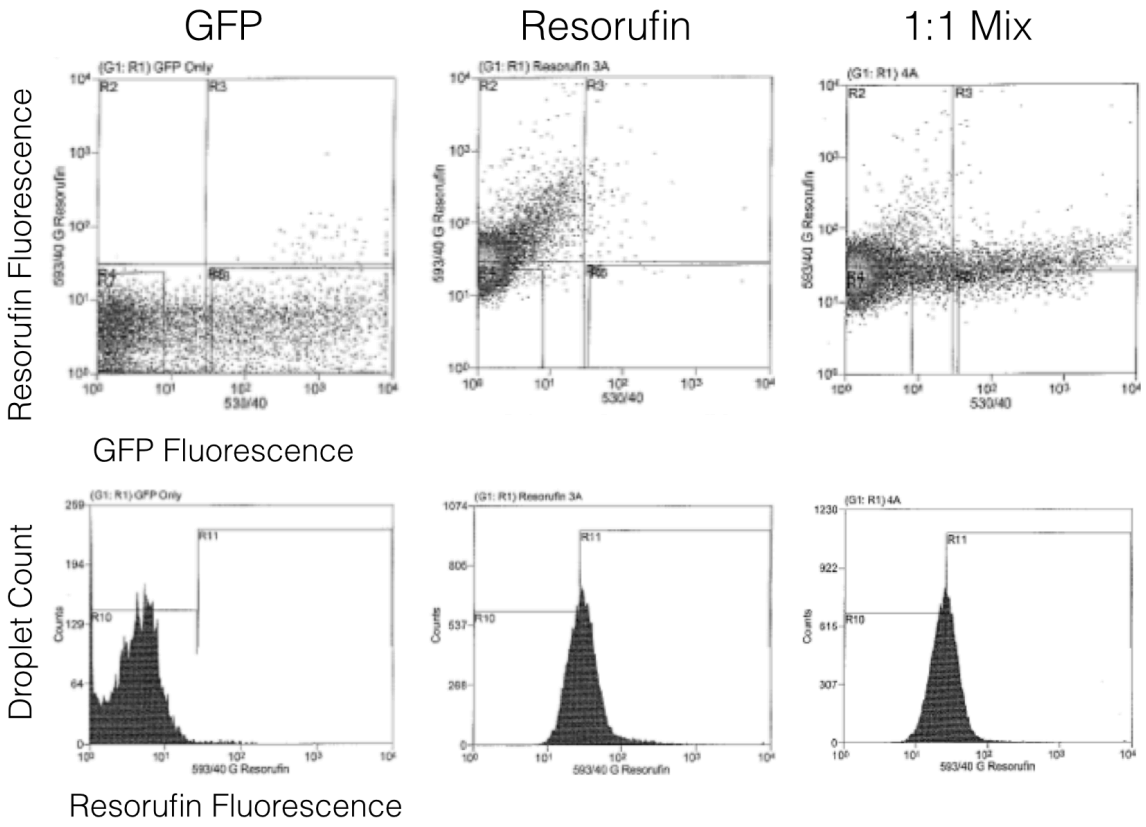


Figure 87. When mixed, two different emulsions appear to be the same instead of registering as two separate populations, indicating droplet cross-talk.

5.8 Conclusions

While I made substantial progress on individual components of this directed evolution strategy, and had several simplified benchmarks in mind, the overall scheme was too complicated to be practical.

I was able to demonstrate many steps successfully, but found that there were a multitude of challenges that I could not overcome. In several instances, I observed that control reactions containing the fluorescent probe Amplex Red (but without a cellulase enzyme) remained colorless before emulsification yet turned pink when emulsified, suggesting that emulsification was somehow activating the molecule that was to serve as an indicator of cellulase activity. Furthermore, despite (at least on one occasion) enriching for a DNA marker of interest (Figure 86), when I mixed emulsions containing different fluorophores, simulating droplets with distinct contents, the resulting emulsion mixture appeared uniform (Figure 87). Thus, the emulsions were not able to maintain their integrity; it appeared that their contents mixed. Re-emulsification of single emulsions

into double emulsions also threatened droplet integrity. For my directed evolution scheme (Figure 76), it was essential that IVC compartments remain distinct in order to accurately associate active enzymes with the genes that encode them.

Following successful proof-of-concept experiments, I had hoped to evolve *Trichoderma reesei* cellulases for greater thermostability, specific activity, and inter-enzyme synergy to aid in making cellulosic biofuels a reality. However, this project was exceptionally ambitious and challenging. In the end, the inventor of IVC himself, Professor Dan Tawfik, advised me in person not to work on this project predicated on IVC technology!

Concluding Remarks

My consistent goal throughout this PhD was to improve cellulase enzymes in order to increase the efficiency of a cellulosic bioenergy process. This, in turn, would result in biofuels that are more cost-competitive with fossil fuels and thus encourage a transition to sustainable energy technologies that ameliorate our effect on climate change.

Engineering cellulase enzymes is particularly difficult for a number of reasons, including:

- Cellulases (especially fungal ones) can be difficult to express, making high-throughput studies challenging.
- Industrially-relevant substrates are solid, crystalline lignocellulosic materials, but assays for activity on solid biomass substrates are challenging, time and labor intensive, and generally of low to moderate throughput.
- Screening for activity on soluble substrate mimics is less difficult, but somewhat irrelevant—enzymatic activity on soluble substrates shows limited correlation to activity on industrially-relevant substrates.[57]
- Cellulases degrade biomass in concert; it is possible that by optimizing an enzyme individually, one may not be preserving or optimizing for activity of the entire suite of enzymes.

I strove to address some of these issues during my PhD here at UC Berkeley:

- I worked only with systems that I believed to enable facile enzyme expression. Unfortunately, this was rarely the case (but that was my intention).
- I corroborated computational studies on Cel7A, generating mutants exhibiting less product inhibition and, in some cases, eliminating sensitivity to cellobiose entirely. (This work revealed a residue of particular interest for future investigations that may eventually lead to a product-tolerant mutant without losses in catalytic activity.)
- I ventured to engineer an LPMO with the hopes of adding another enzyme to a growing cocktail of thermostable, synergistic cellulases for efficient biomass hydrolysis.
- I purified individual enzymes from a fungal enzyme mixture with the intention of studying enzyme synergy using defined enzyme components, and for use as benchmarks to compare with engineered enzymes.
- I attempted to develop a system for high-throughput directed evolution of cellulases on industrially-relevant substrates.

While few of these endeavors were successful, I sincerely hope that my efforts can, in some way, aid our transition to a sustainable energy future.

Finally, the Energy Biosciences Institute (EBI) has been a unique place to work. It provided an opportunity to interact with colleagues from chemistry, molecular biology, chemical engineering, plant biology, etc., and to hear from others who are involved in industry (BP employees). This allowed me to constantly have a real-world perspective on my work, which may be uncommon in academia and is something that I personally find compelling and motivating. The EBI is also distinctive for its comprehensive approach to this important bioenergy challenge; the same institute tackles every part of the cellulosic biofuels process—from the growing of plants in the field, to pretreatment and depolymerization technologies, to microbial conversion of biomass-derived sugars to fuels and chemicals, and finally to economic and environmental analyses of the complete process. It was truly a unique place to work and I feel lucky to have had the opportunity.

References

1. Intergovernmental Panel on Climate Change (IPCC). Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. 151 (IPCC, Geneva, Switzerland, 2014). at <https://www.ipcc.ch/report/ar5/syr/>
2. Sustainable energy. Wikipedia, the free encyclopedia (2015). at <http://en.wikipedia.org/w/index.php?title=Sustainable_energy&oldid=660906788>
3. biomass-carbon-cycle.jpg. at <http://www.wendronbiomass.co.uk/img/biomass-carbon-cycle.jpg>
4. Saxena, R. C., Adhikari, D. K. & Goyal, H. B. Biomass-based energy fuel through biochemical routes: A review. Renew. Sustain. Energy Rev. **13,** 167–178 (2009).
5. Payne, C. M. et al. Fungal Cellulases. Chem. Rev. **115,** 1308–1448 (2015).
6. Rubin, E. M. Genomics of cellulosic biofuels. Nature **454,** 841–845 (2008).
7. The Renewable Fuel Standard: Issues for 2014 and Beyond. Congressional Budget Office at <https://www.cbo.gov/publication/45477>
8. McKendry, P. Energy production from biomass (part 1): overview of biomass. Bioresour. Technol. **83,** 37–46 (2002).
9. Genome Management Information System, Oak Ridge National Laboratory. Lignocellulose.
10. Starch. Wikipedia, the free encyclopedia (2015). at <http://en.wikipedia.org/w/index.php?title=Starch&oldid=661535279>
11. Bioenergy at JGI - The Benefits of Biomass. (2011). at <http://www.jgi.doe.gov/education/bioenergy/bioenergy_1.html>
12. Seidl, V., Seibel, C., Kubicek, C. P. & Schmoll, M. Sexual development in the industrial workhorse Trichoderma reesei. Proc. Natl. Acad. Sci. U. S. A. **106,** 13909–13914 (2009).
13. Reese, E. T. History of the cellulase program at the U.S. army Natick Development Center. Biotechnol. Bioeng. Symp. 9–20 (1976).
14. Chokhawala, H. A. et al. Mutagenesis of Trichoderma reesei endoglucanase I: impact of expression host on activity and stability at elevated temperatures. BMC Biotechnol. **15,** 11 (2015).
15. Beckham, G. T. Family 7 cellobiohydrolase from T. reesei. at <http://www.nrel.gov/biomass/images/beckham_cbh1.jpg>
16. Klein-Marcuschamer, D., Oleskowicz-Popiel, P., Simmons, B. A. & Blanch, H. W. The challenge of enzyme cost in the production of lignocellulosic biofuels. Biotechnol. Bioeng. **109,** 1083–1087 (2012).
17. Davis, R. et al. Process Design and Economics for the Conversion of Lignocellulosic Biomass to Hydrocarbons: Dilute-Acid and Enzymatic Deconstruction of Biomass to Sugars and Biological Conversion of Sugars to Hydrocarbons. (National Renewable Energy Laboratory (NREL), Golden, CO., 2013). at <http://www.osti.gov/scitech/biblio/1107470>
18. Murphy, L. et al. Product inhibition of five Hypocrea jecorina cellulases. Enzyme Microb. Technol. **52,** 163–169 (2013).

19. Bu, L. et al. Probing Carbohydrate Product Expulsion from a Processive Cellulase with Multiple Absolute Binding Free Energy Methods. J. Biol. Chem. **286,** 18161–18169 (2011).
20. Hanson, S. R., Stege, J. T., Cheng, C. & Luginbuhl, P. Variant cbh i polypeptides with reduced product inhibition. (2014). at <http://www.google.com/patents/US20140287471>
21. Silveira, R. L. & Skaf, M. S. Molecular Dynamics Simulations of Family 7 Cellobiohydrolase Mutants Aimed at Reducing Product Inhibition. J. Phys. Chem. B (2014). doi:10.1021/jp509911m
22. Somerville, C. et al. Toward a Systems Approach to Understanding Plant Cell Walls. Science **306,** 2206–2211 (2004).
23. Jalak, J., Kurašin, M., Teugjas, H. & Väljamäe, P. Endo-exo Synergism in Cellulose Hydrolysis Revisited. J. Biol. Chem. **287,** 28802–28815 (2012).
24. Teugjas, H. & Vaeljamaee, P. Product inhibition of cellulases studied with C-14-labeled cellulose substrates. Biotechnol. Biofuels **6,** 104 (2013).
25. Becker, D. et al. Engineering of a glycosidase Family 7 cellobiohydrolase to more alkaline pH optimum: the pH behaviour of Trichoderma reesei Cel7A and its E223S/A224H/L225V/T226A/D262G mutant. Biochem. J. **356,** 19–30 (2001).
26. Gericke, M., Schlufter, K., Liebert, T., Heinze, T. & Budtova, T. Rheological Properties of Cellulose/Ionic Liquid Solutions: From Dilute to Concentrated States. Biomacromolecules **10,** 1188–1194 (2009).
27. Dana, C. M. et al. Biased clique shuffling reveals stabilizing mutations in cellulase Cel7A. Biotechnol. Bioeng. **109,** 2710–2719 (2012).
28. Labbé, S. & Thiele, D. J. [8] Copper ion inducible and repressible promoter systems in yeast. Methods Enzymol. **306,** 145–153 (1999).
29. Dana, C. M. et al. The importance of pyroglutamate in cellulase Cel7A. Biotechnol. Bioeng. **111,** 842–847 (2014).
30. Daniel Gietz, R. & Woods, R. A. in Methods in Enzymology (ed. Fink, C. G. and G. R.) **350,** 87–96 (Academic Press, 2002).
31. Knott, B. C., Crowley, M. F., Himmel, M. E., Ståhlberg, J. & Beckham, G. T. Carbohydrate–Protein Interactions That Drive Processive Polysaccharide Translocation in Enzymes Revealed from a Computational Study of Cellobiohydrolase Processivity. J. Am. Chem. Soc. **136,** 8810–8819 (2014).
32. Nakamura, A. et al. The Tryptophan Residue at the Active Site Tunnel Entrance of Trichoderma reesei Cellobiohydrolase Cel7A Is Important for Initiation of Degradation of Crystalline Cellulose. J. Biol. Chem. **288,** 13503–13510 (2013).
33. Payne, C. M. et al. Glycoside Hydrolase Processivity Is Directly Related to Oligosaccharide Binding Free Energy. J. Am. Chem. Soc. **135,** 18831–18839 (2013).
34. Colussi, F. et al. Probing Substrate Interactions in the Active Tunnel of a Catalytically Deficient Cellobiohydrolase (Cel7). J. Biol. Chem. **290,** 2444–2454 (2015).
35. Bu, L. et al. Product Binding Varies Dramatically between Processive and Nonprocessive Cellulase Enzymes. J. Biol. Chem. **287,** 24807–24813 (2012).

36. Fox, J. M., Levine, S. E., Clark, D. S. & Blanch, H. W. Initial- and Processive-Cut Products Reveal Cellobiohydrolase Rate Limitations and the Role of Companion Enzymes. Biochemistry (Mosc.) **51,** 442–452 (2011).
37. Jalak, J. & Väljamäe, P. Mechanism of initial rapid rate retardation in cellobiohydrolase catalyzed cellulose hydrolysis. Biotechnol. Bioeng. **106,** 871–883 (2010).
38. Andrić, P., Meyer, A. S., Jensen, P. A. & Dam-Johansen, K. Reactor design for minimizing product inhibition during enzymatic lignocellulose hydrolysis: I. Significance and mechanism of cellobiose and glucose inhibition on cellulolytic enzymes. Biotechnol. Adv. **28,** 308–324 (2010).
39. Harris, P. V. et al. Stimulation of Lignocellulosic Biomass Hydrolysis by Proteins of Glycoside Hydrolase Family 61: Structure and Function of a Large, Enigmatic Family. Biochemistry (Mosc.) **49,** 3305–3316 (2010).
40. Vermaas, J. V., Crowley, M. F., Beckham, G. T. & Payne, C. M. Effects of Lytic Polysaccharide Monooxygenase Oxidation on Cellulose Structure and Binding of Oxidized Cellulose Oligomers to Cellulases. J. Phys. Chem. B (2015). doi:10.1021/acs.jpcb.5b00778
41. Busk, P. K. & Lange, L. Classification of fungal and bacterial lytic polysaccharide monooxygenases. BMC Genomics **16,** 368 (2015).
42. Agger, J. W. et al. Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation. Proc. Natl. Acad. Sci. 201323629 (2014). doi:10.1073/pnas.1323629111
43. Isaksen, T. et al. A C4-oxidizing lytic polysaccharide monooxygenase cleaving both cellulose and cello-oligosaccharides. J. Biol. Chem. jbc.M113.530196 (2013). doi:10.1074/jbc.M113.530196
44. Forsberg, Z. et al. Comparative Study of Two Chitin-Active and Two Cellulose-Active AA10-Type Lytic Polysaccharide Monooxygenases. Biochemistry (Mosc.) **53,** 1647–1656 (2014).
45. Hemsworth, G. R., Henrissat, B., Davies, G. J. & Walton, P. H. Discovery and characterization of a new family of lytic polysaccharide monooxygenases. Nat. Chem. Biol. **10,** 122–126 (2013).
46. Forsberg, Z. et al. Cleavage of cellulose by a CBM33 protein. Protein Sci. **20,** 1479–1483 (2011).
47. Kittl, R., Kracher, D., Burgstaller, D., Haltrich, D. & Ludwig, R. Production of four Neurospora crassa lytic polysaccharide monooxygenases in Pichia pastoris monitored by a fluorimetric assay. Biotechnol. Biofuels **5,** 1–14 (2012).
48. Yang, B. & Wyman, C. E. BSA treatment to enhance enzymatic hydrolysis of cellulose in lignin containing substrates. Biotechnol. Bioeng. **94,** 611–617 (2006).
49. Hui, W., Shinichi, K. & Kazuhiro, M. Effect of non-enzymatic proteins on enzymatic hydrolysis and simultaneous saccharification and fermentation of different lignocellulosic materials. Bioresour. Technol. doi:10.1016/j.biortech.2015.04.112
50. Wilson, D. B. Cellulases and biofuels. Curr. Opin. Biotechnol. **20,** 295–299 (2009).

51. Forsberg, Z. et al. Structural and functional characterization of a conserved pair of bacterial cellulose-oxidizing lytic polysaccharide monooxygenases. Proc. Natl. Acad. Sci. **111,** 8446–8451 (2014).
52. Yazdi, M. T., Woodward, J. R. & Radford, A. The cellulase complex of Neurospora crassa: activity, stability and release. J. Gen. Microbiol. **136,** 1313 –1319 (1990).
53. Martinez, D. et al. Genome sequencing and analysis of the biomass-degrading fungus Trichoderma reesei (syn. Hypocrea jecorina). Nat Biotech **26,** 553–560 (2008).
54. Espagne, E. et al. The genome sequence of the model ascomycete fungus Podospora anserina. Genome Biol. **9,** R77 (2008).
55. Gusakov, A. V. et al. Design of highly efficient cellulase mixtures for enzymatic hydrolysis of cellulose. Biotechnol. Bioeng. **97,** 1028–1038 (2007).
56. Merino, S. T. & Cherry, J. in Biofuels (ed. Olsson, L.) **108,** 95–120 (Springer Berlin Heidelberg, 2011).
57. Percival Zhang, Y.-H., Himmel, M. E. & Mielenz, J. R. Outlook for cellulase improvement: Screening and selection strategies. Biotechnol. Adv. **24,** 452–481 (September).
58. Bharadwaj, R. et al. High-throughput enzymatic hydrolysis of lignocellulosic biomass via in-situ regeneration. Bioresour. Technol. **102,** 1329–1337 (2011).
59. Chundawat, S. P. S., Balan, V. & Dale, B. E. High-throughput microplate technique for enzymatic hydrolysis of lignocellulosic biomass. Biotechnol. Bioeng. **99,** 1281–1294 (2008).
60. Tawfik, D. S. & Griffiths, A. D. Man-made cell-like compartments for molecular evolution. Nat. Biotechnol. **16,** 652–656 (1998).
61. Griffiths, A. D. & Tawfik, D. S. Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. EMBO J **22,** 24–35 (2003).
62. Aharoni, A., Amitai, G., Bernath, K., Magdassi, S. & Tawfik, D. S. High-Throughput Screening of Enzyme Libraries: Thiolactonases Evolved by Fluorescence-Activated Sorting of Single Cells in Emulsion Compartments. Chem. Biol. **12,** 1281–1289 (2005).
63. Yuen, C. M. & Liu, D. R. Dissecting protein structure and function using directed evolution. Nat. Methods **4,** 995 (2007).
64. Aharoni, A., Griffiths, A. D. & Tawfik, D. S. High-throughput screens and selections of enzyme-encoding genes. Curr. Opin. Chem. Biol. **9,** 210–216 (2005).
65. Nakano, M. et al. Single-molecule PCR using water-in-oil emulsion. J. Biotechnol. **102,** 117–124 (2003).
66. Miller, O. J. et al. Directed evolution by in vitro compartmentalization. Nat Meth **3,** 561–570 (2006).
67. Rothe, A., Surjadi, R. N. & Power, B. E. Novel proteins in emulsions using in vitro compartmentalization. Trends Biotechnol. **24,** 587–592 (2006).
68. Chu, L.-Y., Utada, A. S., Shah, R. K., Kim, J.-W. & Weitz, D. A. Controllable Monodisperse Multiple Emulsions. Angew. Chem. **119,** 9128–9132 (2007).