Title
Understanding and Engineering Cellulase Binding to Biomass Components

Author
Strobel, Kathryn Lynn

Publication Date
2015

Peer reviewed|Thesis/dissertation

Understanding and Engineering Cellulase Binding to Biomass Components

By

Kathryn Lynn Strobel

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Douglas S. Clark, Chair
Professor Harvey W. Blanch
Professor Matthew B. Francis

Summer 2015

Abstract

Understanding and Engineering Cellulase Binding to Biomass Components

by

Kathryn Lynn Strobel

Doctor of Philosophy in Chemical Engineering

University of California, Berkeley

Professor Douglas S. Clark, Chair

Lignocellulosic biomass is an abundant, low-cost resource for the renewable production of fuels and chemicals. To unlock the potential of lignocellulosic biomass, the cellulose must be broken down into sugars before fermentation to produce ethanol, butanol, or other bio-based products. Unfortunately, lignocellulose is highly resistant to enzymatic degradation, necessitating high enzyme loadings that increase the cost of biofuels. The recalcitrance of biomass stems in part from the presence of lignin, a major component of lignocellulosic biomass. Lignin impedes enzymatic hydrolysis by non-productively binding cellulases and contributing to cellulase denaturation. Despite numerous studies documenting cellulase adsorption to lignin, the structural basis has not been fully elucidated and few attempts have been made to engineer enzymes for reduced lignin affinity.

In this work, we investigate and engineer cellulase adsorption to lignin and the resulting effect on hydrolysis of cellulose. The lignin inhibition of two homologous cellulases, *T. reesei* Cel7A and *T. emersonii* Cel7A, was found to differ significantly. In Chapter 2, we propose that differences in surface charge, stability, and glycosylation patterns may be the driving force/s behind the observed differences in lignin inhibition and we suggest engineering strategies for improving lignin tolerance of Cel7A catalytic domains.

Chapters 3 and 4 detail our efforts to investigate the mechanisms of cellulase lignin adsorption and engineer an enzyme with reduced lignin affinity using site directed mutagenesis of the *T. reesei* Cel7A carbohydrate binding module (CBM) and linker. Mutation of aromatic and polar residues on the planar face of the CBM greatly decreased binding to both cellulose and lignin, supporting the hypothesis that the cellulose-binding face is also responsible for the majority of lignin affinity. Cellulose and lignin affinity of the alanine mutants were highly correlated, indicating similar binding mechanisms for cellulose and lignin. CBM mutations that added hydrophobic or positively charged residues decreased the selectivity toward cellulose, while mutations that added negatively charged residues increased the selectivity. Mutating the linker to alter predicted glycosylation patterns greatly impacted lignin affinity but did not affect cellulose affinity.

Beneficial mutations were combined to generate a mutant with 2.5 fold less lignin affinity and fully retained cellulose affinity. This mutant was not inhibited by added lignin during hydrolysis of Avicel and generated 40% more glucose than the wild type enzyme from dilute acid-pretreated *Miscanthus*. The mutations studied here inform engineering efforts of other homologous CBMs and will hopefully contribute to reducing the cost of biofuels.

The final chapter details the development of a high-throughput selection platform for engineering protease enzymes with new sequence specificity. Proteases are commonly used in research, industry, and medicine, and there is considerable promise for new proteases that could cleave at a user-specified sequence. Positive selection and counter-selection were combined to select a tobacco etch virus protease mutant with new substrate compatibility.

Dedication

To my mother, Jane, who introduced me to chemical engineering by explaining the heat transfer properties of my chicken noodle soup.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my advisor, Professor Doug Clark, for his unwavering support and valuable feedback through my many projects. I also appreciate the space he allowed, which enabled me to explore topics that interested me and become an independent researcher. I am also grateful to the Energy Biosciences Institute for funding my work.

I am very thankful for the wonderful students and postdocs of the Clark lab, without whom I would not have made it through even one year of graduate school. I am especially grateful to Dr. Katherine Pfeiffer for being an outstanding mentor. She has contributed so much to both my scientific and professional development. I would also like to give a special thanks to Harshal Chokhawala for teaching me how to do my first PCR; Sarah Huffer and Elizabeth Schneider for great advice; Kierston Shill, Meera Atreya, and Erin Imsand for fun times in the lab; and Craig Dana and Jerome Fox for patiently passing on their scientific knowledge.

I would also like to acknowledge the many friends, roommates, and family members who supported me through graduate school. I am particularly thankful to Richard Kwant, who contributed many ideas and provided invaluable feedback on my work in addition to joining me on many adventures outside of lab.

Chapter 1: Cellulosic biofuels and lignin inhibition

1.1 The case for renewable fuels

Rising global temperatures are causing changes in weather and climate patterns that present great challenges to our society and environment. The global average temperature increased by more than 1.3°F over the last century and is predicted to rise another 2 to 11.5°F over the next hundred years (1). Changes in the average temperature of the planet can translate to large and potentially dangerous shifts in climate and weather (2).

Most of the warming of the past half-century has been caused by human emissions of greenhouse gases, which have increased significantly over the past century (3). Carbon emissions increased by over 16 fold between 1900 and 2008 and by about 1.5 fold between 1990 and 2008 (2). The majority of these greenhouse gas emissions come from burning fossil fuels to produce energy. Fossil fuels burned for road, rail, air, and marine transportation make up a significant portion of total greenhouse gas emissions. Almost all (95%) of the world's transportation energy comes from petroleum-based fuels, largely gasoline and diesel (2). Replacing these fossil fuels with renewable alternatives, such as biofuels derived from plants, could significantly reduce greenhouse gas emissions. For traditional fuels, carbon is taken from the ground and transferred to the air. In contrast, the plants used to produce biofuels consume carbon dioxide as they grow, offsetting the emissions from burning the fuel derived from them.

1.2 Cellulosic biofuels

A biofuel is simply a fuel derived from biological materials, usually plants. First generation biofuels (ethanol made from sugar, starch, or vegetable oils) are relatively

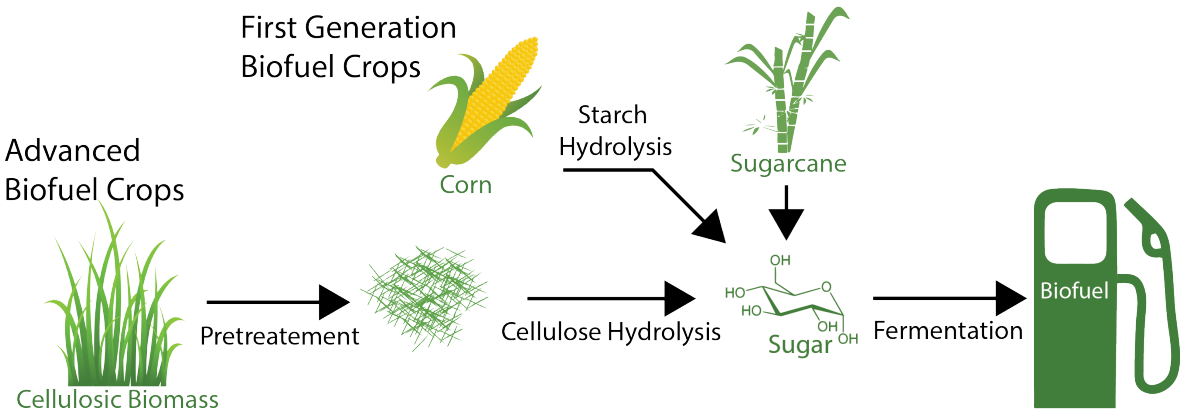

Figure 1.1 Process for producing biofuels from various sources.

straightforward to produce (Figure 1). However, the raw materials for first generation biofuels are plants used as food sources (such as corn, sugar cane, and soybeans), sparking a food versus fuel debate. Advanced biofuels start with non-food crops rich in cellulose, such as agricultural waste or grasses grown specifically for biofuels (4). Cellulose is the most abundant biopolymer on earth and is not used as a food source, which avoids controversies about using food sources as fuel. The drawback, however, is that breaking down the cellulose to fermentable sugars is much more difficult and expensive than obtaining sugar directly from sugarcane or by hydrolyzing starch.

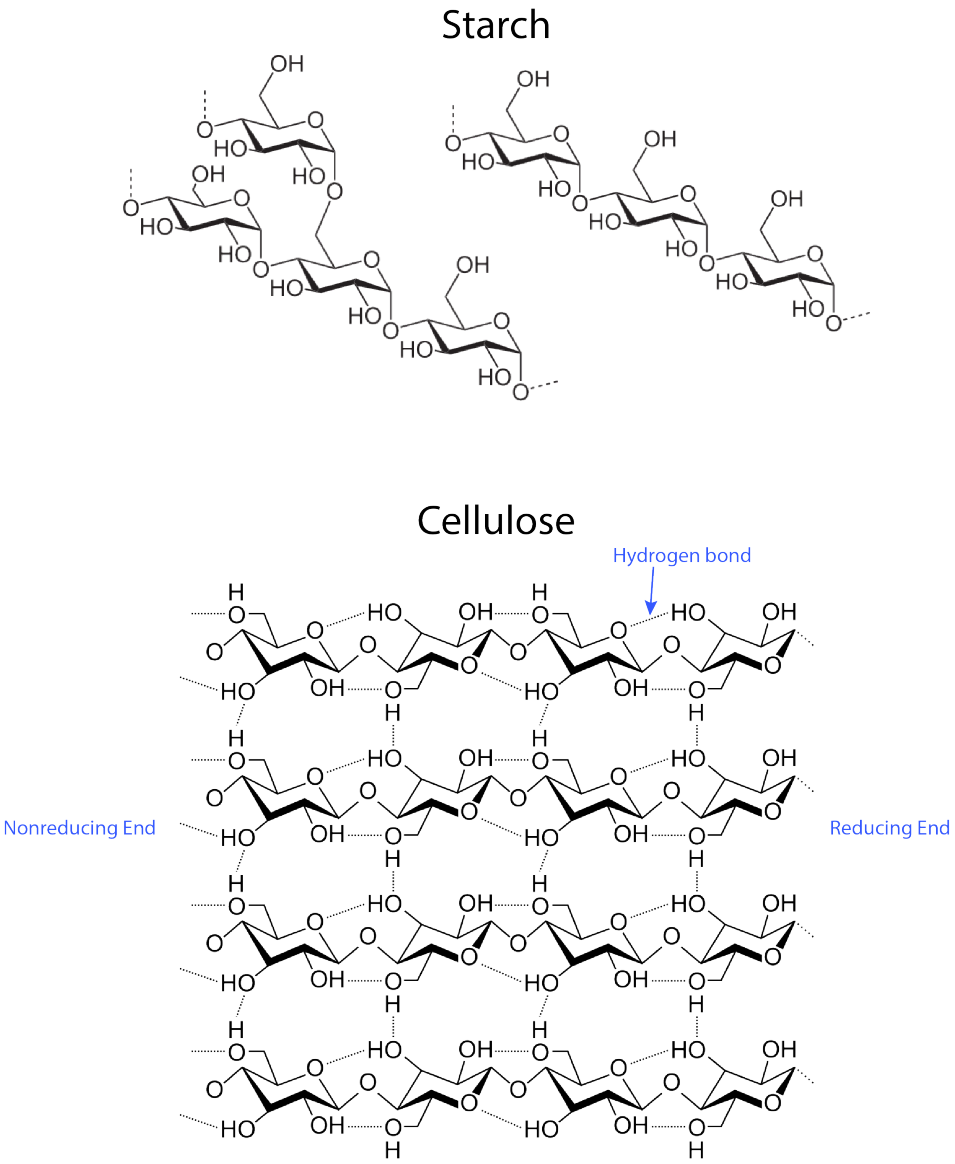

Figure 1.2 The structure of starch and cellulose.
Starch is an alpha1-4 linked polymer of glucose that can contain branches while cellulose is a beta 1-4 linked polymer of glucose.

Advanced biofuel crops (cellulosic biomass) must first be pretreated with mechanical and/or chemical methods to increase the accessibility of the cellulose. Even after pretreatment, cellulose is still much more difficult to hydrolyze than starch. Plants synthesize starch to store sugar as a food source and cellulose to provide mechanical strength. Although they are both polymers of glucose, starch is a branched polymer while cellulose is a linear polymer. In the linear polymer, extensive intra and intermolecular hydrogen bonding, as seen in Figure 1.2, contributes strength and resistance to degradation.

In addition to the intrinsic resistance of cellulose to degradation, advanced biofuel crops contain other polymers (hemicellulose and lignin) that can hider cellulose hydrolysis. The difficulty of generating sugar from cellulosic biomass necessitates expensive processing steps, limiting the application of cellulosic biofuels. Reducing the cost of producing sugar from cellulose is therefore an essential step for replacing fossil fuels with biofuels.

1.3 Biomass pretreatment

Why is pretreatment necessary?

The purpose of pretreatment is to convert biomass from its native form, which is very recalcitrant to hydrolysis with cellulase enzymes, into a form where enzymatic hydrolysis is effective. The best pretreatments improve hydrolysis many fold over the untreated biomass. A number of pretreatment technologies based on physical, chemical and biological methods have been developed that increase the accessibility and digestibility of cellulose. Pretreatments accomplish this by removing other cell wall polymers (lignin and hemicellulose) that otherwise obstruct cellulose and breaking apart the native crystalline structure of cellulose, as illustrated in Figure 1.3.

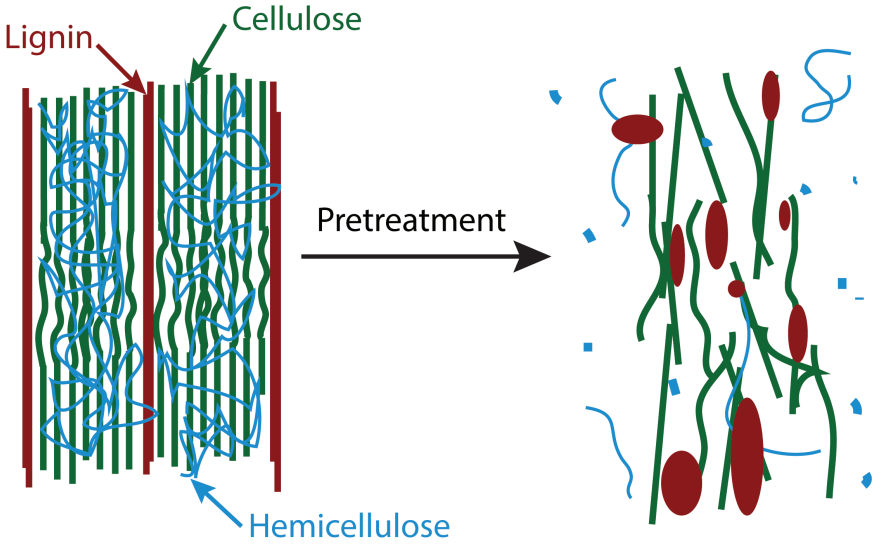

Figure 1.3 Impact of pretreatment on biomass.
Adapted from (5).

Types of pretreatment

The most effective pretreatments preserve and decrystallize cellulose, depolymerize hemicelluloses, form few inhibitors, remove lignin, and have low energy input and operational cost (6). Common pretreatments include dilute acid, ionic liquid, hot water, steam explosion, and ammonia fiber expansion. New pretreatment are also currently being developed.

In dilute acid pretreatment, a solution of sulfuric acid is used to treat the biomass at 160°-220°C for 1-30 minutes (7). Dilute acid pretreatment removes most of the hemicellulose and requires limited chemical inputs and energy, however it also can cause condensation and precipitation of lignin and generate fermentation inhibitors.

For ionic liquid pretreatment, biomass is dissolved in ionic liquid solutions and heated. Water is then added to precipitate the dissolved cellulose. Ionic liquid pretreatment is the best at removing lignin and decrystalizing cellulose, resulting in very fast hydrolysis rates (8). Unfortunately, ionic liquids are incredibly expensive and difficult to recycle, which has limited their use.

Hot water pretreatment is an alternative to caustic chemical strategies that involves exposing lignocellulose to 160-230°C water at high pressures for 10-30 minutes (9). Under these conditions, hemicellulose and amorphous cellulose are partially hydrolyzed. Steam explosion is a variation on hot water pretreatment where the pressure is explosively released, weakening the cellulose fibers and increasing particle surface area. Hot water and steam explosion pretreatments are low cost, mild treatments but require excessive volumes of water.

During ammonia fiber expansion (AFEX), biomass substrates are subjected to 50-100% w/w ammonia/biomass at 40-180°C and 200-1000 psi for 5-45 minutes followed by rapid decompression (9). This process increases accessible surface area, partially hydrolyzes hemicellulose, depolymerizes lignin, and reduces cellulose crystallinity. It is not efficient for biomass sources with high lignin content, however.

1.4 Enzymatic hydrolysis

After pretreatment, a mixture of different enzymes including cellulases, hemicellulases, and oxidative enzymes are used to hydrolyze cellulose to generate fermentable sugars. Cellulose can also be hydrolyzed using strong acids and harsh processing conditions, but these methods are expensive, environmentally hazardous, and produce byproducts that inhibit downstream fermentations. In contrast, enzymatic methods are performed under mild conditions and generate few, if any, inhibitors of fermentation.

Cellulase cocktails

Many organisms have evolved the necessary enzymatic machinery for converting cellulose to soluble sugars for food and energy sources. Given the complexity of plant cell walls, most cellulose-degrading organisms employ a mixture of enzymes with synergistic function to break down polysaccharides (10). The three main classes of enzymes required for cellulose hydrolysis (endoglucanases, exoglucanases, and beta-glucosidase) are illustrated in Figure 1.4. Endoglucanases (EGs) act by making single cuts in the cellulose chain in mostly amorphous regions. Exoglucanases (or cellobiohydrolases, CBHs) complex with one end of a cellulose chain in a crystalline region and process down the length of the chain, depolymerizing cellulose into mostly disaccharide units. CBH1 acts from the reducing end of cellulose chains while CBH2 acts from the non-reducing end. The action of EG and CBH enzymes generates predominately cellobiose, a disaccharide of glucose. In solution, Beta-glucosidase (BG) cleaves cellobiose into glucose for cellular uptake by organisms. Other enzymes that enhance cellulose degradation, polysaccharide monooxyganase (PMO) enzymes, have recently been discovered. These oxidative enzymes likely provide endo-like action in the crystalline regions of cellulose (11).

Cellulases, unlike many commonly studied enzymes, act on an insoluble substrate. The first step in hydrolysis by a cellulase is therefore adsorption to the cellulose surface. Many cellulases contain two parts, or domains: a catalytic domain (CD) responsible for hydrolyzing the bonds in cellulose, and a carbohydrate-binding module (CBM) responsible for binding to the insoluble substrate and increasing the concentration of the CD on the surface of cellulose.

The most well-studied cellulase mixture is produced by the fungus *Trichoderma reesei*. This organism produces a particularly effective mixture of enzymes that are most active around 50°C. It has been engineered to overproduce cellulases, secreting as much as 2.2 g/L (12), which reduces the cost of producing the enzymes required for biofuel production. Cellulases still make up a large portion of biofuel production costs, however, because such large amounts of enzyme are required for efficient hydrolysis.

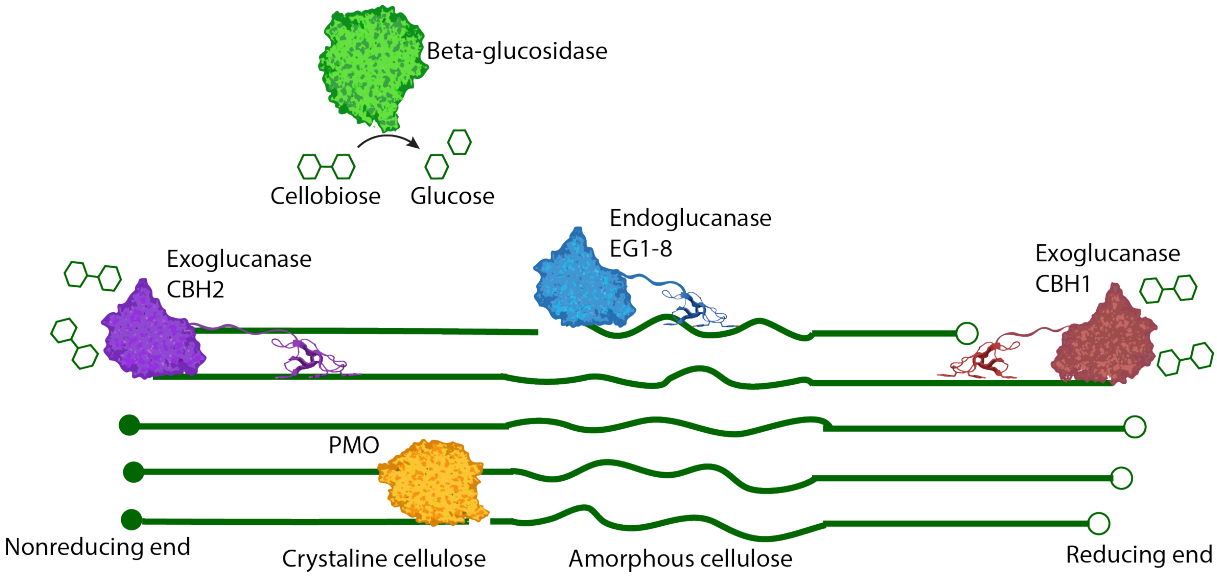

Figure 1.4 Cellulose degradation by a synergistic mixture of enzymes.

Hydrolysis bottlenecks

It is estimated that cellulase enzymes contribute approximately $1.47 per gallon of ethanol to the final cost a consumer would pay at the pump for biofuel (13). This cost could be reduced by decreasing the amount of enzyme required to effectively break down the cellulose in biomass. Large amounts of enzymes are required in part, because the enzymes are inhibited by lignin, a major component of biomass. Lignin can physically block cellulose and non-productively adsorb enzymes, reducing their effective concentration.

1.5 The impact of lignin on enzymatic hydrolysis

Lignin, a polymer of phenylpropane units, is an essential structural component in nearly all plant material (14). It is produced through oxidative coupling of 4-hydroxyphenyl propanoid compounds, generating a cross-linked network lacking a regular repeating structure that is common to many other polymers such as cellulose. An example lignin structure, showing common linkages, is shown in Figure 1.5. In the plant cell wall, lignin acts as a barrier to prevent evaporation and inhibit attack by pests and microbes (15). It comprises 20-30% of the dry mass of grasses and trees and is found primarily in the secondary plant cell wall. It is covalently linked to hemicellulose, which is in turn associated non-covalently with cellulose.

Lignin inhibition

Lignin's detrimental impact on the enzymatic hydrolysis of biomass has been well documented. Two main methods have been employed to study the magnitude and mechanism of lignin inhibition: removing lignin from biomass or adding supplemental lignin to biomass or pure cellulose. Numerous studies have found that removing lignin from biomass increases the final hydrolysis yield (e.g. (16-21)), though the extent of improvement varies with biomass type, pretreatment conditions, and extent of lignin removal. These studies have

implicated two main mechanisms for lignin inhibition: preventing cellulose access and non-productive binding of enzymes, prohibiting them from acting on cellulose. When biomass with lignin is compared to delignified biomass, these two mechanisms cannot be easily resolved.

Isolating lignin's role as a physical barrier from its role in non-productive binding is very difficult. There are few studies, therefore, that focus on the physical barrier mechanism. Mooney et al. found that lignin removal increased the size of pores in Douglas-fir pulp, resulting in more complete hydrolysis due at least in part to removal of a physical barrier (20). Ju et al. quantified surface lignin and found that higher surface lignin lead to lower binding affinity to poplar and lower hydrolysis rates, also implicating lignin acting to impede access to cellulose (19). Ishizawa et al. found that removing a portion of lignin from acid pretreated corn stover improved hydrolysis but removing all of the lignin resulted in decreased digestibility, potentially due to aggregation of cellulose microfibrils and therefore reduced surface area (16).

Kumar et al. studied both mechanisms of lignin inhibition. They compared steam-pretreated softwood (SPS) to delignified SPS and found an approximately linear correlation between delignification and conversion. The authors then added isolated lignin to delignified biomass, enabling measurement of lignin inhibition due to non-productive binding to lignin without physically blocking access to cellulose. Adding the isolated lignin to the delignified SPS reduced hydrolysis, but not to the same extent as hydrolysis was increased with lignin removal, implicating both that both the physical barrier and non-productive binding mechanisms contribute to inhibition (17).

Many authors have sought to decouple lignin's role in non-productively binding enzymes from its role as a physical barrier by added isolated lignin to purified cellulose (e.g. (17,18,23-31)). The magnitude of inhibition due to non-productive binding reported in the literature varies greatly and depends on the enzyme/s, hydrolysis conditions, biomass pretreatment, and lignin isolation method. The variability in inhibition levels due to non-productive binding is discussed further in Chapter 2.

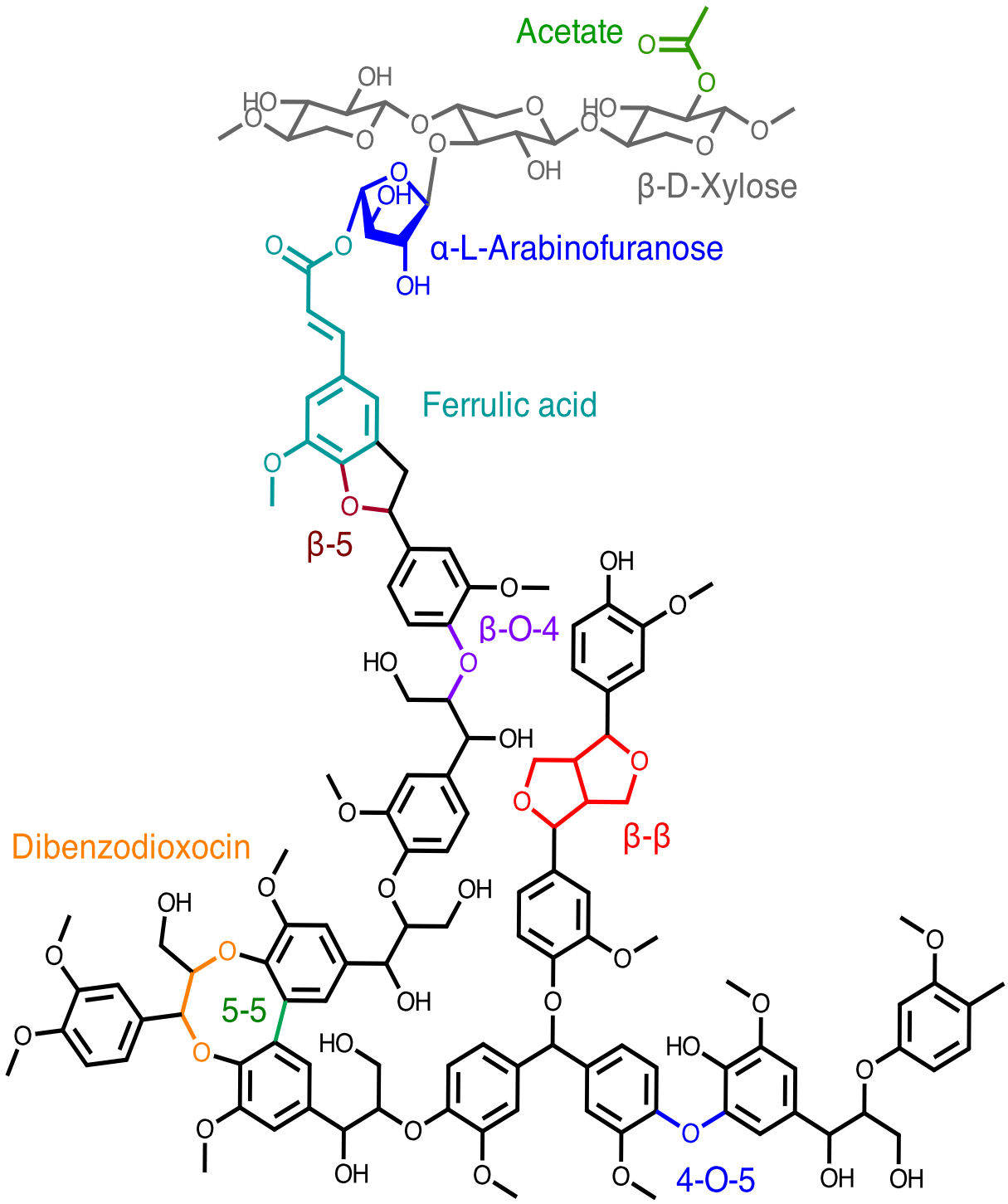

Figure 1.5 Model lignin structure.
Common structural features and bonds, and a lignin-carbohydrate complex with a ferrulic acid linkage are shown. Adapted from (22).

1.6 Current strategies to reduce lignin inhibition

Lignin's detrimental impact on hydrolysis could be fully mitigated by completely removing the lignin during pretreatment. Indeed, this is the only way to eliminate lignin's role as a physical barrier. Unfortunately, pretreatment methods that remove most of the lignin are prohibitively expensive. Engineering efforts have thus focused on alleviating lignin inhibition by decreasing non-productive binding. This can be accomplished by changing the process conditions, modifying the surface of lignin, or engineering the enzymes.

Process conditions for reduced inhibition

Lignin inhibition has been shown to increase with temperature (24). Performing hydrolysis at lower temperature may therefore help to alleviate lignin inhibition. Unfortunately, cellulase catalytic activity is reduced at lower temperatures, which means more enzyme or longer times would be required to achieve the same sugar yields. Also, the risk of contamination by bacteria or other microorganisms increases at lower temperature.

In addition to temperature, lignin inhibition is also affected by pH. Rahikainen et al. found that lignin adsorption decreases as pH increases (32). Lan et al. found that hydrolysis of biomass is most efficient at a pH between 5.2 and 6.2, higher than the pH optimum of the same enzymes when hydrolyzing pure cellulose (pH optimum of 5.0) (33). Lou et al. confirmed that this result was due to reduced non-productive adsorption to lignin (34). Increasing the pH of hydrolysis is therefore a good strategy to reduce lignin inhibition and improve enzyme efficiency. However, if other strategies could be employed to prevent non-productive adsorption to lignin, operating at the enzyme's natural pH optimum would be more effective.

Lignin modification strategies

A number of studies have focused on covalently or non-covalently modifying the lignin surface to reduce enzyme adsorption. Covalent modification strategies, usually achieved during biomass pretreatment, include sulfonation (33,34) and carboxylation (29). Both methods add negative charge to the surface of lignin, rendering it more hydrophilic and less able to adsorb cellulases. Non-covalent modification strategies include blocking the lignin either before or during hydrolysis with exogenous proteins, surfactants, or polymers (35-37). The added materials adsorb strongly to lignin, preventing cellulase adsorption. Although effective for reducing lignin inhibition, all lignin modification methods require additional materials and/or processing steps, adding to the cost of producing biofuels.

Enzyme modification strategies

Very few studies have focused on engineering cellulases to reduce non-productive adsorption. Nordwald and co-workers covalently modified cellulases by succinylation, which introduces extra negative charge to the surface of the enzyme. The resulting cellulases were significantly less inhibited by lignin. Similar to lignin modification strategies, however, cellulase succinylation requires additional materials and processing steps, adding to the cost and complexity of biofuels processes.

1.7 Engineering cellulase Cel7A to reduce lignin inhibition

Engineering enzymes for lignin resistance by changing their amino acid sequence could improve biomass hydrolysis without requiring additional costly materials or processing steps. Once suitable changes in amino acid sequence are identified, the DNA encoding the modified cellulase would be introduced to *T. reesei*, the cellulase-producing organism, such that it replaces the DNA for the native enzyme. Once this is accomplished, the engineered enzyme is always produced by the new strain of *T. reesei* and the biofuel production process proceeds as usual, with no additional steps or materials. The goal of my main dissertation work was to identify cellulase mutants (cellulases with amino acid changes) that display reduced non-productive adsorption to the surface of lignin, while retaining the ability to bind cellulose, which is necessary for their efficient catalytic activity.

The enzyme I choose for engineering is *T. reesei* Cel7A, formally called CBH1. This cellulase is the most abundant enzyme in commercial enzyme mixtures and has the greatest hydrolytic potential (38). In contains a catalytic domain (CD) and a carbohydrate-binding module (CBM), connected by a flexible linker. An illustration of Cel7A adsorbed to lignin is shown in Figure 1.6. Chapters 2-4 detail my work to understand and engineer this enzyme's affinity to cellulose and lignin.

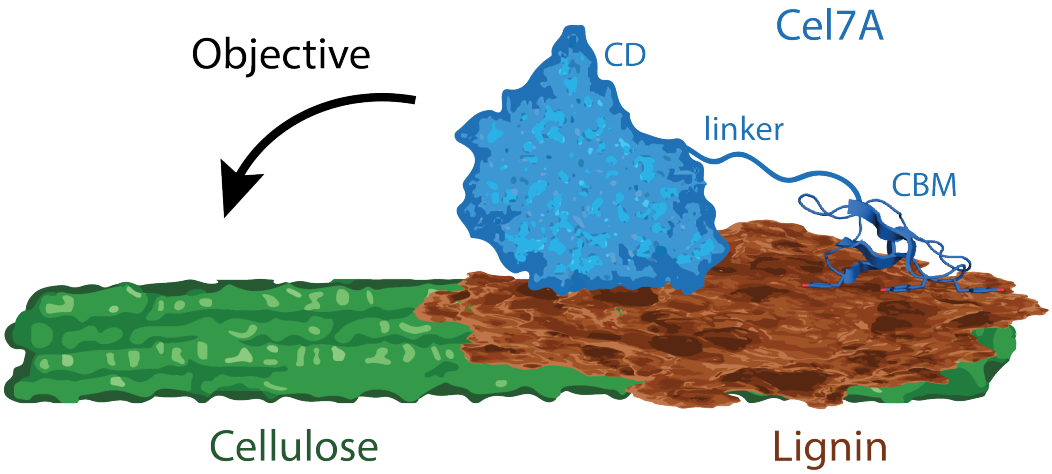

Figure 1.6 Illustration of Cel7A adsorbed to lignin.
The main goal of this work was to increase the selectivity of the enzyme for cellulose over lignin.

1.8 References

1. Barker, T., Davidson, O., Davidson, W., Huq, S., Karoly, D., Kattsov, V., Liu, J., Lohmann, U., Manning, M., and Matsuno, T. (2007) Climate change 2007: Synthesis report. *Valencia; IPPC*
2. P. Forster et al., in Climate Change 2007: The Physical Science Basis. Contribution of Working Group I o the Fourth Assessment Report of the Intergovernmental Panel on limate Change. (Cambridge University Press, Cambridge, United Kingdom, 2007).
3. Choices, N. R. C. C. O. C. (2010) *Advancing the Science of Climate Change: America's Climate Choices*, National Academies Press
4. Sun, Y., and Cheng, J. (2002) Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresour Technol* **83**, 1-11
5. Mosier, N., Wyman, C., Dale, B., Elander, R., Lee, Y. Y., Holtzapple, M., and Ladisch, M. (2005) Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresour Technol* **96**, 673-686
6. Agbor, V. B., Cicek, N., Sparling, R., Berlin, A., and Levin, D. B. (2011) Biomass pretreatment: fundamentals toward application. *Biotechnol Adv* **29**, 675-685
7. Nishiyama, Y., Sugiyama, J., Chanzy, H., and Langan, P. (2003) Crystal structure and hydrogen bonding system in cellulose I(alpha) from synchrotron X-ray and neutron fiber diffraction. *J Am Chem Soc* **125**, 14300-14306
8. Shill, K., Padmanabhan, S., Xin, Q., Prausnitz, J. M., Clark, D. S., and Blanch, H. W. (2011) Ionic liquid pretreatment of cellulosic biomass: enzymatic hydrolysis and ionic liquid recycle. *Biotechnol Bioeng* **108**, 511-520
9. Chundawat, S. P., Beckham, G. T., Himmel, M. E., and Dale, B. E. (2011) Deconstruction of lignocellulosic biomass to fuels and chemicals. *Annual review of chemical and biomolecular engineering* **2**, 121-145
10. Medie, F., Davies, G. J., Drancourt, M., and Henrissat, B. (2012) Genome analyses highlight the different biological roles of cellulases. *Nat Rev Microbiol* **10**, 227-234
11. Phillips, C. M., Beeson, W. T., Cate, J. H., and Marletta, M. A. (2011) Cellobiose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by Neurospora crassa. *ACS Chem Biol* **6**, 1399-1406
12. Montenecourt, B. S., and Eveleigh, D. E. (1977) Preparation of mutants of Trichoderma reesei with enhanced cellulase production. *Appl Environ Microbiol* **34**, 777-782
13. Klein-Marcuschamer, D., Oleskowicz-Popiel, P., Simmons, B. A., and Blanch, H. W. (2012) The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnol Bioeng* **109**, 1083-1087

14. Sjostrom, E. (2013) *Wood chemistry: fundamentals and applications*, Elsevier

15. Vanholme, R., Demedts, B., Morreel, K., Ralph, J., and Boerjan, W. (2010) Lignin biosynthesis and structure. *Plant Physiology* **153**, 895-905

16. Ishizawa, C. I., Jeoh, T., Adney, W. S., Himmel, M. E., Johnson, D. K., and Davis, M. F. (2009) Can delignification decrease cellulose digestibility in acid pretreated corn stover?. *Cellulose* **16**, 677-686

17. Kumar, L., Arantes, V., Chandra, R., and Saddler, J. (2012) The lignin present in steam pretreated softwood binds enzymes and limits cellulose accessibility. *Bioresour Technol* **103**, 201-208

18. Kumar, R., and Wyman, C. E. (2009) Access of cellulase to cellulose and lignin for poplar solids produced by leading pretreatment technologies. *Biotechnology progress* **25**, 807-819

19. Ju, X., Engelhard, M., and Zhang, X. (2013) An advanced understanding of the specific effects of xylan and surface lignin contents on enzymatic hydrolysis of lignocellulosic biomass. *Bioresour Technol* **132**, 137-145

20. Mooney, C. A., Mansfield, S. D., Touhy, M. G., and Saddler, J. N. (1998) The effect of initial pore volume and lignin content on the enzymatic hydrolysis of softwoods. *Bioresour Technol* **64**, 113-119

21. Yoshida, M., Liu, Y., Uchida, S., Kawarada, K., Ukagami, Y., Ichinose, H., Kaneko, S., and Fukuda, K. (2008) Effects of cellulose crystallinity, hemicellulose, and lignin on the enzymatic hydrolysis of Miscanthus sinensis to monosaccharides. *Biosci Biotechnol Biochem* **72**, 805-810

22. Pfeiffer, Katherine. (2014) Quantifying Aspects of Lignin-cellulase Interactions. Ph. D. Dissertation, UC Berkeley.

23. Rahikainen, J. L., Martin-Sampedro, R., Heikkinen, H., Rovio, S., Marjamaa, K., Tamminen, T., Rojas, O. J., and Kruus, K. (2013) Inhibitory effect of lignin during cellulose bioconversion: the effect of lignin chemistry on non-productive enzyme adsorption. *Bioresour Technol* **133**, 270-278

24. Rahikainen, J. L., Moilanen, U., Nurmi-Rantala, S., Lappas, A., Koivula, A., Viikari, L., and Kruus, K. (2013) Effect of temperature on lignin-derived inhibition studied with three structurally different cellobiohydrolases. *Bioresour Technol* **146**, 118-125

25. Rahikainen, J., Mikander, S., Marjamaa, K., Tamminen, T., Lappas, A., Viikari, L., and Kruus, K. (2011) Inhibition of enzymatic hydrolysis by residual lignins from softwood--study of enzyme binding and inactivation on lignin-rich surface. *Biotechnol Bioeng* **108**, 2823-2834

26. Nordwald, E. M., Brunecky, R., Himmel, M. E., Beckham, G. T., and Kaar, J. L. (2014) Charge engineering of cellulases improves ionic liquid tolerance and reduces lignin inhibition. *Biotechnol Bioeng*

27. Nakagame, S., Chandra, R. P., and Saddler, J. N. (2010) The effect of isolated lignins, obtained from a range of pretreated lignocellulosic substrates, on enzymatic hydrolysis. *Biotechnol Bioeng* **105**, 871-879

28. Nakagame, S., Chandra, R. P., Kadla, J. F., and Saddler, J. N. (2011) The isolation, characterization and effect of lignin isolated from steam pretreated Douglas-fir on the enzymatic hydrolysis of cellulose. *Bioresour Technol* **102**, 4507-4517

29. Nakagame, S., Chandra, R. P., Kadla, J. F., and Saddler, J. N. (2011) Enhancing the enzymatic hydrolysis of lignocellulosic biomass by increasing the carboxylic acid content of the associated lignin. *Biotechnol Bioeng* **108**, 538-548

30. Lai, C., Tu, M., Shi, Z., Zheng, K., Olmos, L. G., and Yu, S. (2014) Contrasting effects of hardwood and softwood organosolv lignins on enzymatic hydrolysis of lignocellulose. *Bioresour Technol* **163**, 320-327

31. Berlin, A., Balakshin, M., Gilkes, N., Kadla, J., Maximenko, V., Kubo, S., and Saddler, J. (2006) Inhibition of cellulase, xylanase and beta-glucosidase activities by softwood lignin preparations. *J Biotechnol* **125**, 198-209

32. Rahikainen, J. L., Evans, J. D., Mikander, S., Kalliola, A., Puranen, T., Tamminen, T., Marjamaa, K., and Kruus, K. (2013) Cellulase-lignin interactions-the role of carbohydrate-binding module and pH in non-productive binding. *Enzyme Microb Technol* **53**, 315-321

33. Lan, T. Q., Lou, H., and Zhu, J. Y. (2013) Enzymatic saccharification of lignocelluloses should be conducted at elevated pH 5.2--6.2. *BioEnergy Research* **6**, 476-485

34. Lou, H., Zhu, J. Y., Lan, T. Q., Lai, H., and Qiu, X. (2013) pH-Induced lignin surface modification to reduce nonspecific cellulase binding and enhance enzymatic saccharification of lignocelluloses. *ChemSusChem* **6**, 919-927

35. Yang, B., and Wyman, C. E. (2006) BSA treatment to enhance enzymatic hydrolysis of cellulose in lignin containing substrates. *Biotechnol Bioeng* **94**, 611-617

36. Lou, H., Wang, M., Lai, H., Lin, X., Zhou, M., Yang, D., and Qiu, X. (2013) Reducing non-productive adsorption of cellulase and enhancing enzymatic hydrolysis of lignocelluloses by noncovalent modification of lignin with lignosulfonate. *Bioresour Technol* **146**, 478-484

37. Börjesson, J., Engqvist, M., Sipos, B., and Tjerneld, F. (2007) Effect of poly(ethylene glycol) on enzymatic hydrolysis and adsorption of cellulase enzymes to pretreated lignocellulose. *Enzyme Microb Technol* **41**, 186-195

38. Beckham, G. T., Matthews, J. F., Bomble, Y. J., Bu, L., Adney, W. S., Himmel, M. E., Nimlos, M. R., and Crowley, M. F. (2010) Identification of amino acids responsible for processivity in a Family 1 carbohydrate-binding module from a fungal cellulase. *J Phys Chem B* **114**, 1447-1453

Chapter 2: Toward a greater understanding of lignin inhibition due to non-productive binding

2.1 Abstract

Lignin in commercially relevant biomass sources inhibits cellulase enzymes by physically blocking their access to cellulose and contributing to their non-productive adsorption and deactivation. Numerous studies have observed a correlation between the extent of delignification of biomass and the final hydrolysis yield. Delignifying biomass improves hydrolysis by removing both mechanisms of lignin inhibition. Isolating lignin's role in non-productive adsorption of enzymes has been accomplished by adding supplemental lignin to biomass or pure cellulose. The results of these studies have shown great variability in the extent of inhibition, from an 80% drop in hydrolysis yield to no effect whatsoever.

In this chapter, I first review lignin inhibition studies in the literature and make conclusions (where possible) about factors that increase or decrease lignin inhibition due to non-productive binding. I then present data and analysis on the inhibition of two homologous enzymes: natively expressed *T. reesei* Cel7A and yeast expressed *T. emersonii* Cel7A CD fused to the *T. reesei* linker and CBM. These two enzymes make an excellent case study for the impact of enzyme properties on lignin inhibition because they are highly homologous yet exhibit vastly different levels of inhibition. I end with a few recommendations for enzyme engineering strategies to reduce lignin inhibition.

2.2 Introduction

The variability in lignin inhibition due to non-productive binding stems from many sources. The hydrolysis conditions such as temperature and pH, type of enzymes employed, lignin source, and lignin isolation method all impact the level of inhibition observed. Looking across the literature to compare different studies yields no clear trends to help elucidate the conditions where lignin is most and least inhibitory. This is likely due to the plethora of factors that contribute to the observed level of inhibition that are all simultaneously changed from study to study. Fortunately, informative trends can be seen within individual studies that varied only one factor at a time.

Impact of processing conditions on lignin inhibition

Clear trends have been observed when varying the temperature, pH, and enzyme loading in inhibition studies. Other process conditions such as mixing and timing may impact the level of lignin inhibition; though they have yet to be studied.

Higher temperatures increase inhibition

There is strong evidence from multiple studies that higher temperatures increase lignin's inhibitory effect. Rahikainen et al. showed increasing inhibition with increasing temperature for enzymes with two different Cel7A catalytic domains (1). In another study, Rahikainen et al. observed severe inactivation of lignin-bound enzymes at 45°C but little inactivation at 4°C, suggesting that heat induced denaturation may take place on the surface of lignin at high temperatures (2). Zheng et al. showed that the rate of lignin adsorption for a mixture of cellulases (Accelerase 1000) is significantly slower at 50°C than 4°C, but approximately 10 fold more enzyme is adsorbed to lignin at the higher temperature (3).

Increased pH reduces adsorption and inhibition

The pH of hydrolysis also impacts the degree of lignin inhibition. Rahikainen et al. showed decreased adsorption to lignin but no change in cellulose adsorption as pH was increased (4). Lan et al. showed enhanced hydrolysis of pretreated biomass at elevated pH, and Lou et al. established that this enhancement was due to reduced cellulase-lignin interactions (5,6). Lignin contains carboxylic acids, which can be neutral or negatively charged depending on the pH. At high pH, both cellulases and lignin become more negatively charged, and may therefore repel each other electrostatically, which explains the trends seen in the literature.

Higher enzyme loadings reduce inhibition

The level of lignin inhibition has been observed to vary with enzyme loading. Zheng et al. observed a 10% drop in hydrolysis yield from microcrystalline cellulose (Avicel) in the presence of supplemental lignin at an enzyme loading of 3 FPU/g glucan but no inhibition at 15 FPU/g glucan (3). Kumar et al. found that the inhibitory effect of lignin in pretreated biomass also depended on the enzyme loading. The use of excessive cellulase loadings resulted in near complete cellulose hydrolysis of steam pretreated Douglas-fir, while at low cellulase loadings (<10 FPU/g cellulose) the presence of lignin resulted in poor hydrolysis yields (7). Lignin inhibition is likely reduced at high enzyme loadings because there is enough cellulase to saturate the surface of cellulose regardless of whether or not some enzyme is lost to lignin. At low enzyme loadings, however, the amount of enzyme adsorbed to lignin makes up a greater fraction of the total enzyme, and decreases the amount that is available to bind cellulose.

Impact of enzyme characteristics on lignin inhibition

There are a great number of diverse cellulases of potential interest to the biofuels industry. These cellulases are isolated from Fungi, Bacteria, Achaea, etc. and can have wildly different properties. Cellulase characteristic such as thermal stability, domain structure, and hydrophobicity have all been shown to impact the level of lignin inhibition. Enzyme isoelectric point, glycosylation patterns, and size are other important characteristics but there is not yet sufficient evidence to suggest they impact lignin inhibition.

More thermally stable enzymes exhibit less inhibition

Rahikainen et al. compared the inhibition of Cel7A from two different organisms. The more thermally stable enzyme from *T. emersonii* was more resistant to lignin than the corresponding enzyme from *T. reesei*. The *T. emersonii* enzyme's optimum temperature, T_{opt}, is only 1-2°C higher than *T. reesei's,* but the melting temperature, T_m, is 10°C higher than the *T. reesei* enzyme (8,9). Less stable enzymes are more likely to denaturation on the surface of lignin, greatly increasing the level of lignin inhibition observed.

CBMs increase adsorption and inhibition

Many cellulases consist of two domains: a catalytic domain (CD) and a carbohydrate-binding module (CBM) connected by a flexible linker (10). The CBM is responsible for increasing the effective concentration of the catalytic domain on the surface of its substrate (11) and has also been shown to be responsible for the majority of lignin affinity and inhibition. Cellulases expressed without the CBM, or with the CBM removed, exhibit reduced lignin affinity and inhibition (12-14), highlighting the role of the CBM in lignin inhibition. Although enzymes without a CBM are less inhibited by lignin, they also generate significantly less sugar from biomass than the corresponding enzyme with the CBM. CBMs are therefore still beneficial in hydrolysis reactions, despite their role in lignin inhibition.

Hydrophobicity increases adsorption

A few studies have found that more hydrophobic enzymes are more susceptible to lignin inhibition. In a comparison of *T. reesei* Cel7A and Cel7B, Borjesson et al. showed that Cel7B had higher affinity to lignin, perhaps due to the more hydrophobic surface on the flat face of the CBM together with an additional exposed aromatic residue on the rough face (15). Sammond et al. found that the adsorption of enzymes to lignin surfaces correlates with solvent exposed hydrophobic clusters (16). The presence of a CBM, which increases the hydrophobicity of cellulases, also greatly increases the extent of lignin inhibition.

Impact of lignin source and chemistry on inhibition

Recent advances in NMR and analytical chemistry methods have greatly increased the resolution with which lignin structures can be detected and quantified. Lignin can be analyzed in a dissolved whole cell wall sample without isolating it from other cell wall components (17). Constituent monomer ratio (syringyl/guiacyl), the relative quantities of four major lignin linkages, and quantities of ferrulate and para-coumarate can be resolved from samples of whole biomass or isolated lignin samples (18). These powerful analytical techniques have made it possible to correlate chemical functionality with cellulase adsorption to lignin, despite the complex nature of the substrate. Despite these tools, causality remains elusive, as controlling the quantity of individual functional groups without affecting the lignin structure as a whole remains a challenge.

Lignin monomer ratio impacts affinity but lignin's impact on cellulose accessibility is more important for whole biomass hydrolysis

In one recent study, the structural features of lignin derived from several different biomass sources were quantified and the isolated lignin was tested for cellulase adsorption and inhibition of cellulose hydrolysis (19). Lower S/G ratios were correlated with increased cellulase-lignin adsorption when six different biomass sources were compared, implying that lower S/G ratios should be more inhibitory. However, when the biomass was hydrolyzed, lower S/G ratios were correlated with increased hydrolysis (19). This implies that the differences in the lignin macromolecular structure within the network of cellulose and hemicellulose overshadows the effect of nonproductive adsorption of cellulases to lignin, perhaps due to differences in the physical obstruction of the cellulose.

Another study used mutant eucalyptus to compare the effect of S/G ratio on biomass recalcitrance. As in Gao et al. (19), a negative correlation was seen between S/G ratio and the 24-hour glucose yield. However, after ionic liquid pretreatment, which removes lignin, the glucose yield was identical (20). This result supports the hypothesis that the differences in digestibility of untreated samples with varying S/G ratios are due to differences in enzyme accessibility, which can overshadow differences in lignin affinity. Similar results were seen when synthetic lignins with varying monomer compositions were polymerized in vitro in the presence of cellulose. While the lignin reduced the digestibility of the cellulose, the lignin monomer composition did not impact the level of inhibition (21).

Softwood lignin is more inhibitory than hardwood lignin due to increased hydrophobicity

Other researchers have compared the adsorption of cellulases to lignins extracted from softwood and hardwood biomass samples (22). Their results showed a nearly seven-fold higher maximum adsorptive capacity of organosolv lignin derived from softwoods compared to hardwoods, and found softwood lignin to be inhibitory to Avicel hydrolysis by cellulases. Notable differences between the two lignin samples included higher surface hydrophobicity of the softwood sample, and a more negative surface charge for the hardwood lignin (22).

Biomass pretreatment impacts lignin chemistry

Biomass pretreatment adds a confounding layer to the contributions of lignin chemistry to enzyme adsorption. Pretreatment processes are designed to increase cellulose accessibility, remove lignin, and/or partially hydrolyze the carbohydrate fractions. These processes, which may involve extremes in temperature, pH, and pressure, often modify lignin structure both chemically and physically. This, in turn, affects both the digestibility of the carbohydrate, and the propensity of cellulase to adsorb to the residual lignin. Common effects of pretreatment on lignin structures include cleavage of β-aryl-ether bonds, corresponding increases in condensed structures, cleavages of lignin carbohydrate complexes (ferrulic ester linkages), and changes in S/G ratios.

Condensed lignin structures resulting from pretreatment increase adsorption and inhibition

Nakagame et al. found that increasingly severe steam pretreatments yield biomass that has more accessible cellulose, but lignin that adsorbs more cellulase (19). FTIR and 13C NMR characterization indicated that the lignin in the steam-pretreated substrates became more condensed with increasing severity, suggesting that the cellulases were adsorbed to the lignin by hydrophobic interactions. Ko et al. studied the impact of liquid hot water pretreatment of hardwood and found that increasing the severity of pretreatment resulted in a lignin with a higher propensity to adsorb cellulases. Analysis of the lignin revealed increases in the degree of C5 and C-β condensation, as evidenced by increased glass transition temperature of the lignin and FT-IR analysis (20). Other pretreatments also impact lignin structure and the ability of enzymes to adsorb to the lignin surface.

2.3 Methods

Materials

Microcrystalline cellulose PH-105 (Avicel), Protease from *Streptomyces griseus,* 4-methylumbelliferyl β-D-cellobioside, Tween 20, cellulase from *T. reesei* (Celluclast) and papain were purchased from Sigma. Accelerace 1500 was a gift from Dupont. *A. niger* β-glucosidase was purchased from Megazyme. Acid-pretreated *Miscanthus* was produced by two-step dilute acid pretreatment on the pilot scale by Andritz, Glens Falls, NY. Solids were washed with water to neutral pH and lyophilized before lignin preparation.

Lignin preparation and analysis

Lignin residues were isolated from acid pretreated *Miscanthus* by enzymatic hydrolysis followed by *S. griseus* protease treatment to remove bound enzyme. Pretreated, rinsed, lyophilized biomass was ball milled for 5 min (Kleco ball mill, Visalia, CA). Enzymatic hydrolysis was carried out at 50°C, pH 5.0, with 4 wt% biomass and 0.25 ml Accelerase 1500 per gram of biomass. The reaction mixtures were centrifuged and the supernatant replaced with fresh buffer and enzyme (0.25 ml Accelerase/g original biomass) every 24 hrs. After 7 days of hydrolysis, the solids were separated by centrifugation and washed with acidified water (pH adjusted to 3.5 with HCl to avoid lignin solubilization). Bound protein was removed by an overnight protease treatment at 37°C, pH 7.5, with 2wt% solids and 0.3 mg/ml protease. The solids were then washed with acidified water and heated in a boiling water bath for 60 min to deactivate any remaining protease. The solids were extensively washed with acidified water and freeze-dried. See Chapter 3 for lignin analysis.

Purification of T. reesei Cel7A

Cel7A was purified from Celluclast using a Q Sepharose anion exchange column followed by a MonoQ ion exchange column. Prior to injection to the MonoQ column, the crude Cel7A was treated with 0.1% Tween 20 to disrupt non-specific interactions that were found to exist between Cel7A and Cel7B. Fractions from the MonoQ were checked

for purity by SDS-PAGE and an assay for contaminant endoglucanase activity. Assays contained 2% carboxymethyl cellulose and roughly 10 μM enzyme in 100 mM sodium acetate, pH 4.5. Assays were incubated at 50 °C overnight and endoglucanase activity was assessed by adding 2,4-dinitrosalycilic acid, incubating at 90 °C for 5 minutes, and visually inspecting for the presence of brown color in the reactions. Fractions showing no endoglucanase activity were pooled and buffer exchanged into 50 mM sodium acetate, pH 5.0.

Isolation of the T. reesei CD

The *T. reesei* CD was cleaved from the linker and CBM using the protease papain. Papain was activated by incubating a 28 $mg \cdot ml^{-1}$ solution with 2 mM dithiothreitol and 2 mM ethylenediamine tetraacetic acid in 100 mM sodium phosphate, pH 7.0 for 30 minutes at room temperature. Cleavage reactions contained a w/w ratio of 1:100 papain to Cel7A in 50 mM sodium acetate, pH 5.0 and were incubated overnight at room temperature with stirring. The CD was purified from the protease and CBM using a Q Sepharose anion exchange column.

Te-Tr enzyme production and purification

For Te-Tr chimera and Te CD production, 1 L of SC-Trp medium was inoculated with *S. cerevisiae* containing the Cel7A gene and grown for 3 days at 30°C, 220 rpm. Cultures were spun down at 5,000 x g for 5 min and resuspended in YPD medium supplemented with 500 mM Cu_2SO_4 and cultured for 3 days at 25°C, 220 rpm. Enzyme purification was carried out using a two-step procedure beginning with anion exchange chromatography. Active fractions were combined and further purified by hydrophobic interaction chromatography (HIC). Pure fractions (single band via SDS-PAGE) were combined and buffer exchanged into 50 mM sodium acetate buffer, pH 5.0.

Enzyme quantification

Purified enzymes were quantified by absorbance at 280 nm using the calculated molar extinction coefficient of 83,210 $M^{-1}cm^{-1}$ for the full *T. reesei* enzyme, 78,090 $M^{-1}cm^{-1}$ for the *T. reesei* catalytic domain, 76,240 $M^{-1}cm^{-1}$ for the Te-Tr fusion enzyme and 71,120 $M^{-1}cm^{-1}$ for the Te catalytic domain. Unpurified enzyme and enzyme remaining in the supernatant after binding assays were quantified by their activity on a soluble substrate, 4-methylumbelliferyl β-D-cellobioside (MUC). The specific activity of 10.2 and 0.95 μmols MU $(\mu mol\ enzyme*min)^{-1}$ was measured for enzymes containing the *T. emersonii* CD and *T. reesei* CD, respectively. Assays were conducted in a total volume of 100 μl with 1 mM MUC, pH 5.0, at 45°C for 8 min, followed by a denaturation step at 98°C for 2 min. Sodium hydroxide was added to a final concentration of 0.1 M and the fluorescence was measured on a Paradigm plate reader (Beckman Coulter) with an excitation of 360 nm and emission of 465 nm.

Adsorption isotherms

Adsorption isotherm measurements were performed in 50 mM sodium acetate buffer (pH 5) using 10 mg/ml of Avicel or lignin and initial enzyme concentrations ranging from 0.015-4.0 μM. Experiments were conducted in a total volume of 70 uL in a rotating shaker (300 rpm) at room temperature. An initial time course with the wild type

enzyme was measured to determine the time required for binding to reach equilibrium. Equilibrium was reached after approximately 30 min. Subsequent assays were performed for 1 hr. After equilibration, solids were separated by centrifugation and the supernatant was analyzed for free, active enzyme using MUC. The amount of bound enzyme was calculated from the difference between added enzyme and enzyme remaining in solution. A one-site Langmuir adsorption isotherm (Equation 2) was fit to the binding data by non-linear least squares optimization to determine K_A, Γ_{max}, and their corresponding standard deviations.

$$\Gamma = \frac{\Gamma_{\max} K_A C}{1+K_A C} \qquad (1)$$

Hydrolysis of Avicel supplemented with lignin and Miscanthus

Hydrolysis reactions were performed with 10 mg/ml of Avicel, 10 mg/ml Avicel supplemented with 10 mg/ml of isolated lignin, or 20 mg/ml acid-pretreated *Miscanthus* in 50 mM sodium acetate buffer, pH 5. All reactions were supplemented with 40 U/mL of *A. niger* β-glucosidase. Reactions were carried out in a thermally controlled incubator with end-over-end mixing. Reaction mixtures were heated to 98 °C for 3 minutes to stop the reaction, filtered, and frozen until analysis for glucose by HPLC (Shimadzu).

Predicting the amount of enzyme bound to cellulose during hydrolysis

Equations 3-5 were used to predict the amount of cellulose-bound enzyme in the presence of lignin.

$$\Gamma_C = \frac{\Gamma_{\max,C} K_{A,C} C}{1+K_{A,C} C} \qquad (2)$$

$$\Gamma_L = \frac{\Gamma_{\max,L} K_{A,L} C}{1+K_{A,L} C} \qquad (3)$$

$$C_{total} = C + S_C \Gamma_C + S_L \Gamma_L \qquad (4)$$

where Γ_C and Γ_L are the amounts of enzyme bound per gram of solid to cellulose and lignin, respectively; $K_{A,C}$ and $K_{A,L}$ are the Langmuir binding constants to cellulose and lignin, respectively; $\Gamma_{max,C}$ and $\Gamma_{max,L}$ are the amounts of adsorbed protein at saturation to cellulose and lignin, respectively; C is the concentration of protein remaining free in solution; C_{total} is the initial protein concentration, and S_C and S_L are the concentrations of cellulose and lignin, respectively.

2.4 Results

The importance of Cel7A

Cel7A is the most abundant cellulase in commercial enzyme cocktails and is responsible for the greatest hydrolytic potential. Cel7A comprises up to 40 wt% of the secretome of *Neurospora crassa*, and removing it from the genome results in severe growth reduction on cellulose. Moreover, when Cel7A is removed from the *Trichoderma*

reesei genome, the filter paper activity of its culture filtrate is reduced by 70% (21). Thus, Cel7A is a key enzyme in these and similar cellulase systems and is a practical target for improvement by protein engineering.

Most commercial enzyme mixtures for lignocellulose hydrolysis are produced by *T. reesei*. *T. reesei* enzymes engineered for improved performance can therefore directly replace their wild type counterpart in commercial cocktails. *T. reesei* Cel7A consists of two domains: a catalytic domain (CD) and a carbohydrate binding module (CBM) connected by a flexible, highly glycosylated linker. It is a processive cellobiohydrolase that generates cellobiose as the primary product.

Lignin inhibition of T. reesei Cel7A

T. reesei Cel7A is greatly inhibited by non-productive adsorption to lignin. Figure 2.1 shows the glucose generated over time by the full *T. reesei* enzyme or the isolated CD from Avicel in the presence and absence of lignin. The full enzyme generated 50% less glucose in the presence of added lignin while the CD generated 30% less glucose. Though less inhibited than the full enzyme, the CD generated 5 fold less glucose from pure Avicel and 3.7 fold less glucose from Avicel with added lignin, underscoring the importance of the CBM in cellulose hydrolysis.

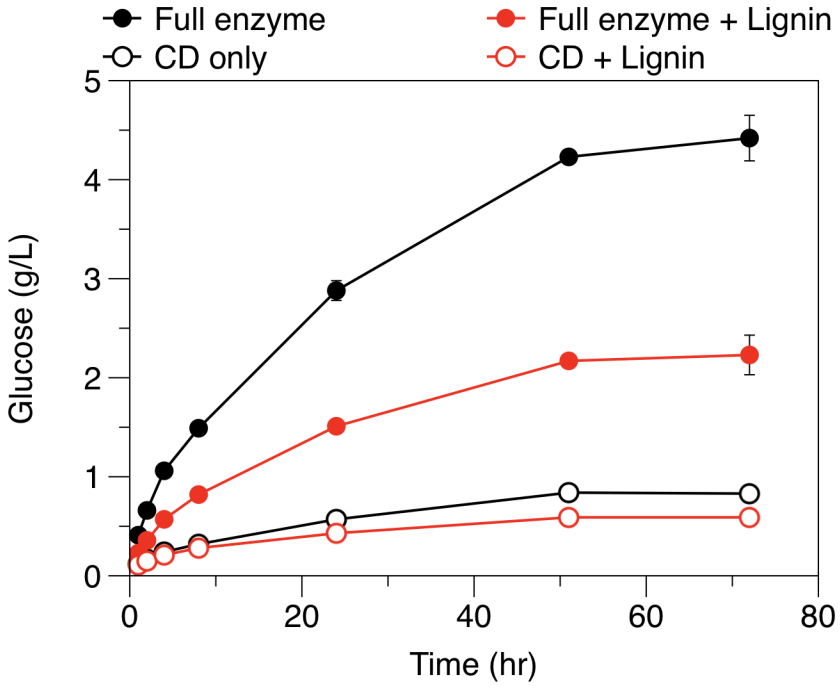

Figure 2.1 Hydrolysis of Avicel and Avicel with added lignin by T. reesei Cel7A. Glucose produced from 10 g/L Avicel or 10 g/L Avicel with 10 g/L lignin. All reactions were supplemented with 40 U/ml *A. niger* β-glucosidase. Reactions were performed at 50°C, pH 5.0, with 1.5 μM Cel7A. Glucose was measured via HPLC. Values presented are means of duplicate samples and errors are standard deviations.

Lignin and cellulose affinity of T. reesei Cel7A

Cellulase binding is often described by a Langmuir adsorption model. This model assumes fully reversible binding, no interactions between proteins on the surface, and a single type of binding site. Although the Langmuir model does not describe the complex mechanisms of cellulase adsorption to cellulose or lignin, it is a useful tool to predict levels of non-productive binding and enables the comparison of different enzymes. The Langmuir isotherm data for the full *T. reesei* enzyme and CD are shown in Figure 2.2. Interestingly, the full enzyme adsorbs to the lignin more than Avicel. In contrast, the isolated CD is more selective for Avicel than lignin.

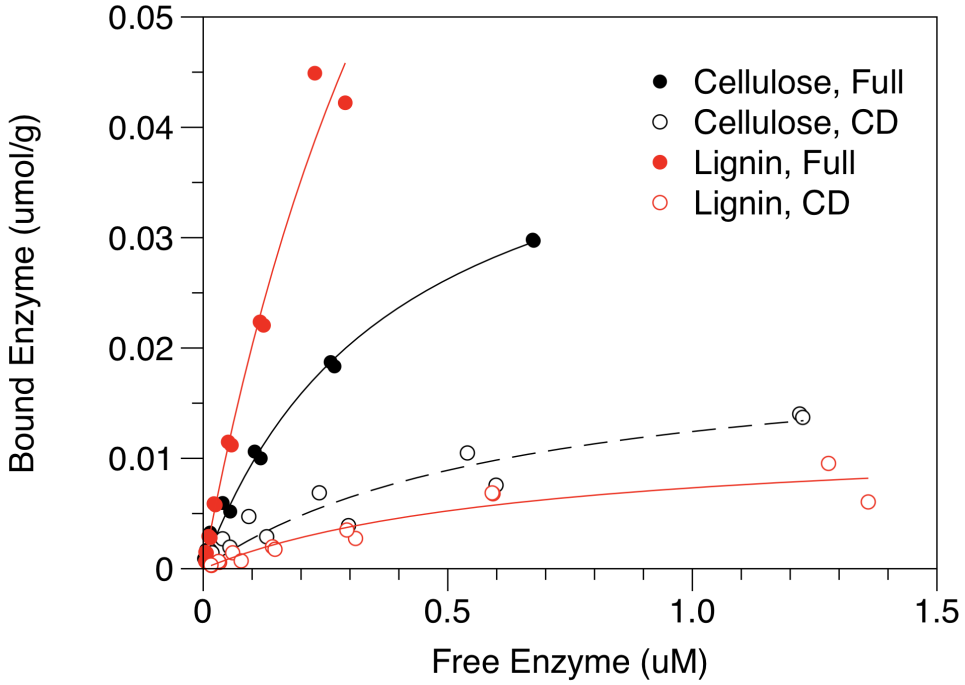

Figure 2.2 Langmuir isotherms of *T. reesei* Cel7A adsorption.
Isotherms were measured with 10 g/L Avicel or lignin at room temperature. The lines show the fit to a one-site Langmuir binding isotherm.

Inhibition of T. reesei compared to T. emersonii Cel7A

T. emersonii Cel7A is highly homologous to *T. reesei* Cel7A except it naturally lacks a linker and CBM. Unlike the *T. reesei* enzyme, it expressed well in *S. cerevisiae*, making it an attractive construct for engineering. We produced *T. emersonii* Cel7A with the *T. reesei* linker and CBM to generate a chimera (Te-Tr chimera) for CBM and linker engineering (detailed in Chapters 3 and 4). Figure 2.3 compares the inhibition of the yeast expressed Te-Tr chimera and Te CD to the natively expressed *T. reesei* full enzyme and CD. The *T. reesei* enzymes are much more inhibited by lignin than the *T. emersonii* enzymes. Under the hydrolysis conditions employed in Figure 2.3, the full *T. reesei* enzyme generates 80% less glucose when lignin is added and the Tr CD generates 40% less. In contrast, the Te-Tr chimera generates 15% less glucose when lignin is added and the Te CD generates the same amount of glucose in the presence and absence of lignin.

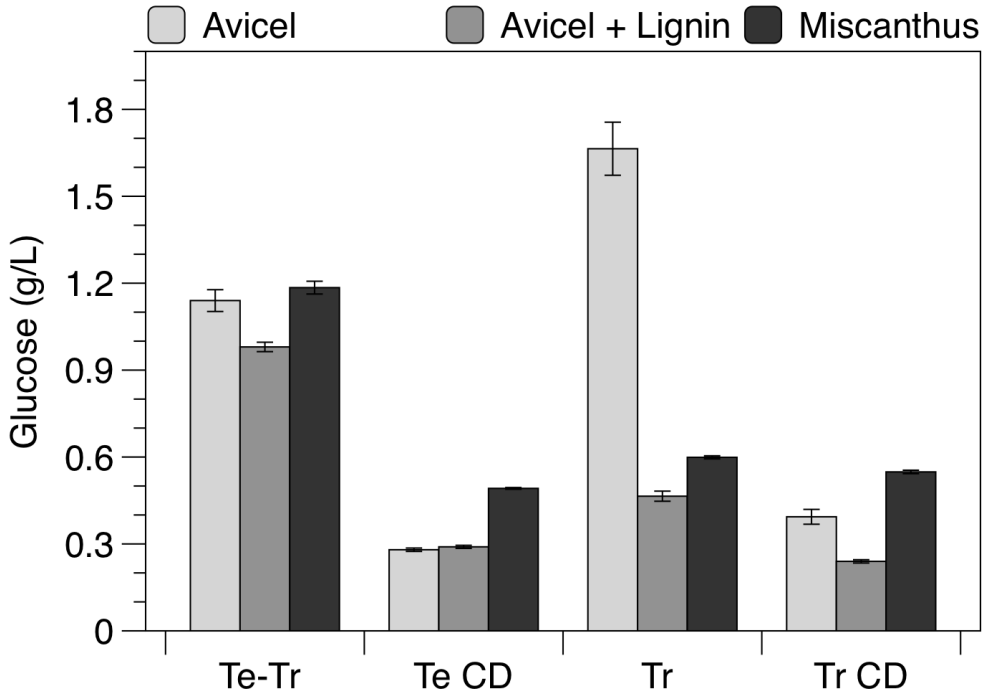

Figure 2.3 Comparison of inhibition for *T. reesei* Cel7A and Te-Tr chimera. Glucose produced after 48 hours with 10 g/L Avicel, 10 g/L Avicel with 10 g/L lignin, or 20 g/L *Miscanthus*. All reactions were supplemented with *A. niger* β-glucosidase. Reactions were performed at 60°C for *T. emersonii* (Te) enzymes and 55°C for *T. reesei* (Tr) enzymes with 0.5 μM Cel7A in Avicel reactions and 1.0 μM Cel7A in Miscanthus reactions. Values presented are means of duplicate samples and errors are standard deviations.

Lignin and cellulose affinity of T. emersonii Cel7A

The Langmuir isotherm data and fits for the *T. emersonii* chimera and CD are shown in Figure 2.4. Both *T. emersonii* constructs adsorb more than the corresponding *T. reesei* construct. This is likely due to the fact that *T. emersonii* Cel7A exists naturally without a CBM and the CD must have higher affinity to compensate. The lignin and cellulose affinity of the Te-Tr chimera are very similar, while the Te CD adsorbs more strongly to cellulose than lignin. The Langmuir parameters for the four enzymes studied are presented in Table 2.1 for comparison.

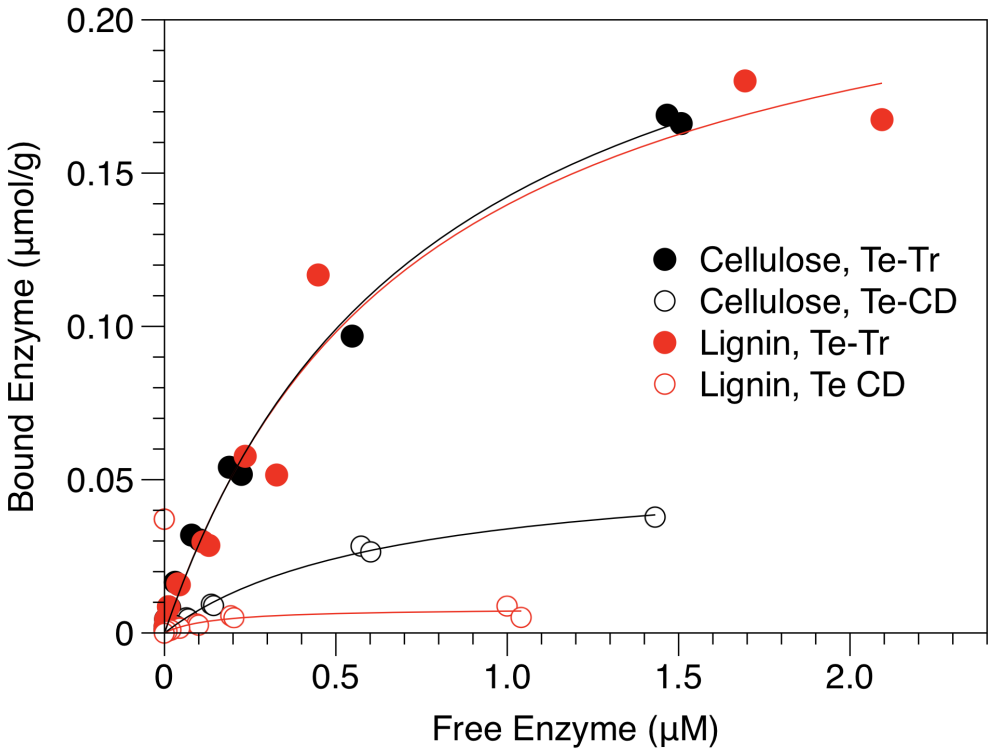

Figure 2.4 Langmuir isotherms of Te-Tr and Te CD adsorption.
Isotherms were measured with 10 g/L Avicel or lignin at room temperature. The lines show the fit to a one-site Langmuir binding isotherm.

Table 2.1 Avicel and lignin Langmuir parameters for TeCel7A and TrCel7A

	Avicel			Lignin		
	Γ_{max} µmol/g	K_A μM^{-1}	α L/g	Γ_{max} µmol/g	K_A μM^{-1}	α L/g
***T. reesei* Cel7A**	0.0466	2.58	0.120	0.137	1.73	0.237
***T. reesei* Cel7A CD**	0.0205	1.54	0.0316	0.0122	1.50	0.0183
***T. emersonii*$_{CD}$- *T. reesei*$_{link-CBM}$**	0.252	1.29	0.325	0.242	1.36	0.329
***T. emersonii* CD**	0.056	1.54	0.0862	0.00815	6.91	0.0563

Predicting the inhibition of TrCel7A and TeCel7A using the Langmuir model

The first step in the hydrolysis of cellulose by cellulases is adsorption to the surface of cellulose. Comparing the amount of cellulose bound enzyme under different conditions should therefore lend insight to the relative rates of hydrolysis. The Langmuir model, with the parameters in Table 2.1, can predict the amount of cellulose bound enzyme for each of the constructs studied under different hydrolysis conditions.

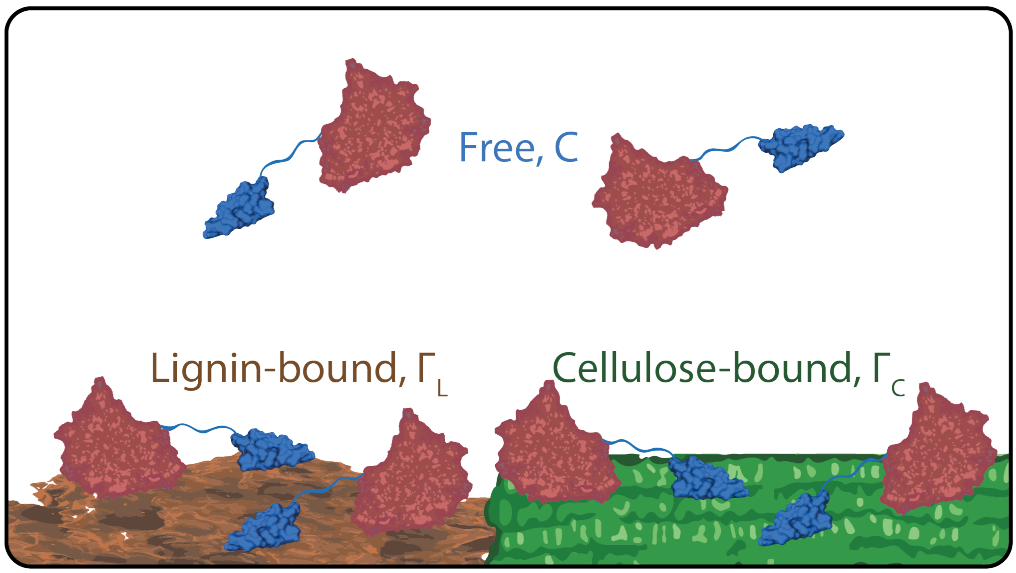

Figure 2.5 Illustration of cellulose hydrolysis in the presence of lignin. Using Langmuir model parameters, the amount of cellulose-bound enzyme, Γ_C, can be calculated and compared for different enzymes and conditions.

The amount of enzyme predicted to be bound to cellulose under the hydrolysis conditions used in Figure 2.3 is shown in Figure 2.6. If we assume the calculated cellulose-bound enzyme is proportional to the hydrolysis yield, the model slightly over predicts the impact of lignin on the Te-Tr chimera and under predicts the impact of lignin on the full *T. reesei* enzyme and CD. The model also predicts that the *T. emersonii* chimera should outperform the full *T. reesei* enzyme in the hydrolysis of pure Avicel, which is not the case. The discrepancy between the model predictions and hydrolysis data may be due to the different temperatures used to fit the Langmuir binding parameters and measure hydrolysis. The Langmuir binding experiments were performed at room temperature, while the hydrolysis experiments were performed at 55 or 60 °C. The Langmuir model also does not account for irreversible binding. The *T. reesei* enzyme is known to display partially irreversible binding to lignin, which could vastly increase the amount of inhibition seen.

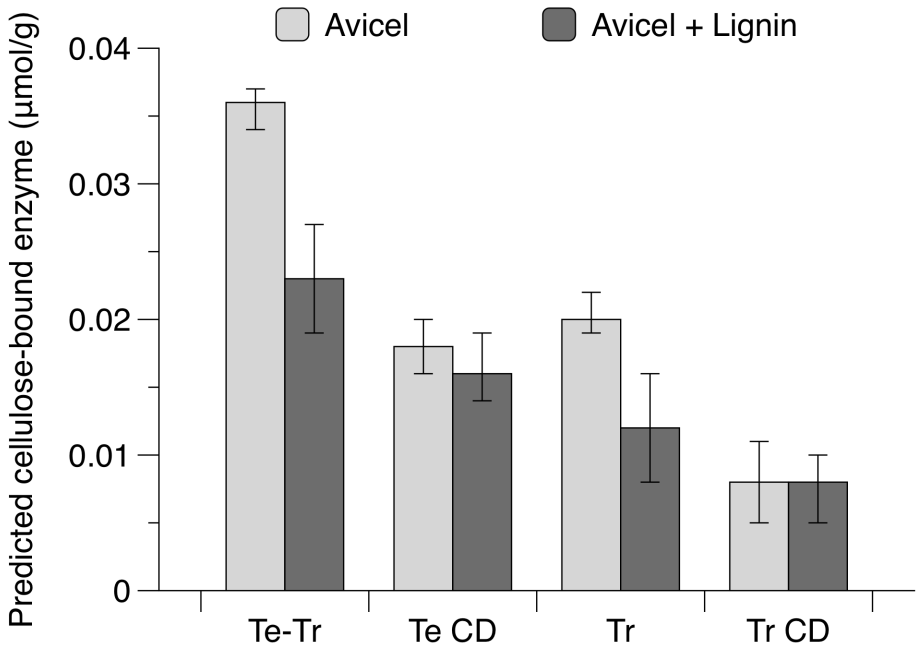

Figure 2.6 Predicted amounts of cellulose-bound enzyme.
Bound enzyme is calculated using the Langmuir parameters in Table 2 for a solution with 0.5 µM total enzyme, 10 g/L Avicel, with or without 10 g/L lignin. Error is presented as the range of values possible given the standard deviation in the Langmuir affinity constant (K_A) and saturation adsorption constant (Γ_{max}).

The Langmuir model predicts that the addition of lignin will not change the amount of enzyme bound to cellulose for both the *T. emersonii* and *T. reesei* CDs. This agrees with the inhibition data for the Te CD, which generates the same amount of glucose regardless of the presence of lignin. It does not explain the large drop in glucose yields when lignin is added to hydrolysis reactions with the Tr CD, however. Irreversible adsorption of the Tr CD, which is not accounted for in the Langmuir model, may explain this result.

2.5 Discussion

The rest of this chapter analyzes the structural and chemical differences in the two highly homologs catalytic domains from *T. reesei* and *T. emersonii*. The goal of this analysis is to understand why the *T. reesei* CD is so much more sensitive to lignin than the *T. emersonii* CD. This understanding can guide rational engineering strategies aimed at reducing the lignin inhibition of *T. reesei* Cel7A.

Structure of TeCel7A vs TrCel7A

The structural alignment of TeCel7A and TrCel7A CD is shown in Figure 2.7. The two structures are nearly identical, with only minor deviations in loop regions. It is

therefore unlikely that differences in structure are responsible for the differences in lignin inhibition.

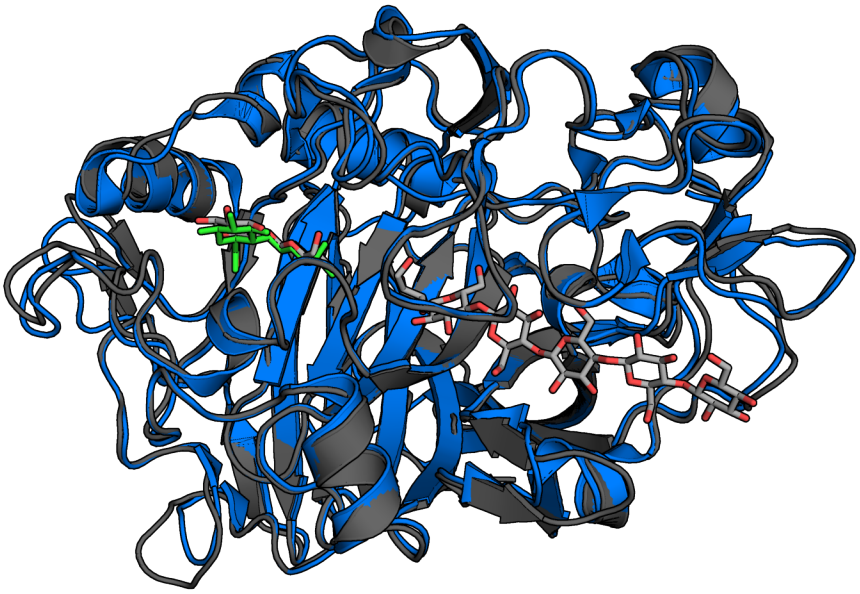

Figure 2.7 Alignment of TeCel7A and TrCel7a.
TeCel7A is shown in blue with a green cellobiose bound. TrCel7A is shown in gray with a gray cellobiose and cellohexaose bound. PBD structures: 3PFX (Te) and 7CEL (Tr). Image made with pymol.

Size, overall charge, and stability of TeCel7A vs TrCel7A

The properties of TeCel7A and TrCel7A are summarized in Table 2.2. The enzymes are very similar in size but differ slightly in their isoelectric point and temperature stability. The *T. reesei* enzyme has a higher isoelectric point and is less negatively charged at pH 5 (the hydrolysis pH) than the *T. emersonii* enzyme. Lignin adsorption decreases with increasing pH (6,22) due to increased repulsive electrostatic interactions between the negatively charged enzymes and carboxylic acid groups in lignin. A difference in enzyme net charge could therefore lead to different levels of lignin inhibition. The *T. emersonii* enzyme's greater number of negative charges may lead to more repulsive electrostatic interactions than are present with the *T. reesei* enzyme, conferring lignin resistance to the Te enzyme.

The *T. emersonii* enzyme is more thermally stable than the *T. reesei* enzyme, as evidenced by a 9°C difference in melting temperatures. This may contribute to the differences seen in inhibition because the less stable enzyme is more likely to denature on the surface of lignin, which will greatly increase the level of inhibition.

Table 2.2 Properties of TeCel7A and TrCel7A.
The molecular weight, pI, and net charge were calculated using the Biosynthesis peptide property calculator. Temperature optimum and melting temperature from (8,9). Hpatch scores were calculated as in (16) and error is presented as the standard deviation of the scores determined using three different crystal structures (*T. reesei*- 7CEL, 1CEL, 4P1J; *T. emersonii*- 3PFZ, 3PFX, 3PFJ)

	MW (kDa)	**pI**	**T_{opt}**	**T_m (C)**	**Hpatch score**	**Net charge (pH 5)**
***T. reesei* Cel7A**	52.3	4.26	60	65	13.4 ± 0.9	-15
***T. reesei* Cel7A CD**	46.3	4.02	60	65	6.9 ± 0.9	-19
***T. emersonii*$_{CD}$-*T. reesei*$_{link-CBM}$ chimera**	53.0	3.95	65	74	12.9 ± 0.3	-21
***T. emersonii* CD**	47.0	3.78	65	74	6.3 ± 0.3	-25

Hydrophobicity of TeCel7A vs TrCel7A

Hydrophobicity has been shown to be a significant contributor to lignin affinity and hydrophobic surface properties have been used to predict the level of lignin adsorption for different enzymes (16). To characterize a protein's hydrophobicity, Lijnzaad et al. developed a method to delineate contiguous hydrophobic patches on a protein surface (23), and Jacak et al. incorporated a similar method into the Rosetta protein design software, adding a scoring function specifically designed to identify larger hydrophobic patches (24). The Rosetta hydrophobic patch (hpatch) score works by assigning a score to each identified patch, with scores increasing exponentially with increasing patch size. Sammond et al. showed a correlation between the hpatch score and lignin adsorption (16).

The hpatch scores are compared for the *T. ressei* and *T. emersonii* Cel7A enzymes in Table 2.2. For comparison, the hpatch score for BSA, which adsorbs very strongly to lignin, is 34.9. The hpatch scores for the CDs are determined by averaging the results determined with three different crystal structures of each protein. The hpatch score for the full protein was calculated by adding the hpatch score of the CD and CBM. There is no significant difference between the hpatch scores of the *T. reesei* and *T. emersonii* enzymes. Therefore, their large differences in inhibition behavior cannot be explained by differences in hpatch score. This result shows that although hpatch scores have been shown to correlate with lignin adsorption (16), they are not able to predict inhibition behavior and therefore are not likely to aid engineering efforts to reduce inhibition.

Location of positive, negative, and aromatic amino acids in TeCel7A vs TrCel7A

Thus far I have examined the differences in bulk properties of the two enzymes. The location of the charged and hydrophobic amino acids also likely impacts lignin affinity. Figures 2.8 shows the structures of the two enzymes with aromatic, positively

charged, and negatively charged residues highlighted. Figure 2.9 shows the structures with each residue colored by hydrophobicity.

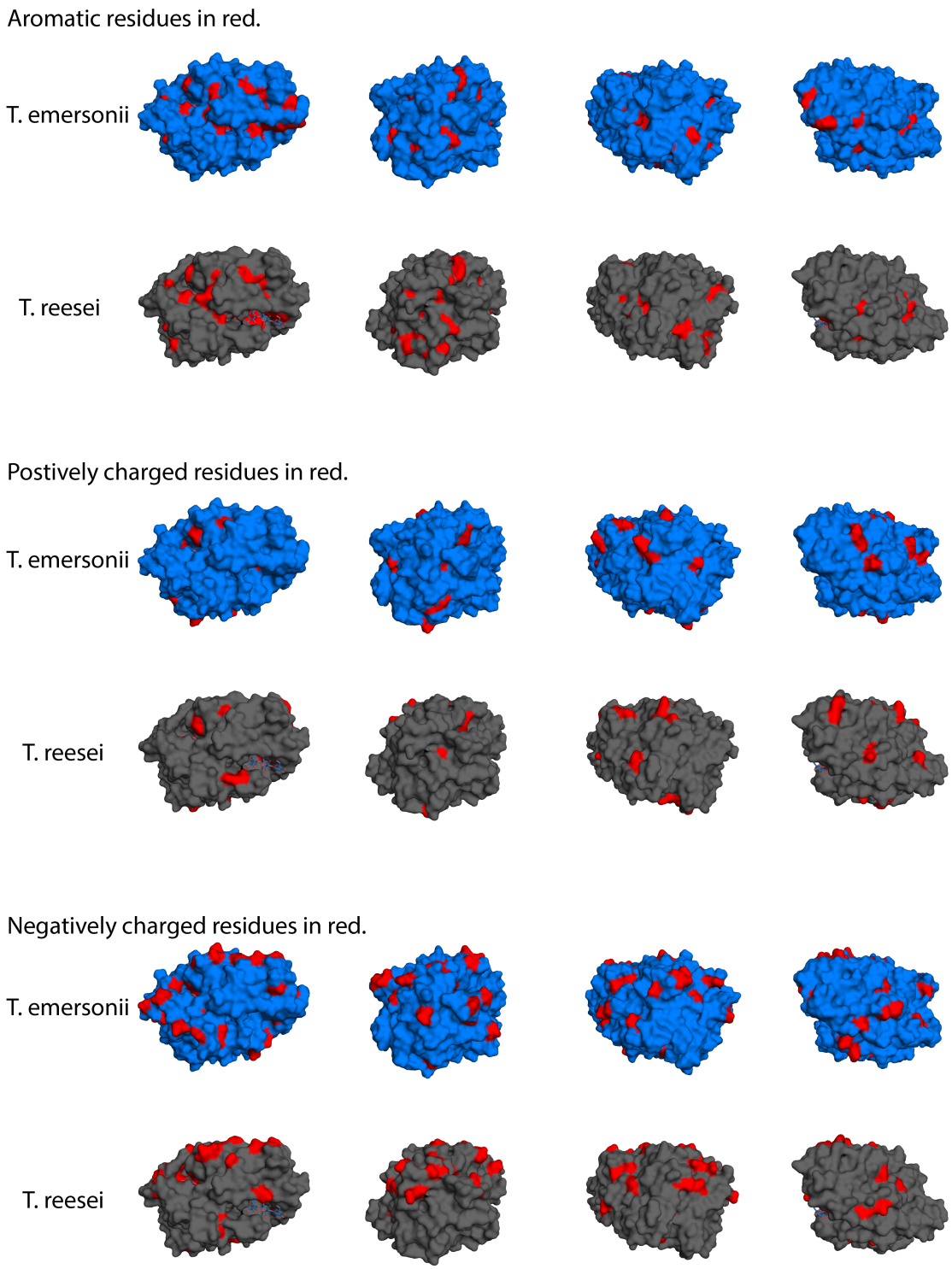

Figure 2.8 Surface exposed aromatic, positive, and negative residues.

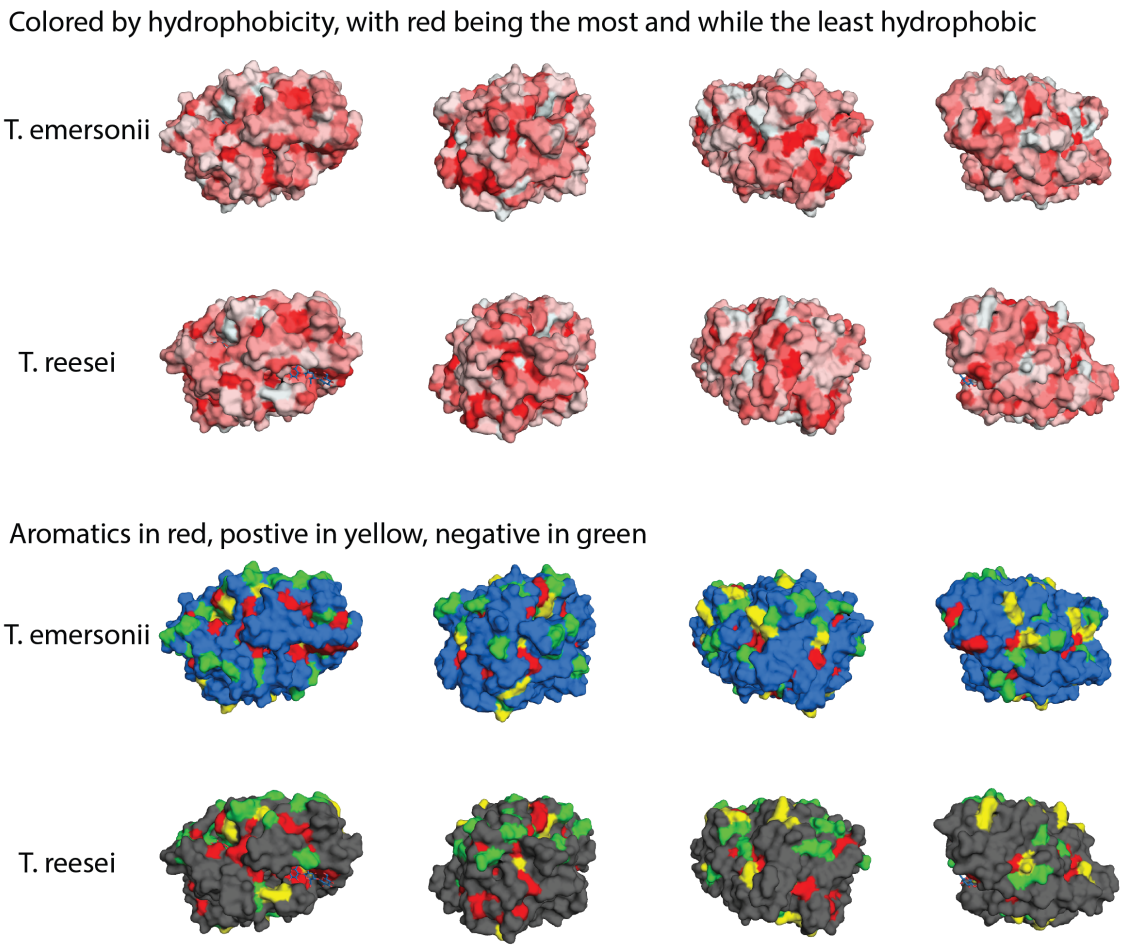

Figure 2.9 Cel7A structures colored by hydrophobicity.

Analyzing the differences in surface exposed aromatic, positive, and negative residues between TeCel7A and TrCel7A may lend insight into changes to make to the TrCel7A CD for lignin resistance. Figure 2.10 shows an alignment of their primary amino acid sequence, highlighting differences in charge and hydrophobicity between the two enzymes, both known to be important for lignin affinity.

The primary sequence alignment highlights 10 differences in hydrophobicity and 7 of them are deemed to be more favorable (less hydrophobic) in the *T. reesei* enzyme. Therefore, it is unlikely that the lignin resistance of TeCel7A compared to TrCel7A results from differences in hydrophobicity. There are 44 highlighted differences in charge, 27 of them favorable for the TeCel7A while only 17 of them are favorable for the TrCel7A. The difference in charge was deemed to be favorable for the enzyme containing a negative charge or missing a positive charge. Mutating the TrCel7A at the locations where TeCel7A has beneficial differences in charge may be a promising strategy to engineer a lignin resistant TrCel7A.

Figure 2.11 shows the locations of the differences that are deemed to be beneficial to the *T. emersonii* enzyme in the structure of the *T. reesei* enzyme. Differences in charge are highlighted in blue while differences in hydrophobicity are highlighted in green.

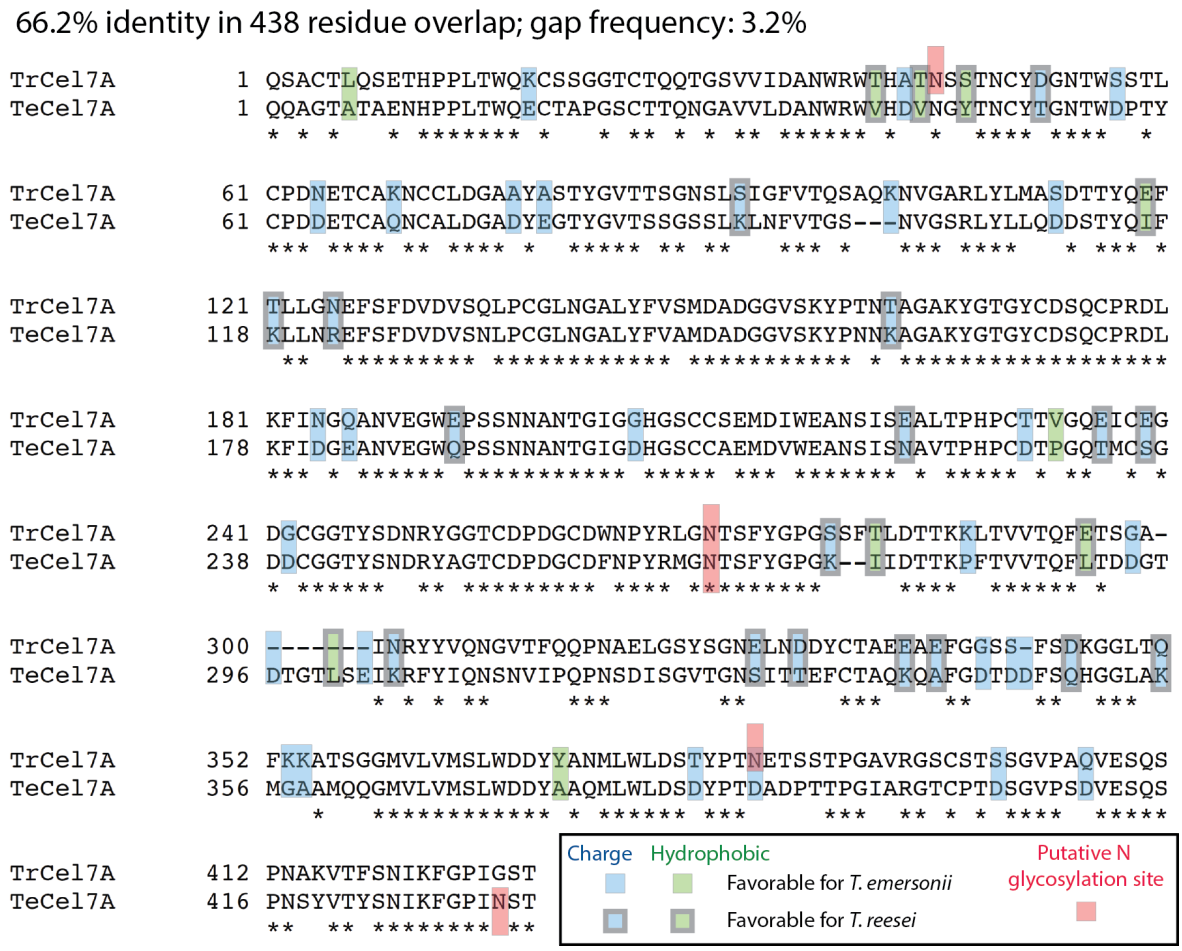

Figure 2.10 Sequence alignment of TrCel7A and TeCel7a.
Differences in charged and hydrophobic amino acids are highlighted.

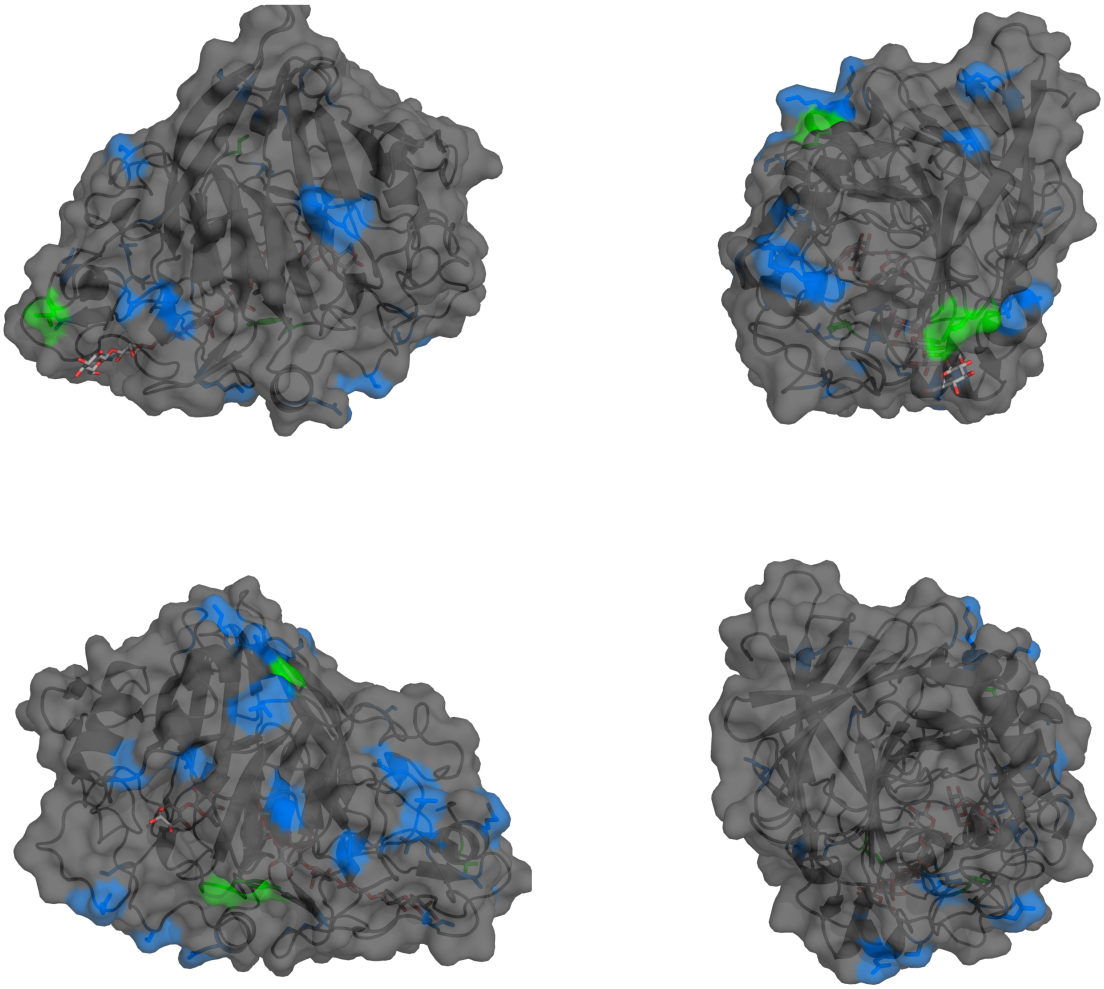

Figure 2.11 *T. reesei* Cel7A CD structure with mutation targets highlighted. Residues that are different and predicted to increase lignin resistance in *T. emersonii* Cel7A are highlighted. Differences in charge are shown in blue and differences in hydrophobicity are highlighted in green.

Glycosylation of TrCel7A vs. TeCel7A

TrCel7A contains three putative N-glysoylation sites: N45, N270, and N384. Stals et al. characterized the predominant N-glycans on TrCel7A expressed in the industrial hyperproducing strains Rut-C30 and RL-P37, the strains that produced the TrCel7A used in our inhibition studies, and found $(ManP)_{0-1}GlcMan_{7-8}GlcNAc_2$ residues (25). TeCel7A contains two putative N-glycosylation sites, N267 and N431, and in the crystal structure two GlcNAc residues have been detected linked to N267 and one to N431 (26). However, the TeCel7A in the crystal structure was produced in the native host, *Talaromyces emersonii*, while the TeCel7A used in our inhibition studies was expressed in *S. cerevisiae*. Therefore, the exact glycan structures on the TeCel7A used in our studies likely do not match the glycans in the crystal structure. O-glycosylation sites are more difficult to predict and characterize than N-glycosylation sites. Consequently, the O-glycosylation sites in TrCel7A and TeCel7A have not been well characterized.

Differences in the glycosylation patterns of TrCel7A and TeCel7A may have caused the observed difference in lignin inhibition, though the impact of glycosylation on lignin affinity has not been well studied. Deglycosylating each cellulase and studying the resulting impact on lignin inhibition would be a good first step toward determining the importance of glycosylation. If glycosylation significantly impacts lignin inhibition, Cel7A could be engineered to contain more or fewer glycosylation sites to confer lignin resistance.

2.6 Conclusions

The *T. reesei* and *T. emersonii* Cel7A enzymes are highly homologous yet exhibit drastically different levels of lignin inhibition. Replacing the native TrCel7A with the TeCel7A in fungal production strains is difficult due to differences in expression levels. Exploring the differences in their properties can inform engineering strategies to improve the *T. reesei* enzyme. The TrCel7A and TeCel7A enzymes have nearly identical structures and hpatch scores, but they differ in thermal stability, surface charge, and glycosylation patterns. Promising strategies for engineering *T. reesei* Cel7A for lignin resistance include 1) engineering for enhanced thermal stability, 2) mutating the residues highlighted in Figures 2.10 and 2.11 to increase the negative charge on the enzyme's surface, and 3) altering the number or pattern of glycosylation sites.

High throughput screening platforms have been developed for engineering more stable cellulases, though expression of *T. reesei* Cel7A in a high throughput host remains challenging. There are 25 residues highlighted in Figure 2.11 for mutation to alter the surface properties of TrCel7A. Mutating each location individually or in groups may lead to variants with improved lignin resistance, and the relatively limited set of mutations is amenable to expression in *T. reesei*. Preliminary studies to determine the impact of glycosylation on lignin inhibition are needed before engineering the glycosylation patterns for improved tolerance to lignin. Nevertheless, glycosylation engineering provides another tool to generate an improved enzyme. The insights presented in this Chapter will hopefully lead to the design of a lignin resistant, industrially relevant Cel7A.

2.7 References

1. Rahikainen, J. L., Moilanen, U., Nurmi-Rantala, S., Lappas, A., Koivula, A., Viikari, L., and Kruus, K. (2013) Effect of temperature on lignin-derived inhibition studied with three structurally different cellobiohydrolases. *Bioresour Technol* **146**, 118-125
2. Rahikainen, J., Mikander, S., Marjamaa, K., Tamminen, T., Lappas, A., Viikari, L., and Kruus, K. (2011) Inhibition of enzymatic hydrolysis by residual lignins from softwood--study of enzyme binding and inactivation on lignin-rich surface. *Biotechnol Bioeng* **108**, 2823-2834
3. Zheng, Y., Zhang, S., Miao, S., Su, Z., and Wang, P. (2013) Temperature sensitivity of cellulase adsorption on lignin and its impact on enzymatic hydrolysis of lignocellulosic biomass. *J Biotechnol* **166**, 135-143
4. Rahikainen, J. L., Evans, J. D., Mikander, S., Kalliola, A., Puranen, T., Tamminen, T., Marjamaa, K., and Kruus, K. (2013) Cellulase-lignin interactions-the role of carbohydrate-binding module and pH in non-productive binding. *Enzyme Microb Technol* **53**, 315-321
5. Lan, T. Q., Lou, H., and Zhu, J. Y. (2013) Enzymatic saccharification of lignocelluloses should be conducted at elevated pH 5.2--6.2. *BioEnergy Research* **6**, 476-485
6. Lou, H., Zhu, J. Y., Lan, T. Q., Lai, H., and Qiu, X. (2013) pH-Induced lignin surface modification to reduce nonspecific cellulase binding and enhance enzymatic saccharification of lignocelluloses. *ChemSusChem* **6**, 919-927
7. Kumar, L., Arantes, V., Chandra, R., and Saddler, J. (2012) The lignin present in steam pretreated softwood binds enzymes and limits cellulose accessibility. *Bioresour Technol* **103**, 201-208
8. Voutilainen, S. P., Puranen, T., Siika-Aho, M., Lappalainen, A., Alapuranen, M., Kallio, J., Hooman, S., Viikari, L., Vehmaanperä, J., and Koivula, A. (2008) Cloning, expression, and characterization of novel thermostable family 7 cellobiohydrolases. *Biotechnol Bioeng* **101**, 515-528
9. Voutilainen, S. P., Murray, P. G., Tuohy, M. G., and Koivula, A. (2010) Expression of Talaromyces emersonii cellobiohydrolase Cel7A in Saccharomyces cerevisiae and rational mutagenesis to improve its thermostability and activity. *Protein Eng Des Sel* **23**, 69-79
10. Payne, C. M., Knott, B. C., Mayes, H. B., Hansson, H., Himmel, M. E., Sandgren, M., Ståhlberg, J., and Beckham, G. T. (2015) Fungal cellulases. *Chem Rev* **115**, 1308-1448
11. Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J* **382**, 769-781

12. Palonen, H., Tjerneld, F., Zacchi, G., and Tenkanen, M. (2004) Adsorption of Trichoderma reesei CBH I and EG II and their catalytic domains on steam pretreated softwood and isolated lignin. *J Biotechnol* **107**, 65-72

13. Rahikainen, J. L., Martin-Sampedro, R., Heikkinen, H., Rovio, S., Marjamaa, K., Tamminen, T., Rojas, O. J., and Kruus, K. (2013) Inhibitory effect of lignin during cellulose bioconversion: the effect of lignin chemistry on non-productive enzyme adsorption. *Bioresour Technol* **133**, 270-278

14. Pfeiffer, K. A., Sorek, H., Roche, C., Strobel, K., Blanch, H. W., and Clark, D. S. (2015) Evaluating endoglucanase Cel7B-lignin interaction mechanisms and kinetics using quartz crystal microgravimetry. *Biotechnol Bioeng*

15. Börjesson, J., Engqvist, M., Sipos, B., and Tjerneld, F. (2007) Effect of poly(ethylene glycol) on enzymatic hydrolysis and adsorption of cellulase enzymes to pretreated lignocellulose. *Enzyme Microb Technol* **41**, 186-195

16. Sammond, D. W., Yarbrough, J. M., Mansfield, E., Bomble, Y. J., Hobdey, S. E., Decker, S. R., Taylor, L. E., Resch, M. G., Bozell, J. J., Himmel, M. E., Vinzant, T. B., and Crowley, M. F. (2014) Predicting enzyme adsorption to lignin films by calculating enzyme surface hydrophobicity. *J Biol Chem* **289**, 20960-20969

17. Mansfield, S. D., Kim, H., Lu, F., and Ralph, J. (2012) Whole plant cell wall characterization using solution-state 2D NMR. *Nat Protoc* **7**, 1579-1589

18. Cheng, K., Sorek, H., Zimmermann, H., Wemmer, D. E., and Pauly, M. (2013) Solution-state 2D NMR spectroscopy of plant cell walls enabled by a dimethylsulfoxide-d 6/1-ethyl-3-methylimidazolium acetate solvent. *Anal Chem* **85**, 3213-3221

19. Nakagame, S., Chandra, R. P., Kadla, J. F., and Saddler, J. N. (2011) The isolation, characterization and effect of lignin isolated from steam pretreated Douglas-fir on the enzymatic hydrolysis of cellulose. *Bioresour Technol* **102**, 4507-4517

20. Ko, J. K., Kim, Y., Ximenes, E., and Ladisch, M. R. (2015) Effect of liquid hot water pretreatment severity on properties of hardwood lignin and enzymatic hydrolysis of cellulose. *Biotechnol Bioeng* **112**, 252-262

21. Phillips, C. M., Beeson, W. T., Cate, J. H., and Marletta, M. A. (2011) Cellobiose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by Neurospora crassa. *ACS Chem Biol* **6**, 1399-1406

22. Rahikainen, J. L., Evans, J. D., Mikander, S., Kalliola, A., Puranen, T., Tamminen, T., Marjamaa, K., and Kruus, K. (2013) Cellulase-lignin interactions-the role of carbohydrate-binding module and pH in non-productive binding. *Enzyme Microb Technol* **53**, 315-321

23. Lijnzaad, P., Berendsen, H. J., and Argos, P. (1996) A method for detecting hydrophobic patches on protein surfaces. *Proteins* **26**, 192-203

24. Jacak, R., Leaver-Fay, A., and Kuhlman, B. (2012) Computational protein design with explicit consideration of surface hydrophobic patches. *Proteins* **80**, 825-838

25. Stals, I., Sandra, K., Devreese, B., Van Beeumen, J., and Claeyssens, M. (2004) Factors influencing glycosylation of Trichoderma reesei cellulases. II: N-glycosylation of Cel7A core protein isolated from different strains. *Glycobiology* **14**, 725-737

26. Grassick, A., Murray, P. G., Thompson, R., Collins, C. M., Byrnes, L., Birrane, G., Higgins, T. M., and Tuohy, M. G. (2004) Three-dimensional structure of a thermostable native cellobiohydrolase, CBH IB, and molecular characterization of the cel7 gene from the filamentous fungus, Talaromyces emersonii. *Eur J Biochem* **271**, 4495-4506

Chapter 3: Structural insights into the lignin affinity of Cel7A carbohydrate-binding module

3.1 Abstract

The high cost of hydrolytic enzymes impedes the commercial production of lignocellulosic biofuels. High enzyme loadings are required in part due to non-productive adsorption to lignin, a major component of biomass. Despite numerous studies documenting cellulase adsorption to lignin, few attempts have been made to engineer enzymes for reduced lignin affinity. In this work, we used alanine-scanning mutagenesis to elucidate the structural basis of the *T. ressei* Cel7A carbohydrate-binding module's (CBM's) lignin affinity. *T. ressei* Cel7A CBM mutants were produced with a *T. emersonii* Cel7A catalytic domain and screened for their binding to cellulose and lignin. Mutation of aromatic and polar residues on the planar face of the CBM greatly decreased binding to both cellulose and lignin, supporting the hypothesis that the cellulose-binding face is also responsible for lignin affinity. Cellulose and lignin affinity of the 31 mutants were highly correlated, indicating that protein engineering to improve the selectivity for cellulose will be challenging. Four mutants with slightly higher cellulose selectivity (Q2A, H4A, V18A, and P30A) did not exhibit improved hydrolysis of Avicel in the presence of supplemental lignin or pretreated *Miscanthus*. Further reduction in lignin affinity while maintaining a high level of cellulose affinity is thus necessary to generate an enzyme with improved hydrolysis capability. This work provides insights into the structural underpinnings of lignin affinity, identifies residues amenable to mutation without compromising cellulose affinity, and informs engineering strategies for family one CBMs.

3.2 Introduction

Lignocellulosic biomass is an abundant, low-cost resource for the renewable production of fuels and chemicals. Unfortunately, lignocellulose is highly resistant to enzymatic degradation, necessitating high enzyme loadings that raise the cost of second-generation biofuels (1). The recalcitrance of biomass stems in part from the presence of lignin. Lignin, a major component of lignocellulosic biomass, inhibits the efficiency and recyclability of cellulase enzymes. Numerous studies have found that removing lignin from biomass increases the final hydrolysis yield (e.g. 2-7), and adding supplemental lignin decreases the final hydrolysis yield (e.g. 8-14). Lignin impedes hydrolysis through two main mechanisms: physically blocking cellulose access (5,6) and non-productively binding enzymes (15), thus lowering the effective enzyme concentration.

Cellulase adsorption to lignin has been well documented for both cellulase cocktails (10,16) and individual enzymes (9,17,18), though the structural basis of the interaction is not fully elucidated. Most fungal cellulases consist of two domains: a catalytic domain (CD) and a carbohydrate-binding module (CBM) connected by a highly glycosylated, flexible linker (19).

The CBM is responsible for increasing the effective concentration of the catalytic domain on the surface of its substrate (20). Fungal CBMs are highly conserved, 30-40 amino acid domains with a planar cellulose-binding surface. Three aromatic amino acids, arranged such that they align with a sequence of every other glucose unit on the surface of cellulose, are crucial for cellulose binding (21). Polar amino acids on the same face of the CBM are also important for binding to cellulose, presumably through the formation of hydrogen bonds (21). Along with binding to cellulose, the CBM is also responsible for the majority of lignin affinity. Fungal cellulases expressed without the CBM, or with the CBM removed, exhibit greatly reduced lignin affinity (17,22,23), highlighting the role of the CBM in lignin adsorption.

Studies of CBM lignin affinity have implicated hydrophobic and electrostatic interaction mechanisms. In a comparison of *T. reesei* Cel7A and Cel7B, Borjesson et al. showed that Cel7B had higher affinity to lignin, perhaps due to the more hydrophobic surface on the flat face of the CBM together with an additional exposed aromatic residue on the rough face (24). Sammond et al. found that the adsorption of enzymes to lignin surfaces correlates with solvent exposed hydrophobic clusters (18). Increasing the carboxylic acid content, and therefore negative charge, of lignin has been shown to decrease the non-productive binding of cellulases and increase the enzymatic hydrolysis of cellulose (12), presumably due to increased repulsive electrostatic interactions between the enzymes and lignin. Adding negative charge to cellulases by succinylation has also been shown to reduce lignin inhibition (25). Improved hydrolysis of cellulose in the presence of lignin at high pH, along with reduced lignin binding at high pH, also demonstrates the importance of electrostatic interactions (26).

Despite the numerous studies documenting cellulase adsorption to lignin, the structural basis has not been fully elucidated and few attempts have been made to engineer enzymes for reduced lignin affinity. Rahikainen et al. showed that mutation of one aromatic amino acid on the flat face of the *T. reesei* Cel7A CBM from tyrosine to alanine reduced lignin affinity, while mutation from tyrosine to tryptophan increased lignin affinity (26). Interestingly, although the cellulose affinity of the alanine mutant was reduced to the level of the catalytic domain alone, the lignin affinity of the mutant remained greater than the catalytic domain, indicating that the mutated enzyme retained considerable lignin affinity.

In this work, we employed extensive alanine scanning mutagenesis of the entire *T. reesei* Cel7A CBM to identify residues involved in lignin adsorption and highlight locations where mutation may increase the selectivity for cellulose. Toward the overall goal of generating a lignin-resistant enzyme, we evaluated the most promising mutants

for hydrolysis in the presence of lignin to investigate the extent to which lignin affinity must be reduced in order to decrease lignin inhibition.

3.3 Materials and Methods

Materials

Microcrystalline cellulose PH-105 (Avicel), Protease from *Streptomyces griseus,* 4-methylumbelliferyl β-D-cellobioside, Tween 20, cellulase from *T. reesei* (Celluclast) and papain were purchased from Sigma. Accelerace 1500 was a gift from Dupont. *A. niger* β-glucosidase was purchased from Megazyme. Acid-pretreated *Miscanthus* was produced by two-step dilute acid pretreatment on the pilot scale by Andritz, Glens Falls, NY. Solids were washed with water to neutral pH and lyophilized before lignin preparation.

Lignin preparation

Lignin residues were isolated from acid pretreated *Miscanthus* by enzymatic hydrolysis followed by *S. griseus* protease treatment to remove bound enzyme. Pretreated, rinsed, lyophilized biomass was ball milled for 5 min (Kleco ball mill, Visalia, CA). Enzymatic hydrolysis was carried out at 50°C, pH 5.0, with 4 wt% biomass and 0.25 ml Accelerase 1500 per gram of biomass. The reaction mixtures were centrifuged and the supernatant replaced with fresh buffer and enzyme (0.25 ml Accelerase/g original biomass) every 24 hrs. After 7 days of hydrolysis, the solids were separated by centrifugation and washed with acidified water (pH adjusted to 3.5 with HCl to avoid lignin solubilization). Bound protein was removed by an overnight protease treatment at 37°C, pH 7.5, with 2wt% solids and 0.3 mg/ml protease. The solids were then washed with acidified water and heated in a boiling water bath for 60 min to deactivate any remaining protease. The solids were extensively washed with acidified water and freeze-dried.

Lignin analysis

Compositional analysis of pretreated biomass and lignin was performed as previously described, with sugar concentrations determined by HPLC (Dionex) (27). Protein content in the solid samples was determined by total nitrogen analysis. The analysis was carried out at the UC Berkeley Microanalytical Facility using a Perkin Elmer 2400 Series II combustion analyzer.

Purification of T. reesei Cel7A

Cel7A was purified from Celluclast using a Q Sepharose anion exchange column followed by a MonoQ ion exchange column. Prior to injection to the MonoQ column, the crude Cel7A was treated with 0.1% Tween 20 to disrupt non-specific interactions that were found to exist between Cel7A and Cel7B. Fractions from the MonoQ were checked for purity by SDS-PAGE and an assay for contaminant endoglucanase activity. Assays contained 2% carboxymethyl cellulose and roughly 10 μM enzyme in 100 mM sodium acetate, pH 4.5. Assays were incubated at 50 °C overnight and endoglucanase activity was assessed by adding 2,4-dinitrosalycilic acid, incubating at 90 °C for 5 minutes, and

visually inspecting for the presence of brown color in the reactions. Fractions showing no endoglucanase activity were pooled and buffer exchanged into 50 mM sodium acetate, pH 5.0.

Isolation of the T. reesei CD

The *T. reesei* CD was cleaved from the linker and CBM using the protease papain. Papain was activated by incubating a 28 $mg \cdot ml^{-1}$ solution with 2 mM dithiothreitol and 2 mM ethylenediamine tetraacetic acid in 100 mM sodium phosphate, pH 7.0 for 30 minutes at room temperature. Cleavage reactions contained a w/w ratio of 1:100 papain to Cel7A in 50 mM sodium acetate, pH 5.0 and were incubated overnight at room temperature with stirring. The CD was purified from the protease and CBM using a Q Sepharose anion exchange column.

Small-scale enzyme production

The *T. emersonii* Cel7A CD or a fusion of *T. emersonii* Cel7A catalytic domain and the *T. reesei* Cel7A linker and CBM was cloned into a high-copy number plasmid (pCu424 (28)) and produced in YVH10 *pmr1*Δ *S. cerevisiae (29)*. Mutations were made by site directed mutagenesis (30). For screening all mutants, 20 mL SC-Trp medium was inoculated with *S. cerevisiae* containing the Cel7A gene and grown for 3 days at 30°C, 220 rpm. Cultures were spun down at 5,000 x g for 5 min and resuspended in SC medium supplemented with 500 mM Cu_2SO_4, then cultured for protein production for 3 days at 25°C, 220 rpm. Supernatant was collected and buffer exchanged into 50 mM sodium acetate, pH 5.0, using Amplicon spin concentrators (Millipore).

Large-scale enzyme production and purification

For large-scale production, 1 L of SC-Trp medium was inoculated with *S. cerevisiae* containing the Cel7A gene and grown for 3 days at 30°C, 220 rpm. Cultures were spun down at 5,000 x g for 5 min and resuspended in YPD medium supplemented with 500 mM Cu_2SO_4 and cultured for 3 days at 25°C, 220 rpm. Enzyme purification was carried out using a two-step procedure beginning with anion exchange chromatography. Active fractions were combined and further purified by hydrophobic interaction chromatography (HIC). Pure fractions (single band via SDS-PAGE) were combined and buffer exchanged into 50 mM sodium acetate buffer, pH 5.0.

Enzyme quantification

Purified enzymes were quantified by absorbance at 280 nm using the calculated molar extinction coefficient of 76,240 $M^{-1}cm^{-1}$ for the fusion enzyme and 71,120 $M^{-1}cm^{-1}$ for the catalytic domain. Unpurified enzyme and enzyme remaining in the supernatant after binding assays were quantified by their activity on a soluble substrate, 4-methylumbelliferyl β-D-cellobioside (MUC). The specific activity of 10.2 μmols MU $(\mu mol\ enzyme*min)^{-1}$ was measured for the purified wild type, catalytic domain, and mutant chimeras. Assays were conducted in a total volume of 100 μl with 1 mM MUC, pH 5.0, at 45°C for 8 min, followed by a denaturation step at 98°C for 2 min. Sodium hydroxide was added to a final concentration of 0.1 M and the fluorescence was measured on a Paradigm plate reader (Beckman Coulter) with an excitation of 360 nm and emission of 465 nm.

Avicel and lignin binding screen

Unpurified, buffer-exchanged mutants were normalized to the same concentration of active Cel7A as determined by activity on a soluble substrate. All mutants contained identical catalytic domains, and therefore should have equivalent specific activity on soluble substrates. Binding to Avicel and lignin was measured in 50 mM sodium acetate buffer, pH 5, using 15 mg/ml of Avicel or lignin and 125 nM enzyme. Experiments were conducted in a total volume of 70 uL in a rotating shaker (300 rpm) at room temperature. An initial time course with the wild type enzyme was measured to determine the time required for binding to reach equilibrium. Equilibrium was reached after approximately 30 min. Subsequent assays were performed for 1 hr. After equilibration, solids were separated by centrifugation and the supernatant was analyzed for free, active enzyme using MUC. The amount of bound enzyme was calculated from the difference between added enzyme and enzyme remaining in solution. The values plotted in Figure 2 are bound enzyme divided by enzyme remaining in solution. At the low enzyme concentrations employed, these values approximate the partition coefficient, α, as defined in Equation 1. The values reported are averages of duplicate experiments and the errors are calculated as standard deviations.

The one-site Langmuir isotherm at low free protein concentrations can be written as

$$\Gamma = \Gamma_{\max} K_A C = \alpha C \qquad (1)$$

where Γ is the amount of adsorbed protein, Γ_{max} is the amount of adsorbed protein at saturation, K_A is the Langmuir affinity constant, C is the concentration of free protein, and α is the partition coefficient.

Adsorption isotherms

Adsorption isotherm measurements were performed in 50 mM sodium acetate buffer (pH 5) using 10 mg/ml of Avicel or lignin and initial enzyme concentrations ranging from 0.015-4.0 μM. Experiments were conducted using the same conditions as the binding screen. A one-site Langmuir adsorption isotherm (Equation 2) was fit to the binding data by non-linear least squares optimization to determine K_A, Γ_{max}, and their corresponding standard deviations.

$$\Gamma = \frac{\Gamma_{\max} K_A C}{1 + K_A C} \qquad (2)$$

Hydrolysis of Avicel supplemented with lignin and Miscanthus

Hydrolysis reactions were performed with 10 mg/ml of Avicel, 10 mg/ml Avicel supplemented with 10 mg/ml of isolated lignin, or 20 mg/ml acid-pretreated *Miscanthus* in 50 mM sodium acetate buffer, pH 5, with 0.5 μM Cel7A. All reactions were supplemented with 40 U/mL of *A. niger* β-glucosidase. Reactions were carried out in a thermally controlled incubator with end-over-end mixing at 60°C for enzymes containing the *T. emersonii* catalytic domain and at 55°C for enzymes containing the *T. reesei* catalytic domain. After 48 hours, reaction mixtures were filtered and analyzed

immediately for glucose by HPLC (Shimadzu). The values reported are means of duplicate samples and the errors are calculated as standard deviation.

Predicting the amount of enzyme bound to cellulose during hydrolysis

Equations 3-5 were used to predict the amount of cellulose-bound enzyme in the presence of lignin.

$$\Gamma_C = \frac{\Gamma_{\max,C} K_{A,C} C}{1 + K_{A,C} C} \quad (3)$$

$$\Gamma_L = \frac{\Gamma_{\max,L} K_{A,L} C}{1 + K_{A,L} C} \quad (4)$$

$$C_{total} = C + S_C \Gamma_C + S_L \Gamma_L \quad (5)$$

where Γ_C and Γ_L are the amounts of enzyme bound per gram of solid to cellulose and lignin, respectively; $K_{A,C}$ and $K_{A,L}$ are the Langmuir binding constants to cellulose and lignin, respectively; $\Gamma_{max,C}$ and $\Gamma_{max,L}$ are the amounts of adsorbed protein at saturation to cellulose and lignin, respectively; C is the concentration of protein remaining free in solution; C_{total} is the initial protein concentration, and S_C and S_L are the concentrations of cellulose and lignin, respectively.

3.4 Results

Isolation and characterization of lignin from pretreated Miscanthus

Lignin used in this study was isolated from acid pretreated *Miscanthus* using a commercial cellulase mixture (Accelerase 1500, Dupont) to ensure that the lignin was as similar as possible to lignin present in an industrial hydrolysis reaction. Excess enzyme and long digestion times were used to remove all accessible carbohydrates. The solid residue was then treated with *S. griseus* protease to remove bound enzymes. The compositions of the starting biomass and resulting isolated lignin are presented in Table 3.1. The Klason lignin, remaining carbohydrates, and protein content are similar to protease treated lignin used in previous studies (9-11,17,26).

Table 3.1 Characterization of pretreated Miscanthus and isolated lignin.

	Glucan (%)	Xylan (%)	Arabinan (%)	Klason lignin (%)	Ash (%)	Nitrogen (%)	BET surface area (m^s/g)
Acid-pretreated *Miscanthus*	55.6	0.54	0.11	33.6	9.73	< 0.2	5.7
Isolated lignin	6.21	0.39	0.13	77.0	14.51	0.34	35.9

Lignin inhibition of Cel7A and isolated catalytic domain

We compared the lignin inhibition of natively expressed *T. reesei* Cel7A to that of the chimera used for engineering in this study, recombinantly expressed *T. emersonii* Cel7A catalytic domain (CD) fused to the *T. reesei* Cel7A linker and carbohydrate-binding module (Te-Tr chimera). Figure 3.1 shows the hydrolysis of Avicel, Avicel in the presence of supplemental lignin, and acid-pretreated *Miscanthus* by each enzyme and the corresponding isolated catalytic domain (CD).

The Te-Tr chimera was less inhibited by lignin than *T. reesei* Cel7A. The addition of lignin to Avicel decreased the glucose generated by approximately 15% for the Te-Tr chimera, and 73% for the full length *T. reesei* enzyme. The *T. emersonii* CD was not inhibited by supplemental lignin, indicating that the inhibiton of the Te-Tr chimera is due entirely to the linker and CBM, our mutagenesis target. The *T. reesei* CD generated 38% less sugar in the presence of lignin. Although both CDs were less inhibited by lignin, they generated significantly less sugar than the corresponding full enzyme.

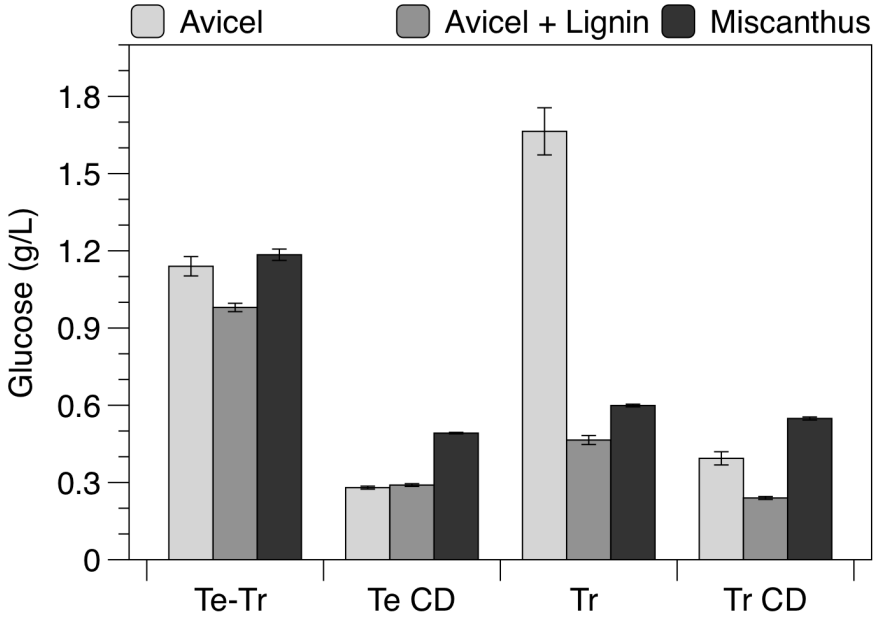

Figure 3.1 Hydrolysis by *T. reesei* (Tr) and *T. emersonii* (Te) Cel7A.
Glucose produced after 48 hours with 10 g/L Avicel, 10 g/L Avicel with 10 g/L lignin, or 20 g/L *Miscanthus*. All reactions were supplemented with *A. niger* β-glucosidase. Reactions were performed at 60°C for *T. emersonii* (Te) enzymes and 55°C for *T. reesei* (Tr) enzymes with 0.5 μM Cel7A in Avicel reactions and 1.0 μM Cel7A in Miscanthus reactions. Values presented are means of duplicate samples and errors are standard deviations.

CBM engineering approach

We used alanine-scanning mutagenesis to identify residues that contribute to lignin affinity. The isolated *T. emersonii* catalytic domain, wild type Te-Tr chimera, and

31 Te-Tr mutant chimeras--each with one position in the CBM changed to alanine--were produced in *S. cerevisiae* strain YVH10 *pmr1*Δ. This strain includes an overexpressed protein disulfide isomerase for increased Cel7A titers and a knockout of the *PRM1* gene for reduced hyperglycosylation (29). Each residue in the CBM was mutated to alanine except cysteine residues involved in disulfide bonds and native alanine residues.

Binding screen

The Cel7A mutants were secreted into defined medium and screened for cellulose and lignin binding without purification. Binding was quantified by measuring the Cel7A activity remaining in the supernatant using a soluble substrate, 4-methylumbelliferyl β-D-cellobioside, which permitted specific detection of Cel7A among other secreted proteins. Each mutant was screened for binding at low enzyme concentration, in the range where the Langmuir isotherm is linear, permitting direct calculation of the partition coefficient from the amount of bound and free enzyme (Equation 1 in Experimental Procedures).

Impact of point mutations on CBM adsorption to Avicel

The cellulose partition coefficient of each mutant is shown in Figure 3.2 (light bars) and the CBM structure is shown in Figure 3.3. The CD alone has significantly less affinity for cellulose than the chimera containing the wild type CBM, as expected. All mutants have an Avicel partition coefficient in the range between the wild type and CD alone. As previously reported, mutating the three tyrosines on the planar face of the CBM (Y5, Y31, and Y32) as well as polar residues on the same face (N29, Q34) to alanine significantly reduced cellulose binding (21). Mutation of Q7, G10, Y13, G15, G22, and L36 to alanine also significantly reduced binding to cellulose (<30% of wild-type affinity). The impact of mutating each residue to alanine on Avicel affinity is shown in the CBM structure in Figure 3.4.

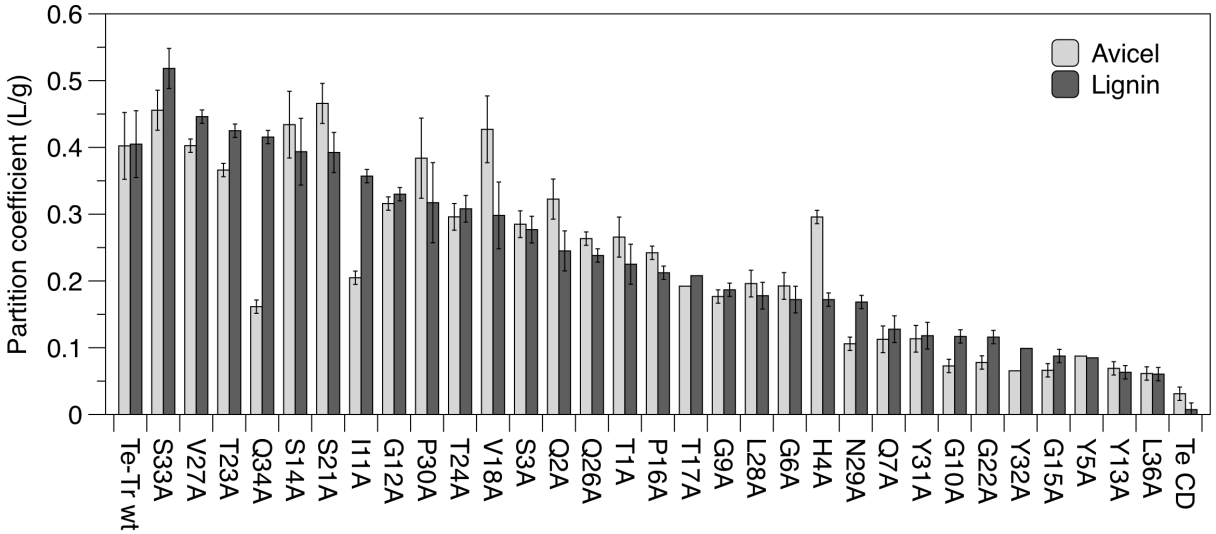

Figure 3.2 Enzyme partition coefficients (α) for Avicel and lignin.
Partition coefficients are calculated from duplicate binding experiments at room temperature and an initial enzyme concentration of 125nM. Values presented are means of duplicate samples and errors are standard deviations.

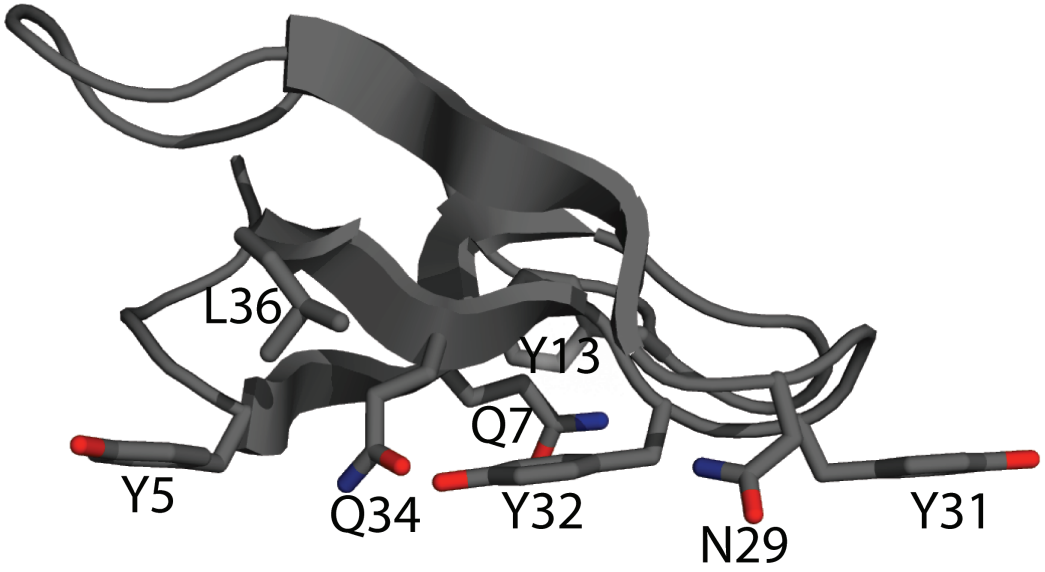

Figure 3.3 *T. reesei* Cel7A CBM structure.
Structure shows side chains important to cellulose or lignin adsorption (38).

Impact of point mutations on CBM adsorption to lignin

The lignin partition coefficient of each mutant is shown in Figure 3.2 (dark bars). The *T. emersonii* CD has negligible lignin affinity when compared to the fusion containing the *T. reesei* linker and CBM. We found that mutation of Y5, Q7, G10, Y13, G15, G22, Y31, Y32, and L36 had the greatest impact on lignin affinity. Y5, Q7, Y31, and Y32 are residues on the cellulose-binding face of the CBM, as seen in Figure 3.3. Although these mutants had significantly reduced lignin affinity compared to the wild type chimera, all mutants retained much greater lignin affinity than the isolated CD. The impact on lignin affinity from mutating each residue to alanine is shown in the CBM structure in Figure 3.4.

Impact of point mutations on CBM selectivity

Our primary aim in engineering CBMs for reduced nonspecific binding is to increase the selectivity of adsorption to cellulose. However, lignin and cellulose affinity of the mutants were highly correlated (R^2=0.92), as shown in Figure 3.5. This indicates that engineering the CBM for reduced lignin affinity but retained cellulose affinity is likely to be very challenging. A few mutants, labeled in Figure 3.5, had slightly greater selectivity toward Avicel, indicated by a higher partition coefficient to Avicel than to lignin. These mutants were chosen for further characterization to determine if their increased selectivity toward cellulose leads to reduced lignin inhibition.

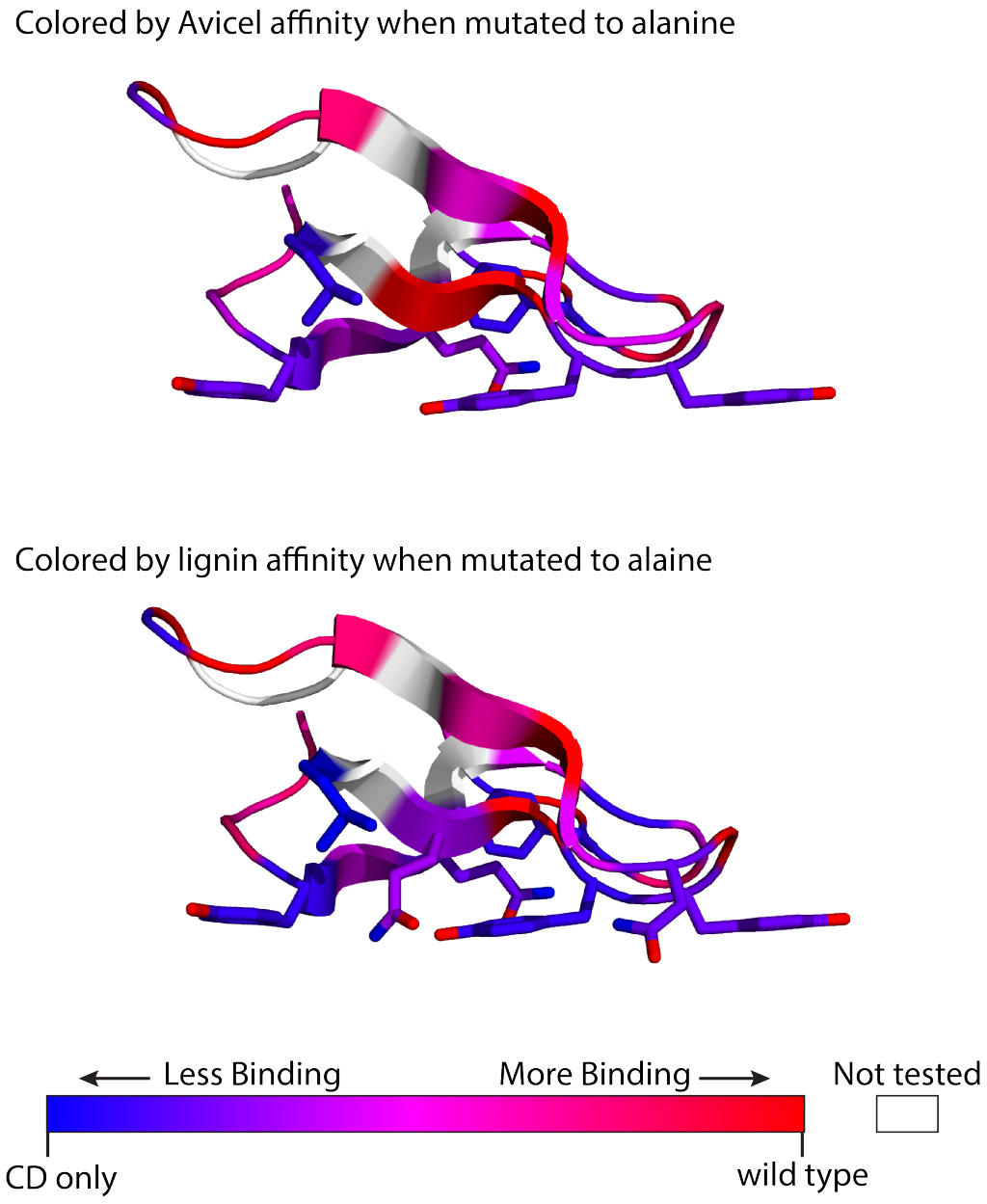

Figure 3.4 CBM structure colored by affinity.
Each residue is colored by the affinity of the enzyme when mutated to alanine. Residues in blue are most important for binding while residues in red could be mutated to alanine with little impact on affinity. Residues in white are native alanine residues or cysteine involved in disulfide bonds, which were not tested.

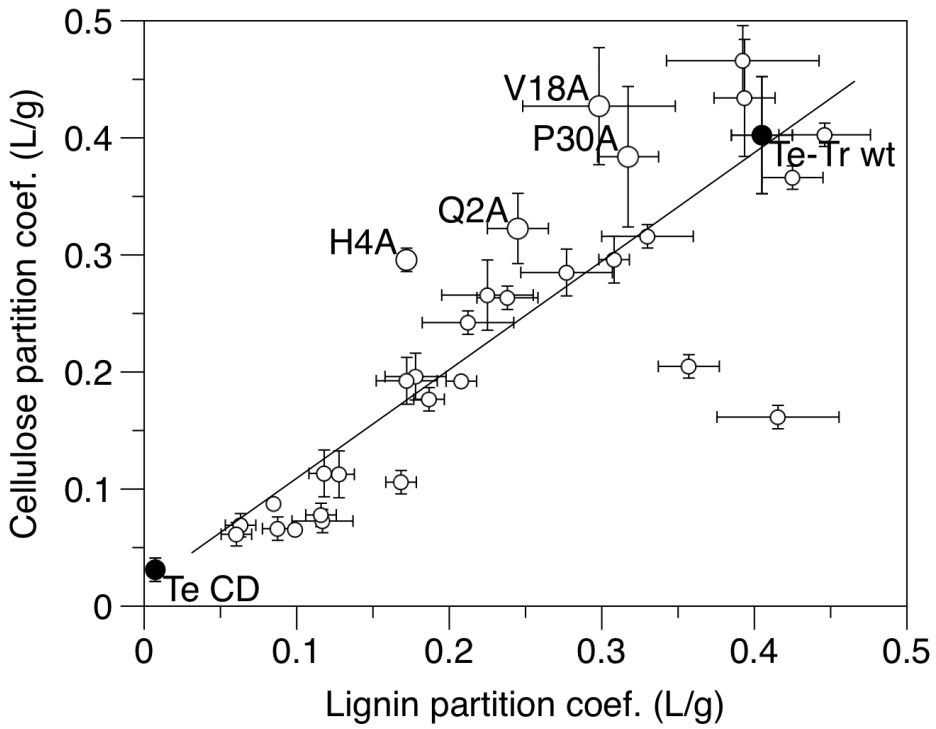

Figure 3.5 Binding selectivity of Te-Tr wild type, CD, and all mutants. Binding selectivity as evidenced by a scatter plot of partition coefficients. Labeled mutants were selected for further characterization. Values presented are means of duplicate samples and errors are standard deviations.

Adsorption isotherms of selected mutants

The wild type enzyme, catalytic domain alone (CD), and several alanine mutants with an increased selectivity toward Avicel (Q2A, H4A, V18A, and P30A) were characterized further. The partition coefficient calculated from the binding screen does not provide enough information to predict binding behavior at higher concentrations, where the Langmuir isotherm is no longer linear. Therefore, full Langmuir isotherms were measured and fit using non-linear least squares regression to determine the adsorption at saturation (Γ_{max}) and the Langmuir adsorption constant (K_A). The isotherms are shown in Figure 3.6 and the model parameters are listed in Table 3.2.

Qualitatively, the mutations decreased lignin adsorption more than Avicel adsorption, as seen in Figure 3.6. The partition coefficients to Avicel derived from the Langmuir isotherms ($\alpha=K_A\Gamma_{max}$) are similar for the wild type and mutant enzymes (Table 3.2), as expected based on data from the binding screen. The saturating amount of enzyme adsorbed to lignin, Γ_{max}, is lower for each of the mutants than the wild type enzyme, though the affinity constants and partition coefficients are within error of the wild-type enzyme.

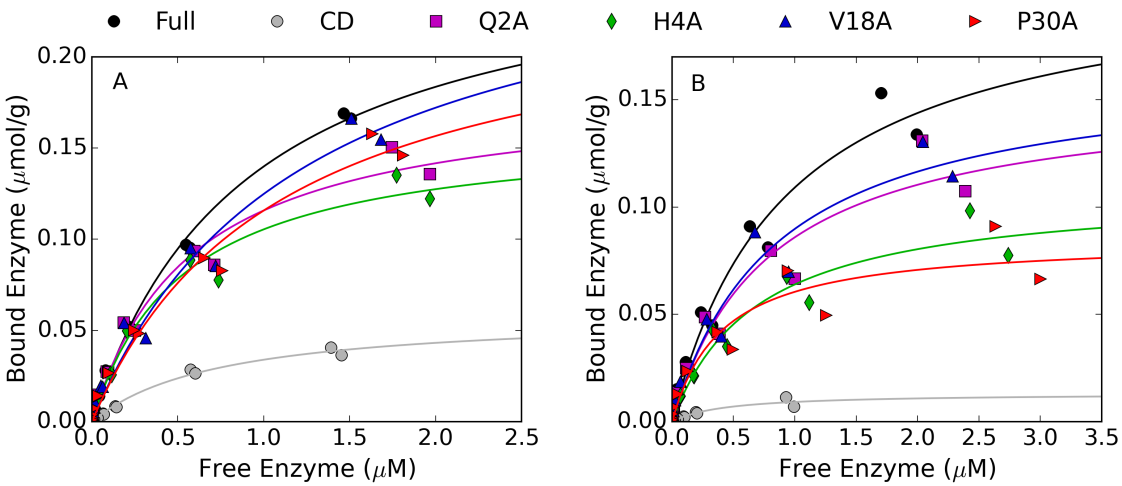

Figure 3.6 Adsorption isotherms to Avicel (A) and lignin (B).
Adsorption measured at room temperature with 10 g/L Avicel or lignin. The lines show the fit to a one-site Langmuir binding isotherm.

Table 3.2 Langmuir adsorption parameters.

	Avicel			Lignin		
	K_A (μM^{-1})	Γ_{max} (μmol/g)	α (L/g)	K_A (μM^{-1})	Γ_{max} (μmol/g)	α (L/g)
Wild type	1.10 ± 0.11	0.27 ± 0.01	0.29 ± 0.03	1.07 ± 0.21	0.21 ± 0.02	0.23 ± 0.05
CD	1.34 ± 0.17	0.059 ± 0.003	0.08 ± 0.01	2.52 ± 0.77	0.013 ± 0.001	0.03 ± 0.01
Q2A	1.74 ± 0.27	0.18 ± 0.01	0.32 ± 0.05	1.26 ± 0.33	0.15 ± 0.02	0.19 ± 0.05
H4A	1.90 ± 0.26	0.16 ± 0.01	0.31 ± 0.05	1.46 ± 0.32	0.11 ± 0.01	0.16 ± 0.04
V18A	0.82 ± 0.21	0.28 ± 0.04	0.23 ± 0.06	1.19 ± 0.30	0.17 ± 0.02	0.20 ± 0.05
P30A	0.92 ± 0.17	0.24 ± 0.02	0.22 ± 0.05	2.49 ± 0.75	0.08 ± 0.01	0.21 ± 0.07

K_A, Langmuir adsorption constant; Γ_{max}, adsorption at saturation; α, partition coefficient ($\alpha = K_A \Gamma_{max}$) indicated with their standard deviation.

Hydrolysis of Avicel in the presence of lignin and acid-pretreated Miscanthus

A central aim of this study was to engineer the Cel7A CBM to have greater specificity for cellulose. Having achieved this goal, we next examined the effect of this change in specificity on cellulose hydrolysis in the presence of lignin. The wild type, CD, and mutant enzymes were tested for their ability to hydrolyze Avicel, Avicel in the presence of lignin, and acid-pretreated *Miscanthus* (Figure 3.7). Although each of the mutants displayed slightly increased selectivity toward binding cellulose, none generated more glucose than the wild-type enzyme under inhibiting or non-inhibiting conditions.

Predicting the amount of enzyme bound to cellulose

We sought to explain the hydrolysis results and predict levels of binding selectivity necessary for hydrolytic improvement by using the Langmuir adsorption model. The first step of cellulose hydrolysis is adsorption to the surface. Predicting the amount of enzyme adsorbed to cellulose can therefore lend insight into the performance of enzymes with different affinities. Using the Langmuir parameters (K_A and Γ_{max}) shown

in Table 3.2, the approximate amount of enzyme bound to Avicel under the hydrolysis conditions employed in this study was calculated for the wild type chimera, CD, and the four selected mutants. This prediction is restricted to hydrolysis of pure cellulose with supplemental lignin, due to the use of Langmuir parameters measured on the pure substances.

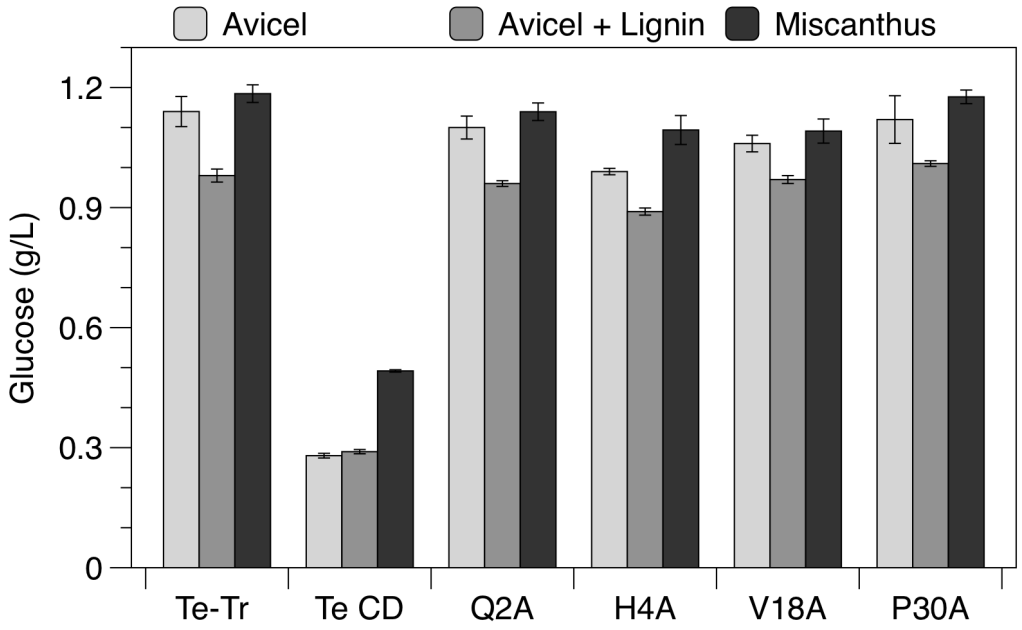

Figure 3.7 Hydrolysis of Avicel with lignin and of acid-pretreated *Miscanthus*. Glucose produced after 48 hours with 10 g/L Avicel, 10 g/L Avicel with 10 g/L lignin, or 20 g/L *Miscanthus*. All reactions were supplemented with *A. niger* β-glucosidase. Reactions were performed at 60°C with 0.5 μM Cel7A in Avicel reactions and 1.0 μM Cel7A in Miscanthus reactions. Values presented are means of duplicate samples and errors are standard deviations.

As shown in Figure 3.8, none of the mutants are predicted to have significantly more enzyme bound to Avicel than the wild type, which explains why none of the mutants generated more glucose during the hydrolysis assays. The catalytic domain is predicted to bind in similar amounts to cellulose in the presence and absence of lignin, as expected given that it performed equally well in hydrolysis assays with and without lignin (Figure 6). The full-length wild type enzyme and each of the mutants are predicted to have between 3% and 55% less enzyme bound to cellulose when lignin is present, which covers the range of lignin inhibition observed in the hydrolysis assays.

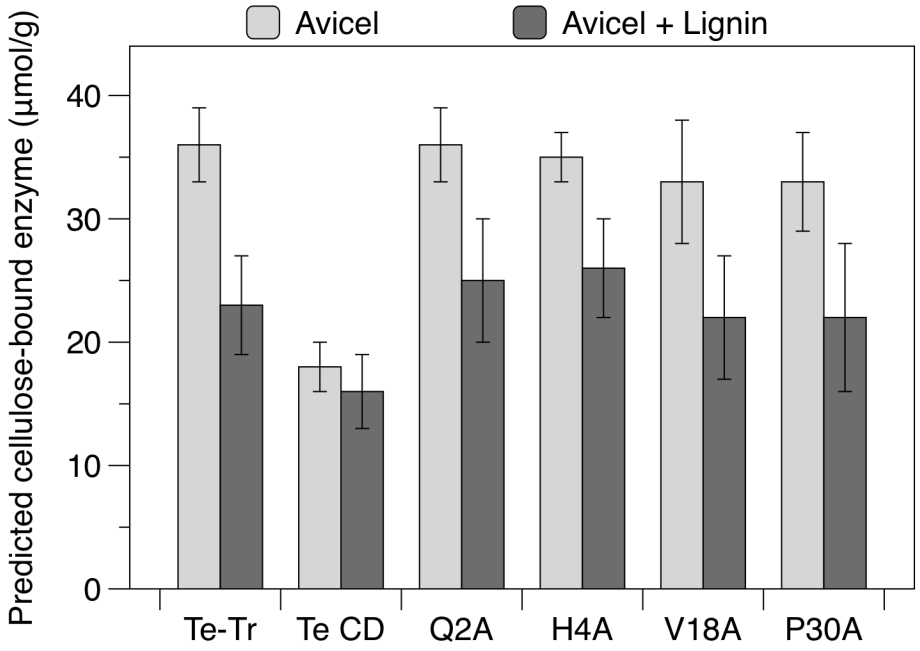

Figure 3.8 Predicted amounts of enzyme bound to Avicel.
Bound enzyme is calculated using the Langmuir parameters in Table 2 for a solution with 0.5uM total enzyme, 10 g/L Avicel, with or without 10 g/L lignin. Error is presented as the range of values possible given the standard deviation in the Langmuir affinity constant (K_A) and saturation adsorption constant (Γ_{max}).

3.5 Discussion

Non-productive adsorption of cellulases to lignin reduces the effective enzyme concentration during biomass hydrolysis, increasing the overall cost of producing biofuels. Although it is well known that the carbohydrate-binding module (CBM) drives this non-productive adsorption, the mechanisms of CBM lignin-affinity have not been fully elucidated. We employed alanine-scanning mutagenesis of the *T. reesei* Cel7A CBM to reveal the structural basis of lignin affinity and identify residues where mutation increases the selectivity for cellulose over lignin. Our studies focused on *T. reesei* Cel7A because it is the most abundant enzyme in many commercial cellulase mixtures. Each CBM mutant was produced with a linker and CD to ensure realistic binding conditions and enable hydrolysis studies. We used the *T. reesei* Cel7A CBM and linker, and the *T. emersonii* Cel7A CD. This CD is homologous to *T. reesei's*, and unlike the *T. reesei* enzyme, expresses well in yeast, making enzyme production and screening tractable.

Although less inhibited by lignin than the native *T. reesei* enzyme, the inhibition of the Te-Tr chimera was entirely due to the CBM, making it a well-suited construct for CBM engineering. Surprisingly, the *T. reesei* Cel7A CD was also inhibited by supplemental lignin. This result is unexpected given the general consensus that the CBM is responsible for the majority of non-productive adsorption to lignin. Engineering both domains of *T. reesei* Cel7A may be required to fully alleviate lignin inhibition. The

highly homologous *T. emersonii* CD was not inhibited by lignin, perhaps due to this enzyme's higher melting temperature (31,32). Engineering the CD for greater thermal stability may therefore be a promising strategy to reduce lignin inhibition.

The wild type Te-Tr chimera, isolated *T. emersonii* CD, and 31 mutants, each with one position in the CBM mutated to alanine were produced and screened for binding to microcrystalline cellulose (Avicel) and lignin. The isolated CD had dramatically reduced affinity for lignin, consistent with previous reports that the CBM is responsible for the majority of lignin binding (17,22).

Mutation to alanine of the three tyrosines (Y5, Y31, and Y32) on the planar face of the CBM significantly reduced cellulose and lignin binding. These tyrosine residues are known to be essential for cellulose affinity (21), though only Y32 has been evaluated previously for lignin binding. According to NMR data, mutation of Y31 and Y32 to alanine does not cause significant structural changes, and mutation of Y5 to alanine causes only minor structural changes (21), indicating that the reductions in binding are due to diminished interactions between these residues and the lignin surface. The three tyrosine residues likely contribute to lignin adsorption through pi stacking interactions with phenolic groups on the lignin surface. This result supports the hypothesis that the flat face of the CBM is responsible for binding not only to cellulose, but to lignin as well.

Mutation of Q7 on the cellulose-binding face of the CBM also greatly reduced the binding to cellulose and lignin. This residue has been predicted to interact with cellulose through hydrogen bonding (33), though its impact on affinity had not been previously assessed experimentally. Loss of these hydrogen-bonding interactions likely causes the decrease in cellulose affinity. It is possible that this residue also hydrogen bonds to the surface of lignin; however, the decrease in affinity to both cellulose and lignin may also be due to structural alterations in the CBM caused by this mutation.

Mutation of an internal tyrosine (Y13) to alanine also significantly decreased affinity for both cellulose and lignin. Y13 has been suggested to undergo a conformational change when the CBM is bound to cellulose, moving away from the protein interior to the wedge surface to establish van der waals contact with the cellulose chain (34,35). Mutation of Y13 to alanine may decrease affinity due to loss of these van der waals contacts, although disruptive structural changes (caused by mutation of a large, internal hydrophobic residue) may also play a important role.

Interestingly, Q34A, which had greatly reduced cellulose affinity, retained lignin affinity similar to that of the wild-type enzyme. This residue is predicted to interact with the surface of cellulose through hydrogen bonding. N29 and I11, also predicted to interact with cellulose through hydrogen bonding (35), also retained a greater fraction of their lignin affinity than their cellulose affinity. These results suggest that hydrogen bonding is more important for binding to cellulose than binding to lignin. The hydrogen-bonding residues at specific positions in the CBM are able to align with the uniformly spaced hydroxyls in crystalline cellulose. In contrast, the free-radical polymerized lignin

structure lacks the precise repeating structures of crystalline cellulose, making specific hydrogen bonding on the CBM surface less likely.

The remaining mutants with severely reduced affinity for cellulose and lignin are G10A, G15A, G22A, and L36A. It is unlikely that these glycine residues interact with the surface of cellulose or lignin. G10 and G22 are completely conserved in an alignment of homologous family one CBMs and G22 is conserved as either a glycine or proline. Therefore, the decreased affinity is most likely due to structural changes caused by mutation to alanine. Mutation of L36, an internal hydrophobic residue, also likely results in structural perturbations that are responsible for decreased affinity. Although these mutants had significantly reduced lignin affinity compared to the wild type chimera, they retained much greater lignin-affinity than the isolated CD. This indicates that the mutated CBMs have residual lignin affinity or that the linker contributes to affinity. Linker mutations that decrease lignin inhibition are described in a patent (36), indicating that the linker does participate in non-productive adsorption to lignin.

Four mutants with increased selectivity for cellulose in the binding screen (Q2A, H4A, V18A, and P30A) were further characterized by constructing full Langmuir isotherms. Although the Langmuir model does not elucidate the complex mechanisms involved in protein adsorption to solid surfaces, the model describes the adsorption data well and is a useful tool for evaluating cellulase adsorption (37).

As expected from the binding screen, the Te-Tr wild type and mutants had similar Langmuir affinity and saturation constants for cellulose. The mutant enzymes exhibit similar lignin binding compared to the wild type enzyme at low enzyme concentration (leading to similar values for K_A) but decreased binding at the higher concentrations, leading to reduced values for Γ_{max}. Given lignin's heterogeneous structure, it is likely that there are multiple binding sites on the lignin surface with different affinities (23). Mutations in the CBM may impact interactions with one type of site more than others, resulting in decreased binding only over certain enzyme concentration ranges. The initial concentrations of enzyme employed here (up to 0.4 μmol (24mg) Cel7A enzyme per gram of cellulose) cover the range of typical enzyme loadings used in hydrolysis of lignocellulosic biomass, and therefore describe the binding behavior under industrially relevant conditions.

Having generated mutants with slightly increased selectivity toward cellulose, we studied their ability to hydrolyze Avicel in the presence of lignin and pretreated *Miscanthus*. Unfortunately, none of the mutants outperformed the wild type enzyme, likely because lignin affinity was not reduced enough to impact hydrolysis. To test this hypothesis, we used the Langmuir model to predict the amount of each enzyme bound to cellulose under the hydrolysis conditions. Because the first step of hydrolysis is adsorption to cellulose, the calculated binding should reflect the relative hydrolytic performance of similar enzymes with different affinities. This analysis assumes that the binding to both Avicel and lignin is reversible, the amount of Avicel and lignin are constant, and that the Langmuir isotherms collected at room temperature accurately describe the binding behavior at 60°C, the hydrolysis temperature. Although limited, this

approach enables comparison of enzymes with different affinities, and estimates the magnitude of change in specificity necessary to generate an enzyme with less lignin inhibition. None of the mutants were predicted to bind in larger quantities to cellulose than the wild-type enzyme, supporting the conclusion that lignin affinity was not decreased enough to impact hydrolysis.

To investigate the required level of lignin affinity reduction necessary to improve hydrolysis, we used the Langmuir model to predict the amount of enzyme bound to cellulose for different ranges of cellulose and lignin affinities. We computationally varied both the cellulose and lignin partition coefficient, calculated the amount of cellulose-bound enzyme, and normalized by the amount of cellulose-bound enzyme predicted for the wild type enzyme to generate the heat map in Figure 3.9. As expected, the maximum amount of cellulose-bound enzyme occurs for a cellulase with zero lignin affinity and the maximum cellulose affinity. A two-fold reduction in lignin affinity without any loss of cellulose affinity is the minimum change necessary to increase the amount of enzyme bound to cellulose compared to the wild type enzyme. If cellulose affinity is not maintained, lignin affinity must be reduced further to generate an improved enzyme.

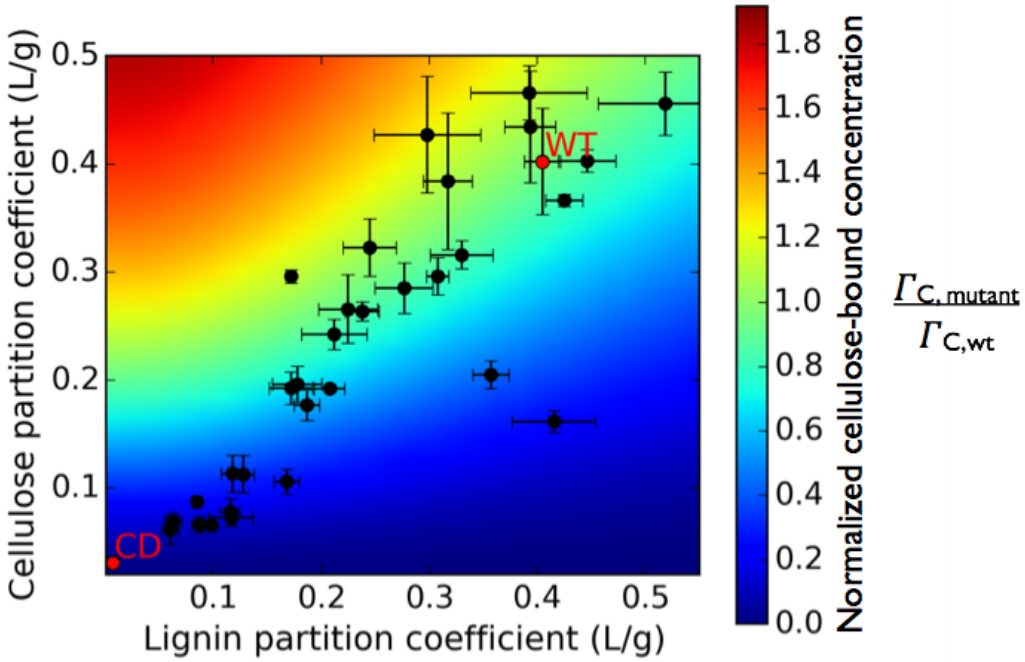

Figure 3.9 Cellulose bound enzyme predictions for different affinities.
A scatter plot of cellulose and lignin partition coefficients with all mutants shown as black data points. The background is colored based on the amount of cellulose bound enzyme for a cellulase with the given affinities normalized by the amount of cellulose bound enzyme with wild type affinity. Cellulose-bound enzyme concentrations were calculated for a mixture with 10 g/L Avicel, 10 g/L lignin, and 0.5 μM enzyme.

A central conclusion from the combined results of this study is that improving cellulose-hydrolyzing activity in the presence of lignin will require a greater reduction in CBM/lignin affinity accompanied by a high level of cellulose binding. The four residues identified here (Q2, H4, V18, P30) are good targets for further mutation. Since hydrophobic interactions have been implicated in lignin adsorption (18,24,26), changing these residues to polar or charged amino acids may produce a more pronounced decrease in lignin adsorption and consequently better hydrolytic performance in the presence of lignin.

Acknowledgements: We thank Craig Dana for assistance with yeast expression, Stefan Bauer for assistance with sample analysis, Richard Kwant for critical reading of the manuscript, and Yuzhang Chen for assistance with protein production. The Energy Biosciences Institute supported this work.

3.6 References

1. Klein-Marcuschamer, D., Oleskowicz-Popiel, P., Simmons, B. A., and Blanch, H. W. (2012) The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnol Bioeng* **109**, 1083-1087
2. Ishizawa, C. I., Jeoh, T., Adney, W. S., Himmel, M. E., Johnson, D. K., and Davis, M. F. (2009) Can delignification decrease cellulose digestibility in acid pretreated corn stover? *Cellulose* **16**, 677-686
3. Kumar, L., Arantes, V., Chandra, R., and Saddler, J. (2012) The lignin present in steam pretreated softwood binds enzymes and limits cellulose accessibility. *Bioresour Technol* **103**, 201-208
4. Kumar, R., and Wyman, C. E. (2009) Access of cellulase to cellulose and lignin for poplar solids produced by leading pretreatment technologies. *Biotechnology progress* **25**, 807-819
5. Ju, X., Engelhard, M., and Zhang, X. (2013) An advanced understanding of the specific effects of xylan and surface lignin contents on enzymatic hydrolysis of lignocellulosic biomass. *Bioresour Technol* **132**, 137-145
6. Mooney, C. A., Mansfield, S. D., Touhy, M. G., and Saddler, J. N. (1998) The effect of initial pore volume and lignin content on the enzymatic hydrolysis of softwoods. *Bioresour Technol* **64**, 113-119
7. Yoshida, M., Liu, Y., Uchida, S., Kawarada, K., Ukagami, Y., Ichinose, H., Kaneko, S., and Fukuda, K. (2008) Effects of cellulose crystallinity, hemicellulose, and lignin on the enzymatic hydrolysis of Miscanthus sinensis to monosaccharides. *Biosci Biotechnol Biochem* **72**, 805-810

8. Rahikainen, J. L., Moilanen, U., Nurmi-Rantala, S., Lappas, A., Koivula, A., Viikari, L., and Kruus, K. (2013) Effect of temperature on lignin-derived inhibition studied with three structurally different cellobiohydrolases. *Bioresour Technol* **146**, 118-125

9. Rahikainen, J., Mikander, S., Marjamaa, K., Tamminen, T., Lappas, A., Viikari, L., and Kruus, K. (2011) Inhibition of enzymatic hydrolysis by residual lignins from softwood--study of enzyme binding and inactivation on lignin-rich surface. *Biotechnol Bioeng* **108**, 2823-2834

10. Nakagame, S., Chandra, R. P., and Saddler, J. N. (2010) The effect of isolated lignins, obtained from a range of pretreated lignocellulosic substrates, on enzymatic hydrolysis. *Biotechnol Bioeng* **105**, 871-879

11. Nakagame, S., Chandra, R. P., Kadla, J. F., and Saddler, J. N. (2011) The isolation, characterization and effect of lignin isolated from steam pretreated Douglas-fir on the enzymatic hydrolysis of cellulose. *Bioresour Technol* **102**, 4507-4517

12. Nakagame, S., Chandra, R. P., Kadla, J. F., and Saddler, J. N. (2011) Enhancing the enzymatic hydrolysis of lignocellulosic biomass by increasing the carboxylic acid content of the associated lignin. *Biotechnol Bioeng* **108**, 538-548

13. Lai, C., Tu, M., Shi, Z., Zheng, K., Olmos, L. G., and Yu, S. (2014) Contrasting effects of hardwood and softwood organosolv lignins on enzymatic hydrolysis of lignocellulose. *Bioresour Technol* **163**, 320-327

14. Berlin, A., Balakshin, M., Gilkes, N., Kadla, J., Maximenko, V., Kubo, S., and Saddler, J. (2006) Inhibition of cellulase, xylanase and beta-glucosidase activities by softwood lignin preparations. *J Biotechnol* **125**, 198-209

15. Berlin, A., Gilkes, N., Kurabi, A., Bura, R., Tu, M., Kilburn, D., and Saddler, J. (2005) Weak lignin-binding enzymes. In *Twenty-Sixth Symposium on Biotechnology for Fuels and Chemicals*, pp. 163-170, Humana Press

16. Machado, D. L., Neto, J. M., Pradella, J. G., Bonomi, A., Rabelo, S. C., and da Costa, A. C. (2014) Adsorption Characteristics of Cellulase and β-glucosidase on Avicel, Pretreated Sugarcane Bagasse and Lignin. *Biotechnol Appl Biochem*

17. Palonen, H., Tjerneld, F., Zacchi, G., and Tenkanen, M. (2004) Adsorption of Trichoderma reesei CBH I and EG II and their catalytic domains on steam pretreated softwood and isolated lignin. *J Biotechnol* **107**, 65-72

18. Sammond, D. W., Yarbrough, J. M., Mansfield, E., Bomble, Y. J., Hobdey, S. E., Decker, S. R., Taylor, L. E., Resch, M. G., Bozell, J. J., Himmel, M. E., Vinzant, T. B., and Crowley, M. F. (2014) Predicting enzyme adsorption to lignin films by calculating enzyme surface hydrophobicity. *J Biol Chem* **289**, 20960-20969

19. Payne, C. M., Knott, B. C., Mayes, H. B., Hansson, H., Himmel, M. E., Sandgren, M., Ståhlberg, J., and Beckham, G. T. (2015) Fungal cellulases. *Chem Rev* **115**, 1308-1448

20. Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J* **382**, 769-781

21. Linder, M., Mattinen, M. -L., Kontteli, M., Lindeberg, G., Ståhlberg, J., Drakenberg, T., Reinikainen, T., Pettersson, G., and Annila, A. (1995) Identification of functionally important amino acids in the cellulose-binding domain of Trichoderma reesei cellobiohydrolase I. *Protein Science* **4**, 1056-1064

22. Rahikainen, J. L., Martin-Sampedro, R., Heikkinen, H., Rovio, S., Marjamaa, K., Tamminen, T., Rojas, O. J., and Kruus, K. (2013) Inhibitory effect of lignin during cellulose bioconversion: the effect of lignin chemistry on non-productive enzyme adsorption. *Bioresour Technol* **133**, 270-278

23. Pfeiffer, K. A., Sorek, H., Roche, C., Strobel, K., Blanch, H. W., and Clark, D. S. (2015) Evaluating endoglucanase Cel7B-lignin interaction mechanisms and kinetics using quartz crystal microgravimetry. *Biotechnol Bioeng*

24. Börjesson, J., Engqvist, M., Sipos, B., and Tjerneld, F. (2007) Effect of poly(ethylene glycol) on enzymatic hydrolysis and adsorption of cellulase enzymes to pretreated lignocellulose. *Enzyme Microb Technol* **41**, 186-195

25. Nordwald, E. M., Brunecky, R., Himmel, M. E., Beckham, G. T., and Kaar, J. L. (2014) Charge engineering of cellulases improves ionic liquid tolerance and reduces lignin inhibition. *Biotechnol Bioeng*

26. Rahikainen, J. L., Evans, J. D., Mikander, S., Kalliola, A., Puranen, T., Tamminen, T., Marjamaa, K., and Kruus, K. (2013) Cellulase-lignin interactions-the role of carbohydrate-binding module and pH in non-productive binding. *Enzyme Microb Technol* **53**, 315-321

27. Shill, K., Padmanabhan, S., Xin, Q., Prausnitz, J. M., Clark, D. S., and Blanch, H. W. (2011) Ionic liquid pretreatment of cellulosic biomass: enzymatic hydrolysis and ionic liquid recycle. *Biotechnol Bioeng* **108**, 511-520

28. Labbé, S., and Thiele, D. J. (1999) Copper ion inducible and repressible promoter systems in yeast. *Methods Enzymol* **306**, 145-153

29. Dana, C. M., Saija, P., Kal, S. M., Bryan, M. B., Blanch, H. W., and Clark, D. S. (2012) Biased clique shuffling reveals stabilizing mutations in cellulase Cel7A. *Biotechnol Bioeng* **109**, 2710-2719

30. Zheng, L., Baumann, U., and Reymond, J. L. (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res* **32**, e115

31. Voutilainen, S. P., Puranen, T., Siika-Aho, M., Lappalainen, A., Alapuranen, M., Kallio, J., Hooman, S., Viikari, L., Vehmaanperä, J., and Koivula, A. (2008) Cloning, expression, and characterization of novel thermostable family 7 cellobiohydrolases. *Biotechnol Bioeng* **101**, 515-528

32. Voutilainen, S. P., Murray, P. G., Tuohy, M. G., and Koivula, A. (2010) Expression of Talaromyces emersonii cellobiohydrolase Cel7A in Saccharomyces

cerevisiae and rational mutagenesis to improve its thermostability and activity. *Protein Eng Des Sel* **23**, 69-79

33. Beckham, G. T., Matthews, J. F., Bomble, Y. J., Bu, L., Adney, W. S., Himmel, M. E., Nimlos, M. R., and Crowley, M. F. (2010) Identification of amino acids responsible for processivity in a Family 1 carbohydrate-binding module from a fungal cellulase. *J Phys Chem B* **114**, 1447-1453

34. Nimlos, M. R., Beckham, G. T., Matthews, J. F., Bu, L., Himmel, M. E., and Crowley, M. F. (2012) Binding preferences, surface attachment, diffusivity, and orientation of a family 1 carbohydrate-binding module on cellulose. *J Biol Chem* **287**, 20603-20612

35. Alekozai, E. M., GhattyVenkataKrishna, P. K., Uberbacher, E. C., Crowley, M. F., Smith, J. C., and Cheng, X. (2013) Simulation analysis of the cellulase Cel7A carbohydrate binding module on the surface of the cellulose Iβ. *Cellulose* **21**, 951-971

36. Scott, B. R., St-Pierre, P., Lavigne, J., Masri, N., White, T. C., and Tomashek, J. J. (2010) Novel lignin-resistant cellulase enzymes. *US 20100221778 A1*

37. Nidetzky, B., Steiner, W., and Claeyssens, M. (1994) Cellulose hydrolysis by the cellulases from Trichoderma reesei: adsorptions of two cellobiohydrolases, two endocellulases and their core proteins on filter paper and their relation to hydrolysis. *Biochem. J* **303**, 817-823

38. Kraulis, J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J., and Gronenborn, A. M. (1989) Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from Trichoderma reesei. A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. *Biochemistry* **28**, 7241-7257

Chapter 4: Engineering Cel7A carbohydrate-binding module and linker for reduced lignin inhibition

4.1 Abstract

Non-productive binding of cellulases to lignin inhibits enzymatic hydrolysis of biomass, increasing enzyme requirements and therefore the overall cost of producing biofuels. This study used site directed mutagenesis of the *T. reesei* Cel7A carbohydrate binding module (CBM) and linker to investigate the mechanisms of cellulase lignin adsorption and engineer an enzyme with reduced lignin affinity. Seven positions in the CBM were mutated to amino acids with varying properties to uncover features that improve cellulose/lignin selectivity. CBM mutations that added hydrophobic or positively charged residues decreased the selectivity toward cellulose, while mutations that added negatively charged residues increased the selectivity. Mutating the linker to alter predicted glycosylation patterns greatly impacted lignin, but not cellulose, affinity. Beneficial mutations were combined to generate a mutant with 2.5 fold less lignin affinity and fully retained cellulose affinity. This mutant was not inhibited by added lignin during hydrolysis of Avicel and generated 40% more glucose than the wild type enzyme from dilute acid-pretreated *Miscanthus*.

4.2 Introduction

The conversion of lignocelluosic biomass to fuels presents an opportunity to greatly reduce greenhouse gas emissions (1). Unfortunately, the high cost of hydrolytic enzymes impedes the development and commercial production of lignocellulosic biofuels. Engineering cellulase enzymes for increased catalytic efficiency and operational stability could greatly reduce this cost by decreasing the amount of enzyme required. Cellulase efficiency is hindered by non-productive adsorption to lignin, a major component of biomass. The addition of lignin to cellulose can reduce hydrolysis yields by as much as 80%, highlighting the detrimental impact of lignin adsorption (2). Cellulase adsorption to lignin and the resulting inhibition has been well documented for numerous different enzymes and lignin sources.

Most fungal cellulases consist of a catalytic domain (CD) and a carhobydrate-binding module (CBM) connected by a flexible, highly glycosylated linker (3). The CBM is responsible for binding to cellulose, increasing the concentration of the enzyme on its substrate (4). The function of the linker is less well elucidated, though it is suggested to also increase the affinity of the enzyme for cellulose (5). Removing the CBM and linker

from fungal cellulases greatly diminishes lignin affinity as well as cellulose affinity, highlighting their important role in adsorption to both cellulose and lignin (6-8).

Hydrophobic interactions are hypothesized to be the predominant mechanism behind non-productive adsorption to lignin. One study found that lignin adsorption correlated with exposed hydrophobic clusters (9) and another study showed greatly diminished lignin affinity upon mutation of surface exposed aromatic residues (see Chapter 3). Electrostatic interactions have also been shown to contribute to lignin adsorption. Increasing the negative charge in lignin (due to increased carboxylic acid content and/or pH) or in cellulases (due to increased pH or charge engineering) has been shown to reduce lignin inhibition, likely due to repulsive electrostatic interactions (10-12).

Thought the CBM is thought to be responsible for the majority of lignin binding, potential interactions of the linker with lignin have yet to be ruled out. A patent describes lignin resistant Cel6A and Cel7A enzymes that have mutations only in the linker region, indicating that this region contributes significantly to lignin adsorption. The mutations decrease the calculated isoelectric point and/or increase the threonine:serine ratio in the linker. Changing the threonine:serine ratio likely impacts the glycosylation patterns of the linker (13). Glycosylation patterns of the CBM are known to affect cellulose affinity (14), but the impact of glycosylation patterns on lignin affinity has not been investigated. The presence of linker-lignin interactions and the impact of glycosylation on lignin affinity warrant further study to increase our understanding of cellulase-lignin interactions and generate lignin resistant enzymes.

Many engineering efforts to reduce lignin inhibition have focused on blocking lignin with additives in order to reduce cellulase-lignin interactions. Addition of supplemental proteins, surfactants, and polymers has been shown to reduce lignin adsorption and improve hydrolysis (15-18). Modification of lignin or cellulases to add negative charge has also been a successful engineering strategy (10,11). Removing the CBM and linker has also be reported to alleviate lignin inhibition (19), however removing the CBM typically results in reduced hydrolysis yields due to diminished cellulose binding.

The most successful strategies to reduce lignin inhibition to date require treatment of the biomass or cellulases with additional materials, increasing the overall cost of producing biofuels and potentially impacting downstream processing. Engineering a mutant cellulase with lower affinity for lignin can achieve the same reduction in lignin inhibition without the extra cost or complication of additives. In this work, we generated 27 Cel7A enzymes with point mutations in the linker or CBM and screened them for binding to lignin and cellulose. Mutations were chosen to test a range of amino acids with different functionality. We combined the mutations that conferred the greatest selectivity toward cellulose and measured their binding affinity and hydrolysis capabilities.

4.3 Materials and Methods

Materials

Microcrystalline cellulose PH-105 (Avicel), Protease from *Streptomyces griseus,* 4-methylumbelliferyl β-D-cellobioside, Bovine serum albumin (BSA), cellulase from *T. reesei* (Celluclast) and papain were purchased from Sigma. Accelerace 1500 was a gift from Dupont. *A. niger* β-glucosidase was purchased from Megazyme. Acid-pretreated *Miscanthus* was produced by two-step dilute acid pretreatment on the pilot scale by Andritz, Glens Falls, NY. Solids were washed with water to neutral pH and lyophilized before lignin preparation.

Lignin preparation

Lignin residues were isolated from acid pretreated miscanthus by enzymatic hydrolysis followed by *S. griseus* protease treatment to remove bound enzyme. Pretreated, rinsed, lyophilized biomass was ball milled for 5 min (Kleco ball mill, Visalia, CA). Enzymatic hydrolysis was carried out at 50°C, pH 5.0, with 4 wt% biomass and 0.25 ml Accelerase 1500 per gram of biomass. The reaction mixtures were centrifuged and the supernatant replaced with fresh buffer and enzyme (0.25 ml Accelerase/g original biomass) every 24 hrs. After 7 days of hydrolysis, the solids were separated by centrifugation and washed with acidified water (pH adjusted to 3.5 with HCl to avoid lignin solubilization). Bound protein was removed by an overnight protease treatment at 37°C, pH 7.5, with 2wt% solids and 0.3 mg/ml protease. The solids were then washed with acidified water and heated in a boiling water bath for 60 min to deactivate any remaining protease. The solids were extensively washed with acidified water and freeze-dried.

Lignin analysis

Compositional analysis of pretreated biomass and lignin was performed as previously described, with sugar concentrations determined by HPLC (Dionex) (20). Protein content in the solid samples was determined by total nitrogen analysis. The analysis was carried out at the UC Berkeley Microanalytical Facility using a Perkin Elmer 2400 Series II combustion analyzer.

Purification of T. reesei Cel7A

Cel7A was purified from Celluclast using a Q Sepharose anion exchange column followed by a MonoQ ion exchange column. Fractions from the Q Sepharose column containing a band for Cel7A on an SDS-PAGE gel were combined and treated with 0.1% Tween20 to disrupt non-specific interactions that were found to exist between Cel7A and Cel7B immediately prior to injection to the MonoQ. Fractions from the MonoQ were checked for purity by SDS-PAGE and an assay for contaminant endoglucanase activity. Assays contained 2% carboxymethyl cellulose and roughly 10 μM enzyme in 100 mM sodium acetate, pH 4.5. Assays were incubated at 50 °C overnight and endoglucanase activity was assessed by adding 2,4-dinitrosalycilic acid, incubating at 90 °C for 5 minutes, and visually inspecting for the presence of brown color in the reactions.

Fractions showing no endoglucanase activity were pooled and buffer exchanged into 50 mM sodium acetate, pH 5.0.

Isolation of the T. reesei CD

The *T. reesei* CD was cleaved from the linker and CBM using the protease papain. Papain was activated by incubating a 28 $mg \cdot ml^{-1}$ solution with 2 mM dithiothreitol and 2 mM ethylenediamine tetraacetic acid in 100 mM sodium phosphate, pH 7.0 for 30 minutes at room temperature. Cleavage reactions contained a w/w ratio of 1:100 papain to Cel7A in 50 mM sodium acetate, pH 5.0 and were incubated overnight at room temperature with stirring. The CD was purified from the protease and CBM using a Q Sepharose anion exchange column.

Small-scale enzyme production

The *T. emersonii* Cel7A CD or a fusion of *T. emersonii* Cel7A catalytic domain and the *T. reesei* Cel7A linker and CBM was cloned into a high-copy number plasmid (pCu424 (21)) and produced in YVH10 *pmr1*Δ *S. cerevisiae (22)*. Mutations were made by site directed mutagenesis (23). For screening all mutants, 20 mL SC-Trp medium was inoculated with *S. cerevisiae* containing the Cel7A gene and grown for 3 days at 30°C, 220 rpm. Cultures were spun down at 5,000 x g for 5 min and resuspended in SC medium supplemented with 500 mM Cu_2SO_4, then cultured for protein production for 3 days at 25°C, 220 rpm. Supernatant was collected and buffer exchanged into 50 mM sodium acetate, pH 5.0, using Amplicon spin concentrators (Millipore).

Large-scale enzyme production and purification

For large-scale production, 1 L of SC-Trp medium was inoculated with *S. cerevisiae* containing the Cel7A gene and grown for 3 days at 30°C, 220 rpm. Cultures were spun down at 5,000 x g for 5 min and resuspended in YPD medium supplemented with 500 mM Cu_2SO_4 and cultured for 3 days at 25°C, 220 rpm. Enzyme purification was carried out using a two-step procedure beginning with anion exchange chromatography. Active fractions were combined and further purified by hydrophobic interaction chromatography (HIC). Pure fractions (single band via SDS-PAGE) were combined and buffer exchanged into 50 mM sodium acetate buffer, pH 5.0.

Enzyme quantification

Purified enzymes were quantified by absorbance at 280 nm using the calculated molar extinction coefficient of 76,240 $M^{-1}cm^{-1}$ for the fusion enzyme and 71,120 $M^{-1}cm^{-1}$ for the catalytic domain. Unpurified enzyme and enzyme remaining in the supernatant after binding assays were quantified by their activity on a soluble substrate, 4-methylumbelliferyl β-D-cellobioside (MUC). The specific activity of 10.2 μM MU $(\mu M \text{ enzyme} * \text{min})^{-1}$ was measured for the purified wild type, catalytic domain, and mutant chimeras. Assays were conducted in a total volume of 100 μl with 1 mM MUC, pH 5.0, at 45°C for 8 min, followed by a denaturation step at 98°C for 2 min. Sodium hydroxide was added to a final concentration of 0.1 M and the fluorescence was measured on a Paradigm plate reader (Beckman Coulter) with an excitation of 360 nm and emission of 465 nm.

Avicel and lignin binding screen

Unpurified, buffer-exchanged mutants were normalized to the same concentration of active Cel7A as determined by activity on a soluble substrate. All mutants contained identical catalytic domains, and therefore should have equivalent specific activity on soluble substrates. Binding to Avicel and lignin was measured in 50 mM sodium acetate buffer, pH 5, using 15 mg/ml of Avicel or lignin and 125 nM enzyme. Experiments were conducted in a total volume of 70 uL in a rotating shaker (300 rpm) at room temperature. An initial time course with the wild type enzyme was measured to determine the time required for binding to reach equilibrium. Equilibrium was reached after approximately 30 min. Subsequent assays were performed for 1 hr. After equilibration, solids were separated by centrifugation and the supernatant was analyzed for free, active enzyme using MUC. The amount of bound enzyme was calculated from the difference between added enzyme and enzyme remaining in solution. The values plotted in Figure 4.2 are bound enzyme divided by enzyme remaining in solution. At the low enzyme concentrations employed, these values approximate the partition coefficient, α, as defined in Equation 1. The values reported are averages of duplicate experiments and the errors are calculated as standard deviations.

The one-site Langmuir isotherm at low free protein concentrations can be written as

$$\Gamma = \Gamma_{\max} K_A C = \alpha C \qquad (1)$$

where Γ is the amount of adsorbed protein, Γ_{max} is the amount of adsorbed protein at saturation, K_A is the Langmuir affinity constant, C is the concentration of free protein, and α is the partition coefficient.

Hydrolysis of Avicel supplemented with lignin and Miscanthus

Hydrolysis reactions were performed with 10 mg/ml of Avicel, 10 mg/ml Avicel supplemented with 10 mg/ml of isolated lignin, 20 mg/ml acid-pretreated *Miscanthus*, or 20 mg/ml acid-pretreated *Miscanthus* supplemented with 20 mg/ml BSA in 50 mM sodium acetate buffer, pH 5, with 0.05 μmol Cel7A/g cellulose. All reactions were supplemented with 40 U/mL of *A. niger* β-glucosidase. For Miscanthus reaction containing BSA, the Miscanthus, BSA, and buffer were mixed at room temperature for 45 min before cellulase addition. Reactions were carried out in a thermally controlled incubator with end-over-end mixing at 60°C. After 48 hours, reaction mixtures were stopped by heating to 98 °C for 3 min, filtered, and analyzed for glucose by HPLC (Shimadzu). The values reported are means of triplicate samples and the errors are calculated as standard deviation.

4.4 Results

Isolation and characterization of lignin from pretreated Miscanthus

Lignin used in this study was isolated from acid pretreated miscanthus using a commercial cellulase mixture (Accelerase 1500, Dupont). Excess enzyme and long digestion times were used to remove all accessible carbohydrates. The solid residue was

then treated with *S. griseus* protease to remove bound enzymes. The compositions of the starting biomass and resulting isolated lignin are presented in Table 1. The Klason lignin, remaining carbohydrates, and protein content are similar to protease treated lignin used in previous studies (2,6,12,24,25).

Table 4.3 Characterization of pretreated *Miscanthus* and isolated lignin

	Glucan (%)	Xylan (%)	Arabinan (%)	Klason lignin (%)	Ash (%)	Nitrogen (%)
Acid-pretreated *Miscanthus*	55.6	0.54	0.11	33.6	9.73	< 0.2
Isolated lignin	6.21	0.39	0.13	77.0	14.51	0.34

Engineering approach

This study focused on reducing the lignin affinity of the *T. reesei* Cel7A linker and CBM. The *T. reesei* Cel7A linker and CBM were produced with the *T. emersonii* Cel7A CD (Te-Tr chimera) in a strain of *S. cerevisiae* with reduced hyperglycosylation. Four residues in the CBM (Q2, H4, V18, and P30) were chosen for mutation based on favorable changes in cellulose/lignin selectivity when previously mutated to alanine (see Chapter 3). Three additional residues (Y5, V27, and L28) were selected for mutation due to their hydrophobic nature. The CBM structure showing the mutated residues is presented in Figure 4.1. The Cel7A mutants were secreted into defined medium, buffer exchanged, and screened for cellulose and lignin binding without purification. Binding was quantified by measuring the Cel7A activity remaining in the supernatant using a soluble substrate, which permitted specific detection of Cel7A among other secreted proteins. Each mutant was screened for binding at low enzyme concentration, in the range where the Langmuir isotherm is linear, allowing direct calculation of the partition coefficient from the amount of bound and free enzyme (Equation 1 in Experimental Procedures).

Impact of point mutations on CBM adsorption to Avicel and lignin

The Avicel and lignin partition coefficient of the wild type chimera, the CD, and each mutant is shown in Figure 4.2. H4V, Y5W, V18R and P30K exhibited increased lignin affinity and V27E displayed increased Avicel affinity compared to the wild type chimera. Many of the mutants showed reduced lignin affinity, though reduced lignin affinity was often accompanied by a corresponding drop in Avicel affinity. None of the mutations reduced lignin binding to the level of the isolated CD.

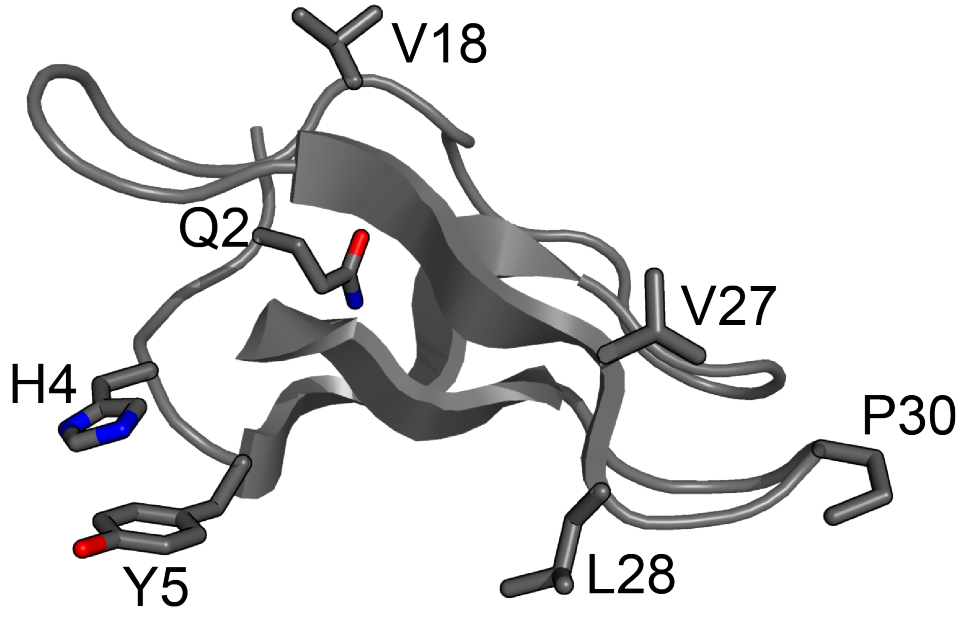

Figure 4.1 *T. reesei* Cel7A CBM structure.
The structure shows side chains mutated in this study (26).

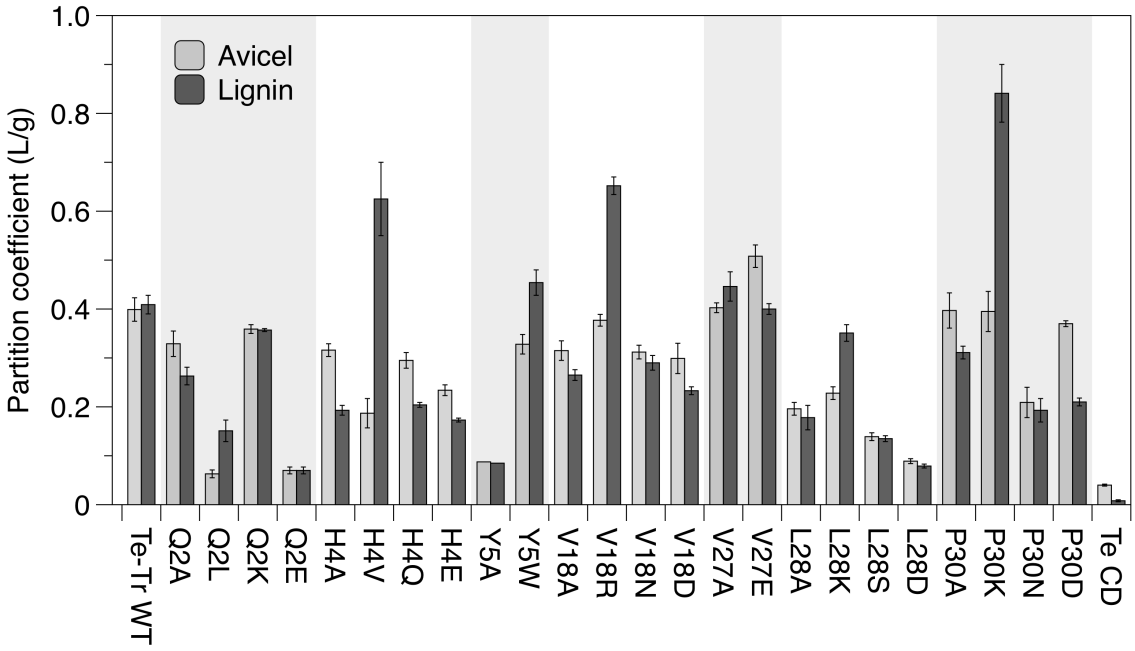

Figure 4.2 Enzyme partition coefficients (α) for Avicel and lignin.
Partition coefficients are calculated from duplicate binding experiments at room temperature and an initial enzyme concentration of 125nM. Values presented are means of duplicate samples and errors are standard deviations.

Impact of point mutations on CBM selectivity

Improving cellulose-hydrolyzing activity in the presence of lignin requires significant reductions in CBM/lignin affinity while maintaining a high level of cellulose binding. It is therefore important to consider the cellulose/lignin selectivity of each mutant. To visualize each mutant's selectivity, a scatterplot of partition coefficients is presented in Figure 4.3. Each data point is colored based on the type of amino acid substituted for the native residue. In general, mutation to positively charged and hydrophobic residues decreased the selectivity toward Avicel. Mutation to negatively charged residues increased the selectivity toward Avicel in some, but not all cases. P30D and V27E had the greatest Avicel selectivity (Avicel partition coefficient divided by lignin partition coefficient) of the mutants tested in this study. P30D reduced lignin binding while maintaining cellulose affinity equivalent to the wild type. V27E did not change lignin binding, but resulted in greater affinity for Avicel than the wild type.

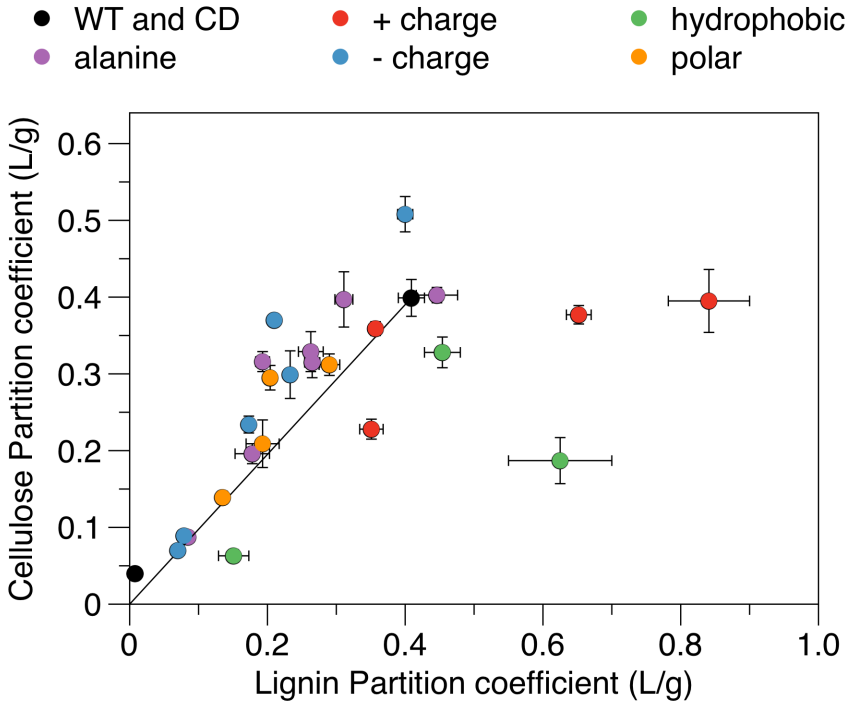

Figure 4.3 Binding selectivity of Te-Tr wild type, CD, and all mutants. Binding selectivity as evidenced by a scatter plot of partition coefficients. Data is colored by type of amino acid substituted. The line is drawn between the wild type chimera and the origin to guide the eye. Values presented are means of duplicate samples and errors are standard deviations.

Impact of linker mutations on Avicel and lignin affinity

We designed and tested three linker mutants (Figure 4.4A) to determine if the linker contributes to lignin affinity and if it can be engineered for reduced lignin affinity. Each mutant was designed to alter the predicted glycosylation patterns when expressed in

S. cerevisiae. The Link1 mutant replaced a positively charged residue with a threonine in addition to adding new predicted glycosylation sites. The Link2 and Link3 mutants reduced the number of predicted glycosylation sites. The partition coefficients to Avicel and lignin, along with the selectivity, can be compared in Figure 4.4B. All three mutants retained Avicel affinity similar to the wild type chimera. Link2 and Link3 (predicted to have reduced glycosylation) exhibited increased affinity for lignin while Link1 (predicted to have increased glycosylation) displayed reduced affinity for lignin.

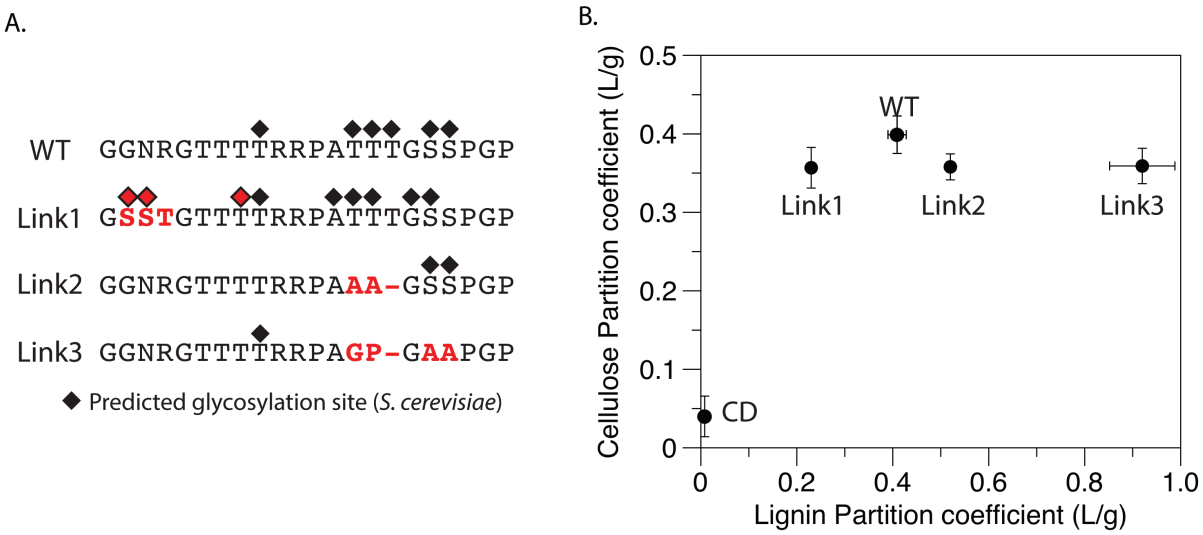

Figure 4.4 Linker mutant sequences and selectivity.
A) Sequences of linker mutants with mutations highlighted in red and predicted O-glycosylation site marked with a diamond. Glycosylation prediction from (27).
B) Binding selectivity of Te-Tr wild type, CD, and linker mutants, as evidenced by a scatter plot of partition coefficients. Values presented are means of duplicate samples and errors are standard deviations.

Selectivity of combined mutations

To generate an enzyme with the greatest reductions in lignin affinity while maintaining cellulose affinity, the three mutations that conferred the greatest cellulose selectivity (V27E, P30D, and Link1) were combined. The partition coefficients to Avicel and lignin of the combined mutants along with the single mutants for comparison are shown in Figure 4.5. Combining the two mutations that decreased lignin affinity (P30D and Link1) resulted in further reductions in lignin adsorption, though the effects were not fully additive. Combining the mutation that increased cellulose affinity (V27E) with P30D or Link1 resulted in mutants with both increased cellulose affinity and decreased lignin affinity. The triple mutant exhibited the highest selectivity for cellulose.

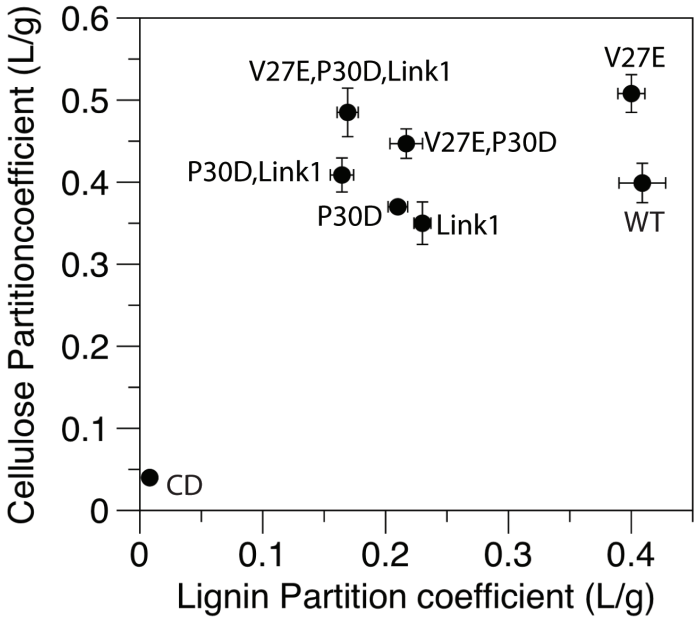

Figure 4.5 Binding selectivity of combined mutants.
Binding selectivity as evidenced by a scatter plot of partition coefficients. Values presented are means of duplicate samples and errors are standard deviations.

Hydrolysis of Avicel with supplemental lignin and Miscanthus

Having generated mutants with increased specificity toward cellulose, we next tested their ability to hydrolyze Avicel in the presence of lignin. Figure 4.6A shows the glucose generated by the wild type chimera, the *T. emersonii* CD, the 6 best mutants, the natively expressed *T. reesei* enzyme, and the *T. reesei* CD. Figure 4.6B shows the ratio of inhibited (with lignin) to non-inhibited activity. The wild type chimera generated 20% less glucose in the presence of lignin while the *T. emersonii* CD was not inhibited at all by added lignin. The *T. reesei* enzyme generated 85% less glucose and the *T. reesei* CD generated 65% less glucose in the presence of lignin. The best mutant, the triple mutant (V27E, P30D, Link1), was not inhibited by lignin, generating the same amount of glucose regardless of the presence of lignin.

The best mutants were also tested for their ability to hydrolyze a commercially relevant feedstock, dilute acid-pretreated *Miscanthus* (Figure 4.7A). The Te-Tr wild type and mutant chimera enzymes performed significantly better than the *T. reesei* enzyme. The best mutant, the triple mutant, generated 40% more glucose than the wild type Te-Tr chimera. To investigate how much of the improvement in hydrolysis was due to reduced lignin binding, we also tested the ability of each enzyme to hydrolyze *Miscanthus* that have been pretreated with BSA to block lignin adsorption (Figure 4.7B). The BSA treatment increased the glucose generated by the Te-Tr wild type chimera 1.4 fold and the *T. reesei* enzyme 5 fold. BSA did not improve the activity of the triple mutant. With the addition of BSA, the wild type Te-Tr chimera, the best mutant, and the *T. reesei* enzyme produced approximately equivalent amounts of glucose.

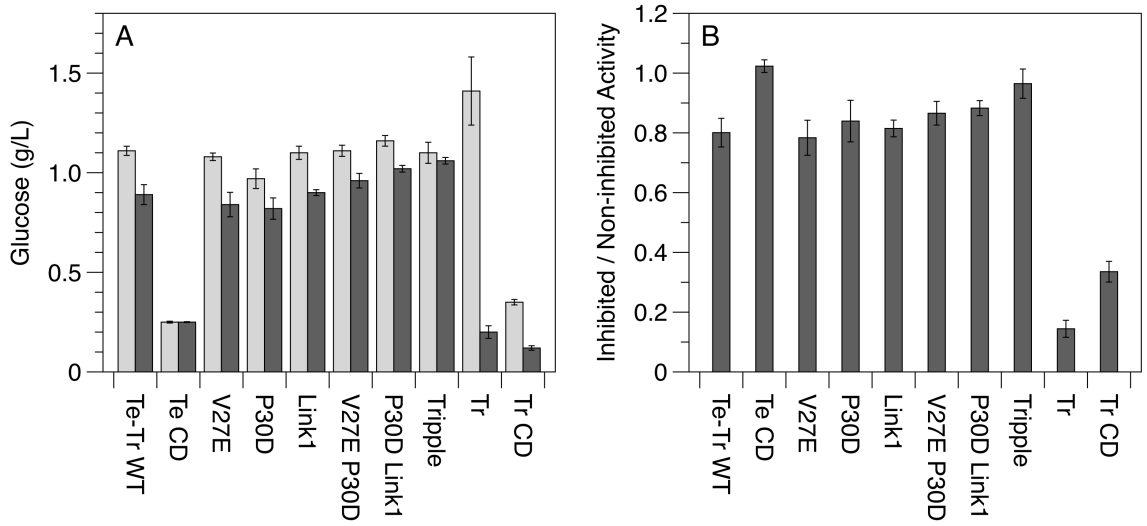

Figure 4.6 Hydrolysis of Avicel and Avicel with added lignin.
A) Glucose produced after 48 hours with 10 g/L Avicel or 10 g/L Avicel with 10 g/L lignin. All reactions were supplemented with *A. niger* β-glucosidase. Reactions were performed at 60°C with 0.05 μmol Cel7A per gram of cellulose. Values presented are means of triplicate samples and errors are standard deviations.
B) Ratio of glucose produced from Avicel with added lignin to glucose produced without added lignin.

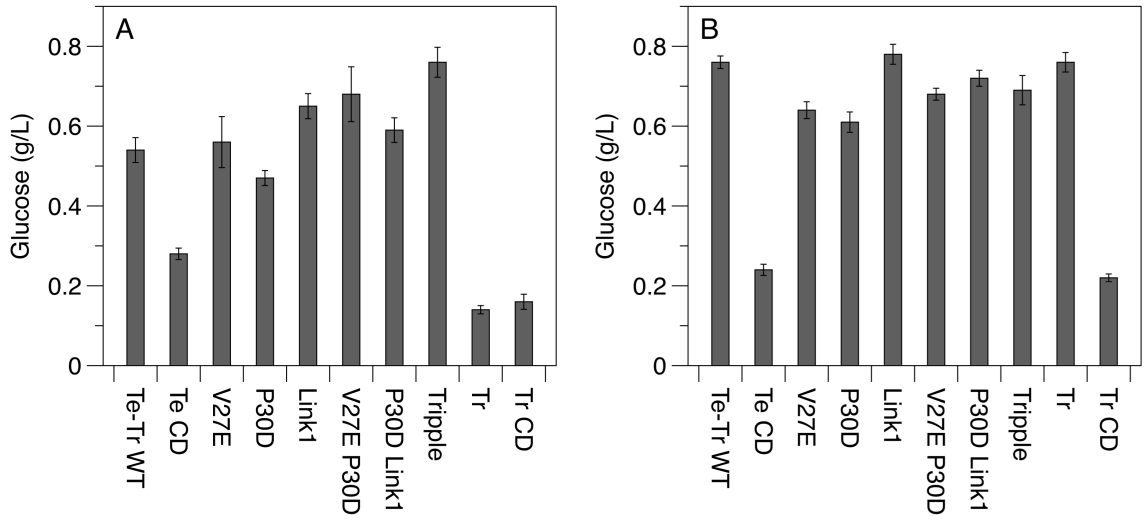

Figure 4.7 Hydrolysis of acid pretreated *Miscanthus*.
Glucose produced after 48 hours with 20 g/L acid-pretreated *Miscanthus* without (A) or with BSA pretreatment (B). All reactions were supplemented with *A. niger* β-glucosidase. Reactions were performed at 60°C with 0.05 μmol Cel7A per gram of cellulose. Values presented are means of triplicate samples and errors are standard deviations.

4.5 Discussion

Non-productive adsorption of cellulase to lignin inhibits the hydrolysis of lignocellulose biomass, increasing the cost of producing biofuels. Despite the numerous studies documenting cellulase adsorption to lignin, few attempts have been made to engineer enzymes for reduced lignin affinity. We used site-directed mutagenesis of the *T. reesei* Cel7A CBM and linker to generate an enzyme with reduced lignin affinity and elucidate trends that can inform rational engineering efforts of other CBMs and linkers. We chose to engineer the CBM and linker from *T. reesei* Cel7A because it is the most abundant enzyme in commercial cellulase mixtures and is greatly inhibited by lignin (28) (see Chapter 2). The *T. reesei* Cel7A linker and CBM mutants were expressed with a homologous Cel7A catalytic domain (CD) from *T. emersonii* to enable efficient expression in *S. cerevisiae*.

Four residues in the CBM were chosen for mutation based on favorable changes in cellulose/lignin selectivity when previously mutated to alanine (see Chapter 3) and an addition three residues where chosen to test mutations to hydrophobic residues. Each residue was mutated to alanine and 1-3 other amino acids chosen to vary the charge or hydrophobicity at that position. We found that mutating residues to hydrophobic or positively charged amino acids increased the selectivity toward lignin, providing strong evidence that both hydrophobic and electrostatic interactions contribute to lignin binding. Substituting negatively charged residues at the CBM positions studied increased the selectivity in most, but not all cases. Scanning through the CBM with aspartic or glutamic acid mutations could be a fruitful strategy for finding mutants with even greater cellulose selectivity.

To investigate the impact of linker glycosylation on lignin affinity, three mutated linkers were designed to alter the predicted glycosylation patterns when expressed in *S. cerevisiae* (27). Removing predicted glycosylation sites from the linker lead to increased lignin affinity while adding predicted glycosylation sites (and simultaneously mutating a positively charged residue) decreased lignin affinity. If the O-linked saccharides contain mannosyl phosphate, altering glycosylation sites can add or remove negative charges from the linker, modifying electrostatic interactions between the enzyme and lignin. All three linker mutants had identical affinity for cellulose, indicating that glycosylation of the linker does not impact cellulose binding. Although the linker has been shown to bind to cellulose (5) and CBM glycosylation patterns have been shown to impact cellulose affinity (14), to our knowledge, this is the first study to example linker glycosylation as a factor in affinity.

Confirming the glycosylation patterns of the linker and extending the number of linker mutations tested is needed to increase our understanding of cellulase-lignin affinity and is also a promising strategy for generating a lignin resistant mutant. Our best linker mutant simultaneously removed a positive charge and added predicted glycosylation sites. Further experimentation is required to deconvolute the mechanism by which these mutations reduce lignin binding.

Combining linker and CBM mutations that increase cellulose specificity lead to further increases in selectivity. The most selective mutant (triple mutant-V27E, P30D, Link1) exhibited a 2.5 fold reduction in lignin affinity compared to the wild type chimera, though the mutant retained considerable lignin affinity compared to the isolated catalytic domain. The increased selectivity of the triple mutant translated to significantly less lignin inhibition during hydrolysis. The triple mutant generated approximately the same amount of sugar from Avicel in the presence and absence of lignin. It also generated 40% more sugar than the wild type chimera from acid-pretreated *Miscanthus,* a commercially relevant feedstock. When BSA was added to Miscanthus prior to cellulase addition, the difference between the mutant and the wild type enzyme disappeared. Since BSA improves hydrolysis by blocking lignin adsorption (15), this result provides further evidence that the triple mutant improves hydrolysis through reduced non-productive adsorption to lignin.

The *T. emersonii* CD - *T. reesei* linker and CBM chimera used for mutagenesis in this study is less inhibited by lignin than the native *T. reesei* Cel7A enzyme. We therefore expect the beneficial mutations in the CBM (V27E and P30D) to translate to equal or even greater improvements in lignin resistance when transferred to the native *T. reesei* enzyme. The best mutant chimera with no mutations in the linker, P30D-V27E, generated 25% more sugar from Miscanthus than the wild type chimera. The benefits of the exact linker mutations identified in this study may not translate to equal improvements when expressed in *T. reesei* due to differences in O-linked glycosylation patterns between *T. reesei* and *S. cerevisiae*. Nevertheless, expressing and screening a limited set of linker mutations designed to remove positive charges and increase glycosylation in *T. reesei* should yield improved mutants.

In addition to Cel7A, lignin adsorption decreases the efficiency of other cellulases in commercial enzyme mixtures from *T. reesei* (6,18). Since the other main cellulases contain a homologous CBM, our work engineering the Cel7A CBM and linker can potentially be extended to reduce the lignin affinity of the other enzymes. Figure 4.8 shows an alignment of homologous CBMs with each residue colored by the degree of conservation. The first 5 sequences are the CBMs of *T. reesei* enzymes. The P30D mutation we identified is naturally present in Cel7B (EGI), and Cel6A (CBH2). The V27E mutation is not naturally present in any of the *T. reesei* enzymes, but is present in homologous CBMs from *P. chrysosporium*. Adding these mutations to the CBMs of the other main *T. reesei* cellulases may increase their selectivity for cellulose and reduce lignin inhibition.

The linker sequences from four main *T. reesei* cellulases are shown in Figure 4.9. The linkers vary in length from 25 (Cel7A and B) to 41 residues (Cel6A). All of the linkers are rich in proline, serine, and threonine and each contains between two and four positively charged arginine residues. The Cel6A linker is particularly interesting due to the C-terminal end having little glycosylation, a positively charged residue, and two hydrophobic residues. Given these attributes, I expect the Cel6A linker to greatly contribute to lignin affinity and inhibition. Studying the lignin affinity and inhibition of a

cellulase with this linker and various mutant versions of it could lend more insight into linker-lignin interactions. Engineering the linkers of the main *T. reesei* enzymes by removing the positively charged residues and adding glycosylation sites may reduce their affinity for lignin and result in less inhibition.

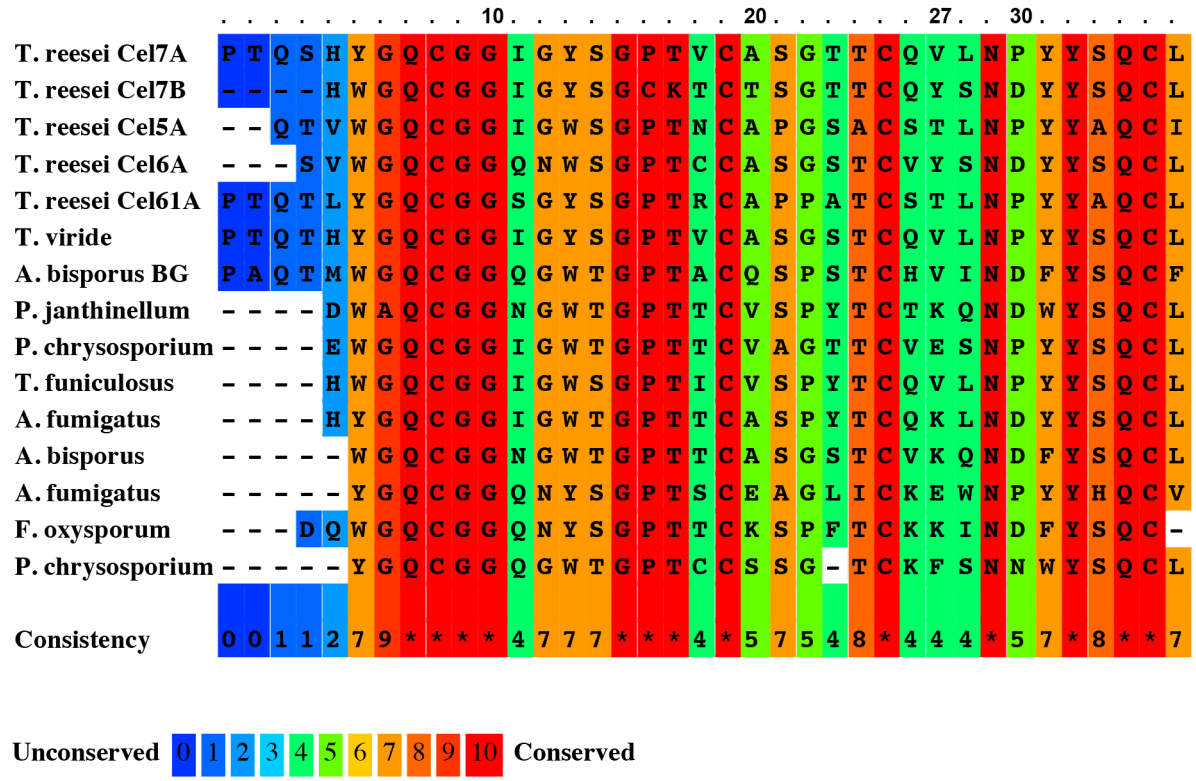

Figure 4.8 Alignment of family-one CBMs.
Each residue is colored based on conservation.

Figure 4.9 *T. reesei* linker sequences with putative O-glycosylation sites.

Acknowledgements: We thank Stefan Bauer for assistance with sample analysis and Yuzhang Chen for assistance with protein production. This work was supported by the Energy Biosciences Institute.

4.6 References

1. Caspeta, L., and Nielsen, J. (2013) Economic and environmental impacts of microbial biodiesel. *Nature biotechnology* **31**, 789-793
2. Rahikainen, J., Mikander, S., Marjamaa, K., Tamminen, T., Lappas, A., Viikari, L., and Kruus, K. (2011) Inhibition of enzymatic hydrolysis by residual lignins from softwood--study of enzyme binding and inactivation on lignin-rich surface. *Biotechnol Bioeng* **108**, 2823-2834
3. Payne, C. M., Knott, B. C., Mayes, H. B., Hansson, H., Himmel, M. E., Sandgren, M., Ståhlberg, J., and Beckham, G. T. (2015) Fungal cellulases. *Chem Rev* **115**, 1308-1448
4. Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J* **382**, 769-781
5. Payne, C. M., Resch, M. G., Chen, L., Crowley, M. F., Himmel, M. E., Taylor, L. E., Sandgren, M., Ståhlberg, J., Stals, I., Tan, Z., and Beckham, G. T. (2013) Glycosylated linkers in multimodular lignocellulose-degrading enzymes dynamically bind to cellulose. *Proc Natl Acad Sci U S A* **110**, 14646-14651
6. Palonen, H., Tjerneld, F., Zacchi, G., and Tenkanen, M. (2004) Adsorption of Trichoderma reesei CBH I and EG II and their catalytic domains on steam pretreated softwood and isolated lignin. *J Biotechnol* **107**, 65-72
7. Rahikainen, J. L., Martin-Sampedro, R., Heikkinen, H., Rovio, S., Marjamaa, K., Tamminen, T., Rojas, O. J., and Kruus, K. (2013) Inhibitory effect of lignin during cellulose bioconversion: the effect of lignin chemistry on non-productive enzyme adsorption. *Bioresour Technol* **133**, 270-278
8. Pfeiffer, K. A., Sorek, H., Roche, C., Strobel, K., Blanch, H. W., and Clark, D. S. (2015) Evaluating endoglucanase Cel7B-lignin interaction mechanisms and kinetics using quartz crystal microgravimetry. *Biotechnol Bioeng*
9. Sammond, D. W., Yarbrough, J. M., Mansfield, E., Bomble, Y. J., Hobdey, S. E., Decker, S. R., Taylor, L. E., Resch, M. G., Bozell, J. J., Himmel, M. E., Vinzant, T. B., and Crowley, M. F. (2014) Predicting enzyme adsorption to lignin films by calculating enzyme surface hydrophobicity. *J Biol Chem* **289**, 20960-20969
10. Nakagame, S., Chandra, R. P., Kadla, J. F., and Saddler, J. N. (2011) Enhancing the enzymatic hydrolysis of lignocellulosic biomass by increasing the carboxylic acid content of the associated lignin. *Biotechnol Bioeng* **108**, 538-548

11. Nordwald, E. M., Brunecky, R., Himmel, M. E., Beckham, G. T., and Kaar, J. L. (2014) Charge engineering of cellulases improves ionic liquid tolerance and reduces lignin inhibition. *Biotechnol Bioeng*

12. Rahikainen, J. L., Evans, J. D., Mikander, S., Kalliola, A., Puranen, T., Tamminen, T., Marjamaa, K., and Kruus, K. (2013) Cellulase-lignin interactions-the role of carbohydrate-binding module and pH in non-productive binding. *Enzyme Microb Technol* **53**, 315-321

13. Scott, B. R., St-Pierre, P., Lavigne, J., Masri, N., White, T. C., and Tomashek, J. J. (2010) Novel lignin-resistant cellulase enzymes. *US 20100221778 A1*

14. Taylor, C. B., Talib, M. F., McCabe, C., Bu, L., Adney, W. S., Himmel, M. E., Crowley, M. F., and Beckham, G. T. (2012) Computational investigation of glycosylation effects on a family 1 carbohydrate-binding module. *J Biol Chem* **287**, 3147-3155

15. Yang, B., and Wyman, C. E. (2006) BSA treatment to enhance enzymatic hydrolysis of cellulose in lignin containing substrates. *Biotechnol Bioeng* **94**, 611-617

16. Kumar, L., Arantes, V., Chandra, R., and Saddler, J. (2012) The lignin present in steam pretreated softwood binds enzymes and limits cellulose accessibility. *Bioresour Technol* **103**, 201-208

17. Lou, H., Wang, M., Lai, H., Lin, X., Zhou, M., Yang, D., and Qiu, X. (2013) Reducing non-productive adsorption of cellulase and enhancing enzymatic hydrolysis of lignocelluloses by noncovalent modification of lignin with lignosulfonate. *Bioresour Technol* **146**, 478-484

18. Börjesson, J., Engqvist, M., Sipos, B., and Tjerneld, F. (2007) Effect of poly(ethylene glycol) on enzymatic hydrolysis and adsorption of cellulase enzymes to pretreated lignocellulose. *Enzyme Microb Technol* **41**, 186-195

19. Várnai, A., Siika-Aho, M., and Viikari, L. (2013) Carbohydrate-binding modules (CBMs) revisited: reduced amount of water counterbalances the need for CBMs. *Biotechnol Biofuels* **6**, 30

20. Shill, K., Padmanabhan, S., Xin, Q., Prausnitz, J. M., Clark, D. S., and Blanch, H. W. (2011) Ionic liquid pretreatment of cellulosic biomass: enzymatic hydrolysis and ionic liquid recycle. *Biotechnol Bioeng* **108**, 511-520

21. Labbé, S., and Thiele, D. J. (1999) Copper ion inducible and repressible promoter systems in yeast. *Methods Enzymol* **306**, 145-153

22. Dana, C. M., Saija, P., Kal, S. M., Bryan, M. B., Blanch, H. W., and Clark, D. S. (2012) Biased clique shuffling reveals stabilizing mutations in cellulase Cel7A. *Biotechnol Bioeng* **109**, 2710-2719

23. Zheng, L., Baumann, U., and Reymond, J. L. (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res* **32**, e115

24. Nakagame, S., Chandra, R. P., and Saddler, J. N. (2010) The effect of isolated lignins, obtained from a range of pretreated lignocellulosic substrates, on enzymatic hydrolysis. *Biotechnol Bioeng* **105**, 871-879

25. Nakagame, S., Chandra, R. P., Kadla, J. F., and Saddler, J. N. (2011) The isolation, characterization and effect of lignin isolated from steam pretreated Douglas-fir on the enzymatic hydrolysis of cellulose. *Bioresour Technol* **102**, 4507-4517

26. Kraulis, J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J., and Gronenborn, A. M. (1989) Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from Trichoderma reesei. A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. *Biochemistry* **28**, 7241-7257

27. Hamby, S. E., and Hirst, J. D. (2008) Prediction of glycosylation sites using random forests. *BMC Bioinformatics* **9**, 500

28. Rahikainen, J. L., Moilanen, U., Nurmi-Rantala, S., Lappas, A., Koivula, A., Viikari, L., and Kruus, K. (2013) Effect of temperature on lignin-derived inhibition studied with three structurally different cellobiohydrolases. *Bioresour Technol* **146**, 118-125

Chapter 5: Simultaneous selection and counter-selection for the directed evolution of proteases in E. coli using a cytoplasmic anchoring strategy

5.1 Forward

This chapter details a collaborative protein-engineering project I worked on predominantly during my second and third year of graduate school (2011-2013). This project aimed to develop a high-throughput selection platform for engineering protease enzymes. Zac Carrico started the project and completed the proof of concept experiments. I took over the work when Zac graduated and completed the selection experiments.

While the rest of my dissertation work involved rational protein-engineering approaches, this work involved designing and implementing directed evolution techniques. Both rational and random protein-engineering strategies have been very successful at generating novel proteins with improved properties, and it is increasingly clear that a combination of approaches is the best way forward. As our understanding of structure-function relationships improve, it becomes possible to launch more focused attacks on particular regions of the protein landscape, and so optimize evolution experiments. In this chapter, I detail our work using this combined approach to evolve a protease for new sequence specificity.

5.2 Abstract

With the goal of generating new enzymes that can cleave after custom sequences, this article describes a selection strategy for evolving proteases with desirable characteristics. Positive selection and counter-selection are combined to select for and against specified cleavage sequences simultaneously. Cleavage of the positive selection sequence permits *E. Coli* growth, and cleavage of the counter-selection sequence slows growth. Growth occurs when cleavage of the positive selection sequence releases β-lactamase into the periplasm where it can confer antibiotic resistance. The counter-selection traps β-lactamase in the cytoplasm, preventing antibiotic resistance and growth. Thus, proteases with a preference for the positive selection sequence relative to the counter-selection sequence grow more rapidly. This system was used to select a tobacco etch virus (TEV) protease mutant with new substrate compatibility.

5.3 Introduction

Proteases are commonly used in research, industry, and medicine, with applications ranging from non-specific protein degradation by commercial detergent additives to the highly specific cleavage of peptide sequences involved in apoptotic cascades and prodrugs (Li et al. 2013; Putt et al. 2006). There is considerable promise for new proteases that could cleave at a user-specified sequence, as these enzymes could be used to target specific proteins and toxins *in vivo*. This potential has encouraged several labs to invent screens and selections for evolving proteases for new substrate selectivity. The screens have used fluorescence (Kostallas and Samuelson 2010; Varadarajan et al. 2008; Yi et al. 2013), growth (Kohler 2003), or colorimetric changes (Renicke et al. 2013; Sellamuthu et al. 2008) to detect new proteases *in vitro*, and the selections have used antibiotic resistance (Sandersjoo et al. 2014; Verhoeven et al. 2012), translation (Dickinson et al. 2014), or non-toxicity (O'Loughlin et al. 2006) to identify microorganisms expressing proteases with new sequence specificity.

The screens for new proteases offer greater versatility than selections, but they typically cost more and take longer to investigate large numbers of mutants. However, most selections lack simultaneous positive and counter-selections, limiting their ability to identify proteases with reduced activity toward the original sequence while improving activity for a new one. To address this possibility, this paper evaluates a combined selection and counter-selection strategy for the directed evolution of proteases. We demonstrate the effectiveness of this selection strategy for the identification of a tobacco etch virus protease (tevp) mutant with new substrate selectivity.

5.4 Materials and methods

General procedures and materials

Unless otherwise noted, all chemicals and reagents were obtained from commercial sources and used without further purification. All restriction enzymes were purchased from New England Biolabs, Inc. (NEB). All oligonulceotides were purchased desalted from Integrated DNA Technologies, Inc. Or valuegene, Inc. Water used in biological procedures or as a chromatography solvent was deionized using a nanopure purification system (Barnstead, USA).

E. Coli strains and expression plasmids

Full gene sequences are provided in the supplementary information. *E. Coli* BL21-Pro (Clontech Laboratories, Inc.) Was used for western blot experiments. All other experiments used DH10B (NEB). The β-lactamase ("BL") sequence optimized for expression in *E. Coli* was purchased from DNA 2.0, Inc. In the pjexpress vector with the ycbk twin-arginine signal sequence gene 5' of the BL gene and the canonical TEV protease cleavage sequence (P6-ENLYFQG-P1') appended to the 3' end of the BL gene. The β-galactosidase ("BG") gene cloned from pgoal17 (Addgene) was ligated to the 3' end of the canonical TEV protease cleavage sequence. The twin-arginine ycbk signal sequence was used in place of the wildtype BL signal sequence because it permits

periplasmic export of folded proteins (Natale et al. 2008). The counter-selection had the canonical tevp cleavage sequence P6-ENLYFQG-P1' between ycbk and BL. The same linker with a methionine in place of the P1 glutamine was used in all constructs compared to the counter-selection because the methionine results in much less cleavage while maintaining identical protein lengths (Dougherty et al. 1988). Prk603 encodes S219D tevp and was purchased from Addgene. Prk603 was used for all selection experiments. Prk793 expressing MBP-fused S219V tevp was purchased from Addgene and was used in overexpressing and purifying tevp as previously reported (Austin et al. 2009). The abbreviation tevp used throughout the paper refers to S219D tevp for selection experiments or S219V tevp for peptide hydrolysis experiments.

Culture growth for isolation of cytosolic and periplasmic proteins and SDS-PAGE analysis

BL21-Pro was transformed with prk603 (tevp) and pjexpress (BL-P6-E-BG or BL-P6-T-BG; see results section for definitions). They were grown in low-salt LB (10 g tryptone, 5 g nacl, 5 g yeast extract, 1 L water) containing 35 μg/ml kanamycin and 25 μg/ml zeocin at 30 °C overnight. A 150 μL sample of the culture was added to 15 ml of low-salt LB and gently mixed. A 10 ml sample was taken from the 15 ml culture and IPTG was added to 40 μM. A 5 ml sample was taken from this and anhydrotetracyclin was added to 100 ng/ml. The samples were grown for 13.5 h and then centrifuged for 10 min at 4,700 rpm.

Isolation of cytoplasmic and periplasmic proteins

The pelleted cultures were resuspended in 8 ml of buffer (30 mm Tris-Cl ph 8, 20% sucrose, 1 mm EDTA, 0.5 mg/ml lysozyme) chilled to 4 °C. This was left on ice for 10 min and then the centrifugation step was repeated. The pellet was then resuspended in 8 ml of 4 °C 5 mm $mgso_4$ and left on ice for 10 min before repeating the centrifugation. The supernatant containing periplasmic proteins was concentrated to 350 μL using a 10 kda Amicon spin concentrator before SDS-PAGE. The cell pellet was resuspended in 1 ml of phosphate buffered saline (PBS) before SDS-PAGE. Tris-hcl precast 10-20% SDS-PAGE Ready Gel (Bio-Rad Laboratories, Inc.) Were used. Two gels were prepared identically: one for Coomassie staining and one for western blotting.

Western blotting

Antibodies used for western blotting were mouse anti-BL (1:500; Abcam plc.), anti-groel (1:2,000; Assay Designs, Enzo Life Sciences, Inc.), and anti-mouse HRP-conjugate (1:2,000, EMD Millipore Corp.). Western Lightning (perkinelmer, Inc.) Was used for visualization.

Growth comparison

Culture growth was quantified by OD600 measured in a 96 well plate. The difference in times at which OD600 reached 0.14 in the absence of carbenicillin was subtracted from the times for growth in the presence of carbenicillin to account for slight variations (<1 h) in the onset of growth.

Site-saturation mutagenesis

TEV protease residues N171, N176, and Y178 were mutated to the NNK codon (N = any nucleotide; K = G or T), which encodes all 20 amino acids, by overlap extension PCR (Williams et al. 2014). The PCR product was ligated into prk603 and transformed by electroporation into DH10B cells containing pjexpress along with the P6-T or P6-P selection sequence. Cells were allowed to recover in 1 ml SOC media for one hour then diluted 40-fold into 5 ml low-salt LB containing 35 μg/ml kanamycin, 25 μg/ml zeocin, 40 μM IPTG, 100 ng/ml atet, and 100 μg/ml carbenecillin. A positive control culture without carbenecillin and a negative control culture with IPTG were also included. A 2.5 μL sample from the positive control culture was plated to calculate the number of transfected *E. Coli* to ensure that the transfection produced >$1x10^5$ colonies, thus ensuring >95% coverage of the sequence space (Reetz and Carballeira 2007). Cultures were then grown at 30°C until they reached an OD600 > 1. A 4 ml sample was centrifuged, purified with a Zymoclean PCR Purification Kit following the manufacturer provided instructions, and sequenced.

In vivo fusion protein cleavage quantification

Cultures were grown until an OD600 of 0.4-0.8 and BL-BG expression was induced with 40 μM IPTG. The cultures were lysed after 16 h by heating at 100 °C in SDS-PAGE loading buffer for 10 min. The lysate was centrifuged and the supernatant was used for SDS-PAGE.

In vitro peptide cleavage quantification

All hydrolysis reactions were allowed to proceed for 30 min at 30 °C in 50 mm Tris-hcl, 5 mm DTT, 0.5 mm EDTA, ph 8. The P6-E peptide (T*E*NLYFQSGTRR) was purchased from anaspec Inc. At >95% purity. The P6-P peptide (T*P*NLYFQSGTRR) was purchased from neobiolab. High performance liquid chromatography (HPLC) was performed on an Agilent 1100 Series HPLC System (Agilent Technologies, USA) using a Phenomenex Gemini C18 column. Tyrosine fluorescence peak integration was used to quantify hydrolysis (excitation 274 nm, emission monitored at 303 nm). A gradient of 0.1% TFA/water and 0.1% TFA/acetonitrile was used as the HPLC mobile phase. Graphpad Prism software was used to fit the results to the Michaelis-Menten equation, and the subsequent curve was used to calculate the kinetic parameters. Wild-type TEV protease concentrations were 0.06 μM and 2.4 mm for the P6-E and P6-P peptide substrates, respectively. The N171D TEV protease concentration was 0.21 μM and 0.96 μM for the P6-E and P6-P peptide substrates, respectively. These enzyme concentrations were chosen because each produced less than 15% substrate hydrolysis after 30 min at all substrate concentrations. After incubating the enzyme and substrate at 30 °C for 30 min, the reaction was quenched by heating to 80 °C for 10 min.

5.5 Results

The results are divided into two sections. The first section evaluates the selection system with tevp and includes the sub-sections entitled *Positive selection* and *Combined positive and counter-selection*. The second section tests the selection, evaluates a selected mutant, and includes the sub-sections entitled *Selection of tevp mutants with new P6 selectivity* and *N171D tevp evaluation*.

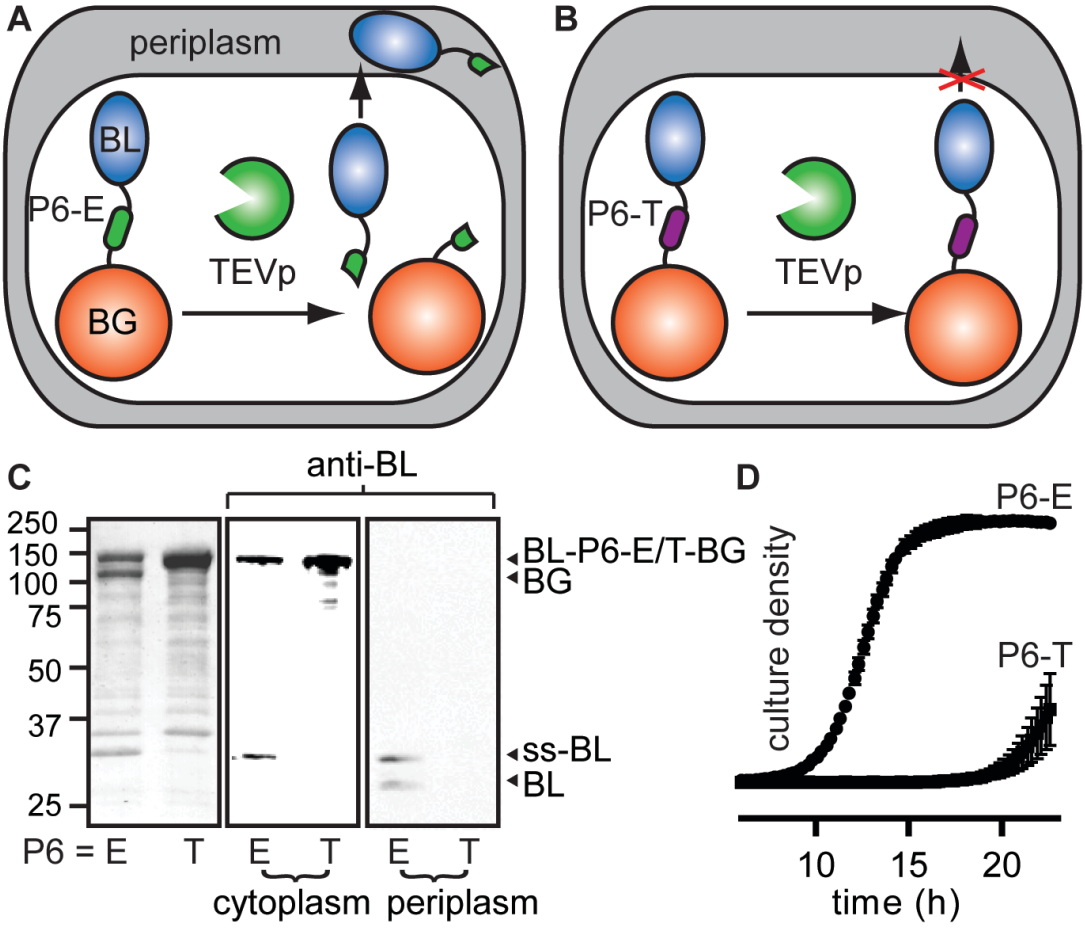

Figure 5.1 Positive protease selection strategy.
(A) TEVp cleaves the linker with Glu (E) at position 6 ("P6-E") in the cleavage sequence, releasing β-lactamase (BL) to be transported into the periplasm. This permits growth in carbenicillin media. (B) TEVp does not cleave the linker with Thr at P6 ("P6-T"), thus preventing BL transport into the periplasm. These cells will not grow in carbenicillin containing media. (C) SDS-PAGE analysis of cultures expressing TEVp and BL-P6-E/T-BG showing cleavage and transport of BL with and without its signal sequence (ss). Left: Coomassie stain of cell lysate. Middle: western blot for BL in the cytoplasmic fraction. Right: western blot for BL in periplasmic fraction. (D) Growth of

the E. coli expressing sTEVp with BL-P6-E/T-BG. Standard error of the mean was used for error bars (n = 3).

Positive selection

We first created a positive selection favoring cells expressing proteases that cleave a desired sequence (Fig. 1A). Central to the selection is the location dependent functionality of β-lactamase ("BL")—it must be in the periplasm to confer carbenicillin resistance (Stanley et al. 2002; Tommassen et al. 1985). BL expressed alone is transported to the periplasm(Neu 1968), but fusing it to β-galactosidase ("BG") anchors it in the cytoplasm due to increased bulk (Tommassen et al. 1985). The linker between BL and BG thus becomes the key to the positive selection: BL is transported into the periplasm where it can prevent antibiotic activity only if the linker is cleaved by the protease candidate (Fig. 1A). Otherwise, BL remains trapped in the cytoplasm (Fig. 1B) and its host *E. Coli* cell is unable to grow in carbenicillin-containing media.

Two initial fusion proteins were made to test the positive selection strategy. They differed only by the P6 residue in the tevp cleavage sequence: P6-*E*NLYFQ↓G-P1' (cleavage occurs between Q and G). P6-E is the canonical tevp cleavage sequence residue at P6, while P6-T changes the linker to one that is poorly cleaved by tevp (Dougherty et al. 1988). As expected, the P6-E linker was cleaved by tevp (Fig. 1C, left panel), permitting BL transport into the periplasm (Fig. 1C, right panel), as observed by two bands in the periplasmic western blot. The upper band was assigned as BL with its ycbk signal sequence intact ("ssbl"), and the lower band was assigned as the protein lacking the signal sequence. This indicated that the rate of ssbl periplasmic transport outpaced peptidase activity, resulting in BL with intact signal sequences in the periplasmic fraction. P6-T was cleaved to a reduced extent, and no BL was observed in the periplasm by western blot. Correspondingly, P6-E produced more rapid growth than P6-T in *E. Coli* cultured in carbenicillin (Fig. 1D). This difference in growth rates forms the basis of the positive selection. Our next step before selecting protease mutants was to incorporate a counter-selection.

Combined positive and counter-selection

A counter-selection was combined with the positive selection to lower protease selectivity for an undesired sequence (the canonical sequence in this case). The counter-selection ("CS") was introduced via a second linker between BL and its periplasmic signal sequence (Fig. 2A). Thus, a protease that only cleaves the positive selection linker between BL and BG should grow in carbenicillin. However, if the protease *also* cleaves the CS linker, BL remains trapped in the cytoplasm and the cells will not grow (Fig. 2B).

To test this idea we used the canonical tevp cleavage sequence for the CS and the P6-T cleavage sequence for the selection linker between BL and BG. In Fig. 2C, the first two lanes have no CS linker, and the last lane has the CS linker and a P6-T linker between BL and BG (abbreviated as "CS_T"). In this last lane there is a clear absence of an intact signal sequence (top band). This signal sequence cleavage is thought to be the cause of the halted growth shown in Fig. 2D for CS_T relative to P6-E or P6-T without the CS linker.

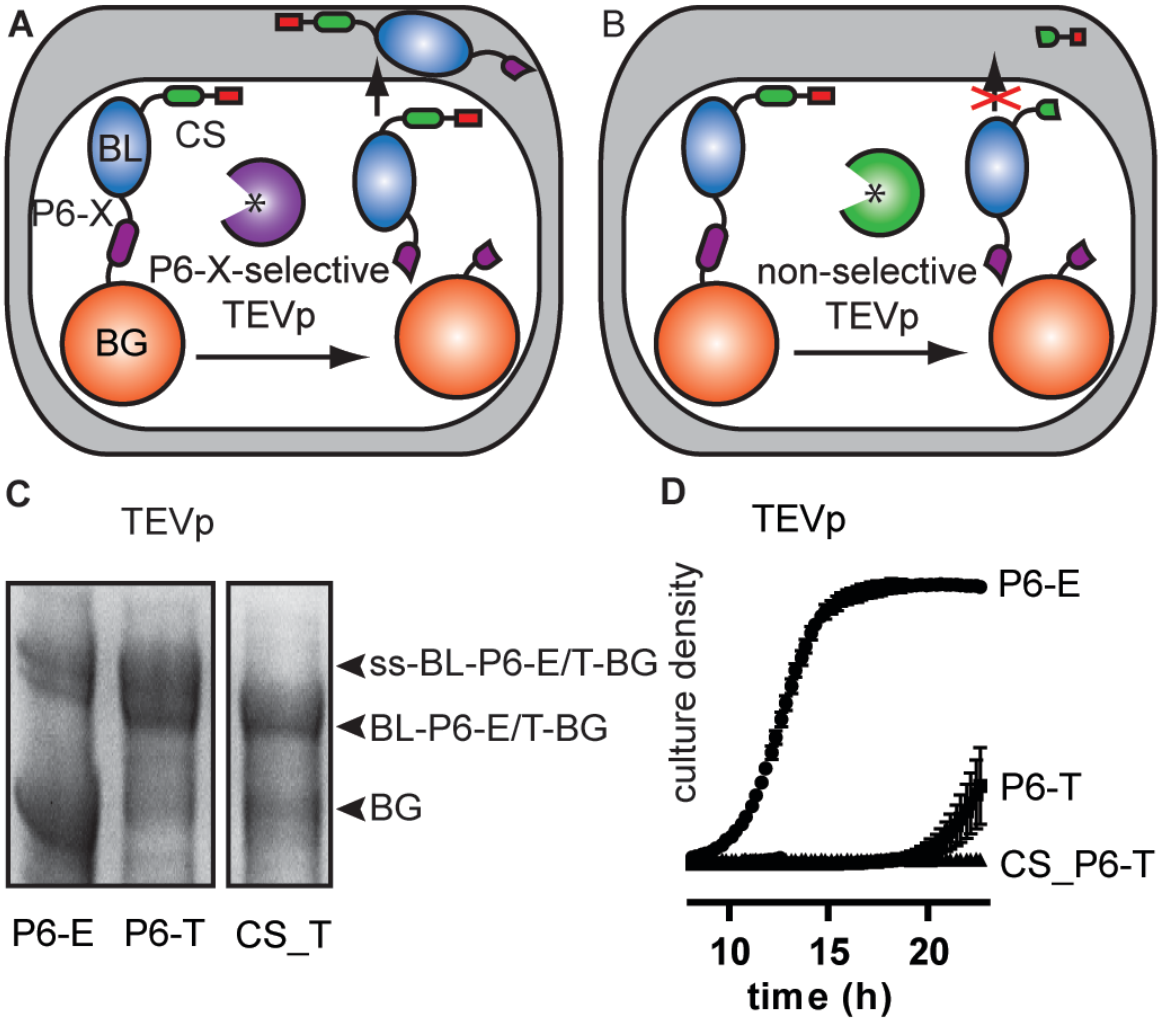

Figure 5.2 Combined positive and counter selections for protease evolution.
(A) A mutant TEVp with X-selectivity (where X is any P6 amino acid other than Glu) cleaves the linker between BL and BG. This P6-X-selective TEVp does not cleave the P6-E in the counter-selection (CS) linker between the signal sequence (ss) and BL. (B) A non-selective TEVp cleaves the positive selection linker and the CS linker, preventing periplasmic export of BL and thus E. coli division. (C) Coomassie stained SDS-PAGE culture expressing TEVp with BL-P6-E/T-BG ("P6-E" or "P6-T") in the first two lanes, or the same protein but with CS sequence: CS-BL-P6-T-BG ("CS_P6-T"), in the right-most lane. (D) Growth curve of cultures expressing wt TEVp and P6-E or CS_P6-T. Standard error of the mean was used for error bars (n = 3).

Selection of tevp mutants with new P6 selectivity

Having demonstrated the effectiveness of the CS strategy, we next wanted to test the combined selection and counter-selection with libraries of tevp to generate mutants with new selectivity at P6. The crystal structure of peptide-bound tevp was used to select residues for mutation (Phan et al. 2002). Tevp residues N171, N176, and Y178 were selected because they are predicted to form hydrogen bonds with P6-E (Fig. 3A). A mutant library was generated by mutating all of these residues to NNK codons (N = A, T, G, or C; K = G or T), introducing all 20 amino acids into each of these positions. The objective of this was to identify a mutation with selectivity for a residue other than glutamate at P6. To accomplish this, the CS sequence was kept as P6-E and the positive selection sequences used were P6-T or P6-P: P6-(*T* or *P*)NLYFQ↓G-P1' (Fig. 3B). Threonine and proline were chosen because they create changes in charge and structure and are known to be poorly cleaved by tevp (Dougherty et al. 1988).

The library was produced and screened for growth in the presence of carbenicillin. The surviving cells were then sequenced in the tevp gene region that had been mutated. The top sequencing chromatogram in Fig. 3B is from the culture containing substrate CS_T grown in the absence of carbenicillin, and shows the expected NNK codons. When this same transfected culture was grown in the presence of carbenicillin, it selected the N171D and N176T tevp mutant shown in the second chromatogram. The libraries combined with CS_T and CS_P independently selected the N171D mutation, so this tevp variant was used in subsequent experiments. Position 178 reverted to the wild type (Y) in both cases, and two different residues (T and N) were selected in the 176 site.

N171D tevp evaluation

N171D tevp was evaluated for its ability to cleave substrates *in vivo* and *in vitro*. These substrates used in these experiments lacked any counter-selection sequence. The *in vivo* SDS-PAGE results in Fig. 4A show the amount of cleavage by N171D tevp for substrates with P6-E, T, or P linkers. Band densitometry results are presented in Fig. 4B, showing that substrates with P6-E and P6-P are cleaved comparably. P6-T produces the least cleavage product. To quantify the kinetics of N171D tevp for P6-E and P6-P, we used a peptide substrate: P7-T(*E* or *P*)NLYFQ↓SRR-P4' (Fig. 4C). Curve fitting was used to determine the kinetic constants and the values are summarized in Table 1.

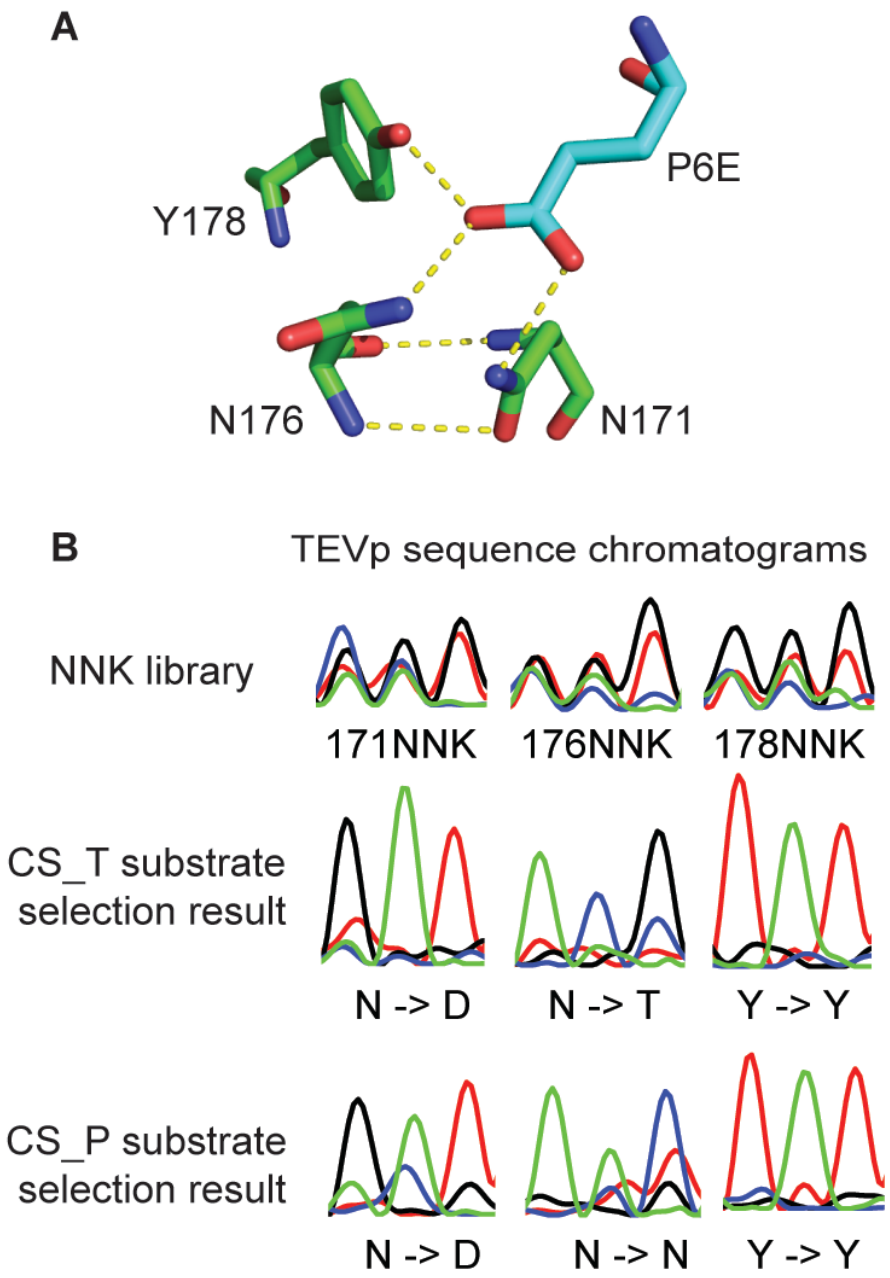

Figure 5.3 Evolution of mutant TEV proteases.
(A) Predicted hydrogen bonding between TEVp residues (green) and substrate site P6E (torquoise). Modeling was done with Pymol using crystal structure PDB reference 1LVB. (B) Sequence chromatograms from site-saturation NNK (N = A, T, G, or C, and K = G or T; NNK encodes all 20 amino acids) mutant libraries at positions 171, 176 and 178. The top sequencing chromatogram resulted from a culture grown in the absence of carbenicillin and the bottom two were obtained from cultures that grew in carbenicillin containing media. Chromatogram colors: black = G; blue = C; red = T; green = A.

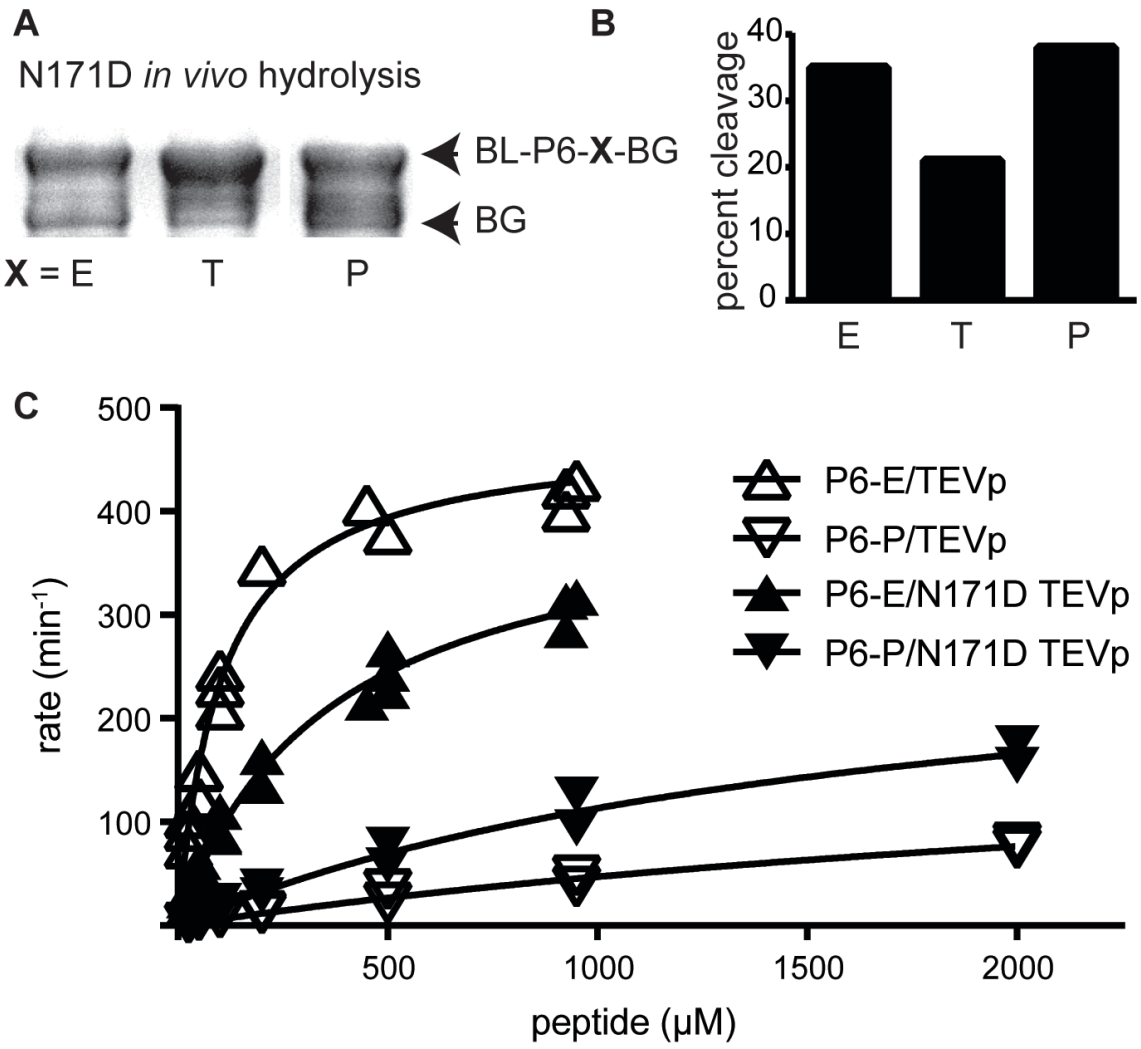

Figure 5.4 Characterization of new TEVp activity and selectivity.
(A) SDS-PAGE analysis of E. coli cultures expressing N171D TEVp and a fusion protein (BL-X-BG) containing a TEVp cleavage sequence linker, where P6-X = E, T, or P. (B) Percent cleavage calculated from SDS-PAGE densitometry measurements. (C) Peptide hydrolysis plots of wt or N171D TEVp with P6-E or P6-P peptides.

Table 5.1 Peptide hydrolysis kinetics for wt and N171D TEVp

P6	TEVp	kcat (min^{-1})	Km (μM)	kcat/Km (min^{-1} μM^{-1})
E	wt	475 +/- 15	104 +/- 10	4.59 +/- 0.47
P	wt	202 +/- 45	3289 +/- 1063	0.06 +/- 0.02
E	N171D	420 +/- 21	360 +/- 42	1.17 +/- 0.15
P	N171D	306 +/- 41	1700 +/- 403	0.18 +/- 0.05

5.6 Discussion

Herein we describe a new selection for the directed evolution of proteases in *E. Coli*. It was evaluated using tevp and mutations to its canonical cleavage sequence: P6-ENLYFQ↓G-P1'. The combined use of positive and counter-selection yielded a new mutant tevp-substrate combination.

For the positive selection, a P6-E tevp cleavage sequence linker was used as a positive control and a P6-T linker served as a negative control. We monitored BL-P6-E/T-BG cleavage and appearance of BL in the periplasm (Fig. 1C). BL-P6-E-BG was cleaved by wild type tevp and the cytoplasmic and periplasmic fractions both contained BL. As expected, BL-P6-T-BG was poorly cleaved and no BL was observed in the periplasmic fraction. Fig. 1D shows that the BL-P6-E-BG expressing culture has a much shorter time to the onset of growth in carbenicillin relative to the BL-P6-T-BG. This result validated the cytoplasmic anchoring strategy for selecting proteases able to cleave the linker between BL and BG. To improve the selection further, we subsequently incorporated a counter-selection strategy.

The counter-selection (CS) is a cleavage sequence linker between BL and its periplasmic signal sequence (Fig. 2A and B). It contains the tevp canonical cleavage sequence and reduces growth of *E. Coli* if that sequence is cleaved. It was combined with the P6-T selection sequence to create "CS_T", which was used in subsequent selection experiments. The SDS-PAGE image in Fig. 2C shows overlapping bands at the top of the gel. The upper band is likely the signal sequence bearing fusion protein (ssbl-P6-(E or T)-BG) and the lower band lacks the signal sequence. The signal sequence linker appears to be completely cleaved for CS_T substrates, but there is significantly less cleavage for fusion proteins lacking the CS sequence. There is a clear reduction in growth for CS_T relative to T without the CS sequence (Fig. 2D), supporting the intended function of the CS region.

A mutant tevp library was created to identify mutants with increased specificity for new residues at P6. To achieve this, we mutated tevp residues predicted to form hydrogen bonds with P6-E in the canonical cleavage sequence (Fig. 3A). Tevp sites N171, N176, and Y178 were mutated to NNK codons, which collectively encode all 20 amino acids. This library of 8,000 possible candidates was transfected into *E. Coli* containing CS_T or CS_P selection substrates. When the transfected cultures were grown in the absence of carbenicillin, the mutant libraries retained the expected NNK codon mixture at N171, N176 and Y178 without bias. However, when the same transfected library was grown with carbenicillin, a strongly biased population was obtained with the N171D mutation in tevp. The CS_P selection also selected N171D, preserving wildtype residues N176 and Y178. The N171D mutation selection may result from compensation of the negative charge P6-E would normally provide. Neither selection observed mutations to Y178, indicating its mutation is detrimental to activity. Based on these results, N171D tevp was purified and characterized.

In vivo and *in vitro* cleavage assays were used to evaluate the activity of N171D tevp. An *in vivo* cleavage assay without the CS sequence was used to simplify interpretation while recreating the selection environment. Cleavage was measured by SDS-PAGE densitometry (Fig. 4A and B) to determine the order of cleavage for the different P6 mutants: P > E > T. The possibility that N171D tevp might prefer P6-P to P6-E was further investigated using peptide substrates, the results of which are shown in Figure 4C and Table 1. Clearly the peptide with P6-E is still preferred over P6-P for both tevp and N171D tevp. This observation conflicts with the *in vivo* data shown in Fig. 4A and B. For the peptides, N171D tevp has approximately 25% of the catalytic efficiency for P6-E and twice the efficiency for P6-P relative to the initial tevp.

What might explain the discrepancy between the *in vitro* and *in vivo* results? The ratio of tevp to substrate is expected to be greater *in vivo* relative to *in vitro* and, although exact concentrations are unknown, this could decrease the ability to compare protease efficiencies accurately *in vivo*. If this is the case, this selection strategy may be limited to identifying proteases that can cleave new sequences in living cells. If proteases with increased *in vitro* selectivity for proteins are desired, a potential solution is to use a selection system in which tevp is expressed only after substrate expression (rather than through the weak but constitutive expression of tevp in the current system). Future experiments will explore this possibility, as well as the applicability of this strategy to proteases other than tevp. The strategy will also be applied towards selecting for increased stability and reduced toxicity.

In summary, this report describes a new protease evolution strategy that uses cytoplasmic anchoring to regulate cell growth based on desired and undesired cleavage events. It is minimally complex so as to reduce artifacts and decrease experimental costs. By including a selection and counter-selection in the same fusion protein, we were able to identify a tevp mutant with increased selectivity for mutant sequences and decreased selectivity for the canonical recognition sequence. Future experiments will explore the capacity of the selection to evolve other protease/substrate combinations.

Acknowledgements
These studies were generously supported by the DOD Breast Cancer Research Program (BC061995) and the NCI SPORE in Breast Cancer (P50-CA58207). Zac Carrico was supported by the Berkeley Chemical Biology Graduate Program (NRSA Training Grant 1 T32 GMO66698). Kathryn Strobel was supported by the Energy Biosciences Institute. We would like to thank Brian Austin and Dr. David Waugh for the generous gift of the prk603 and BL21-Pro and Dr. Carolyn Bertozzi for pgoal17. We would also like to thank Prof. Danielle Tullman-Ercek for helpful discussions on the Tat pathway.

5.7 References

Austin BP, Nallamsetty S, Waugh DS. 2009. Hexahistidine-tagged maltose-binding protein as a fusion partner for the production of soluble recombinant proteins in Escherichia coli. Methods Mol Biol 498:157-72.

Dickinson BC, Packer MS, Badran AH, Liu DR. 2014. A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations. Nat Commun 5:5352.

Dougherty WG, Carrington JC, Cary SM, Parks TD. 1988. Biochemical and Mutational Analysis of a Plant-Virus Polyprotein Cleavage Site. Embo Journal 7(5):1281-1287.

Kohler F. 2003. A yeast-based growth assay for the analysis of site-specific proteases. Nucleic Acids Res 31(4):e16.

Kostallas G, Samuelson P. 2010. Novel fluorescence-assisted whole-cell assay for engineering and characterization of proteases and their substrates. Appl Environ Microbiol 76(22):7500-8.

Li Q, Yi L, Marek P, Iverson BL. 2013. Commercial proteases: present and future. FEBS Lett 587(8):1155-63.

Natale P, Bruser T, Driessen AJ. 2008. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane--distinct translocases and mechanisms. Biochim Biophys Acta 1778(9):1735-56.

Neu HC. 1968. The surface localization of penicillinases in Escherichia coli and Salmonella typhimurium. Biochem Biophys Res Commun 32(2):258-63.

O'Loughlin TL, Greene DN, Matsumura I. 2006. Diversification and specialization of HIV protease function during in vitro evolution. Molecular Biology and Evolution 23(4):764-772.

Phan J, Zdanov A, Evdokimov AG, Tropea JE, Peters HK, 3rd, Kapust RB, Li M, Wlodawer A, Waugh DS. 2002. Structural basis for the substrate specificity of tobacco etch virus protease. J Biol Chem 277(52):50564-72.

Putt KS, Chen GW, Pearson JM, Sandhorst JS, Hoagland MS, Kwon JT, Hwang SK, Jin H, Churchwell MI, Cho MH and others. 2006. Small-molecule activation of procaspase-3 to caspase-3 as a personalized anticancer strategy. Nat Chem Biol 2(10):543-50.

Renicke C, Spadaccini R, Taxis C. 2013. A tobacco etch virus protease with increased substrate tolerance at the P1' position. Plos One 8(6):e67915.

Sandersjoo L, Kostallas G, Lofblom J, Samuelson P. 2014. A protease substrate profiling method that links site-specific proteolysis with antibiotic resistance. Biotechnol J 9(1):155-162.

Sellamuthu S, Shin BH, Lee ES, Rho SH, Hwang W, Lee YJ, Han HE, Il Kim J, Park WJ. 2008. Engineering of protease variants exhibiting altered substrate specificity. Biochemical and Biophysical Research Communications 371(1):122-126.

Stanley NR, Sargent F, Buchanan G, Shi J, Stewart V, Palmer T, Berks BC. 2002. Behaviour of topological marker proteins targeted to the Tat protein transport pathway. Mol Microbiol 43(4):1005-21.

Tommassen J, Leunissen J, van Damme-Jongsten M, Overduin P. 1985. Failure of E. Coli K-12 to transport phoe-lacz hybrid proteins out of the cytoplasm. Embo J 4(4):1041-7.

Varadarajan N, Rodriguez S, Hwang BY, Georgiou G, Iverson BL. 2008. Highly active and selective endopeptidases with programmed substrate specificities. Nat Chem Biol 4(5):290-4.

Verhoeven KD, Altstadt OC, Savinov SN. 2012. Intracellular detection and evolution of site-specific proteases using a genetic selection system. Appl Biochem Biotechnol 166(5):1340-54.

Williams EM, Copp JN, Ackerley DF. 2014. Site-saturation mutagenesis by overlap extension PCR. Methods Mol Biol 1179:83-101.

Yi L, Gebhard MC, Li Q, Taft JM, Georgiou G, Iverson BL. 2013. Engineering of TEV protease variants by yeast ER sequestration screening (YESS) of combinatorial libraries. Proc Natl Acad Sci U S A 110(18):7229-34.