BINDING INTERACTIONS OF FAMILY 1 CARBOHYDRATE BINDING MODULES WITH CELLULOSE ALLOMORPHS

By

VIBHA NARAYANAN

A thesis submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Master of Science

Graduate Program in Chemical and Biochemical Engineering

Written under the direction of

Dr. Shishir Chundawat

And approved by

_____

_____

_____

New Brunswick, New Jersey

[October, 2017]

ABSTRACT OF THE THESIS

Binding interactions of family 1 carbohydrate binding modules with cellulose allomorphs

By Vibha Narayanan

Thesis Director:

Dr. Shishir Chundawat

Conversion of inedible lignocellulosic plant biomass, composed primarily of carbohydrate polymers like cellulose, into biofuels can quench the ever-growing societal demands for sustainable and renewable forms of energy. Naturally occurring microbial enzymes called cellulases, that are extracellularly secreted by industrial workhorse fungi such as *Trichoderma reesei*, can aid in the deconstruction of lignocellulosic feedstock to soluble sugars (for further upgrading to fuels) using benevolent enzymatic hydrolysis based processes unlike high thermochemical severity based acid-catalyzed processes. The most abundant cellulase expressed by *T. reesei* has a two-domain structure consisting of a Carbohydrate Binding Module (CBM) and a Catalytic Domain (CD) linked by a glycosylated linker peptide that catalyzes cellulose hydrolysis to cellobiose. This CBM was classified as the first cellulose – binding protein family (or CBM1) of its type since it was the first CBM to be discovered and belongs to family 1 (or CBM1). CBMs are auxiliary carbohydrate-active enzyme (CAZyme) protein domains that aid in the binding of enzymes to carbohydrate substrates as well as improve the activity of the catalytic domain through mechanisms that are still far from being fully understood. Cellulose exists

naturally in plant cell walls as self-assembled microfibrils and has a defined crystalline allomorphic structure called cellulose-I. The tight packing of cellulose fibrils lowers accessibility to cellulases, thus severely limiting the rate of cellulose enzymatic hydrolysis. Pretreatment of native cellulose with anhydrous liquid ammonia can restructure the hydrogen bonds network to form an unnatural allomorph called cellulose-III that is more easily digestible by some families of cellulolytic enzymes. Wild-type family 1 CBMs and their engineered mutants' characterization was the main focus of this study. These proteins were expressed fused on the C-terminus to a Green Fluorescent Protein (GFP), purified using a two-step purification process and their adsorption to cellulose-I and cellulose-III were systematically studied. Effect of physical parameters such as pH and ionic strength on CBM binding to both allomorphs of cellulose was also studied. This study helps understand the role of conserved amino residues on the flat binding face of CBM1 that impact cellulase binding to native and unnatural cellulose allomorphs. This would ultimately impact cost-effective conversion of cellulosic biomass to biofuels or biochemicals using CAZymes in an industrial biorefinery.

ACKNOWLEDGEMENTS

I would firstly like to express my deepest gratitude to my thesis advisor, Dr. Shishir Chundawat, Assistant Professor, Department of Chemical and Biochemical Engineering, Rutgers University for presenting me with the opportunity to pursue this research project. I would also like to thank him for his prompt and timely guidance, invaluable suggestions, and his weekly discussions throughout the course of this thesis work. I would also like to thank our collaborator Dr. Brian Fox at University of Wisconsin-Madison for providing the pEC-GFP-CBM plasmid and Dr. Chundawat for providing the pEC-GFP-CBM1 mutants used in this study. The work reported in sections 2.1.1 and 2.1.2 of this research thesis, was carried out by Dr. Chundawat. I would like to express my gratitude to the NSF Award (#1236120 and #1604421) for partially supporting this research work.

I will also be forever grateful to the Department of Chemical and Biochemical Engineering, Rutgers University, and its staff for being supportive at every step along the way. I also cannot be more thankful for having the best laboratory members. I am especially grateful to Bhargava Nemmaru and Chandrakanth Bandi for their constant support and never ending patience.

Lastly, this work would have been nearly impossible without the moral support of my parents and their continued faith in my abilities. I will also take this moment to thank my roommate for making sure I was fed properly to be able to achieve my thesis goal.

TABLE OF CONTENTS

Abstract of the thesis **ii**

Acknowledgements **iv**

Chapter 1. Introduction **1**

1.1 Lignocellulosic biofuels via biochemical conversion platform 1

1.2 Carbohydrate Binding Module (CBM) 4

1.3 Cellulose ultrastructures 5

1.4 Structure-function characterization of *T. reesei* Cel7A family 1 CBM 8

1.5 Objectives of thesis project 11

Chapter 2. Expression and purification of GFP-CBM1 and its mutants **12**

2.1 Materials and Methods 12

2.1.1 Cloning of *Trichoderma reesei* CBM1 gene 12

2.1.2 CBM1 Y5 Mutagenesis 13

2.1.3 Cell culturing and protein expression 13

2.1.4 Cell pellet lysis and protein purification 14

2.2 Results 16

2.3 Discussion 20

Chapter 3. Characterization of GFP-CBM1 adsorption to cellulose-I and cellulose-III **23**

3.1 Materials and Methods 23

3.1.1 Partition coefficient of CBM1 mutants to cellulose allomorphs 23

3.1.2 Effect of pH and ionic strength on CBM1 adsorption 24

3.1.3 Small scale growth and screening of CBM1 mutants 25

3.2 Results 26

3.2.1 Partition coefficient of CBM1 mutants to cellulose allomorphs 26

3.2.2 Effect of pH and ionic strength on CBM1 adsorption 29

3.2.3 Small scale growth and screening of CBM1 mutants 30

3.3 Discussion 30

3.3.1 Partition coefficient of CBM1 mutants to cellulose allomorphs 30

3.3.2 Effect of pH and ionic strength on CBM1 adsorption 33

3.3.3 Small scale growth and screening of CBM1 mutants 36

Chapter 4. Conclusions and Future Work 38

LIST OF TABLES

Table 1: Protein yield calculated per gram of wet cell pellet obtained from culture volume..... 16

Table 2: Comparison of the percentage of the pooled eluents obtained from the small scale binding assay with Cellulose-I and the percentage of purity obtained from SDS PAGE analysis. 19

Table 3: Amino acid sequence for the CBM1 Y5 Wildtype and its mutants. 21

Table 4: Partition coefficients analyzed with the standard error calculated from OriginLabs software for binding of the Wildtype and CBM1 Y5 mutants. 31

Table 5: Results of the pH depend adsorption studies with Cellulose-I conducted for each mutant at four different pH values..... 33

Table 6: Dependence of ionic strength and pH on adsorption studies to Cellulose-I was analyzed for all the mutants. 100 mM NaCl was maintained as the effective salt concentration in the wells of the microplate. 34

LIST OF FIGURES

Figure 1: Comprehensive overview of the biochemical conversion of lignocellulosic feedstock into fuels and chemicals. Figure taken from (Payne et al., 2015) 2

Figure 2: Model of TrCel7A acting on single cellulose chain bound to crystalline cellulose surface. Cel7A consists of a large catalytic domain and a smaller carbohydrate-binding module (CBM) connected by a glycosylated linker. Figure taken from (Chundawat, Beckham, et al., 2011). 3

Figure 3: Cross-sectional (top) and top-view (bottom) structures of two crystalline allomorphs of cellulose found in nature - Cellulose Iβ (left) and Cellulose Iα (right). Figure taken from (Payne et al., 2015). 6

Figure 4: Cellulose III, an unnatural allomorph of cellulose is obtained upon treatment with ammonia. Cross-sectional and top views are shown here. Figure taken from (Payne et al., 2015). 7

Figure 5: Carbohydrate binding module from Tr.Cel7A with highlighted Y5, Y31, Y32, Q34 and N29 residues is shown. This structure (1CBH) was obtained from the PDB database and the residues were highlighted using PYMOL. 8

Figure 6: The final plasmid construct for pEC-GFP-CBM1 shown using Geneious software. 12

Figure 7: An overview of the large-scale purification process of the GFP-CBM1 wildtype and its mutants. 15

Figure 8: Chromatogram of the IMAC purification process obtained from UNICORN software. As per the legend on the top left corner of the chromatogram, the blue curves depicting UV absorbance are crucial in providing a visual aid to the purification occurring within the column. 17

Figure 9: Chromatogram for Hydrophobic Interaction Chromatography (HIC) showing a gradient elution for the Y5N mutant over a period of three hours. As it can be noted, as % elution buffer increases, conductivity decreases. Fractions elute off the column with increasing hydrophobic proteins. From small scale binding assay results, the final peak observed contains the protein of interest. 19

Figure 10: Coomassie stained SDS PAGE Gel (BioRad) analysis shows pure GFP CBM 1 Protein in the last lane (Lane no. 10). Lane 2 is the IMAC B eluent. Lane 3-7 indicates the presence of GFP and GFP CBM1. The band exists in the 37 kDa Molecular weight region when compared to the standard. 20

Figure 11: Partition coefficient curve analysis on OriginLabs Software for the GFP CBM1 Wildtype binding to cellulose I..... 18

Figure 12: Partition Coefficient curve analysis for GFP CBM1 Wildtype and cellulose III using OriginLab Software. 18

Figure 13: A comprehensive screening of the Y5 mutants was done and the percentage of bound protein to cellulose-I is shown in the table at the bottom of this graph..... 37

Chapter 1. Introduction

1.1 Lignocellulosic biofuels via biochemical conversion platform

The ever-increasing demand of sustainable, renewable, and clean energy has arrived at a solution in the form of lignocellulosic plant biomass. In the near future, conversion of biomass into fuel used for transportation can play a critical role in reducing greenhouse gas (GHG) emissions that could decelerate climate change. Inedible plant biomass is composed primarily of cell walls that contain sugar polymers (e.g., cellulose and hemicellulose) and aromatic polymers (e.g., lignin). Cellulose (a β-1,4-glucose linked linear polysaccharide) is the most abundantly organic molecule on the planet and serves a key structural/functional role in plants. The covalently linked glucose monomers in cellulose provide an abundant and renewable feedstock for the production of biofuels based on the biochemical conversion platform as outlined in Figure 1 (da Costa Sousa, Chundawat, Balan, & Dale, 2009).

Cellulosic biomass can be hydrolyzed into monomeric sugars by using concentrated acids, however, the final sugar yields were low due to extensive degradation (Saeman, 1945). With the discovery of naturally occurring microbial enzymes such as cellulase/hemi-cellulase, it is now possible to achieve near theoretical hydrolysis yields of sugars from biomass. However, in order to increase enzyme accessibility to embedded plant polysaccharides and thus enhance overall hydrolysis rate, it becomes necessary to pretreat the biomass using a mild thermochemical process (Chundawat, Beckham, Himmel, & Dale, 2011). Various thermochemical pretreatment methods include steam explosion, ammonia fiber expansion (AFEX), dilute acid, and ionic-liquid pretreatment (Wi et al., 2015). This improves digestibility of the fiber as well as helps increase enzymatic hydrolysis rates. Enzymatic hydrolysis of the pretreated biomass

depolymerizes the intact polymer into soluble hexose (C6) and pentose (C5) sugars. Fermentation of these sugars in yeast or bacteria are a viable option to convert the sugars into biofuels like ethanol.

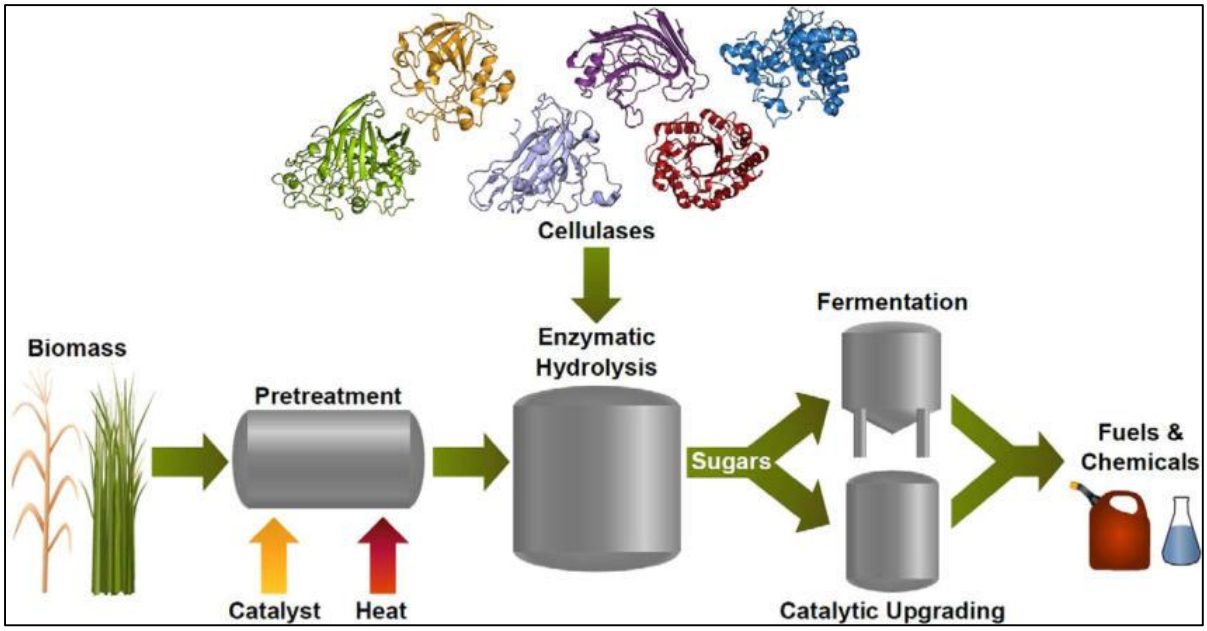

Figure 1: Comprehensive overview of the biochemical conversion of lignocellulosic feedstock into fuels and chemicals. Figure taken from (Payne et al., 2015)

Many microorganisms have evolved the necessary enzymatic machinery to breakdown cellulosic biomass into a food or energy source. Of all the studied cellulolytic microorganisms, fungi are one of the most efficient degraders of biomass into soluble sugars that were discovered in the 1940's by the Natick Army Research Lab. The filamentous fungus, *Trichoderma reesei*, is a fine example of a eukaryotic species that secretes a highly synergistic cocktail of Carbohydrate-Active enZymes (CAZymes) called cellulases that deconstruct lignocellulosic biomass. Cellulases are predominantly glycosyl hydrolases (GH) and can be classified into two

major classes (based on the relative abundance of secreted enzymes) necessary for cellulose solubilization – Endoglucanases and Cellobiohydrolases (or Exoglucanases). Endoglucanases are non-processive cellulases that cleave the cellulose chains present mostly in the amorphous and more easily accessible regions, while cellobiohydrolases are cellulases that bind to the reducing or non-reducing chain end of cellulose chains and aid in the polymers processive depolymerization or hydrolysis. The CBH1 exocellulase derived from *Trichoderma reesei*, now renamed as *Tr*Cel7A, has a bifunctional domain organization consisting of a Carbohydrate Binding Module (CBM) and a large catalytic domain (CD) connected by an O-glycosylated linker peptide as shown in Figure 2. *Tr*Cel7A is the most abundant cellulase secreted by the fungi and that plays a critical role in the deconstruction of cellulose into soluble sugars like cellobiose.

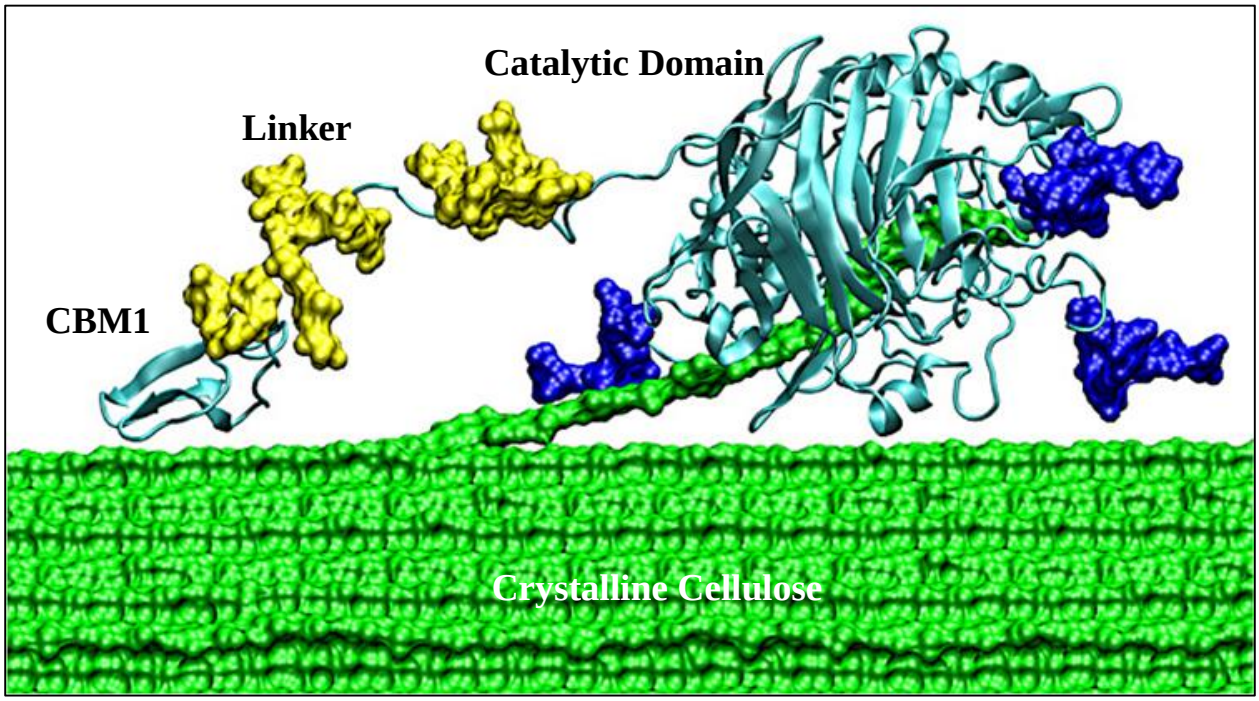

Figure 2: Model of *Tr*Cel7A acting on single cellulose chain bound to crystalline cellulose surface. Cel7A consists of a large catalytic domain and a smaller carbohydrate-binding module (CBM) connected by a glycosylated linker. Figure taken from (Chundawat, Beckham, et al., 2011).

1.2 Carbohydrate Binding Module (CBM)

The first family of Carbohydrate Binding Modules (CBMs) was first identified from *T. reesei.* CBMs were initially called CBDs or Cellulose Binding Domains as these were initially thought to only bind to cellulose. However, with the subsequent discovery of other protein modules that bind to carbohydrates other than cellulose, this family of protein modules was renamed to Carbohydrate Binding Modules or CBMs (Payne et al., 2015). CBMs are defined as contiguous amino acid sequences within a larger protein sequence that can fold into a discrete module. They are generally appended to a CD or Catalytic Domain and display binding affinity towards distinct carbohydrates on which the CDs have preferential activity. The role of CBMs is to bind to a targeted carbohydrate and direct the catalytic machinery towards the substrate. However, CBMs are themselves devoid of any catalytic activity. They are thought to not undergo any major structural changes upon binding to the insoluble carbohydrate ligand. However, due to the lack of available crystal structures of CBMs bound to insoluble substrates, this hypothesis is tough to validate. Furthermore, the topography of the binding site is preformed to complement the shape of the target carbohydrate. This is achieved by the presence of specific amino acid chains that facilitate binding. There are currently 81 families of CBMS based on the amino acid sequences reported in the CAZy database (www.cazy.org). CBMs are also classified into three main superstructure fold types based on their 3D structures and functional similarities. Type A CBMs have binding sites that are planar that create a flat platform to bind to the surface of crystalline cellulose and chitin. Type B CBMs binds to amorphous glycan chains with four or more monosaccharide units. Type C CBMs can bind only to the terminus of glycans as their binding sites are short pockets that can accommodate only shorter sugar chains (www.cazypedia.org). Carbohydrate Binding Module Family 1 is classified under Type A CBMs and are typically

connected to a catalytic core domain by a linker peptide that may be glycosylated as shown in Figure 2. Both domains have been shown to bind to cellulose, but the binding affinity of the catalytic core domain is significantly lesser than that of the CBM. While the activity of the CD and the full-length enzyme on soluble glycan substrates are similar, the activity of the catalytic core domain towards insoluble glycan polymers like cellulose is severely restricted that reaffirms the functional role of CBMs.

1.3 Cellulose Ultrastructure

Cellulose is a β-1,4-linked polysaccharide consisting of several hundred to thousand β-D-glucose units linked together to form the polymer backbone chain that self-assembled to form semi-crystalline microfibrils. The polysaccharide exhibits strong covalent bonding that proves to be a minor hurdle for cellulolytic enzymes during degradation of cellulose. However, the main roadblock is the tight packing of the polymer chains within the cellulose microfibrils via strong non-covalent bonding that limits enzyme activity. The self-aggregation of the polymer chains also lowers accessibility of the glycosidic bonds to enzymatic attack. Cellulose chains can naturally assemble into distinct polymorphic or crystalline states such as Cellulose Iβ or Cellulose Iα depending on its source (Figure 3). The main differences between these two polymorphs can be attributed to different hydrogen bonding patterns and interlayer stacking of the cellulose chains (Brown, 2004). Both polymorphs are devoid of any inter sheet hydrogen bonds but have a strong inter-meshed network of non-covalent bonding within each individual layers/sheets of cellulose as inter-chain or/and intra-chain hydrogen bonds.

While cellulose Iβ and cellulose Iα elementary microfibril crystals (composed of 24 or 36 cellulose chains each) are naturally formed by cellulose synthase enzyme complexes in plant,

algal, and bacterial cell walls, other allomorphic forms of cellulose can be produced from native cellulose by various forms of thermochemical methods.

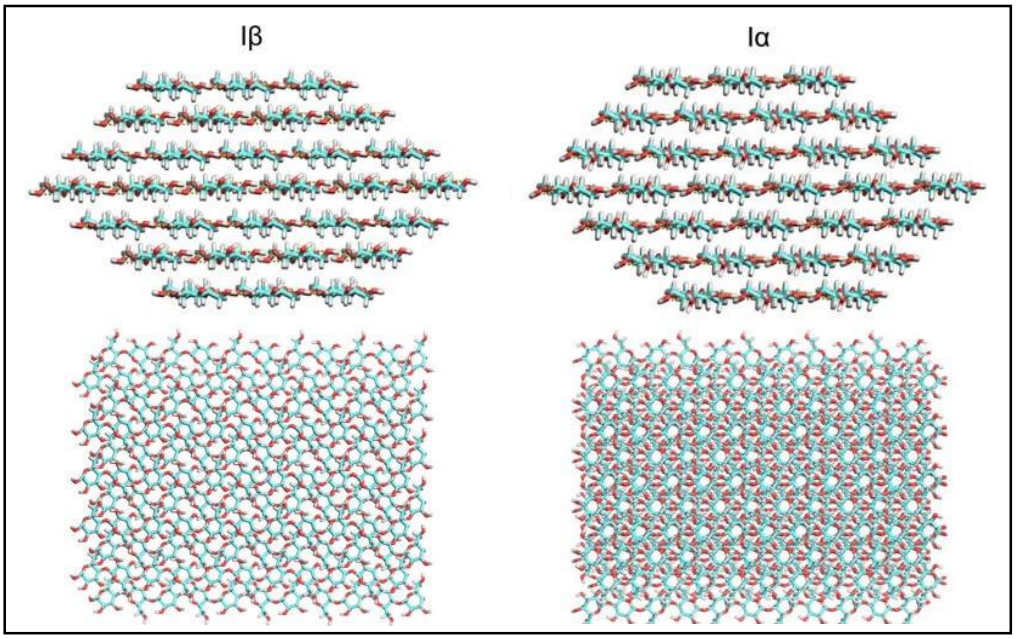

Figure 3: Cross-sectional (top) and top-view (bottom) structures of two crystalline allomorphs of cellulose found in nature - Cellulose Iβ (left) and Cellulose Iα (right). Figure taken from (Payne et al., 2015).

Treatment of native cellulose I crystals with anhydrous liquid ammonia allows formation of an unnatural crystalline allomorph of cellulose called cellulose III (Payne et al., 2015). Unlike cellulose-I, cellulose III has staggered layers that have hydrogen bonds within each individual chain as well as between two parallel chains as depicted in Figure 4. The conversion of cellulose-I to cellulose-III using ammonia pre-treatment was initially performed using AFEX pretreatment method using low water loading (Chundawat, Bellesia, et al., 2011). The Extractive Ammonia (EA) pre-treatment process occurs at low moisture levels unlike its predecessor, AFEX pretreatment. The EA process removes lignin effectively while also producing cellulose-III (da Costa Sousa et al., 2016). The pretreatment process with ammonia decreases the number of intra layer hydrogen bonds and increases the number of inter layer hydrogen bonds. This arrangement

of the hydrogen bonds allows for an increased accessibility of the surface of cellulose-III to some families of cellulases (Chundawat, Bellesia, et al., 2011). Since a higher enzymatic hydrolysis rate was observed, a higher binding of cellulases to this unnatural allomorph of cellulose was expected to be observed. However, while enzymatic hydrolysis rates were significantly higher, the partition coefficient of cellulases to Cellulose-III was lower than that of Cellulose-I (Gao et al., 2013). This reduced binding can be attributed to a low affinity of the CBM 1 to the relatively more hydrophilic surface of the Cellulose-III structure based on molecular dynamic simulations of the substrate. In addition, additional molecular simulations of CBM1 bound to the surface of cellulose III have further revealed that the Y5 position of CBM1 likely has steric clashes that also causes reduction in binding (unpublished data from Dr. Gnanakaran/Dr. Chundawat). This hypothesis forms the basis of the binding studies of the CBM1 wildtype and its Y5 mutants with cellulose-III performed in this research work.

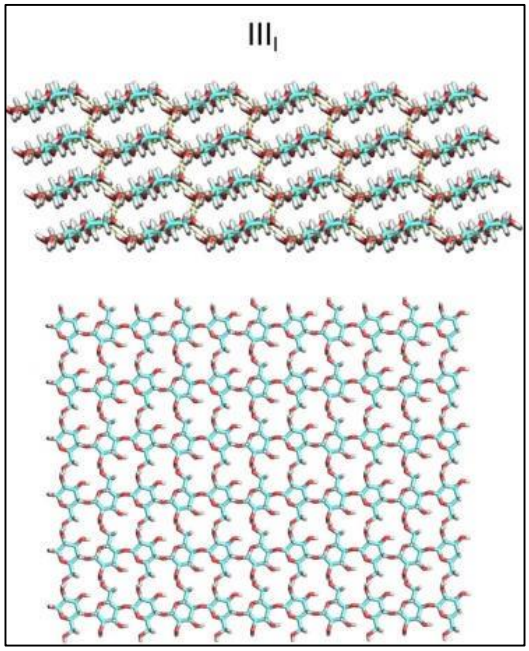

Figure 4: Cellulose III, an unnatural allomorph of cellulose is obtained upon treatment with ammonia. Cross-sectional and top views are shown here. Figure taken from (Payne et al., 2015).

1.4 Structure-function characterization of *T. reesei* Cel7A family 1 CBM

The CBM of *Tr*Cel7A consists of 36 amino acid residues (pdb crystal structure 1CBH). The CBM 1 folds into a wedge shaped structure formed by small irregular triple-stranded β-sheets (Figure 5). The CBM sequence also contains four cysteine residues that form disulfide bonds to hold the structure together. The structure of CBM 1 has two distinct faces of which one is flat/planar and the other is rough/irregular, wherein the flat face contains multiple aromatic and polar residues deemed critical for binding to cellulose. The flat face of the binding module has five conserved residues – Y31, N29, Y32, Q34 and Y5 (Driscoll, Gronenborn, Beress, & Clore, 1989). The aromatic tyrosines at positions 5, 31 and 32 are strictly conserved. Structures of protein and soluble carbohydrate complexes suggest that aromatic residues are frequently involved in binding to ligands through hydrophobic interactions.

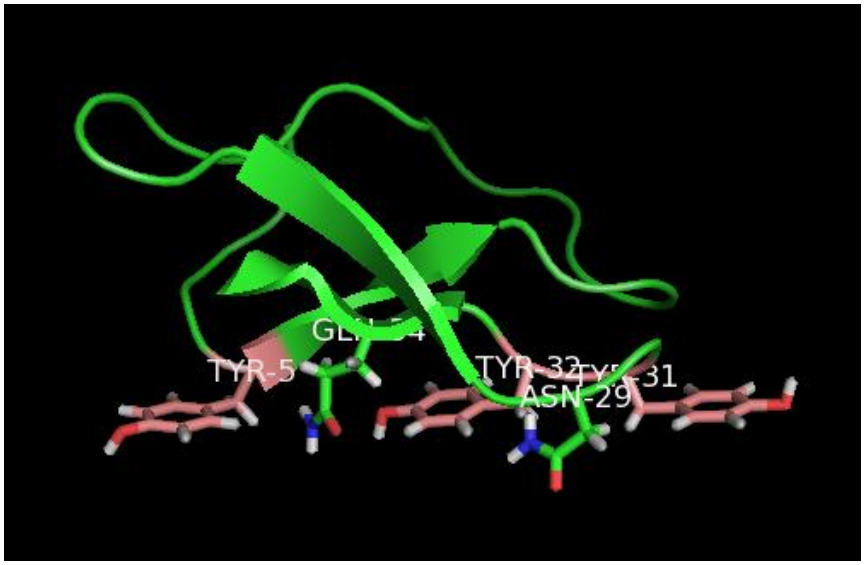

Figure 5: Carbohydrate binding module from *Tr*.Cel7A with highlighted Y5, Y31, Y32, Q34 and N29 residues is shown. This structure (1CBH) was obtained from the PDB database and the residues were highlighted using PYMOL.

From the structure of CBM 1 as shown in Figure 5, it can be observed that the three tyrosine residues are aligned along the planar face of the peptide. The binding of CBM 1 to the surface of

cellulose has been attributed to the stacking of aromatic residues against the faces of the sugar rings along a glucan chain assisted by hydrogen-bonding interactions (Mattinen et al., 1997). The periodicity of the aromatic rings (and other residues) and the periodicity of the glucose rings along the surface of the crystalline cellulose is highly complementary. The aromatic residues as well as the polar residues on the flat face of the CBM 1 form strong hydrogen bonds to surface cellulose chains. A recent molecular simulation study has shown that for the CBM – cellulose interactions van der Waals interactions are important for CBM binding to cellulose and is driven by its hydrophobicity (Beckham et al., 2010). Electrostatic forces are also observed in the function of the CBM after binding which are attributed to its sliding processivity. The interaction of CBM1 or Type A CBMs as a whole to crystalline cellulose is associated with positive entropy change that are the critical thermodynamic forces drive binding of CBMs to the insoluble ligands (Boraston, Bolam, Gilbert, & Davies, 2004).

The position of the side chains on the flat face of the CBM 1 formed by the tyrosine residues are functionally important in the binding of the CBM to cellulose (Linder, Mattinen, et al., 1995). The primary interaction between the Y5 residues and the alcohol rings on cellulose are also observed with the other conserved residues on the flat face – Y31, Y32, N29 and Q34 (Beckham et al., 2010). The asparagine and glutamine residues are commonly involved in interactions between proteins and carbohydrates. This could suggest possible interactions between the N29 and Q34 conserved residues through hydrogen bonding. Site directed mutagenesis have been performed to study the precise effect these particular residues have on the folding of the protein and the binding of the peptide with cellulose (Mattinen et al., 1997). Alanine was substituted in place of the tyrosine residues at positions 5, 31 and 32 as well as the Asparagine-29 and Glutamine-34 and structural characterization was done for these mutants. The

substitution of the Tyr-5 with alanine resulted in an overall loss of compactness of structure which led to the flat face of the CBM being less accurately defined (Linder, Mattinen, et al., 1995; Mattinen et al., 1997). However, the engineered Y31A mutant saw only minor changes in the structure of the flat face in comparison to the wildtype structure. In case of the Y32A mutant, the space between the Q34 and N29 residues was observed to have reduced as well as a tilting of the rings of the tyrosine residues out of the flat planar surface (Mattinen et al., 1997). Hence, the alanine mutation of the Tyr-32 affected both the planarity and the periodicity of the rings on the flat face of the CBM. The mutations to the glutamine and asparagine residues show only slight conformational changes (Linder, Mattinen, et al., 1995).

The Y5 residue can thus, be hypothesized as important for the structural integrity of the N terminus of the peptide. The position of the tyrosine residue in the fifth position in the wild-type CBM is in a type II turn usually occupied by a glycine and is an unfavorable conformation stabilized by a possible histidine – tyrosine interaction (Loewenthal, Sancho, & Fersht, 1992). This makes the N terminus of the CBM particularly sensitive to mutations. The Y31 and Y32 residues are not important to structural integrity of the backbone of the protein. However, they are functionally important in the binding of the CBM to the crystalline face of cellulose. The Y5 residue is also thought to interact with adjacent cellulose chains distinct from the Y31 and Y32, which could explain the reduced binding for unnatural cellulose allomorphs like cellulose III that have a step-like surface that may cause steric hindrances in binding to CBM1 Y5 residue (Gao et al., 2013).

1.5 Objectives of thesis project

There are two main objectives of this thesis project;

1. Validate the hypothesis that reduced binding of Trichoderma cellulases to cellulose III is driven primarily by lower binding of the family 1 CBMs.

2. Systematically explore the effect of mutations at the conserved tyrosine residue (Y5) position of CBM 1 on binding cellulose I versus cellulose III.

To achieve these two objectives, various single site mutations to the Y5 residue of CBM1 (fused to green fluorescent protein or GFP on its C-terminus) were made to study the effect it had on the binding of the GFP-CBM fusion protein to cellulose. Substitutions of the tyrosine residue by phenylalanine, histidine, tryptophan, asparagine, and alanine were primary targets for expression in Rossettagami strain of *Escherichia coli* that were subsequently purified using an in-house optimized two-step purification process. Next, the partition coefficients of the CBM mutants were determined to cellulose using a microplate-based binding assay. While, alanine is expected to reducing binding drastically (as shown previously), the role of other residues such as phenylalanine is not clear and have been chosen to study previously uncharacterized roles in the binding mechanism to cellulose III. Comparative binding studies of the various single point CBM1 mutants with cellulose I and cellulose III is the main focus of this project, that has never before been systematically studied for position 5 (Arola & Linder, 2016).

Chapter 2. Expression and purification of GFP-CBM 1 and its mutants

2.1 Materials and Methods

2.1.1 Cloning of *Trichoderma reesei* CBM1 gene

The CBM 1 gene (from *Trichoderma reesei*) was synthesized by Genscript and fused with a Green Fluorescent Protein (GFP) with 8x-histidine tags at N-terminus using the pEC-GFP-CBM plasmid as described previously (Lim, Chundawat, & Fox, 2014). Briefly, the wild-type CBM 1 DNA fragment (or mutagenized genes) was inserted between *AflII* and *BamHI* by restriction digestion using the two enzymes followed by ligation with T4 DNA ligase (New England Biolabs). The ligation mixture was transformed into *E. cloni* 10G chemically competent cells (Lucigen). The mixture obtained from transformation was plated on Kanamycin resistant LB Agar plates. The colonies with the correct insert size were identified using colony PCR and sequences confirmed using NcoI forward and T7 terminator primers (Figure 6).

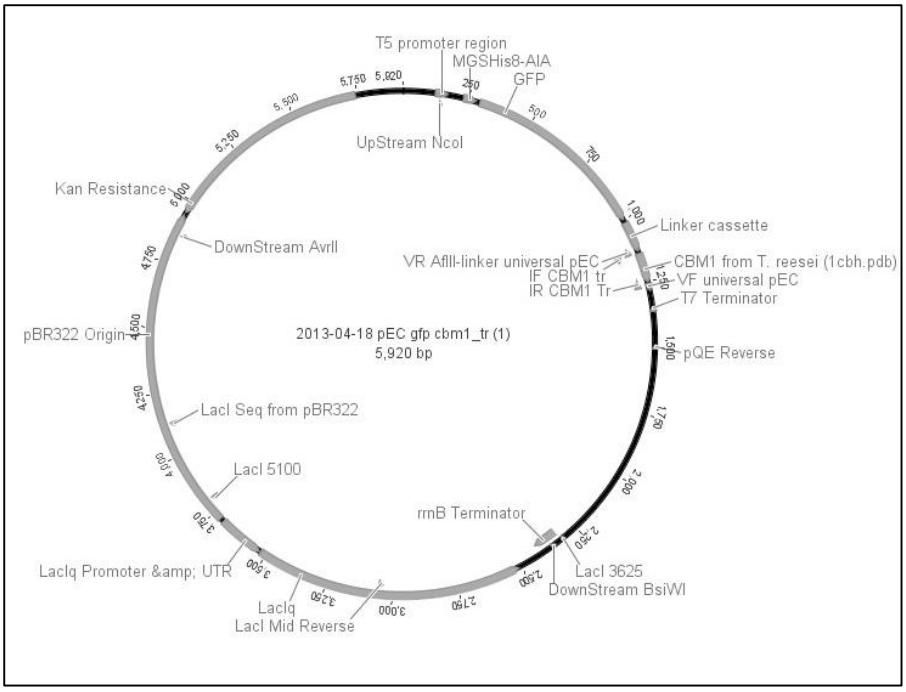

Figure 6: The final plasmid construct for pEC-GFP-CBM1 shown using Geneious software.

Positive colony hits were then transformed into the Rossettagami expression strain and stored for future use in 20% glycerol at -80 degree Celsius. The transformed sequence with verified plasmid was named pEC-GFP-CBM1 and encodes a 5' Histidine tag, followed by a *gfp,* a linker sequence and *cbm 1* gene as shown in Figure 6.

2.1.2 CBM1 Y5 mutagenesis

Site-saturation CBM1 mutants for position Y5 were generated by Life Technologies (Invitrogen & Applied Biosystems) and sub-cloned into the pEC-GFP-CBM1 plasmid. Plasmids were then transformed into the Rossettagami expression strain and stored for future use in 20% glycerol at -80 degree Celsius.

2.1.3 Cell culturing and protein expression

Rossettagami glycerol stocks stored in -80 degree Celsius were used to inoculate LB cultures and grown overnight. Kanamycin was added to avoid contamination and the cultures were kept in the incubator at 37 degrees Celsius for 16 hours. Next, 500 mL TB+G Studiers auto-induction media, as reported previously (Lim et al., 2014), was inoculated in 1 L shake flasks with 5% (v/v) overnight grown starter cultures at 37 degrees Celsius for 4 hours. Briefly, TB+G media is an auto-induction media with the following composition: 1.2% tryptone, 2.4% yeast extract, 2.3% KH_2PO_4, 12.5% K_2HPO_4, 0.375% aspartate, 2 mM MgSO4, 0.8% glycerol, 0.015% glucose, and 0.5% α-lactose. As soon as the optical density (OD) value in the shake flasks reached the exponential stage, the temperature in the incubator is turned down to 25 degrees Celsius and the cultures were grown for 24 hours (Lim et al., 2014). Cell pellets were finally harvested by centrifuging the liquid cultures at 7500 RPM for 20 minutes. The wet cell pellets were then weighed and stored at -80 degrees Celsius.

2.1.4 Cell pellet lysis and protein purification

The stored wet cell pellets at -80 degrees Celsius were lysed using 15 mL cell lysis buffer (20 mM Sodium Phosphate, 500 mM NaCl, 20% (v/v) glycerol; pH 7.4), 200 μL Protease Inhibitor Cocktail (1 μM E-64 from Sigma #E3132, 0.5 mM Benzamidine from Calbiochem #199001, and 1mM EDTA tetrasodium dihydride) and 15 μL Lysozyme (Sigma Aldrich, USA) for every 3 grams of wet cell pellet. The cell lysis mixture was then sonicated using Misonix Sonicator 3000 at an output level of 4.5, for a 10 second pulse on time and a 30 second pulse off time for a period of 5 minutes never allowing the temperature to exceed 7-8 degrees Celsius using an ice bath. The cell lysate was then centrifuged at 20000 RPM for 1 hour at 4 degrees Celsius. The supernatant was collected and filtered through 0.2 μM filter (Fisherbrand).

The GFP-CBM1 wildtype and its Y5 based mutants were purified using a two-step purification method using an automated AKTA FPLC system (GE Healthcare). The first stage of the purification process used IMAC (Immobilized Metal Affinity Chromatography) and a 5 mL HisTrap FF Crude column (GE Healthcare). The column was first washed using IMAC B solvent (100 mM MOPS, 500mM Imidazole, 500 mM NaCl; pH 7.4) and then re-equilibrated with an equal volume of IMAC A (100 mM MOPS, 10 mM Imidazole, 500 mM NaCl; pH 7.4). Next, the cell lysate was loaded onto the column, washed using 5%:95% IMAC B: IMAC A to remove non-specifically bound proteins, and the bound His-tagged protein was next eluted off the column using 100% IMAC B as solvent. This eluent collected was found to contain both proteolyzed his-tagged GFP and his-tagged GFP-CBM1. Instead of using cellulose affinity purification that is laborious due to unavailability of pre-packed columns and low protein recovery yields (Sugimoto, Igarashi, & Samejima, 2012), the collected eluent was further purified using conventional Hydrophobic Interaction Chromatography (e.g., phenyl or butyl

sepharose) using the AKTA FPLC. The difference in hydrophobicity of the cleaved protein is used to separate intact GFP-CBM1 from the mixture. Firstly, the IMAC B eluent is desalted into the HIC start buffer (1 M Ammonium Sulfate, 50 mM Sodium Phosphate; pH 7) using a HiPrep 26/10 Desalting column with a column volume of 53 mL. The desalted protein is then loaded onto a HiTrap butyl sepharose (5 mL) column (GE Healthcare) at slow flow rates after equilibrating the column with the start buffer. Using a linear gradient over a period of three hours and flowrate of 0.5 ml/min, wherein the concentration of the binding buffer decreases and the concentration of the elution buffer (50 mM Sodium Phosphate; pH 7.0) increases, multiple eluents were collected. This was followed by a microtube based binding assay to cellulose-I wherein the eluents and 100 mg/mL of cellulose-I was added in a 1:1 (v/v) ratio. This allowed for a rough 5-7 mg protein/ gm glucan loading. After mixing the tubes well, it was centrifuged in the Eppendorf Microcentrifuge at 13500 RPM for 5 minutes. The supernatants were transferred to an opaque 96 well plate and the fluorescence was observed based on 480 nm excitation, 512 nm emission and a 495 nm cutoff using Molecular Devices Spectramax M5E microplate reader.

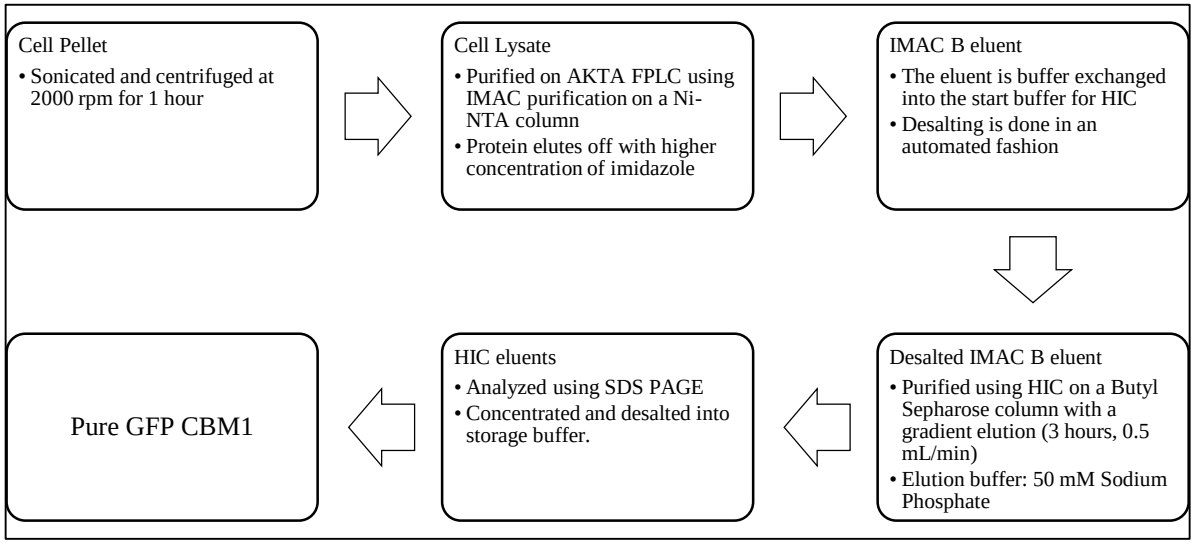

Figure 7*:* An overview of the large-scale purification process of the GFP-CBM1 wildtype and its mutants.

Using SDS PAGE analysis, the eluent containing the correct molecular weight range was identified as GFP-CBM1 and pooled together with other such similar eluents. This pooled sample is then concentrated using Amicon centrifugal concentrators (10 kDa) at 5000 RPM for 30 minutes. The concentrated protein is then desalted into 10 mM MES at pH 6.5 using a PD-10 gravity column and aliquoted for storage. Concentration of the protein was estimated using Molecular Devices Spectramax M5e microplate reader at UV 280 nm (using appropriate molecular weight and extinction coefficients for the various CBM mutants) and flash frozen to be stored at -80 degrees Celsius.

2.2 Results

The pEC-GFP-CBM1 transformed into *E. coli* Rossettagami strains and cells were grown in non-inducing LB medium till an optical density value of 0.5 – 0.7. Auto-induction for 24 hours was found to be optimal to yield maximum protein expression. This was done for each of the pEC-GFP-CBM1 Y5 mutants and the amount of cell pellet obtained for each mutant was noted to calculate protein yields at each purification step.

Protein Type	**Protein yield (in mg) per gm of starting wet cell pellet**	**Grams of cell pellet per liter of Culture Volume**
Wildtype CBM1	**0.27**	**9.78**
Y5N	**0.51**	**7.96**
Y5A	**0.74**	**7.89**
Y5F	**0.26**	**9.63**
Y5H	**0.10**	**9.35**
Y5W	**0.24**	**10.17**

Table 1: Protein yield calculated per gram of wet cell pellet obtained from culture volume.

The his-tagged GFP-CBM1 was first purified from the cell lysate using IMAC with a Ni-NTA resin pre-packed in a 5 mL HisTrap FF crude column using an AKTA FPLC. The HisTrap column has a maximum binding capacity of 200 mg of tagged protein. A single run on the HisTrap FF crude 5 mL column allows for the loading of a maximum of 40 mL of cell lysate after which bleeding of the protein from the column is observed. The AKTA FPLC runs with the help of the UNICORN software from which chromatograms for the various runs are obtained. From the chromatogram (Figure 8), it can be observed that the his-tagged proteins of interest along with IMAC B elutes off the column in one single peak of UV absorbance. The high UV absorbance observed in the protein loading stage is associated to the background *E. coli* proteins not binding to the column due to the lack of histidine tags. The chromatograms from the UNICORN software further aided in identifying when the protein elutes off the desalting column. The HiPrep 26/10 desalting column has a maximum sample loading volume of 15 mL. The protein in the start buffer is eluted off in two peaks. The IMAC B buffer contents elutes off second as a smaller peak while the protein itself elutes first.

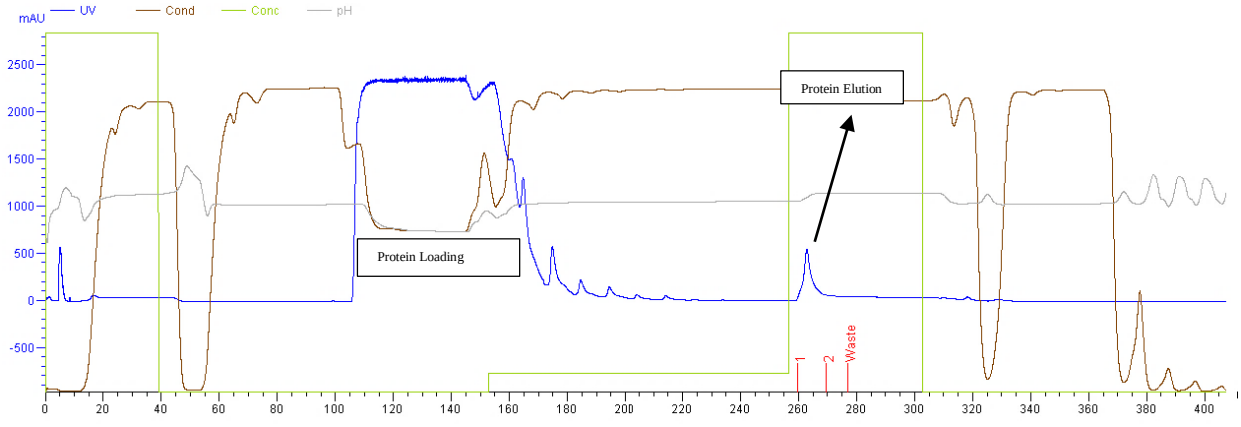

Figure 8: Chromatogram of the IMAC purification process obtained from UNICORN software. As per the legend on the top left corner of the chromatogram, the blue curves depicting UV absorbance are crucial in providing a visual aid to the purification occurring within the column.

Hydrophobic Interaction Chromatography was chosen as the second step of purification. Previous to selecting this method of chromatography, cellulose affinity purification was conducted. However, the final protein was not of desired purity (data not shown). The start buffer for the Hydrophobic Interaction Chromatography step was chosen from the Hoffmeisters series of salts. Choosing the start buffer is the most important parameter in this process. Using various "salting out" salts promotes interactions between the ligand in the column and the protein. Based on an increasing "salting out" effect, the start buffer of a high salt concentration was chosen. Decreasing the salt concentration reduces the hydrophobic interactions between the ligand and the protein and helps elute off the protein with the higher hydrophobicity which in this case is the GFP-CBM1. A decreasing salt concentration is integrated into the process in the form of a linear gradient over a 3-hour time period at a flow rate of 0.5 mL/min. The slow flow rate coupled with the gradient length provides chromatograms with better resolution between observed peaks and also helps avoid co-elution of the two proteins. From the chromatogram (Figure 9) and accompanying small-scale binding assays to observe binding to cellulose I, it is noted that GFP-CBM1 elutes off at the end of the linear gradient. The small-scale binding assays are done for each eluent of each mutant. The percentage bound of protein is determined by the drop in fluorescence on comparing wells with the eluent and the wells where eluents are bound to cellulose-I. The eluent that contains just GFP or a mixture of GFP and GFP CBM1 shows a lower binding to cellulose as GFP by itself does not bind to cellulose. However, this does not ring true for all the mutants as some of them do not bind as well as the wild type CBM1. The microtube based binding assay results of the eluents that contain the final purified GFP-CBM1 proteins alone to cellulose are shown in Table 2.

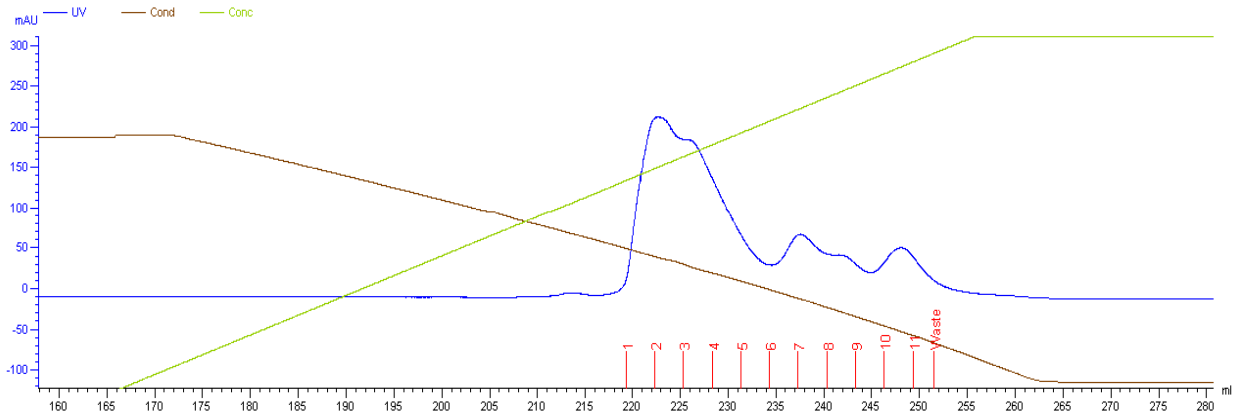

Figure 9: Chromatogram for Hydrophobic Interaction Chromatography (HIC) showing a gradient elution for the Y5N mutant over a period of three hours. As it can be noted, as % elution buffer increases, conductivity decreases. Fractions elute off the column with increasing hydrophobic proteins. From small scale binding assay results, the final peak observed contains the protein of interest.

Protein Type	% Binding of the pooled eluents	% Final Purity of the pooled HIC eluents
GFP-CBM1 Wildtype	90	96
GFP-CBM1 Y5N	71	97
GFP-CBM1 Y5A	66	95
GFP-CBM1 Y5F	85	99
GFP-CBM1 Y5H	50	98
GFP-CBM1 Y5W	58	93

Table 2: Comparison of the percentage of the pooled eluents obtained from HIC with Cellulose-I and the percentage of purity obtained from SDS PAGE analysis.

The eluents are however, correctly identified using SDS PAGE analysis. The lanes that contain bands only in the 37 kDa region are the ones that consist purely of GFP CBM1 (Figure 10). These eluents were pooled together and concentrated. The concentrated protein was then desalted into the storage buffer of choice (10 mM MES; pH 6.5). The concentration of the protein is estimated and the yield of protein obtained for all the mutants and the wild type GFP CBM1 is calculated per gm of original wet cell pellet purified.

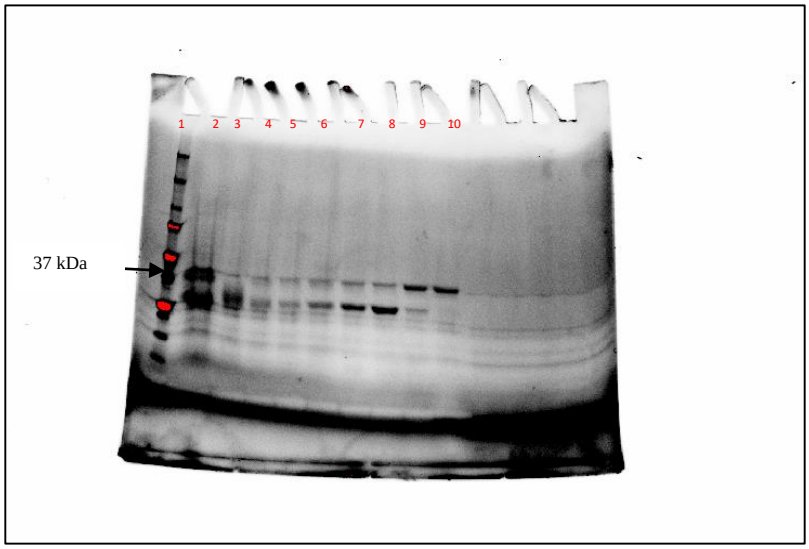

Figure 10: Coomassie stained SDS PAGE Gel (BioRad) analysis shows pure GFP CBM 1 Protein in the last lane (Lane no. 10). Lane 2 is the IMAC B eluent. Lane 3-7 indicates the presence of GFP and GFP CBM1. The band exists in the 37 kDa Molecular weight region when compared to the standard.

2.3 Discussion

Family 1 CBMs from fungi *T. reesei* have three conserved tyrosine residues on the flat face of the CBM1. The tyrosine residue at the fifth position is considered the most important residue due to its structural importance as well as affect its binding to cellulose I. Site directed mutagenesis at the Y5 position is critical to study the effect this residue on binding of CBM1 to cellulose. Substituting the tyrosine with alanine is one of the most commonly studied mutations. However, in this study we were interested in systematically studying the effect of all possible mutations at the Y5 position on the structure and function of CBM1 and thus, generated all 19 mutants. Each mutant was then broadly categorized based on whether the substituted amino acid is aromatic, aliphatic, charged, and polar uncharged. Based on this categorization, five amino acids were chosen (Table 1) to substitute the tyrosine at the Y5 position –Alanine (Nonpolar, aliphatic),

Histidine (positively charged), Tryptophan and Phenylalanine (Nonpolar, aromatic) and Asparagine (polar, uncharged).

Gene Name	CBM1 Translated Gene Sequence
CBM1_TrCel7A_wild type	TQSHYGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL
CBM1_TrCel7A_Y5N	TQSHNGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL
CBM1_TrCel7A_Y5A	TQSHAGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL
CBM1_TrCel7A_Y5F	TQSHFGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL
CBM1_TrCel7A_Y5H	TQSHHGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL
CBM1_TrCel7A_Y5W	TQSHWGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL

Table 3: Amino acid sequence for the CBM1 Y5 Wildtype and its mutants.

The GFP fused CBM1 facilitates rapid detection of protein as well as carrying out confocal fluorescence microscopy. The presence of the GFP also aids in simplifying the purification process. As the name suggests, Green Fluorescent Protein is highly fluorescent (i.e., Excitation at 488 nm and Emission at 512 nm) which makes it easy to visually track the protein as it is loaded onto the columns as well as when it elutes off as various fractions and is collected. The chromatograms obtained from the UNICORN software used to run the AKTA FPLC is a clear indication of the advantage of having a GFP fused to the family 1 CBM.

The Immobilized Metal Affinity Chromatography process was selected as a purification step to separate the histidine tagged GFP CBM1 from the cell lysate. The histidine side chains show the highest interaction with immobilized metal ion matrices such as nickel-nitriloacetic acid (Ni^{2+}-NTA). This allows for the retention of his-tagged proteins on the column matrix. These proteins are then eluted by adding high concentrations of imidazole to the buffer (Bornhorst & Falke, 2000). The his-tagged proteins obtained after this step however is a mixture of two proteolysis products- GFP and the intact full length protein – GFP CBM1. The percentage

of the full length protein is estimated to be around 50% of the obtained pure histidine tagged protein. This estimation required a second step of purification – Hydrophobic Interaction Chromatography. There are multiple resins that can be used to separate the two proteolyzed products based on the degree of the difference of hydrophobicity – Phenyl Sepharose, Butyl Sepharose, and Octyl Sepharose. Each resin media caters to different hydrophobic interaction needs. The phenyl sepharose resin promotes very strong hydrophobic interactions which makes elution of the protein difficult even with minimal salt concentration (data not shown). Using the butyl sepharose resin media, which promotes medium interactions seemed ideal for the protein in question. The microtube based assay results allowed for the pooling in of only certain eluents based on how much binding affinity to cellulose-I was observed. For Y5 mutants, where a clear inference couldn't be made, SDS PAGE analysis showed which eluents still contained a mixture of the proteolyzed products and which eluents contained the full length GFP CBM1.

The results for the yield calculation for the GFP CBM1 Wildtype and the various Y5 mutants are not uniform for all the proteins. With each mutant, there is a difference in the protein yield, although the growth conditions were uniform across the mutants and the wildtype. While the Y5N and the Y5A mutants showed the highest yields of protein, the Y5 wildtype and the other aromatic mutations of the CBM game low protein yields. However, additional growth experiments need to be done to come to a certain conclusion.

Chapter 3 Characterization of GFP-CBM1 adsorption to cellulose-I and cellulose-III

The current work has a central focus on five select mutants, their expression and purification strategies and the adsorption studies of the GFP CBM 1 Y5 Wildtype and mutants on cellulose-I and cellulose-III. The methods used to derive the partition coefficients and the results drawing a comparison between the assays with both the allomorphs are explained in detail. However, a more comprehensive study can be done to screen mutants and their adsorption to cellulose I based on a small scale growth and simultaneous expression of all nineteen mutants (Lesley, 2009). Using this data would allow to choose mutants that show higher affinity to cellulose and perform partition coefficient assays on the same mutants with both forms of cellulose. Studies must also be done to investigate the effect of pH and ionic strength on the binding between CBM and cellulose. Preliminary studies done in both aspects and their results have been included here.

3.1 Materials and Methods

3.1.1 Partition Coefficient Assays of CBM1 mutants to cellulose allomorphs

Avicel PH-101 (Sigma Aldrich) based microcrystalline cellulose I, derived from plant biomass, was used for the partition coefficient binding assays of the wildtype CBM and its mutants. Avicel cellulose I was pretreated with liquid ammonia to obtain cellulose III (Pretreatment conditions of 90 degrees Celsius, 30 minutes, 1000 psi, ratio of ammonia to cellulose 6:1) and was kindly prepared by Dr. Leonardo Sousa at Michigan State University. Binding assays were performed with four replicates each, in 300 μL volume based 96 well round bottomed clear polypropylene microplates (USA Scientific), at increasing protein concentrations using the solid depletion method as described elsewhere (Abbott & Boraston, 2012). Briefly, 25 μL of 100

mg/mL of Cellulose – I slurry was added to the microwells or 100 μL of 100 mg/mL Cellulose-III was added for respective substrate binding assays. Protein dilutions were prepared in 10 mM MES at pH 6.5 to give rise to effective protein concentration in the well ranging from 10 to 120 μg/mL. Bovine serum albumin (BSA) was added to each reaction volume to maintain an effective concentration of 2.5 mg/mL in each microwell. The addition of BSA helps prevent non-specific binding of the protein to the well walls. Also, 1 M MES was spiked into the BSA solution to ensure an effective concentration of 10 mM MES in the final well volume. The control wells didn't contain cellulose were replaced with deionized water at equivalent volumes. The microplates were then shaken in an end-over-end fashion using a VWR Hybridization oven at 5 RPM for 60 minutes at room temperature. Unshaken controls, with the same composition as the shaken controls were used to account for protein denaturation and were left on the laboratory bench during the incubation period. As soon as the incubation period ended, the microplates were centrifuged at 2000 rpm for 2 minutes using an Eppendorf centrifuge. 100 μL of the supernatant from the wells was transferred to a 96 well opaque/black microplate to estimate protein concentration based on GFP fluorescence at the following settings: 480 nm excitation, 512 nm emission and a 495 nm cutoff using Molecular Devices Spectramax M5E microplate reader.

3.1.2 Effect of pH and ionic strength on adsorption to Cellulose-I

The assays were set up for one protein loading of 5 mg/gm glucan for the five mutants at four different pHs – 5.5, 6, 6.5 and 7.0. Binding assays were performed in 300 μL clear, round bottom, 96 well microplates in triplicates for each pH and an equal number of controls were maintained. 25 μL of cellulose-I slurry was added to the wells along with 50 μL of a BSA mixture which has 1M MES of the four different pHs spiked into it to maintain a 50 mM

effective concentration in the well. Similar assays were carried out for the wildtype protein, but an effect well concentration of 10 mM MES. The amount of protein and deionized water added into the wells was adjusted to maintain a 5 mg protein/gm glucan loading and make up the total volume to 200 μL. The microplates were then incubated for 1 hour in an end over end fashion using a VWR Hybridization oven at room temperature. Another set of microplate based experiments were run wherein 1 M NaCl was additionally spiked into the BSA and MES mixture to achieve a 100 mM NaCl effective concentration in the wells. These microplates were also incubated in the above stated fashion. 100 μL of the supernatant was then transferred into an opaque 96 well microplate to read GFP fluorescence at the following settings: 480 nm excitation, 512 nm emission and a 495 nm cutoff using Molecular Devices Spectramax M5E microplate reader.

3.1.3 Small scale growth and screening of GFP CBM1 Y5 Wildtype and mutants

The Rossettagami glycerol stocks stored at -80 degrees Celsius was used to inoculate 500 μL of LB culture in 2 mL deep well, autoclaved 96 well plates (Zerbs, Frank, & Collart, 2009). Each strain was grown in triplicates for 16 hours in the incubator at 37 degrees Celsius. Kanamycin was added to each well at a 1:1000 ratio of Kanamycin to culture volume. 1 mL TB+G culture was then inoculated with 5% starter culture volume as triplicates in another 2 mL deep well, autoclaved 96 well plate and placed in the incubator at 37 degrees Celsius for 4 hours to reach an OD value in the exponential stage. As soon as the OD value was reached, the temperature in the incubator was turned down at 25 degrees Celsius and the cultures were grown overnight for 24 hours.

The 2 mL deep well plate was then centrifuged at 3900 RPM for 15 minutes and the supernatant was discarded. The cell pellets were then lysed with a 1:1 mixture of cell lysis buffer and B-PER Bacterial Protein Extraction Reagent (Thermo Scientific) to make up a total volume of 200 µL. The plate was then centrifuged at 3900 RPM for 20 minutes and the supernatant was then transferred into 300 µL 96 well microplates. A small scale assay to determine percentage bound of protein in cell lysate to Cellulose-I was performed. A 5 mg protein/gm glucan loading based on GFP fluorescence was maintained across the microplate and incubated for 1 hour in an end over end fashion using a VWR Hybridization oven at room temperature. SDS PAGE was used to analyze the percentage of protein bound and left in the supernatant results by using the intensity of the band that falls in the 37 kDa region (Shih et al., 2002). Both controls and the supernatants from the cellulose binding assay were run in the gels and analyzed using both stain free and coomassie staining methods.

3.2 Results

3.2.1 Partition Coefficient Assays of CBM1 mutants to cellulose allomorphs

The data obtained from the fluorescence values are used to measure the partition coefficient of CBM protein to cellulose based on the Langmuir adsorption model for protein concentration ranges in the linear binding range. Reversibility of binding studies of the CBM protein to both allomorphs of cellulose will be conducted in the future (Preliminary studies by Dr. Chundawat and others have shown that CBM 1 reversibly binds to cellulose). A standard curve was obtained as a plot between fluorescence and the effective protein concentration in the control wells using the unshaken standards. Using the slope from this plot, the protein concentration in the wells containing cellulose-I and cellulose-III can be easily deduced. An isotherm of concentration of

Bound protein (μM/gmol of Cellulose) versus Free protein (μM) was generated and the partition coefficient was determined as the slope of the isotherm line. The partition coefficient, α is defined as

$$\alpha = \frac{\mathrm{N}}{K_d}$$

where, N is the total number of binding sites and K_d is the dissociation constant.

A nonlinear regression model can be used to estimate the number of available binding sites on the surface of the ligand using a conventional one or two-site Langmuir type binding model (Gilkes et al., 1992),

$$[B] = \frac{[N_1][F]}{K_{d1}+[F]} + \frac{[N_2][F]}{K_{d2}+[F]}$$

where, [B] is the bound protein concentration, [F] is the free protein concentration, $[N_1]$ and $[N_2]$ refer to the density of protein binding sites on the cellulose and K_d refers to the dissociation constant. This research work does not include findings of the affinity constant and binding capacity due to low protein yield constraints. However, the initial linear range partition coefficient experiments will lead to determining the apparent relative affinity constants for distinct cellulose allomorphs in the near future.

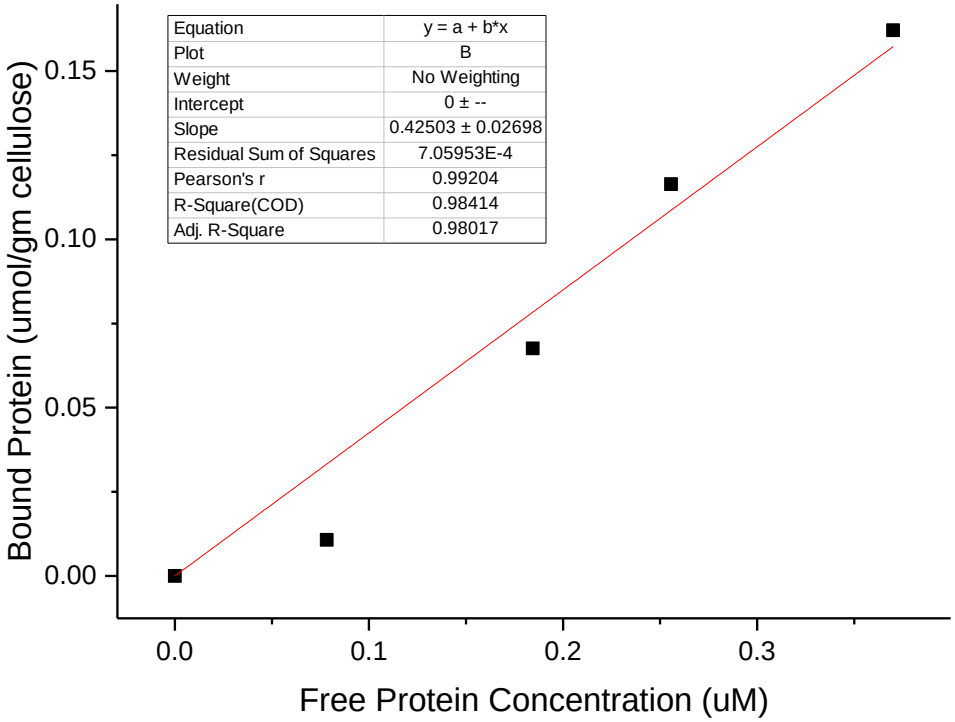

Figure 11: Partition coefficient curve analysis on OriginLabs Software for the GFP CBM1 Wildtype binding to cellulose I

Preliminary data analysis was done in Microsoft Excel while the partition coefficient and the standard error obtained for the data set was calculated using OriginLabs software. The partition coefficients are obtained for all the five mutants and the wildtype CBM1 using the same analysis strategy.

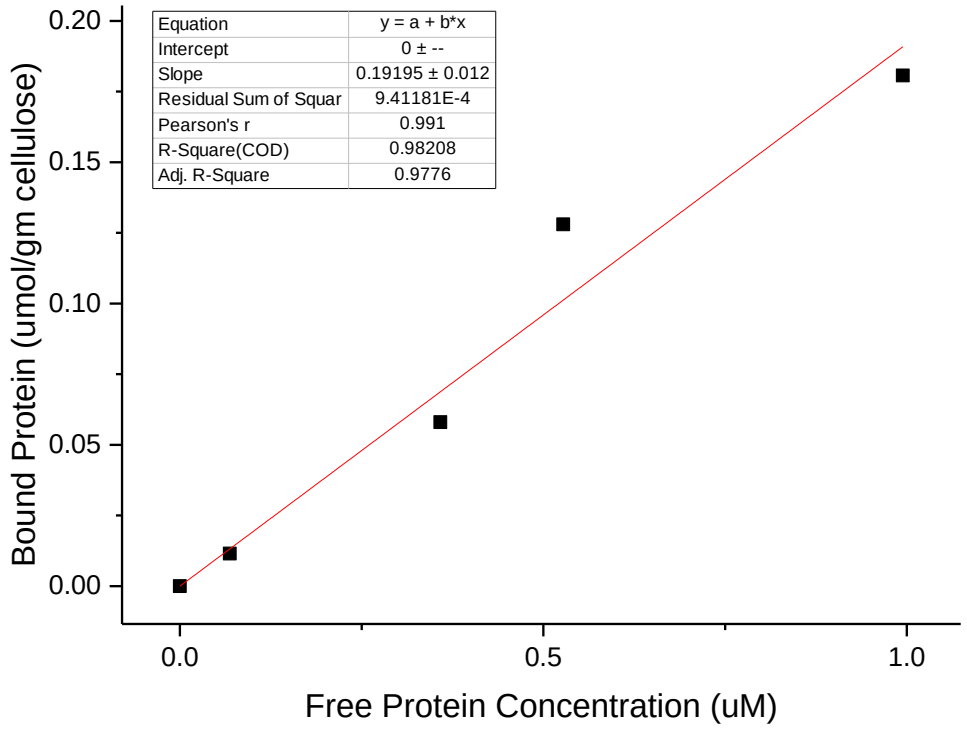

Figure 12: Partition Coefficient curve analysis for GFP CBM1 Wildtype and cellulose III using OriginLab Software.

3.2.2 Effect of pH and ionic strength on adsorption to Cellulose I

The microplate assays were run to draw comparisons between effects observed in a data set where the wells were spiked with a mixture of BSA and MES at different pH values. Another set of assays were run to draw comparisons with similar compositions, but with an additional amount of NaCl spiked into the wells.

Preliminary data analysis as well as error based analysis was performed on Microsoft Excel.

3.2.3 Small scale screening and growth of mutants

Small scale expression and growth was done in a similar fashion as the large scale production of protein. The analysis was the main focus of this study. 10 µL samples of the supernatant from the cellulose binding assay and the controls were added into Eppendorf PCR tubes with 10 µL sample buffer (95% Laemmli buffer, 5% β – Mercaptoethanol) and denatured in the Eppendorf Mastercycler Nexus Gradient (Conditions: 95 degrees Celsius; 5 minutes, 10 degrees Celsius; 10 minutes). After running the gels at 200 V for 40 minutes, they were analyzed in the BioRad Gel Doc Imager using both Stainfree and Coomassie methods. The bands pertaining to the 37 kDa region in the standard were highlighted and the volume intensity was used to calculate percentage bound of the protein present in the cell lysate.

3.3 Discussion

3.3.1 Partition Coefficient Assays of CBM1 mutants to cellulose allomorphs

The adsorption of CBMs on the surface cellulose-I and cellulose-III are of interest to predict the productive binding capacity of the full-length *Tr*Cel7A enzymes that will impact hydrolysis yields. The structure of CBM1 and the conserved residues on the flat face are relevant to the binding to cellulose. A quantitative study on how single point mutations to the Y5 residue affect this binding could lead us to a better understanding of the significance of the tyrosine in that particular position (Reinikainen et al., 1992). However, binding of CBM1 to cellulose-III should also yield some interesting structural insights about how CBMs interact with distinct cellulose morphologies. The comparison between the partition coefficients of CBM1 with both forms of cellulose is the main focus of this study. From previous studies, it has been observed that adsorption of full-length cellulases (like *Tr*Cel7A) on the surface of cellulose-III is lower in

comparison to cellulose-I (Gao et al., 2013). The results of the binding assays of the CBM 1 wildtype with cellulose-III clearly confirm the hypothesis that the reduced binding of the Cel7A to cellulose-III can be attributed to a lower affinity of the CBM to the allomorph. The lower affinity is associated with the likely presence of a more hydrophilic binding surfaces on cellulose-III while CBM1 is thought to bind more tightly to hydrophobic surfaces (Nimlos et al., 2012). However, it has never before been confirmed that reduced binding of CBM is likely driving the reduced binding of the full-length cellulase to cellulose III.

Protein Type	**Partition Coefficient for Cellulose I (x 10^{-4} L/gm)**	**Partition Coefficient for Cellulose III (x 10^{-4}L/gm)**
Wildtype	**4250 ± 270**	**1920 ± 130**
Y5N	**130 ± 10**	**50 ± 3**
Y5A	**340 ± 20**	**50 ± 3**
Y5F	**830 ± 80**	**130 ± 10**
Y5H	**250 ± 10**	**20 ± 0**
Y5W	**160 ± 10**	**70 ± 1**

Table 4: Partition coefficients analyzed with the standard error calculated from OriginLabs software for binding of the Wildtype and CBM1 Y5 mutants.

From the partition coefficient results of CBM1 Y5 Wildtype and its mutants (Table 4), it is noted that the partition coefficient for wildtype is the highest in comparison to the partition coefficients observed for the Y5 mutants. Lower binding of all the mutants in comparison to that of the wildtype can be due to the effect the mutation has on the structural stability of the CBM. It is also important to take into consideration that these single point mutations could cause misfolding of the protein which could again explain the low partition coefficients. Among the mutants, the Y5F shows higher binding towards both cellulose I and cellulose III. This mutant was deemed interesting to be studied due to the presence of the hydroxyl side chain on the ring of the phenylalanine. While the phenylalanine mutant shows the highest binding among the mutants to

cellulose I, it is observed that the asparagine mutant gave the lowest partition coefficient. However, with respect to cellulose III, the Y5H mutant is observed to have the smallest partition coefficient in comparison to the other mutants and the wildtype. The CBM1 wildtype binds to the face of cellulose due to the exact positioning of the specific side chains on the flat face – Y5, Y31 and Y32. While, substituting the tyrosine at the fifth position with alanine, which is nonpolar and aliphatic, could affect the structural compactness of the CBM1 which can explain the reduced binding observed as well. Both structural stability issues and protein electrostatic repulsion could be deemed responsible for the reduced binding affinity of the mutant CBM to the surface of cellulose. The mutants - Y5A, Y5N and Y5W, show low binding affinity to both allomorphs of cellulose. The phenylalanine and tryptophan residues were important to this study to observe the effect of hydrophobic residues on the binding to CBM1. While, the higher hydrophobicity of the aromatic ring on the tryptophan was expected to show a higher affinity to cellulose, (Linder, Lindeberg, Reinikainen, Teeri, & Pettersson, 1995) the results obtained here are somewhat puzzling. It is unclear if the substitution of the tyrosine with the tryptophan causes the protein to be structurally unstable and not bind as tightly to cellulose as the phenylalanine mutant. Furthermore, this also points to the need of a better understanding that the pH at which these experiments are carried out could have on binding of the CBM to the surface of cellulose. While pH could be one factor affecting the results, the final strength of the buffer spiked into these wells could also be a major factor. Further experiments determining the effect of pH and ionic strength on the binding of the CBM to cellulose are necessary to draw definite conclusions from the binding.

These experiments pertain only to a few Y5 mutants based on the amino acid that substituted the tyrosine. Further mutants can be selected based on small scale screening and binding

experiments to cellulose-I (Lesley, 2009). The results can aid in studying mutants that may otherwise be ignored.

3.3.2 Effect of pH and ionic strength on adsorption to Cellulose I

pH based adsorption studies were performed with the presence of an effective concentration of 100 mM NaCl in the wells. The same studies were performed without any external addition of NaCl as well for the same range of pH. These pH dependent adsorption studies reveal the effect that electrostatic repulsion within the protein bound to the surface of the cellulose could have on the binding of the CBM to the cellulose (Reinikainen, Teleman, & Teeri, 1995).

	Percentage of Bound Protein			
Protein Type	**pH 5.5**	**pH 6.0**	**pH 6.5**	**pH 7.0**
Y5 Wildtype*	**84 ± 2**	**72 ± 3**	**68 ± 1**	**N.D**
Y5N	**58 ± 8**	**13 ± 11**	**29 ± 2**	**24 ± 6**
Y5A	**51 ± 4**	**29 ± 10**	**31 ± 1**	**27 ± 8**
Y5F	**44 ± 1**	**54 ± 2**	**47 ± 3**	**61 ± 4**
Y5H	**41 ± 2**	**55 ± 3**	**18 ± 6**	**16 ± 5**
Y5W	**56 ± 3**	**29 ± 5**	**28 ± 10**	**18 ± 0**

Table 5: Results of the pH depend adsorption studies with Cellulose-I conducted for each mutant at four different pH values. *The wildtype assays were done at a well concentration of 10 mM MES for pH 5.5, 6.0 and 6.5. (N.D – Not Determined)

From the results of the pH dependent studies alone (Table 5), a consistent trend seems to appear with three of the mutants showing higher binding to Cellulose-I at pH 5.5. This pH is closest to the isoelectric point of the protein which is estimated to be around 6.0 (estimated using expasy protparam tool). At the isoelectric point, the overall charge of the molecules is lowest. Apart from having minimal charge, electrostatic repulsion forces between the protein and the cellulose are also at their minimum. The adsorption of the Y5A and Y5N mutant reduces at higher pHs. This trend is also observed for the tryptophan residue where the highest binding is observed at

pH 5.5 and subsequently decreases with increasing pH. Tryptophan is a hydrophobic amino acid residue due to which the binding of the CBM to cellulose at pH 5.5 could be attributed to strong hydrophobic interaction due to a minimal net charge. However, as pH increases, the net charge is disturbed and electrostatic repulsion overshadows the hydrophobic forces and shows a reduced binding to cellulose. Interestingly, the affinity of the Phenylalanine mutant seems to stay in the same range across all four pH values, whereas, the Histidine mutant shows an increase in binding to cellulose from pH 5.5 to 6.0 and then rapidly declines as the pH further increases. The pKa value of the histidine side chain is around 6 and hence, a higher binding can be explained at pH 6.0. At a pH higher than the pKa, the side chain is deprotonated and thus, has an overall negative charge. This repels the slightly negatively charged surface of cellulose and thus, the binding is detrimentally affected. While the behavior of the Y5F mutant is albeit puzzling, the Y5H mutant's behavior could be attributed to the effect of other charged residues on the flat face of the CBM and their interaction with cellulose.

	Percentage of Bound Protein			
Protein Type	**pH 5.5**	**pH 6.0**	**pH 6.5**	**pH 7.0**
Y5 Wildtype*	**70 ± 2**	**64±3**	**70±3**	**N.D**
Y5N	**21±1**	**27±2**	**31±10**	**21±1**
Y5A	**16±5**	**31±1**	**21±1**	**20±3**
Y5F	**65±4**	**48±1**	**42±4**	**72±4**
Y5H	**38±11**	**30±8**	**13±1**	**12±2**
Y5W	**26±5**	**27±5**	**14±1**	**16±1**

Table 6: Dependence of ionic strength and pH on adsorption studies to Cellulose-I was analyzed for all the mutants. 100 mM NaCl was maintained as the effective salt concentration in the wells of the microplate. *The wildtype assays were done at a well concentration of 10 mM MES for pH 5.5, 6.0 and 6.5. (N.D- Not Determined)

The effect of ionic strength has been studied by adding high concentrations of salts such as NaCl or stabilizing salts (Kyriacou, Neufeld, & MacKenzie, 1988). Addition of such salts is said to increase the affinity of the CBM to cellulose (Table 6). The Y5N mutant shows an increased binding to Cellulose-I at all the pH values except 5.5, in comparison to the assay without the addition of NaCl. At pH 5.5, which is close to the isoelectric point of the protein, the addition of the salt is observed to be hindering binding. The binding also seems to increase linearly with an increase in pH. The alanine, histidine and tryptophan mutants however, show decreased affinity to cellulose in the presence of NaCl. This result could differ if another salt with a higher ionic strength such as magnesium sulfate or sodium sulfate was to be added (Reinikainen et al., 1995) instead or if a higher concentration of NaCl were to be used. For both the histidine and tryptophan mutants, a similar trend is observed where binding affinity increases from pH 5.5 to 6.0 and then reduces as the pH increases. The most interesting result was observed with phenylalanine which showed distinctly higher binding to cellulose at pH 7.0 in the presence of NaCl. The binding reduces from pH 5.5 to 6.5 and then reaches almost the same binding observed at pH 5.5. A possible explanation for Y5F mutant's higher affinity to cellulose at pH 7.0 could be that the presence of salt ions shields the electrostatic forces present.

The effect of pH and ionic strength on the Y5 wildtype was also studied, albeit at a different effective concentration of MES. Although a reduction in its affinity to cellulose is observed with reducing pH, it isn't a significant difference in comparison with the mutants. The results discussed in the pH and ionic strength dependence study need more in depth investigation. This could allow for more comprehensive partition coefficient assays which are carried out on the basis of the results obtained in the pH dependence study. Taking into account all these necessary

physical parameters while setting up the partition coefficient assays, can provide some much needed clarity on the binding affinities of the CBM1 mutants to Cellulose-I and cellulose-III.

3.3.3 Small scale screening and growth of mutants

A small scale screening of the mutants and a preliminary binding assay to cellulose-I at a fixed protein loading can highlight specific mutants that show a higher binding affinity to cellulose. This can aid in conducting a more comprehensive study extending into mutants that aren't investigated typically in most CBM mutagenesis research studies reported in the literature. These select mutants can then be used to perform adsorption studies on both forms of cellulose. From the results obtained, three mutants were shown to have higher affinity to cellulose in comparison to the others – Proline, Threonine and Methionine (Figure 13). Proline is a special case of amino acid with a cyclic side chain. The distinctive structure of proline gives it extraordinary conformational rigidity when compared with other amino acids which could also be an underlying reason in the high binding affinity to cellulose-I. While threonine is a polar and uncharged residue, methionine is a hydrophobic amino acid residue with a sulfur side chain. All three amino acid residues in place of the other mutations show a higher binding to cellulose-I. Another point of interest is that each of the observed three residues all belong to different categories of amino acids.

This small scale screening study described in this thesis can be fine-tuned and used as a screening method. Further research needs to be done to refine this preliminary investigation into small scale screening and adsorption studies.

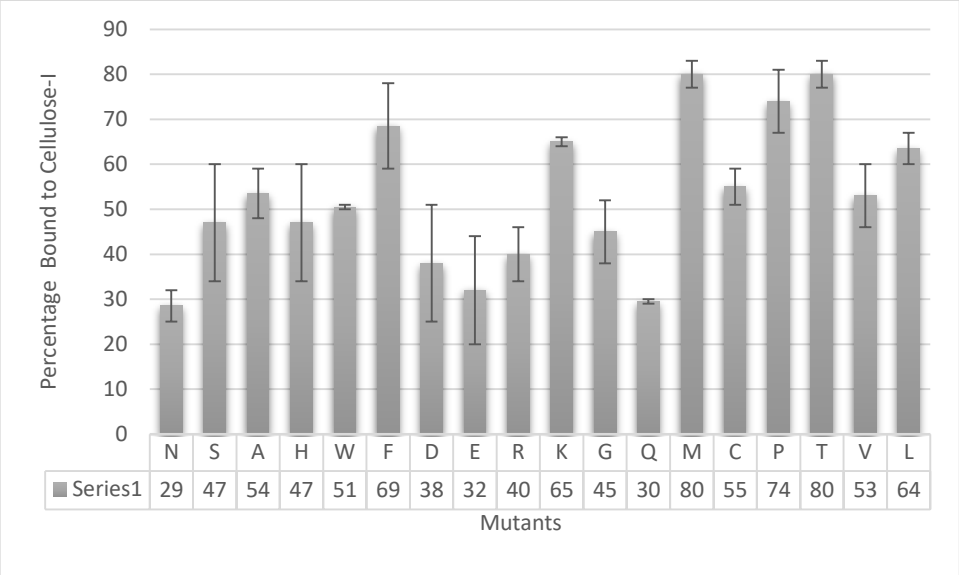

Figure 13: A comprehensive screening of the Y5 mutants was done and the percentage of bound protein to cellulose-I is shown in the table at the bottom of this graph. Error bars indicate one standard deviation from reported mean value. Biological replicate experiments were carried out on two separate days.

Chapter 4 Further Studies and Conclusions

Three decades of research on Family 1 CBMs have unearthed information about the structure of the CBM, the functionality of the CBM in the process of cellulose degradation, and the mechanism of its adsorption to carbohydrates. Studies have been performed detailing the effect of single point mutations on the structure of CBM and its affinity to the surface of cellulose. The effect of physical parameters affecting binding of the CBM to cellulose-I have allowed to develop conditions of optimal affinity and catalytic activity of the appended catalytic domain. Structural changes to cellulose-I such as the ammonia pretreatment process that results in the formation of an unnatural allomorph called cellulose-III is another alternative substrate being explored to help enhance cellulosic biofuel production. The change in the hydrogen bonding interactions in cellulose-III allow for a higher accessibility for some cellulase enzymes to the surface of the polysaccharide. However, the enhanced cellulose accessibility does not always guarantee an increase in binding of the enzyme to the substrate surface. During the course of this thesis, a lower binding of CBM1 to cellulose-III has also been observed. This provides direct evidence for the reduced binding of Cel7A to cellulose-III that has been reported in the literature.

Here, we further studied in detail the effect of single point mutations on the adsorption of GFP CBM1 to two allomorphs of cellulose – Cellulose-I and Cellulose-III, in comparison to the wildtype CBM1. The single point mutants chosen for this study were based on previous studies and the type of amino acid residues. The Y5 residue on the flat face of the CBM has been of interest to researchers due to its impact on the structure of CBM1. It has been hypothesized that the interactions of the H4 residue with the Y5 residue give CBM1 its stability. Therefore, future studies on making mutations at the H4 position, in tandem with the Y5 position, must be considered as well. It must be taken into consideration that the histidine residue is not a

conserved amino acid residue and might tolerate additional mutations at this position unlike the Y5 position.

Using a high throughput screening method to select additional Y5 mutants to study based on binding to cellulose-I will aid in rapidly identifying novel mutants. Refining this method can aid in selecting mutants that show a high affinity to cellulose-I or cellulose-III, which in turn can then be expressed, grown and purified on a larger scale. Development of an optimal purification strategy to separate the protein of interest which was yet another focus of this thesis. A two-step purification process with Immobilized Metal Affinity Chromatography separating out the proteins with histidine tags from the cell lysate and Hydrophobic Interaction Chromatography to separate proteins based on the difference in degree of hydrophobicity was necessary to separate the proteolyzed products for the GFP CBM1 Y5 Wildtype and its mutants. The purification strategy can be fine- tuned to achieve higher protein yields. Further experiments can be conducted to study the effect of parameters such as – resin media, flow rate, concentration of the buffers and gradient length.

The adsorption studies of the CBM1 mutants need to be conducted for the non-linear binding isotherm range of protein concentrations as well. This will allow to gain a deeper understanding pertaining to the number of binding sites available on the surface of the ligand as well as determine the affinity constant (Guo & Catchmark, 2013). From the current studies, the phenylalanine mutant shows a higher affinity to both forms of cellulose in comparison to other mutants. This mutation can be explored as a viable option to use for full length cellulase to break down cellulose-III. Studies should also be conducted to exhibit reversibility of binding of the CBM to cellulose allomorphs. Prior to conducting the adsorption studies, the optimal pH and ionic concentration needs to be determined to get accurate results.

Lastly, circular dichroism studies need to be conducted to understand the structural stability of the mutated CBM 1 proteins. Single point mutations can also be carried out for a residue on the non-planar face of the CBM1 to study any possible effect on binding to the surface of cellulose. In conclusion, the road to discovering the full extent of the functionality of the CBM1 and its impact on cellulosic biomass still has a long way to go.

APPENDIX

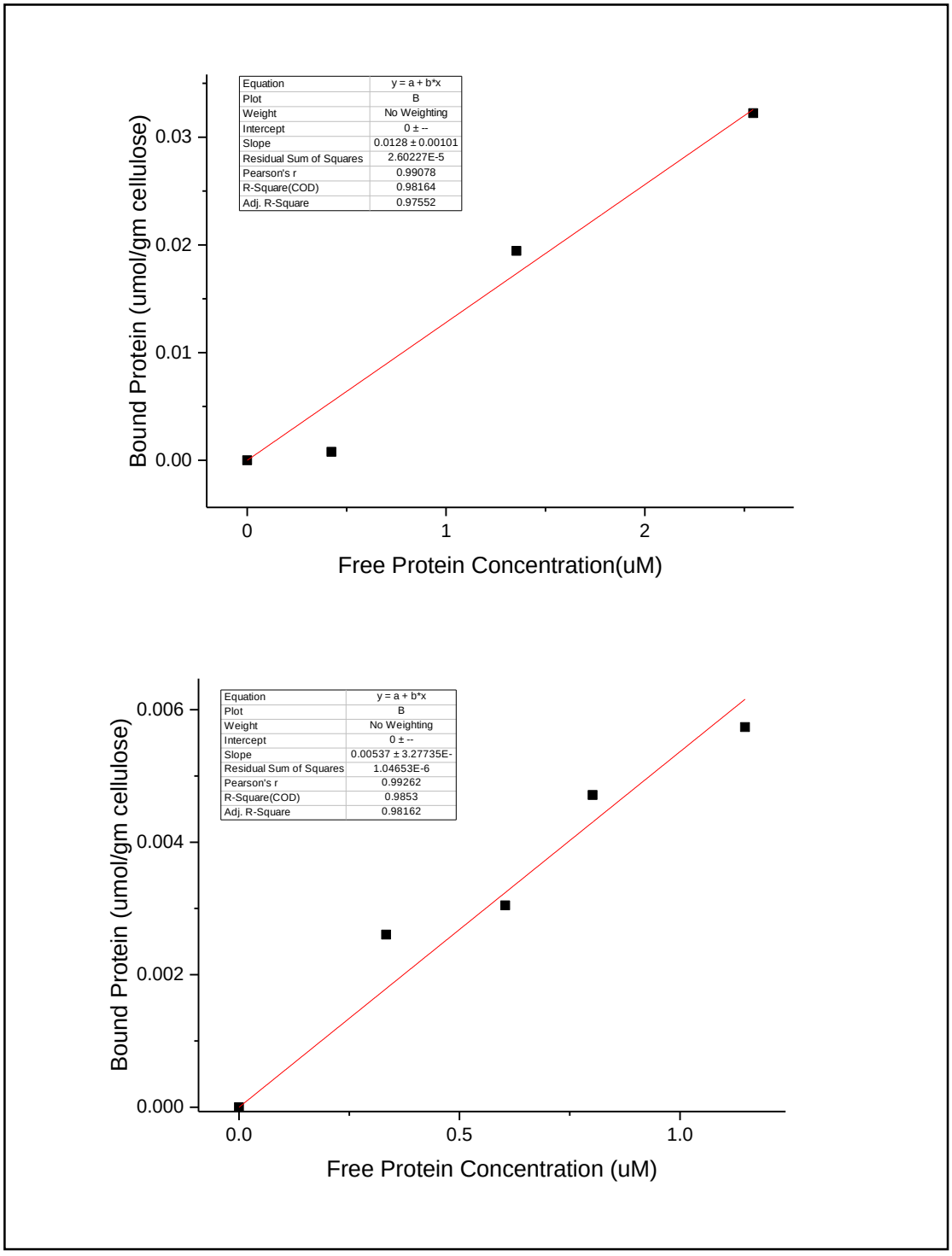

Figure A1: Partition Coefficients for GFP CBM1 Y5N with (i) Cellulose-I (Top) (ii) Cellulose-III (Bottom)

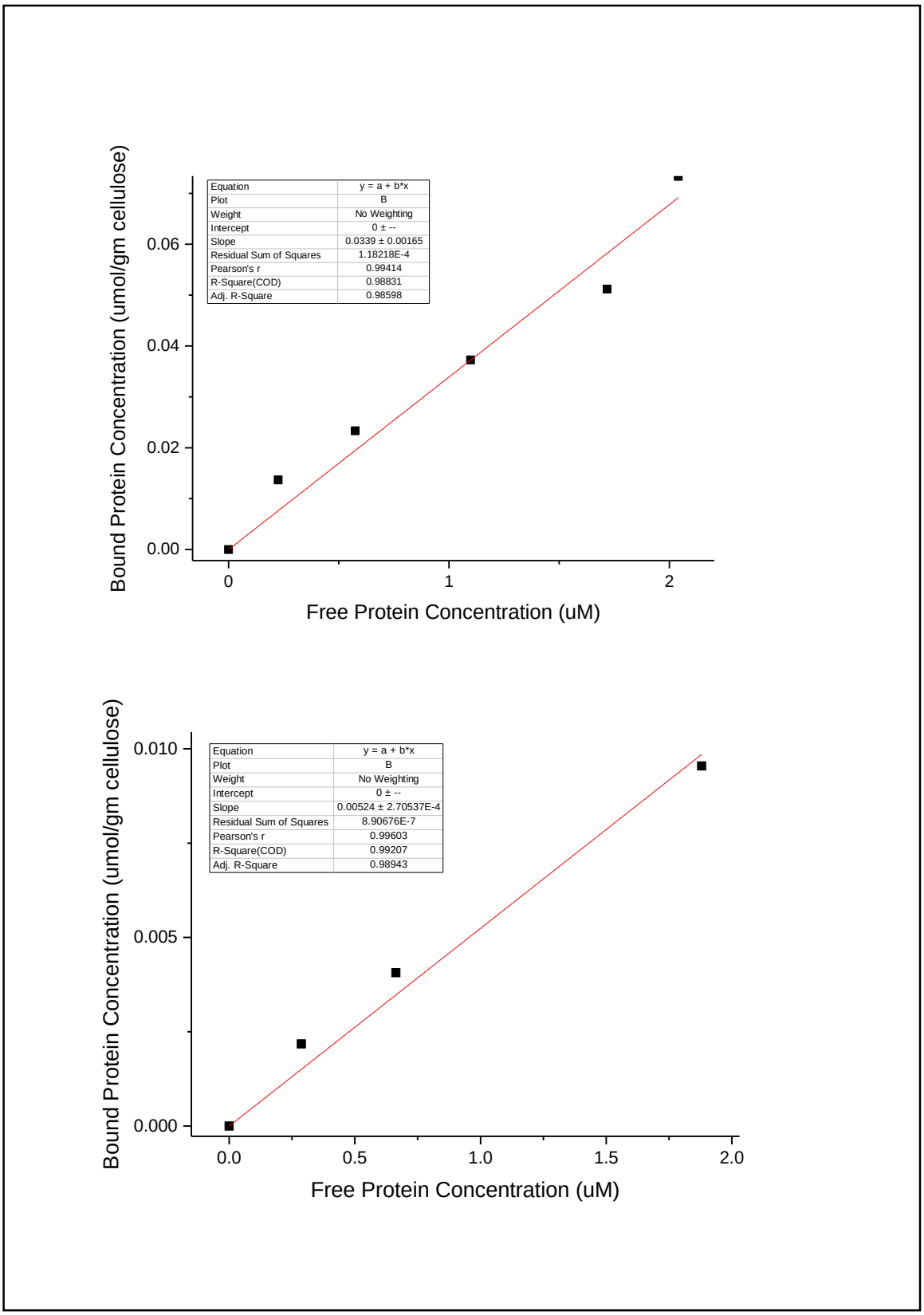

Figure A2: Partition Coefficient of GFP CBM1 Y5A with (i) Cellulose-I (ii) Cellulose-III

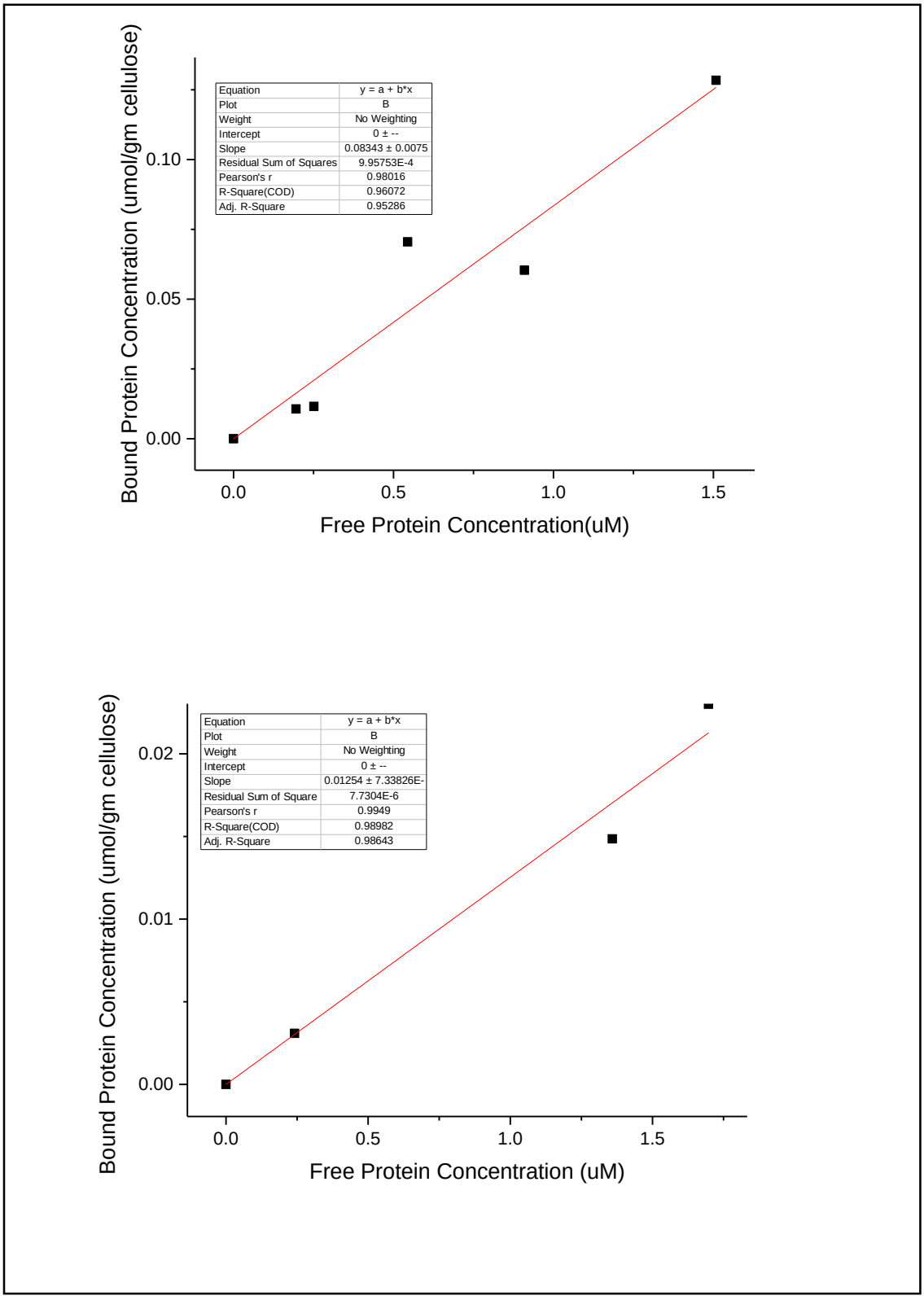

Figure A3: Partition Coefficient for GFP CBM1 Y5F with (i) Cellulose-I and (ii) Cellulose-III

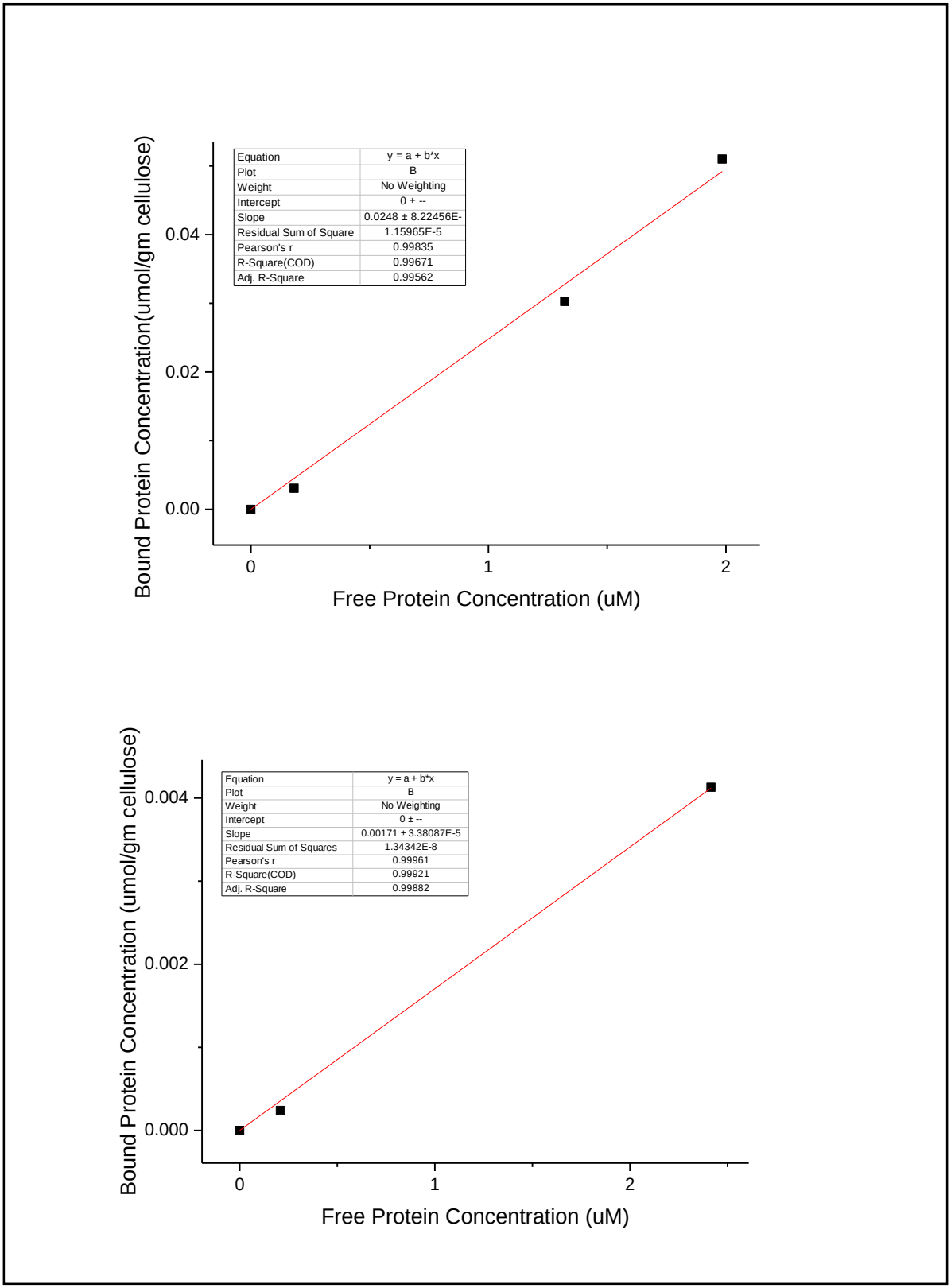

Figure A4: Partition Coefficient of GFP CBM1 Y5H with (i) Cellulose-I and (iii) Cellulose-III

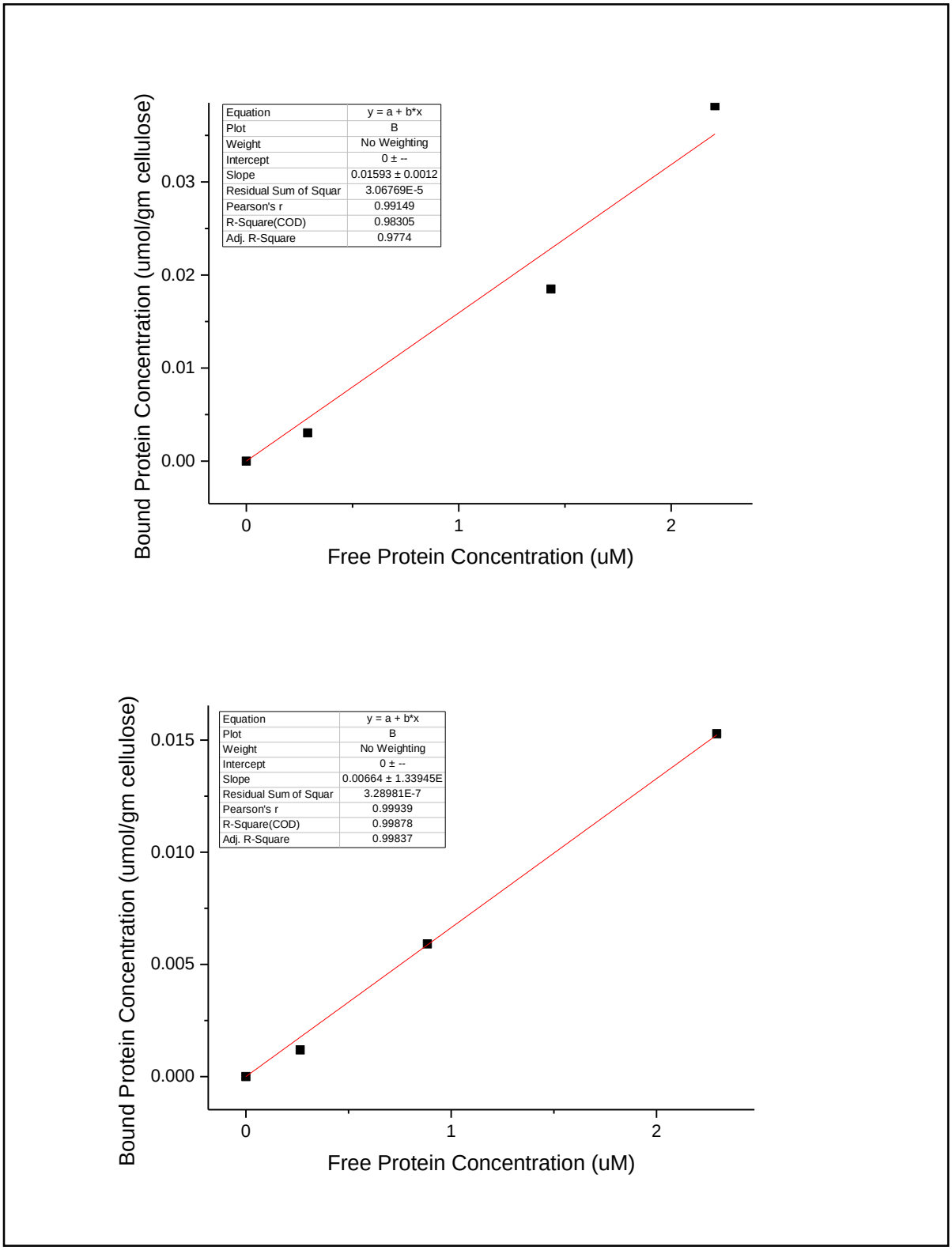

Figure A5: Partition Coefficient of GFP CBM1 Y5W with (i) Cellulose-I and (iii) Cellulose-III

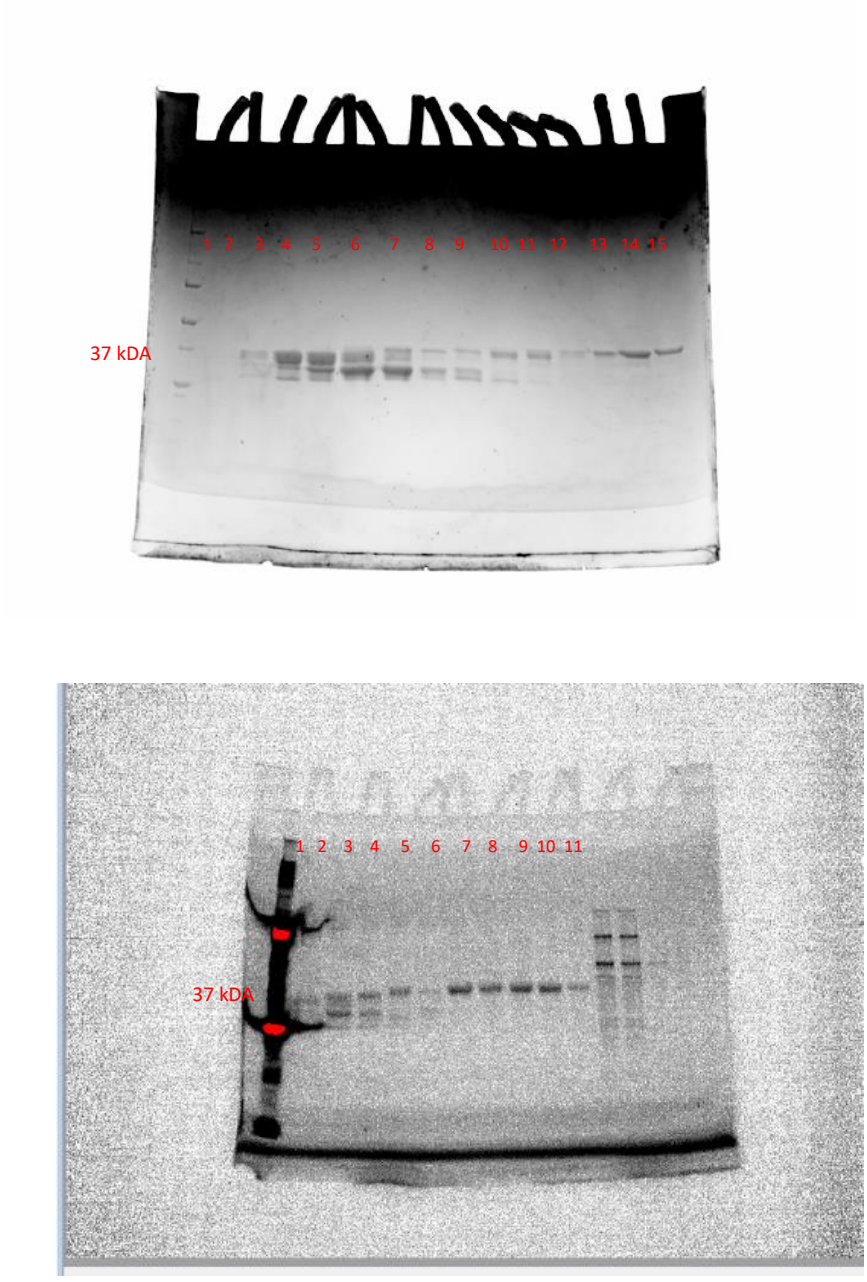

Figure A6: Coomassie stained SDS PAGE Analysis of (i) GFP CBM1 Y5N (Lanes 12-15) (ii) GFP CBM1 Y5A (Lanes 7-11).

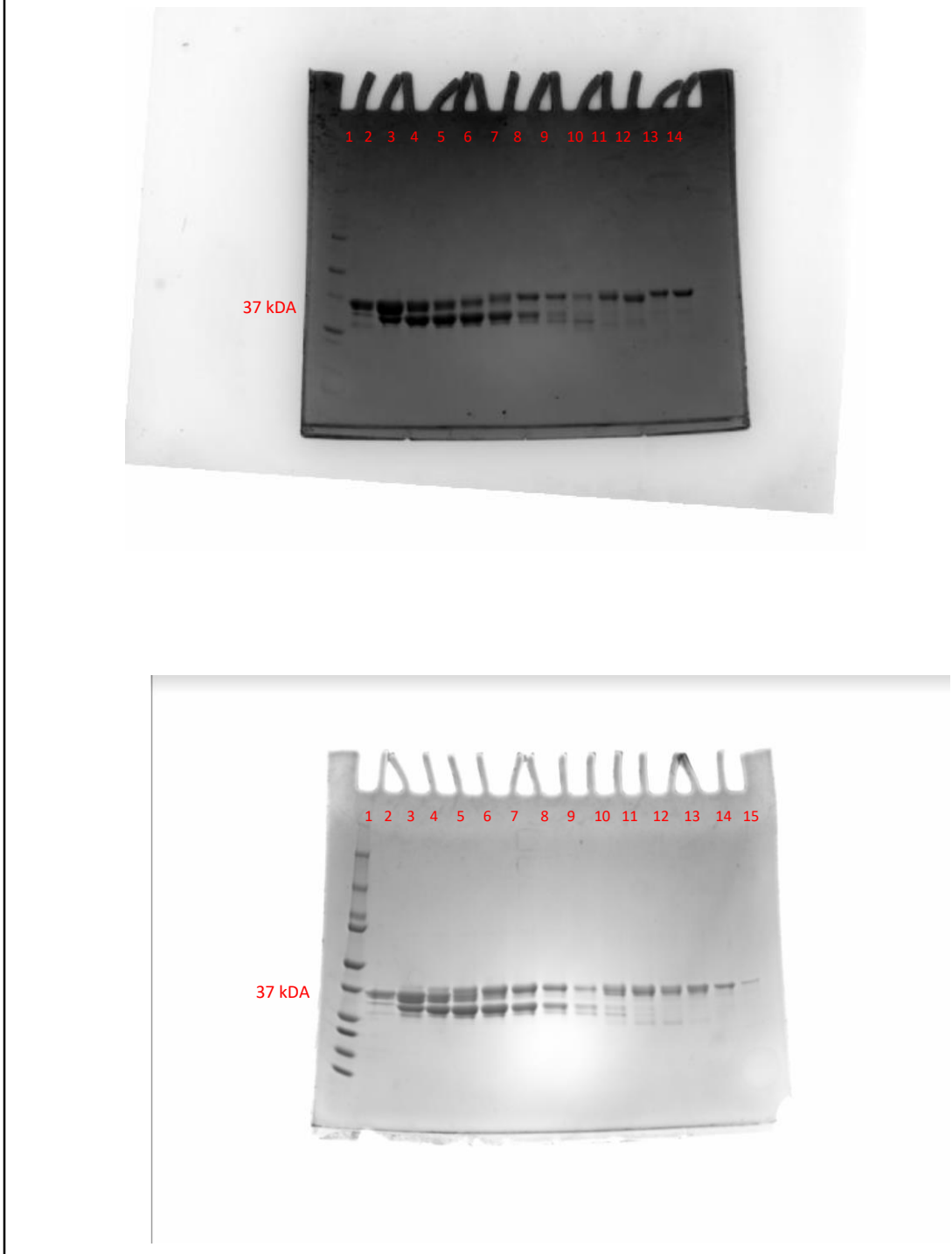

Figure A7: Coomassie stained SDS PAGE Analysis of (i) GFP CBM1 Y5F (Lanes 11-14) (ii) GFP CBM1 Y5 W (Lanes 11-15).

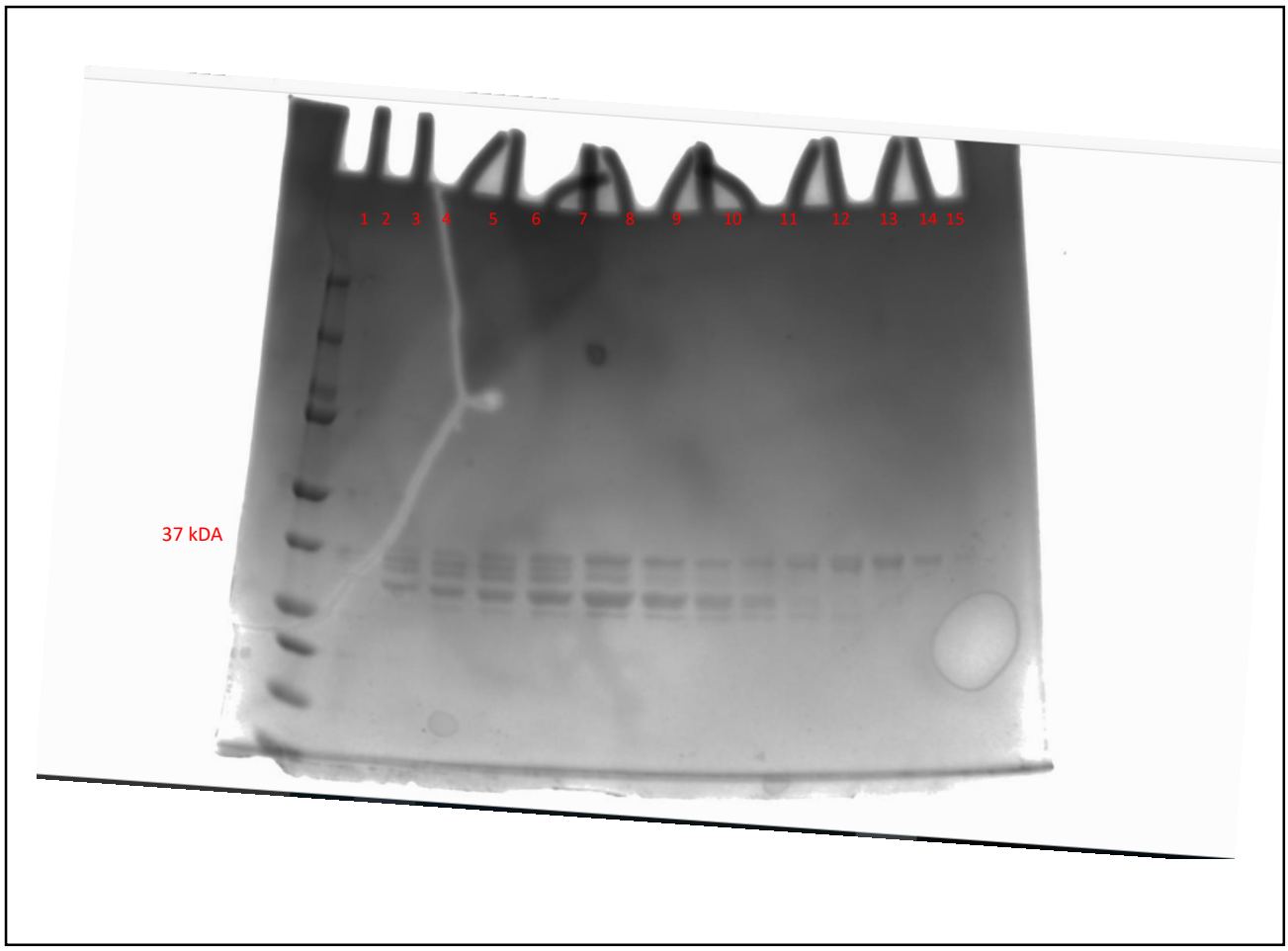

Figure A8: Coomassie stained SDS PAGE Analysis for GFP CBM1 Y5H (Lanes 11-15).

REFERENCES

Abbott, D. W., & Boraston, A. B. (2012). *Quantitative approaches to the analysis of carbohydrate-binding module function*. *Methods in Enzymology* (1st ed., Vol. 510). Elsevier Inc. https://doi.org/10.1016/B978-0-12-415931-0.00011-2

Arola, S., & Linder, M. B. (2016). Binding of cellulose binding modules reveal differences between cellulose substrates. *Scientific Reports*, *6*, 35358. https://doi.org/10.1038/srep35358

Beckham, G. T., Matthews, J. F., Bomble, Y. J., Bu, L., Adney, W. S., Himmel, M. E., … Crowley, M. F. (2010). Identification of amino acids responsible for processivity in a family 1 carbohydrate-binding module from a fungal cellulase. *Journal of Physical Chemistry B*, *114*(3), 1447–1453. https://doi.org/10.1021/jp908810a

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

Bornhorst, J. A., & Falke, J. J. (2000). [16] Purification of Proteins Using Polyhistidine Affinity Tags. *Methods in Enzymology*, *326*, 245–254. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2909483/

Brown, R. M. (2004). Cellulose structure and biosynthesis: What is in store for the 21st century? *Journal of Polymer Science Part A: Polymer Chemistry*, *42*(3), 487–495. https://doi.org/10.1002/pola.10877

Chundawat, S. P. S., Beckham, G. T., Himmel, M. E., & Dale, B. E. (2011). Deconstruction of Lignocellulosic Biomass to Fuels and Chemicals. *Annual Review of Chemical and Biomolecular Engineering*, *2*, 121–45. https://doi.org/10.1146/annurev-chembioeng-061010-114205

Chundawat, S. P. S., Bellesia, G., Uppugundla, N., Da Costa Sousa, L., Gao, D., Cheh, A. M., … Dale, B. E. (2011). Restructuring the crystalline cellulose hydrogen bond network enhances its depolymerization rate. *Journal of the American Chemical Society*, *133*(29), 11163–11174. https://doi.org/10.1021/ja2011115

da Costa Sousa, L., Chundawat, S. P., Balan, V., & Dale, B. E. (2009). "Cradle-to-grave" assessment of existing lignocellulose pretreatment technologies. *Current Opinion in Biotechnology*, *20*(3), 339–347. https://doi.org/10.1016/j.copbio.2009.05.003

da Costa Sousa, L., Jin, M., Chundawat, S. P. S., Bokade, V., Tang, X., Azarpira, A., … Balan, V. (2016). Next-generation ammonia pretreatment enhances cellulosic biofuel production. *Energy Environ. Sci.*, *9*, 1215–1223. https://doi.org/10.1039/C5EE03051J

Driscoll, P. C., Gronenborn, A. M., Beress, L., & Clore, G. M. (1989). Determination of the three-dimensional solution structure of the antihypertensive and antiviral protein BDS-I from the sea anemone Anemonia sulcata: a study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. *Biochemistry*, *28*(5), 2188–98. https://doi.org/10.1021/bi00444a016

Gao, D., Chundawat, S. P. S., Sethi, A., Balan, V., Gnanakaran, S., & Dale, B. E. (2013).

Increased enzyme binding to substrate is not necessary for more efficient cellulose hydrolysis. *Proceedings of the National Academy of Sciences*, *110*(27), 10922–10927. https://doi.org/10.1073/pnas.1213426110

Gilkes, N. R., Jervis, E., Henrissat, B., Tekant, B., Miller, R. C., Warren, R. A. J., & Kilburn, D. G. (1992). The adsorption of a bacterial cellulase and its two isolated domains to crystalline cellulose. *Journal of Biological Chemistry*, *267*(10), 6743–6749.

Guo, J., & Catchmark, J. M. (2013). Binding specificity and thermodynamics of cellulose-binding modules from trichoderma reesei Cel7A and Cel6A. *Biomacromolecules*, *14*(5), 1268–1277. https://doi.org/10.1021/bm300810t

Kyriacou, A., Neufeld, R. J., & MacKenzie, C. R. (1988). Effect of physical parameters on the adsorption characteristics of fractionated Trichoderma reesei cellulase components. *Enzyme and Microbial Technology*, *10*(11), 675–681. https://doi.org/10.1016/0141-0229(88)90059-2

Lesley, S. A. (2009). Chapter 41 Parallel Methods for Expression and Purification. *Methods in Enzymology*, *463*, 767–785. https://doi.org/http://dx.doi.org/10.1016/S0076-6879(09)63041-X

Lim, S., Chundawat, S. P. S., & Fox, B. G. (2014). Expression, purification and characterization of a functional carbohydrate-binding module from Streptomyces sp. SirexAA-E. *Protein Expression and Purification*, *98*, 1–9. https://doi.org/10.1016/j.pep.2014.02.013

Linder, M., Lindeberg, G., Reinikainen, T., Teeri, T. T., & Pettersson, G. (1995). The difference in affinity between two fungal cellulose-binding domains is dominated by a single amino acid substitution. *FEBS Letters*, *372*(1), 96–98. https://doi.org/10.1016/0014-5793(95)00961-8

Linder, M., Mattinen, M.-L., Kontteli, M., Lindeberg, G., Ståhlberg, J., Drakenberg, T., … Annila, A. (1995). *Identification of functionally important amino acids in the cellulose-binding domain of Trichoderma reesei cellobiohydrolase I. Protein science : a publication of the Protein Society* (Vol. 4). https://doi.org/10.1002/pro.5560040604

Loewenthal, R., Sancho, J., & Fersht, A. R. (1992). Histidine-aromatic interactions in barnase. *Journal of Molecular Biology*, *224*(3), 759–770. https://doi.org/http://dx.doi.org/10.1016/0022-2836(92)90560-7

Mattinen, M. L., Kontteli, M., Kerovuo, J., Linder, M., Annila, A., Lindeberg, G., … Drakenberg, T. (1997). Three-dimensional structures of three engineered cellulose-binding domains of cellobiohydrolase I from Trichoderma reesei. *Protein Science : A Publication of the Protein Society*, *6*, 294–303. https://doi.org/10.1002/pro.5560060204

Nimlos, M. R., Beckham, G. T., Matthews, J. F., Bu, L., Himmel, M. E., & Crowley, M. F. (2012). Binding preferences, surface attachment, diffusivity, and orientation of a family 1 carbohydrate-binding module on cellulose. *Journal of Biological Chemistry*, *287*(24), 20603–20612. https://doi.org/10.1074/jbc.M112.358184

Payne, C. M., Knott, B. C., Mayes, H. B., Hansson, H., Himmel, M. E., Sandgren, M., … Beckham, G. T. (2015). Fungal cellulases. *Chemical Reviews*, *115*(3), 1308–1448.

https://doi.org/10.1021/cr500351c

Reinikainen, T., Ruohonen, L., Nevanen, T., Laaksonen, L., Kraulis, P., Jones, T. A., … Teeri, T. T. (1992). Investigation of the function of mutated cellulose-binding domains of Trichoderma reesei cellobiohydrolase I. *Proteins: Structure, Function, and Bioinformatics*, *14*(4), 475–482. https://doi.org/10.1002/prot.340140408

Reinikainen, T., Teleman, O., & Teeri, T. T. (1995). Effects of pH and high ionic strength on the adsorption and activity of native and mutated cellobiohydrolase I from Trichoderma reesei. *Proteins: Structure, Function, and Bioinformatics*, *22*(4), 392–403. https://doi.org/10.1002/prot.340220409

Saeman, J. F. (1945). Kinetics of Wood Saccharification - Hydrolysis of Cellulose and Decomposition of Sugars in Dilute Acid at High Temperature. *Industrial & Engineering Chemistry*, *37*(1), 43–52. https://doi.org/10.1021/ie50421a009

Shih, Y.-P., Kung, W.-M., Chen, J.-C., Yeh, C.-H., Wang, A. H.-J., & Wang, T.-F. (2002). High-throughput screening of soluble recombinant proteins. *Protein Science : A Publication of the Protein Society*, *11*, 1714–1719. https://doi.org/10.1110/ps.0205202

Sugimoto, N., Igarashi, K., & Samejima, M. (2012). Cellulose affinity purification of fusion proteins tagged with fungal family 1 cellulose-binding domain. *Protein Expression and Purification*, *82*(2), 290–296. https://doi.org/10.1016/j.pep.2012.01.007

Wi, S. G., Cho, E. J., Lee, D.-S., Lee, S. J., Lee, Y. J., & Bae, H.-J. (2015). Lignocellulose conversion for biofuel: a new pretreatment greatly improves downstream biocatalytic hydrolysis of various lignocellulosic materials. *Biotechnology for Biofuels*, *8*(1), 228. https://doi.org/10.1186/s13068-015-0419-4

Zerbs, S., Frank, A. M., & Collart, F. R. (2009). *Chapter 12 Bacterial Systems for Production of Heterologous Proteins*. *Methods in Enzymology* (1st ed., Vol. 463). Elsevier Inc. https://doi.org/10.1016/S0076-6879(09)63012-3