IMPACT OF CARBOHYDRATE BINDING MODULES ON TRANSGLYCOSYLATION EFFICIENCY IN THE GLYCOSIDE HYDROLASE FAMILY 5 PROTEINS

By

MADHURA KASTURE

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 P. S. Chundawat

And approved

_____

_____

_____

New Brunswick, New Jersey

[January, 2020]

ABSTRACT OF THE THESIS

Impact of Carbohydrate Binding Modules on Transglycosylation Efficiency In The Glycoside Hydrolase Family 5 Proteins

By MADHURA KASTURE

Thesis Director:

Dr. Shishir Chundawat

Oligosaccharides are significant in many biological processes such as modulating protein folding, cell interactions, and binding and are widely used in prebiotics and pharmaceutical industries. This rising popularity has accelerated the research for their sustainable source of generation. Enzymatic synthesis offers an alternative over standard chemical synthesis to produce these oligosaccharides with higher specificity and product purity. Transglycosylases are a class of glycoside hydrolases (GH) present in nature and are ideal biocatalysts for this process. Since these are few in number and limited in substrate specificity, over the past couple of years, extensive studies and advancements have been made to engineer glycoside hydrolases to reduce the hydrolytic activity of these enzymes to perform transglycosylation to produce oligosaccharides.

This thesis investigates the transglycosylation activity of glycoside hydrolases family five that has tethered carbohydrate-binding module (CBM) CBM3a. CBMs improve the catalytic efficiency by binding to the polysaccharide chain, making the substrate more

accessible to the enzyme. The catalytic nucleophile of GH5 enzymes was substituted for alanine, glycine, and serine amino acids. Activity assay in the presence of an activated soluble sugar was performed to obtain a product profile for the enzymes. With the data from the activity assays, a kinetic model was developed. A structure-functional analysis was attempted to explain the results. It was observed that few enzymes with a mutated catalytic nucleophile show transglycosylation like activity in the presence of an appended CBM3a, suggesting that the presence of CBM3a affected the mechanism of the enzyme activity other than just increasing the local concentration of the substrate at the catalytic pocket. In summary, this study provides results that led further confidence in the active role of the CBM3a in the transglycosylation activity of the GH5 enzymes and understanding the mechanism.

ACKNOWLEDGEMENTS

I would like to convey my sincere gratitude to the following persons who have aided me in accomplishing my thesis for my graduate program. I would like to thank my thesis advisor, Dr. Shishir Chundawat, for his guidance and sincere support throughout the project. I would like to thank Dr. Henrik Pedersen and Dr. Benjamin Schuster for being part of my thesis committee. I am especially thankful to Chandra Kanth Bandi for being an exceptional mentor, for being supportive and patient throughout the project.

Additionally, I would like to thank Dr. Brian Fox from the University of Wisconsin Madison for gifting the GH5 genes. I am also very grateful for having great laboratory colleagues who have encouraged me all through this time. Finally, I would like to appreciate my parents for their continued support during my graduate studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

GH: Glycosyl hydrolase

GS: Transglycosidase

pNP: Para Nitrophenyl group

WT: Wild-Type

PIPE: Polymerase Incomplete Primer Extension (PIPE) Cloning Method

SLIC: Sequence- and Ligation-Independent Cloning

PCR: Polymerase Chain Reaction

SDM: Site directed mutagenesis

PDB: Protein Data Bank

SDS: Sodium Dodecyl Sulfate

IB: Inclusion Bodies

B-PER II: Bacterial-Protein Extraction Reagent (2X)

PIC: Protein Inhibitor Cocktail

IPTG: Isopropyl β-D-1-thiogalactopyranoside

FPLC: Fast Protein Liquid Chromatography

IMAC: Immobilized Metal Affinity Chromatography

MES: 2-(N-Morpholino)ethane Sulfonic acid

TLC: Thin Layer Chromatography

DNS: 3,5-Dinitrosalicylic acid

Tm: Melting Temperature (°C)

MSA: Multiple Sequence Alignment

MAFFT: Multiple Alignment using Fast Fourier Transform

CHAPTER 1. INTRODUCTION

1.1 Background

Glycoside Hydrolase (GH) or Glycosidases are enzymes that catalyze the hydrolysis of the glycosidic bond between carbohydrates and other secondary moieties such as carbohydrate, a protein or lipid. The GHs have been categorized into diverse families by the Carbohydrate-Active Enzymes (CAZy) database [1]. This categorization of these GHs is based on their amino acid sequence similarities, which show conserved structures, catalytic mechanisms, and active site residues but could differ in substrate specificity. Glycosidases are present in a wide range of organisms such as archaea and bacteria to animals and plants to facilitate the breaking of complex sugars to simple sugars for survival. Most of these enzymes are tethered to auxiliary domains such as carbohydrate-binding modules (CBMs), which improve their catalytic efficiency through substrate recognition and binding [2].

The application of these enzymes is visible in various industries based on the functions that they perform. Amylases are enzymes that assist in the hydrolysis of starch to produce sugars. These enzymes usual belong to GH family 14 and are used for the fermentation of beer and production of flour/baking additives [3]. The paper and pulp industry use xylanases, which belong to GH families 10 and 11, to break down the hemicellulose, which is xylan into xylose [4]. GH family 29 are fucosidases that catalyze the release of fucose sugars from N-linked glycans and human milk oligosaccharides (HMO)[5]. The application of glycoside hydrolases for biomass degradation in the bioenergy industry has been vastly studied. Lignocellulosic biomass consists of cellulose, hemicellulose, and

lignin and requires multiple glycosidases such as cellulases, xylanases, mannanases, etc. to function synergistically to convert them to soluble sugars for fermentation [6]. Cellulases break the β(1-4) linkage of cellulose to release cellobiose and glucose that can be used by microorganisms to produce biofuel. Three kinds of cellulases, such as endocellulases, exocellulases, and glucosidases, are present in nature, which is classified in a multitude of GH families like families 1, 5, 6, 7 [13].

Glycosyl hydrolases catalyze the hydrolysis reaction through either a retaining or an inverting mechanism. Glycosidases that retain the anomeric configuration of the substrate in the yielded product by following a double displacement reaction using a nucleophilic and acid/base residue in the enzyme active site are called retaining glycosidases. In comparison, the enzymes that invert the configuration of the anomeric center of the substrate via a single displacement mechanism with catalytic acid/base residues are called inverting glycosidases. The mechanism of retaining glycosidases, as described by Koshland, is illustrated in Figure 1. The catalytic nucleophile attacks the anomeric carbon while simultaneously, the catalytic acid-base protonates the glycosidic oxygen of the leaving group forming an enzyme-substrate complex. During hydrolysis, a water molecule is activated by the catalytic acid/base residue generating hydroxyl group, which attacks the intermediate complex to form the hydrolyzed product. In addition to hydrolysis, glycosidases can catalyze the synthesis of sugar polymers using a transglycosylation mechanism where the intermediate enzyme-substrate complex is attacked by another sugar molecule instead of water. Transglycosylation mechanisms have been adopted by many glycosyl-hydrolases to enable the utilization of substrates in nature. This property of transglycosylation has been exploited in various studies to

synthesize sugars, which are somewhat challenging to produce using conventional chemical synthesis techniques.

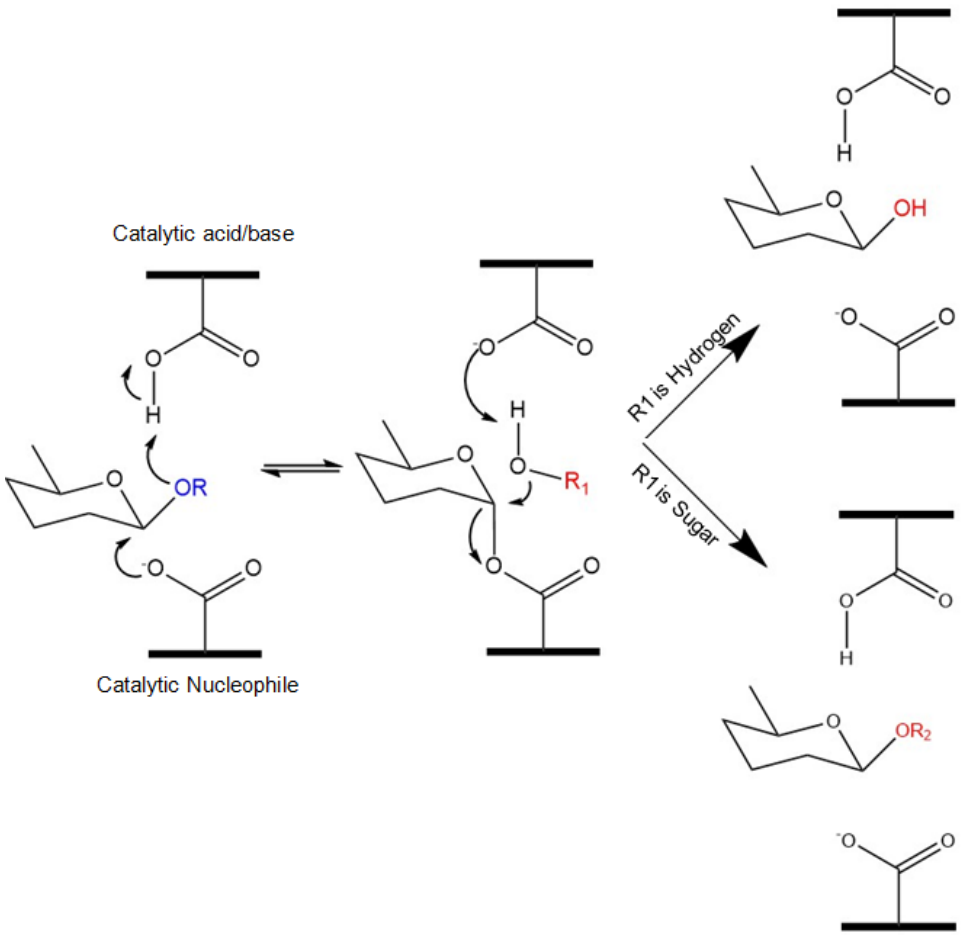

Figure 1: Retaining mechanism of hydrolysis vs. transglycosylation reactions for GH

However, the presence of a catalytic nucleophile makes the products of transglycosylation prone to further hydrolysis, which reduces the yields of synthesized products. This drawback of transglycosidases has enabled the generation of a new class of mutant GHs called glycosynthases (GSs) [7]. Glycosynthases are mutant glycosyl hydrolases that have their catalytic nucleophile mutated to a small amino acid such as alanine, serine, and glycine. The mutation kills the activity of the GH as it cannot form a

substrate-enzyme intermediate. These inactive GSs, when supplemented with an activated donor substrate that mimics the enzyme-substrate complex and an acceptor sugar, will synthesize stable oligosaccharides, as seen in Figure 2 [8].

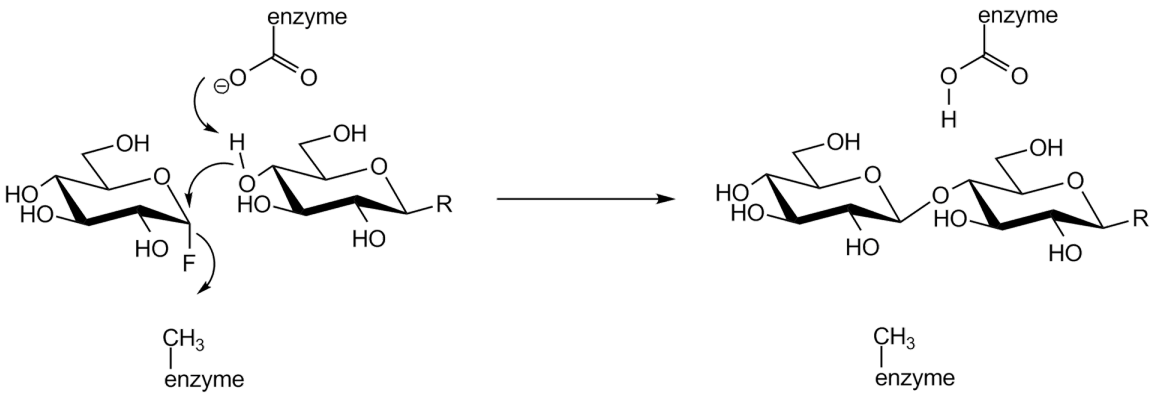

Figure 2: Glycosynthase Mechanism Using a Donor and Acceptor Sugar

In this mechanism, a donor and an acceptor molecule are required to carry out the reaction. The acceptor sugar is deprotonated by the catalytic acid-base, which then attacks the anomeric carbon on the donor sugar is a glycosyl-fluoride. The fluoride is the leaving group why the sugars form a glycosidic bond and form a polysaccharide.

The auxiliary domains, such as carbohydrate-binding modules (CBMs), enhance the hydrolytic activity of GHs [14], [15]. CBMs improve the catalytic efficiency by either (i) localizing the enzyme near the substrate, thus increasing the local substrate concentration near the catalytic domain. (ii) disrupting the polysaccharide chain making the substrate more accessible to the enzyme. In most cases, CBMs have been studied to impact the hydrolysis of insoluble sugars. However, their influence on the activated soluble sugars has not been studied before.

Previous studies in our lab focused on studying the impact of CBM on the transglycosylation efficiency of a GH family 5 enzymes from Clostridium thermocellum (CelE) in the presence of an activated sugar (pNP-cellobiose). As a part of this study, several nucleophilic mutants of CelE were made to kill the activity of the enzyme. Interesting, it was observed that the activity of nucleophilic mutant of CelE on an activated soluble substrate (pNP-cellobiose) was recovered when tethered to CBM. Figure 3 shows the formation of the cellotriose and cellotetraose for the CelE-CBM3a-E316G, which has the catalytic nucleophile of CelE mutated to glycine. Detailed biochemical assays, structural, and computational studies were performed to understand the mechanistic nature of the CBM recovered activity. It was concluded that CBM3a helps to dock the substrate in the active site of CelE and provides additional stacking interaction for a novel SNi-like mechanism as opposed to a retaining mechanism of wild type enzyme.

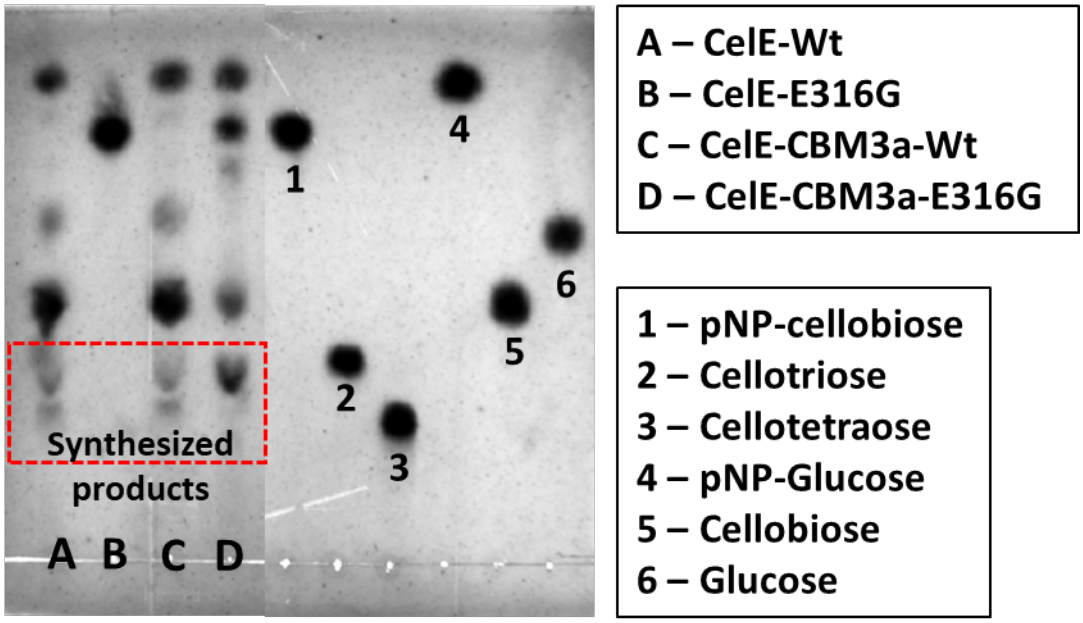

Figure 3: Thin layer chromatography image showing the formation of synthesized products using nucleophilic mutant of CelE-CBM3a in the presence of pNP-cellobiose

The SNi-like mechanism is generally adopted by a different, and related class of enzymes called glycosyltransferases (GTs) for synthesizing sugars. This observed change in the activity of the enzyme in the presence of CBM was intriguing, and we hypothesize that this effect can be observed in other phylogenetically related enzymes.

1.2 Objectives

In this work, enzymes that belong to GH family 5 subfamily 4 and phylogenetically related to CelE were chosen to see to test if the observed effect of CBM3a can be expanded to other sub-families. This thesis tries to cover the following objectives:

1. The role of the CBM3a on the catalytic efficiency of the GH5 family proteins
2. Build a kinetic model based on the observed activity to relate the transglycosylation and hydrolytic activities.

To achieve these two aims, the DNA constructs were generated for the expression of the protein. Next, the protein growth and expression conditions were optimized, and the proteins were purified. The substrate and pH for the activity were then optimized. Lastly, activity assays were performed, and the products formed were analyzed to the efficiency of the CBM3a on the GH5 proteins.

CHAPTER 2. CLONING AND GENERATION OF CONSTRUCTS

The GH5 enzymes that were gifted by Dr. Brian Fox from the University of Wisconsin Madison were present in the pEU vector and were obtained from a cell-free system. Cell-Free protein synthesis and expression do not require biological machinery like a bacterial host to express and produce the protein [9]. In the lab, we do not use the cell-free expression to express the protein, so the GH5 enzymes needed to be transformed into the pEC vector so that the engineered E. coli cells can accept it [16].

Name	GH family	UniProt	Gene Accession	Bacteria
gene1	GH5_4	Q53302	AAB19708.1	*Ruminococcus flavefaciens*
gene2	GH5_unknown	I0XAI8	WP_009104139	*Treponema* sp. JC4
gene3	GH5_4	D2KFI9	ACZ98591.1	*Cellulosilyticum ruminicola*
gene4	GH5_unknown	Not found	2211254A	*Ruminococcus albus*
gene5	GH5_4	H2BCH6	ADU86907.1	*Uncultured bacterium*
gene6	GH5_4	D2K7Z0	ACZ54907.1	*Uncultured bacterium*
gene7	GH5_4	Q97DK6	AAK81397.1	*Clostridium acetobutylicum*
gene8	GH5_4	B7FBJ7	CAL91972	*Polyplastron multivesiculatum*
gene9	GH5_unknown	B0MR54	EDR99710.1	*Eubacterium siraeum DSM 15702*
gene10	GH5_4	Q006P4	ABJ52796.1	uncultured *Butyrivibrio sp*
gene11	GH5_4	A9LDT2	ABW39338.1	*Uncultured bacterium*
gene12	GH5_4	P28623	AAA23233.1	*Clostridium cellulovorans*
gene13	GH5_4	H9KVH3	AAR65335.1	*Paenibacillus pabuli*
gene14	GH5_2	O85465	AAC19169.1	*Salipaludibacillus agaradhaerens*
gene15	GH5_unknown	I5AUN0	EIM57503.1	*[Eubacterium] cellulosolvens 6*

Table 1: Summary of the gene accession number for all 15 Enzymes

Throughout this thesis, we refer to the 15 enzymes as the sample names mentioned in the table. In this chapter, we discuss the various approaches undertaken to generate the required wild type constructs and their respective mutants.

2.1 GH5 Gene Constructs

The standard molecular cloning technique that was used for the construction of the GH5 genes from the pEU vector into the pEC vector was Restriction Enzyme (RE) cloning [10]. The genes of interest that were cloned were through the restriction sites AsiSI and BamHI. The pEU vector is used for cell-free protein expression. For this project, since we used E. coli as our system for protein expression, it was essential to transfer the gene of interest into the vector that is compatible with the system we use.

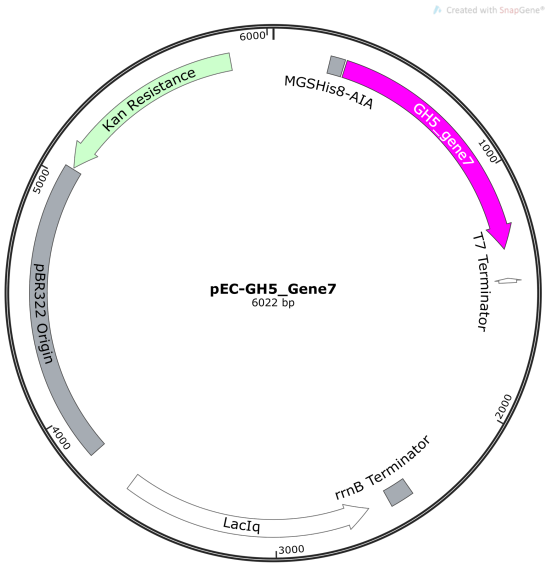

Figure 4: SnapGene representation of the GH5 Gene7 plasmid

2.1.1 Materials and methods

For this experiment, 1 µg of the GH5 gene fused with an 8x-Histidine tag at the N-terminus of the sequence in the pEU vector and the plasmid in pEC (pEC-CelE) vector was used which also has a high plasmid copy number. 1 µl of restriction enzymes that cut at the restriction sites AsiSI and BamHI were added. The reaction was carried out in the presence of 4 µl of 1X cutsmart buffer (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM Magnesium acetate, 100 µg/ml BSA, pH 7.9 at 25°C). The total volume of the reaction was made up of 40 µl using PCR water. Digestion was carried out for an hour at 37°C. The digested mixture was then run through a 0.7% agarose gel for 45 min at 120V to isolate the insert and the vector. The insert and vector were extracted from the agarose gel and purified using the IBI scientific Gel Extraction kit. 1 µl of T4 DNA ligase was used to ligate insert and vector in various vector to insert ratios (1:1, 1:2, and 1:3) to check for optimal ligation. Around 50 ng of the vector was used for the ligation in 2 µl T4 DNA Ligase buffer. The total reaction volume is made up of 20 µl and was incubated for 16 hours at 16°C. The ligated constructs were transformed in 50 µl of E.cloni Z-competent cells by heat shock for 45 seconds at 42°C and then transferred to 250 µl of SOB media. 100 µl of ligation mixture was plated on LB agar plates with the appropriate antibiotic after the mixture was incubated for 2hours at 37°C. The LB agar plated with the streaked ligation mixtures were left in the oven for 14-16 hours at 37°C.

For each of the constructs, colony screening was performed to check for successful ligation. Multiple colonies were picked using a sterile pipette tip and streaked on a fresh LB agar plate. The plate was collected after streaks were grown for 16 hours at 37°C. Using a sterile pipette tip, a small part of the streak colony it picked and dissolved in PCR

water by thoroughly mixing it. Forward and reverse primers were added along with 2x Phusion master mix the colony mix, and a PCR is performed. Gel electrophoresis on the PCR mixtures confirmed if the ligation for the selected colonies were successful.

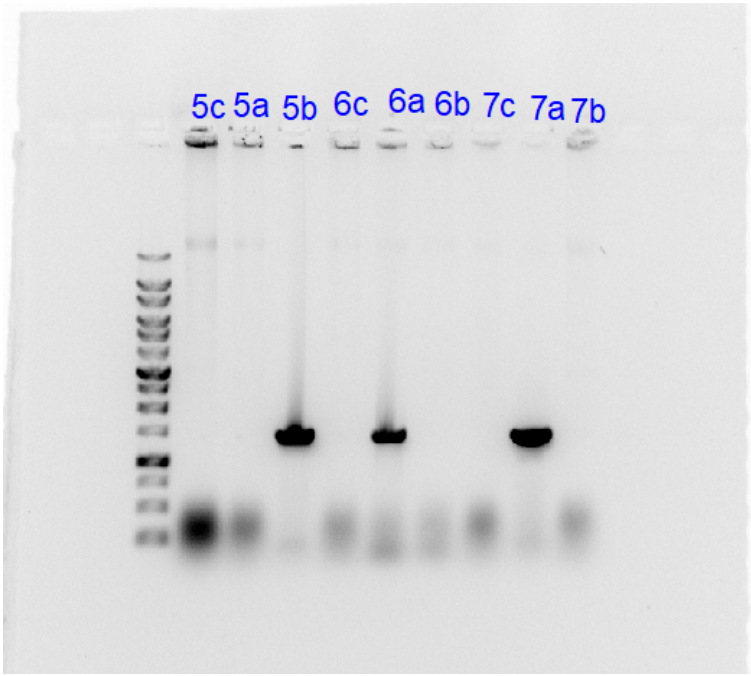

Figure 5: Example of Colony screening for GH5 gene5 (5a, 5b,5c), gene6 (6a, 6b, 6c) and gene7 (7a, 7b, 7c) where a, b and c are different colonies of the same construct

DNA was extracted using the IBI scientific Plasmid extraction kit and sequenced by Genscript to confirm the sequences using NcoI forward and T7 reverse primers. Once the sequences were confirmed (Appendix A). The concentration of the DNAs was measured, and the plasmid was transformed into Ecloni, RG2, and BL21 competent cells for protein expression and purification.

2.2 GH5-CBM3a Gene Constructs

The other cloning method that was used for the generation of the pEC-GH5-CBM3a constructs was Polymerase Incomplete Primer Extension (PIPE) Cloning Method along with sequence- and ligation-independent cloning (SLIC) [11]. The constructs were generated by PIPE/SLIC, as RE cloning was not successful. The primers were designed (Appendix B) for each of the 15 wild type GH5 enzymes. The vector used was pEC-CelE-CBM3a.

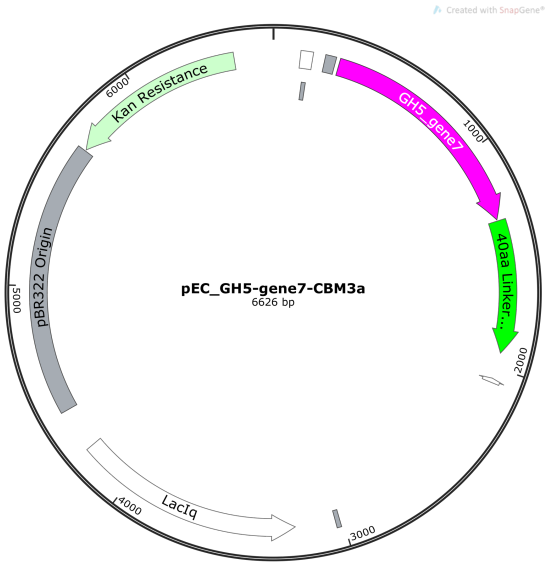

Figure 6: SnapGene representation of the GH5 Gene7 with CBM3a plasmid

2.2.1 Materials and Methods

First step of the PIPE was to extract the require DNA segment and amplify the product. A PCR reaction is set up for in PCR tubes for the primer extension step. The PCR reaction for the vector and the insert contained 20ng for the DNA, along with 10 µM of forward and reverse primer each. The 2X master mix is used as the buffer, and the total volume

was made up to 20 µl using PCR water. The setup was kept on an ice bath to ensure there is no degradation of the product. The reaction mixture was then placed in a thermocycler. The cycle was as follows. The initial denaturation of the DNA was done at 98°Cfor 30 seconds. Next, the primers were annealed to the DNA fragments for 30sec at the respective primer melting temperatures (Tm). The extension time for the DNA was calculated based on the number of base pairs in the DNA. The final extension time was 5mins for 72 °C to ensure that the complete DNA was formed. Once the reaction was complete, a small amount of the PCR product was taken for gel electrophoresis to confirm PCR products. The amplified DNA fragment was separated and identified using a 0.7% agarose gel and ethidium bromide staining, which is visible under the UV light. The confirmed PCR product was purified or cleaned using the IBI scientific kit. Simultaneously, the digestion of the products was performed to help remove the methylation of the DNA strand. The concentration for the purified product was measured on Molecular Device Spectramax M5E microplate at 260 and 280 nm. According to the weight of the product, the insert and vector were mixed in specific ratios for ligation.

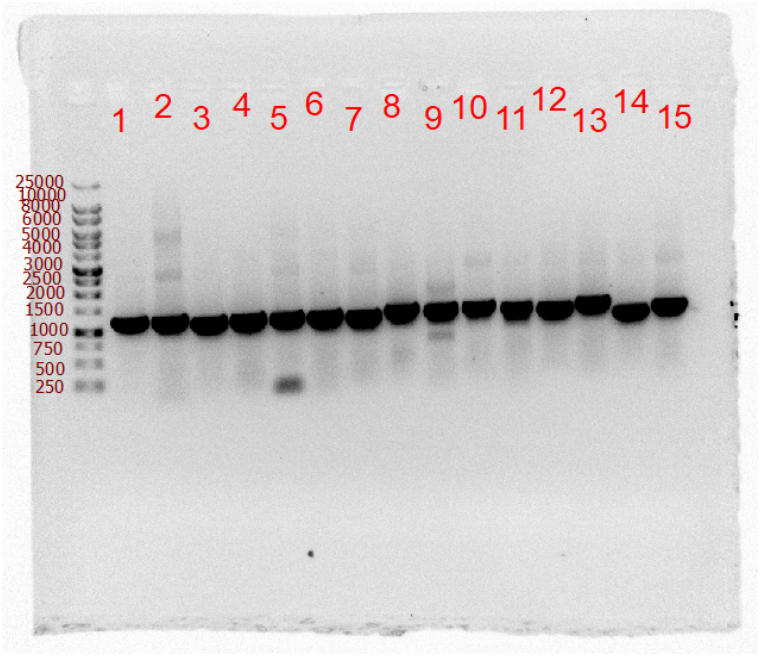

Figure 7: Example of Gel Electrophoresis for the Inverts for the gene1 (1) to gene15 (15)

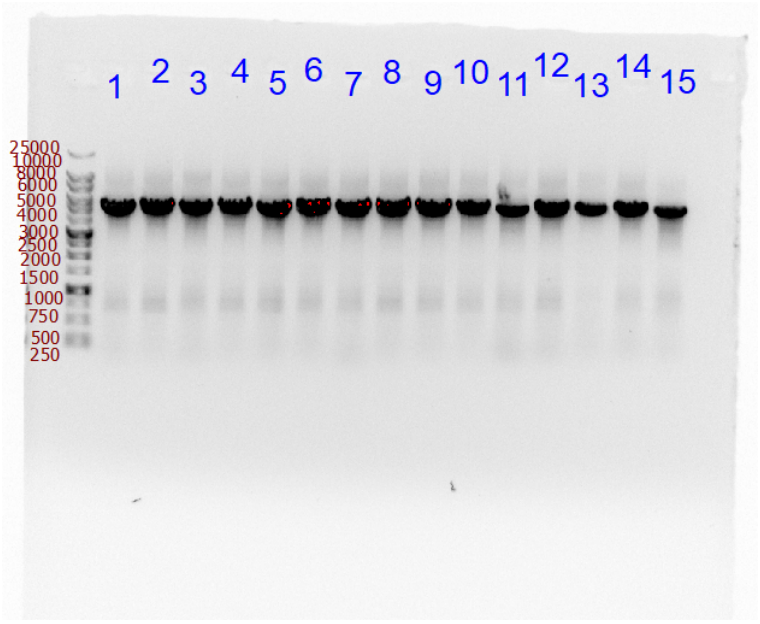

Figure 8: Example of Gel Electrophoresis for the Vectors (pEC_CelE_CBM3a) for gene1 (1) to gene15 (15)

SLIC uses T4 DNA polymerase, an exonuclease, to generate the overhangs for the insert and the vector and ligates the two. 0.5 µl ofT4 DNA polymerase was added to the insert and vector in the presence of the 2 µl NEB 2.1 buffer to help with the correct binding of the insert and vector of the plasmid. Incubate the tubes at 25°C for 5 min. The ligated products were then transformed into ECloni cells using the heat shock method and then incubated for 2 hours at 37°C in SOC. The transformation mixture was plated on LB agar plates that have the antibiotic that the vector is resistant to ensure a specific product. The smeared plates are left for 16-18hours in the over at 37°C. Once the plate was recovered the next day, a few of the colonies were chosen for plasmid extraction. The picked colonies were inoculated in the same volume of LB media and grown overnight for about 16hours. The plasmid extraction was performed using the IBI scientific kit for miniprep. The extracted plasmid was sequenced to confirm the correct enzyme sequence. The correct DNA was measured was then transformed into BL21 and RG2 cells.

2.3 GH5 Gene and GH5-CBM3a Gene Mutants

Further, the wild type enzymes were to be mutated at the catalytic nucleophile to alanine, glycine, and serine. As most of the enzymes did not have solved structures, the catalytic nucleophile and the acid/base were unknown to perform the mutations. Thus, a small bioinformatics study was performed to identify the catalytic nucleophile. All the existing GH5 subfamily 4 enzymes were downloaded. A couple of solved structures were identified in a lot of the proteins. For these proteins, the nucleophile and acid/base were annotated. All the enzymes were aligned on Geneious software using the MAFFT alignment.

From the alignment, we could observe the conserved residues of the enzymes. The catalytic nucleophile and the acid/base should be the conserved residues as those residues are part of the active site. Thus, the nucleophiles were identified. The primers were designed (Appendix C) on Geneious software and then sent to IDT. The site-directed mutagenesis PCR was then set up to create the mutants of the respective enzyme and the enzyme CBM3a constructs. The PCR thermocycler settings were run for 25 cycles.

PRC Step		Temp	Duration (mins)
1	Initial denaturation	98°C	30
2	Denaturation	98°C	10
3	Annealing	Tm	30
4	Extension	72°C	210
5	Final Extension	72°C	300
6	Hold	10°C	∞

Table 2: PCR Thermocycler Settings

The PCR product completion was checked by gel electrophoresis. The product was then cleaned up using the IBI scientific kit. The concentration of the DNA was measured using the spectra drop instrument and stored at -80 °C.

2.4 Conclusion

The GH5 wild type genes were effectively transferred from the pEU vector to the pEC Vector. The GH5-CBM3a constructs were successfully generated as well using the PIPE/SLIC method. The serine mutations for the GH5-CBM3a for gene2 and gene3 were unsuccessful and were skipped from further experiments. The remaining DNA was appropriate, labeled, and stored.

CHAPTER 3. PROTEIN EXPRESSION AND OPTIMIZATION

3.1 Small Scale Expression for Optimization

Now that the constructs were generated, the protein expression conditions had to be determined. Since most of the proteins were unsolved structures, the expression conditions had to be established. Additionally, the substrate on which they were active was not established. Thus, crude activity assays with various substrates were performed. The expression conditions were optimized by changing the pH of the activity assay to ensure that we have maximum activity.

3.1.1 Materials and Methods

The small-scale protein expression was performed in sterile 96 deep well plates. Expression conditions were verified and optimized by permutation and combinations of cells, media, and temperature. The following table shows the setup.

Experiment set	Media	Cells	Temperature
1	**LB Media + IPTG**	**BL21**	**16 °C**
2			**25 °C**
3		**RG2**	**16 °C**
4			**25 °C**
5	**Auto-induction Media (TB+g)**	**BL21**	**16 °C**
6			**25 °C**
7		**RG2**	**16 °C**
8			**25 °C**

Table 3: Experimental Setups for growth condition optimization by varying the media, competent expression cells and expression temperatures

For the 1 ml small-scale protein expression of the wild type enzymes, a starter culture 5% of the total expression volume was prepared in LB media. The starter culture, which was kept overnight for 16 hours at 37°C, was transferred into LB media and TB+g with the necessary antibiotics. The LB media cultures were grown at 37°C till they reached an OD of 0.4-0.8. The protein expression for the LB media was induced by added IPTG, and then culture kept at 25°C and 16°C. For the TB+g, the protein expression was auto induced after all the glucose was consumed. This was estimated to be after 6 hours from

transferring the starter culture into the larger culture. The temperature of the larger cultures was lowered to 25°C and 16°C. The cultures that were kept at 25°C were collected after 24 hours, and the 16°C were collected after 48hours. The grown cultures were centrifuged at 3900 rpm for 15 mins at 4°C. The supernatant was removed, and the cell pellets at the bottoms were collected. The cell pellets were resuspended in DI water and centrifuged again to remove any remaining salts or media. After discarding the supernatant, the cell pellets were lysed using Bacterial Protein Extraction Reagent (B-PER). The bottom of the deep well plates is rubbed to give it a mechanical shock. The lysed pellets were centrifuged again to ensure that the cell debris was collected at the bottom, and the protein is suspended in the supernatant. The supernatant was collected and stored for further analysis.

3.2 Crude Activity Assays for Substrate Identification

Further, for substrate identification, we performed activity assays using pNP-cellobiose, pNP-xylose, pNP-mannose. 50 µl of 4mM substrate was added to 30 µl of the protein from the small-scale expression. The plate was placed on the thermomixer for 16 hours at 40°C. The reason 40°C was considered has the activity assay temperature as most of the proteins were found in the bacteria present in the gut or rumen of yak. The activity was arrest by added 70 µl of 0.1M NaOH, and the pNP release for the activity assay was recorded.

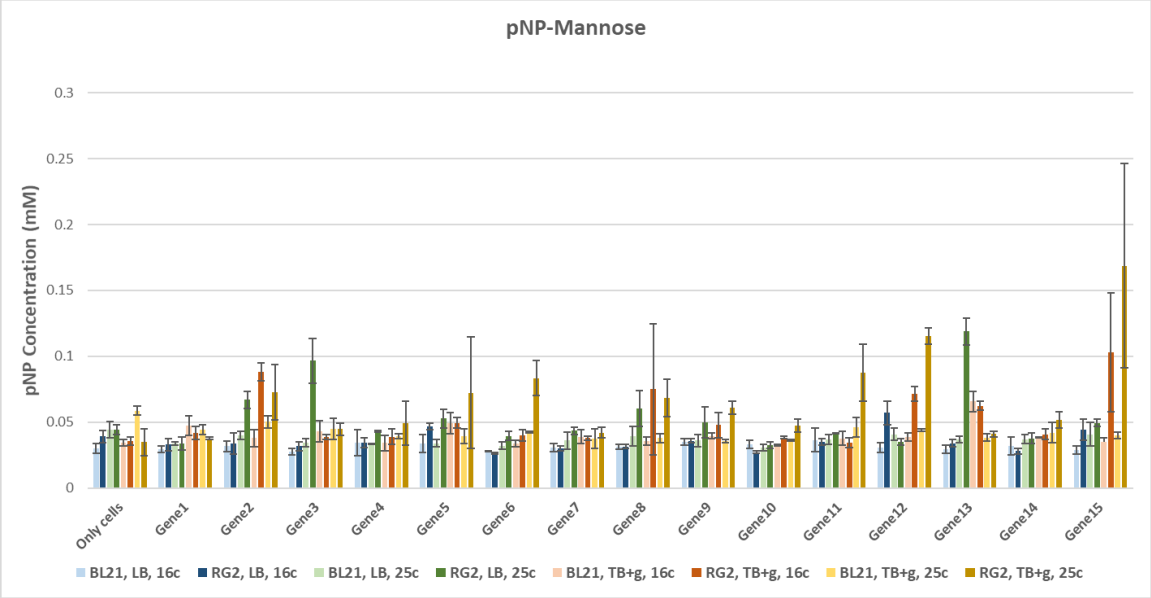

Figure 9: Crude Activity Assay for varying expression conditions using pNP-Mannose as substrate for all the 15 wild type enzymes without CBM

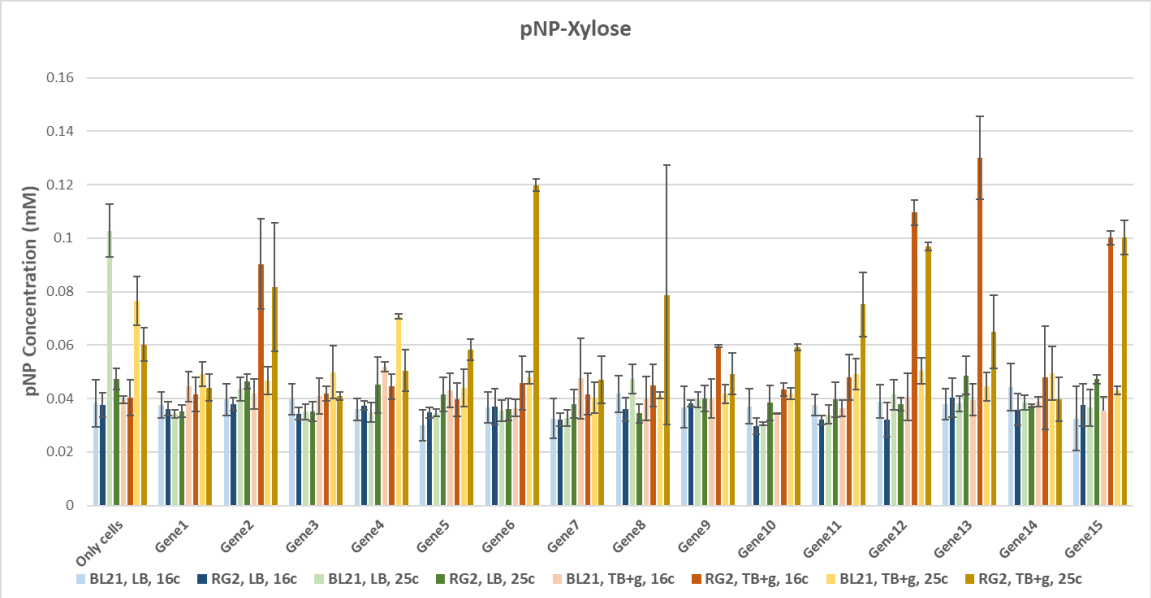

Figure 10: Crude Activity Assay for varying expression conditions using pNP-Xylose as substrate for all the 15 wild type enzymes without CBM

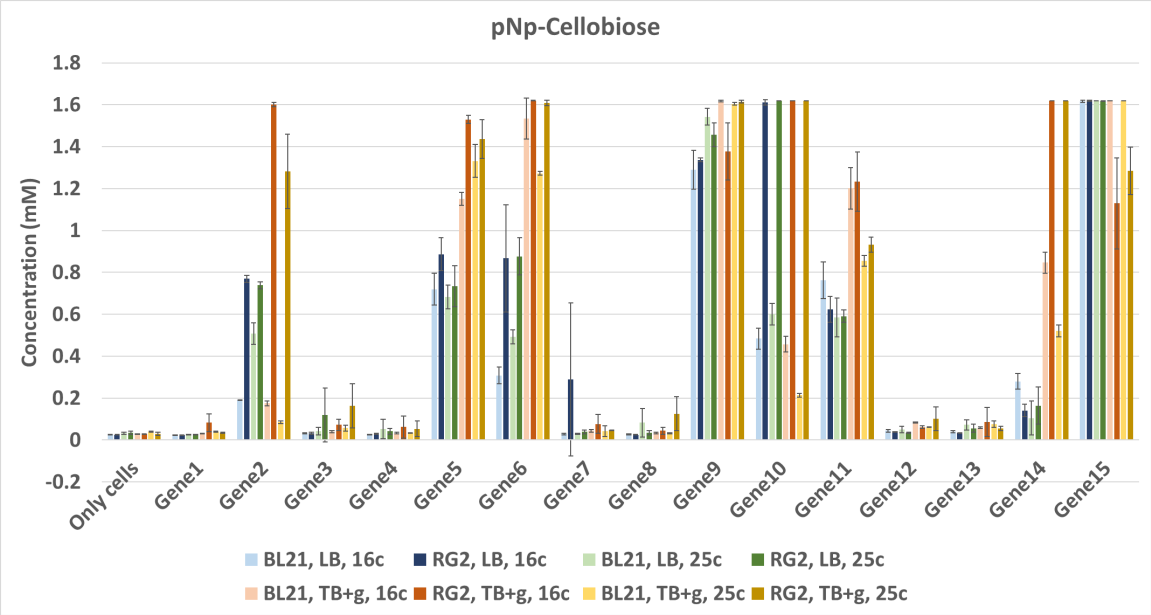

Figure 11: Crude Activity Assay for varying expression conditions using pNP-Cellobiose as substrate for all the 15 wild type enzymes without CBM

The activity results for the wild type enzymes showed that most of the enzymes had significant activity on the pNP-cellobiose substrate as compared to the other substrates. The max pNP release from the pNP-Mannose activity assay was about 0.2 mM. For the pNP-xylose, the maximum activity observed was close to 0.13 mM. From the pNP-cellobiose graph, we observed that, on average, the pNP released on an average was higher than the previous two substrates.

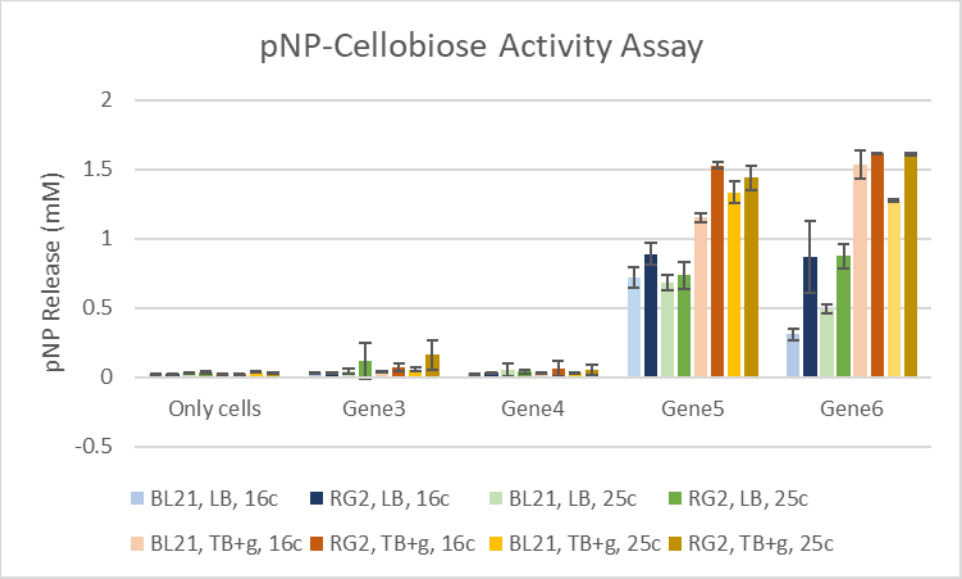

Figure 12: Closer overlook of pNP-Cellobiose for the varying expression conditions

A closer look at the pNP-Cellobiose activity assay shows us the comparison of the expression condition for the culturing. Gene4 shows close to no activity while gene3 expresses some protein in RG2, TB+g, and 25 °C condition. From Gene5 and gene6, the proteins were expressed the best in RG2 cell and TB+ media at 16°C and 25 °C. Comparing all these expression conditions, the proteins seemed to show activity in RG2, TB+g, and 16 °C.

However, there were a few proteins that still did not show significant activity for any of the conditions in which they were grown to express. Thus, the activity assay with pNP-glucose as a substrate was also performed to verify if these proteins depicted activity. Nevertheless, there was no significant activity, as seen by the graph that was observed, and thus cellobiose was used as the primary substrate for further studies.

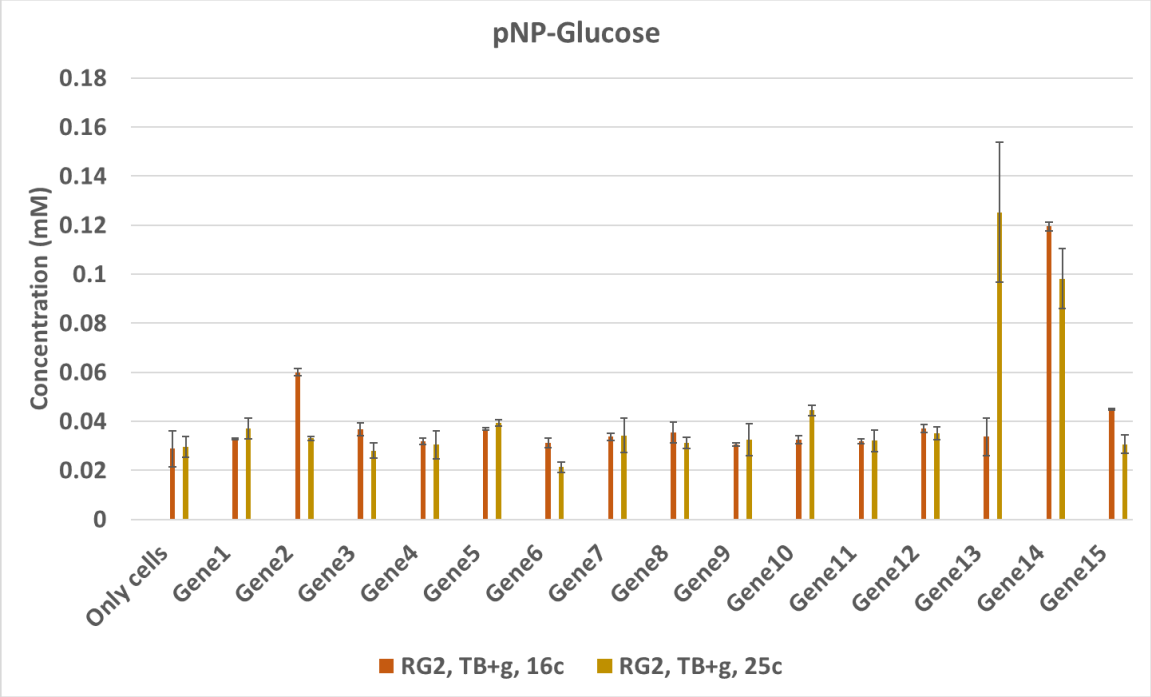

Figure 13: Crude Activity Assay for varying expression conditions using pNP-Glucose as substrate for all the 15 wild type enzymes without CBM

The pNP-cellobiose activity graph was used as a basis to establish the expression conditions as it gave us the best activity. We observed that for the varying cell, media, and temperature conditions on average, RG2 cells with TB+g media at 25°C was the best expression condition for the wild type enzyme without the CBM3a.

3.3 Dot Blot

Now to ensure proper protein expression was occurring, a dot blot was performed for the chosen optimal conditions of RG2 cells with TB+g media at 25°C. Dot blot is a technique that detects proteins. This was done by directly applying the proteins on a nitrocellulose membrane. Initial, the membrane was prepared by applying a blocking buffer that contains BSA. The BSA ensured that there is no non-specific binding of the proteins on the membrane. The proteins were then immobilized on the membrane. Once the protein

was bound to the membrane, the primary antibody was added. The primary antibody had anti histag that binds to the histidine tag on the protein. Soon after, the secondary antibody was added and incubated. The secondary antibody bound to the primary antibody. The secondary antibody carried the tag that allows visualization of the protein. The image of the membrane was acquired. The intensity of the dots represents the amount of the protein that was expressed.

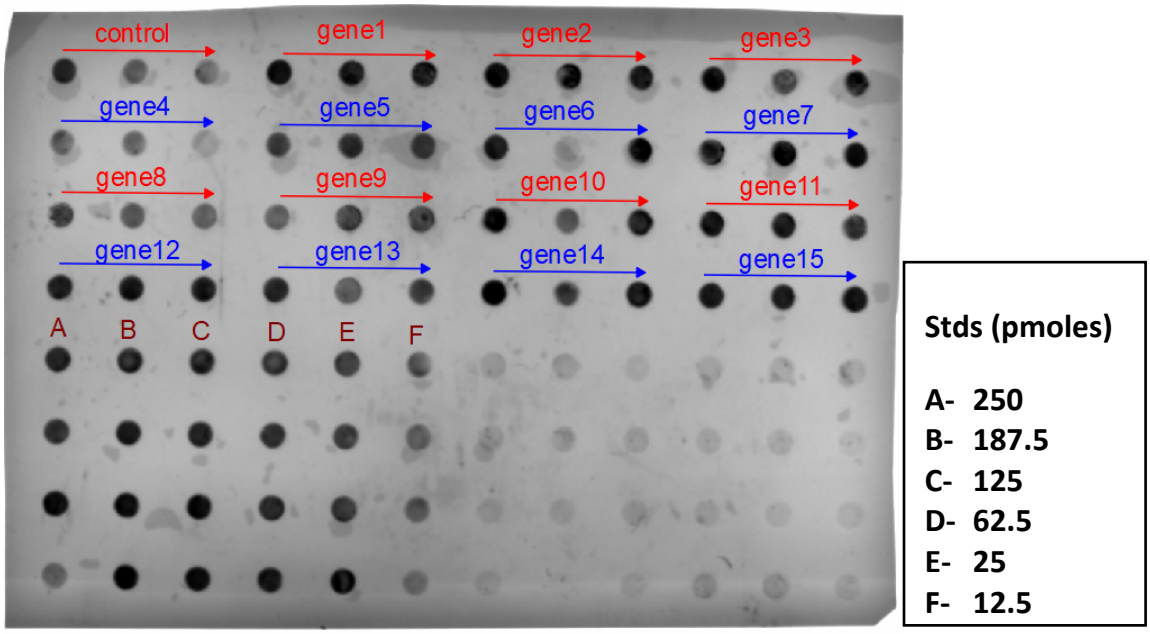

Figure 14: Dot Blot image for proteins expression of the wild type enzymes without CBM at RG2 and TB+g, 25 °C

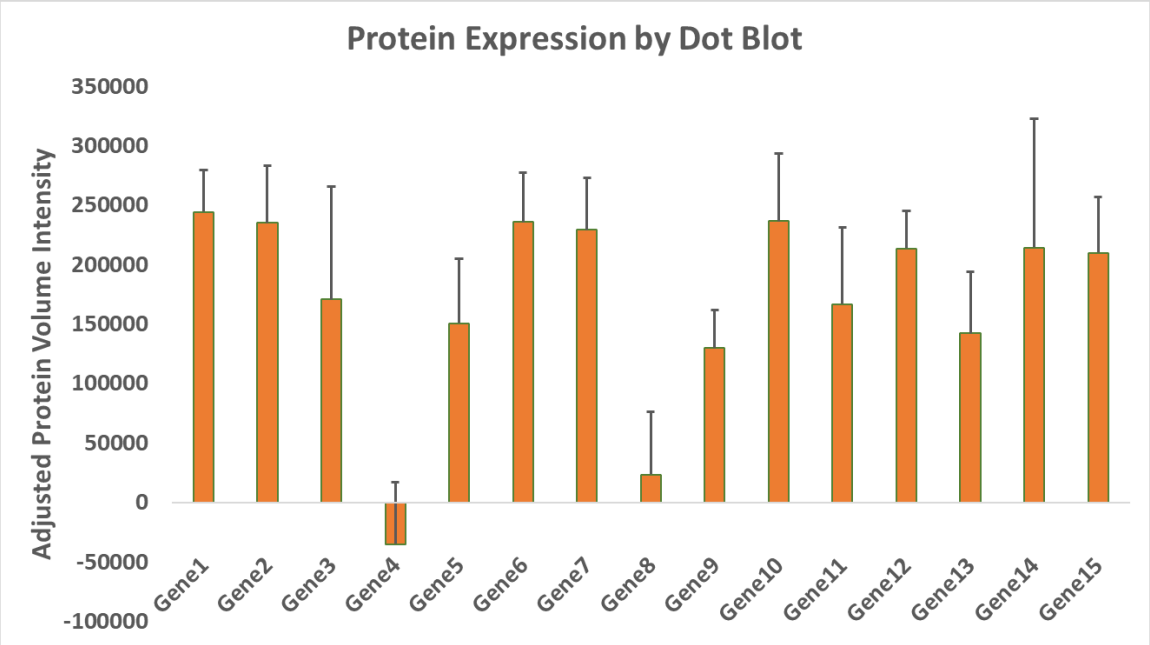

Figure 15: Adjusted Protein Volume intensity for the Dot Blot image for the 15 wild type enzymes without CBM

From the dot blot results and the activity assay results, we noticed that the wild type for gene4 and gene8 had no activity at all as well has no protein expression. From further analysis from the NCIB data bank, we observed that the enzymes were obtained from mammalian cells and probably did not express well in the bacterial system. Thus, the two proteins were dropped from further large-scale expression and purification.

3.4 Purifications of Proteins

Once the expression conditions were established, the genes that were transformed into RG2 cells were grown in from the large-scale protein expression. The large-scale cultures were harvested, and the cell pellets were obtained for protein purification. Affinity based chromatography was used to purify the proteins.

3.4.1 Materials and Methods

The starter culture prepared was 5% of the larger culture. The starter culture was inoculated with the respective transformed RG2 cells in LB media were prepared the night before the larger-scale protein expression. The starter cultures were grown for 14-16hours and then transferred to a 100ml TB+g media with the respective antibiotics. The larger cultures were incubated at 37°C for 6hours, and then the temperature was reduced to 25°C. TB+g is an auto-induction media. Once the glucose in the media was wholly consumed, the lac operon kicks in as the cell must break down lactose, and thus protein expression starts. The time taken to consume all the glucose in the media was around 6 hours. The temperature was reduced to 25°C to ensure that the protein that was expressed folded appropriately. The cultures are incubated for 24hours before the cells were harvested. The cultures were centrifuged for 15 mins at 10000 rpm, and the cell pellet was collected and stored at -80°C for further purifications.

In the case of the GH5 CBM3a genes, the amount of cell pellet that was collected from a 100ml culture was not significant enough. Thus, the 500ml cultures were grown for these constructs. The protein gel images after electrophoresis showed significant cleaving of the protein after purification. This could be due to the improper folding and expression of the proteins as the CBM3a domain is large. Thus, the temperature during protein expression was reduced to 16°C to ensure that we would not observe further cleaving. For this temperature, the cell pellets were harvested 48 hours later and preserved at -80 °C.

The collected cell pellet was lysed in cell lysis buffer by adding 5 µl lysozyme for 1 gm of pellet and 67 µl protease inhibitor cocktail (1 µM E-64 Sigma Aldrich E3132, 0.5 mM Benzamidine and 1 mM EDTA) of the pellet. Protease hydrolyze the peptide bond between amino acids in a polypeptide chain of protein, thus cleaving the proteins. Protease inhibitor cocktail has a combination of chemicals that, when added, inhibits the activity of the protease and ensures that the protease does not degrade the protein while the cell is being lysed. To speed up the lysis process, the cell pellet was sonicated (Misonix Sonicator 3000, 1.6 mm diameter microtip). The sonicated mixture was then centrifuged at 10000rpm at 4°C for an hour. The cell debris was accumulated at the bottom of the tube while the protein was suspended in the supernatant. The supernatant was then transferred into another falcon tube. The protein was purified using affinity-based chromatography. Since we were dealing with many proteins, a high throughput method was utilizing the beads was used. For this method, nickel-iron beads were used with a magnetic rack to separate the tagged proteins present in the supernatant. The beads were initially washed with IMAC A (100 mM MOPS, 500 mM NaCl, 10 mM Imidazole, pH 7.4) buffer to equilibrate the beads to ensure the binding of the his-tags to the beads. The cell lysate was then incubated with the beads at 4°C on the shaker incubator for the tagged proteins to bind to the beads. The beads were then washed with 95% IMAC A and 5% IMAC B (100 mM MOPS, 500 mM NaCl, 500 mM Imidazole, pH 7.4) solution to remove the non-specific protein that would have bound to the beads along with the specific protein. The protein of interest was then eluted with IMAC B, which has more imidazole that helps detach the his-tagged protein from the beads as it has more affinity to the bead as compared to the tag. Purified his-tagged proteins were desalted into 10 mM

MES, pH 6.5, and purity was characterized using SDS-PAGE. Figure 16 represents protein gel for a few of the proteins that were purified. In the figures, 1A signifies mutants for GH5 gene1 with an alanine mutant at the catalytic nucleophile.

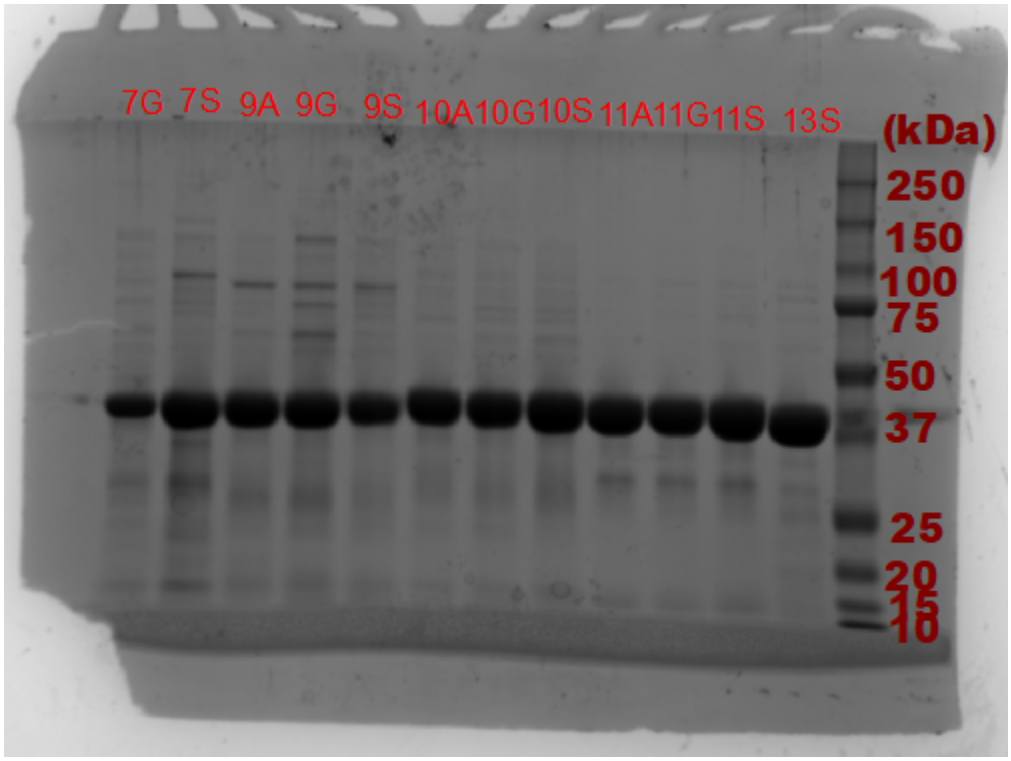

Figure 16: Sample Protein Gel image of the mutants for the few GH5 genes where the number stands for the corresponding gene number and the A, S, G stands for the mutated nucleophile to alanine, serine and glycine respectively

Name	Conc (mg/ml)
GH5_gene1_wt	0.68
GH5_gene1_E325A	0.44
GH5_gene1_E325G	0.32
GH5_gene1_E325S	0.33
GH5_gene2_wt	7.48
GH5_gene2_E324A	0.34
GH5_gene2_E324G	0.97
GH5_gene2_E324S	1.46
GH5_gene3_wt	4.63
GH5_gene3_E642A	1.06
GH5_gene3_E642G	1.65
GH5_gene3_E642S	2.98
GH5_gene5_wt	3.16
GH5_gene5_E310A	1.21
GH5_gene5_E310G	1.51
GH5_gene5_E310S	1.85
GH5_gene6_wt	3.05
GH5_gene6_E362A	0.21
GH5_gene6_E362G	0.35
GH5_gene6_E362S	0.88
GH5_gene7_wt	4.45
GH5_gene7_E643A	0.40
GH5_gene7_E643G	1.52
GH5_gene7_E643S	4.94
GH5_gene9_wt	3.20
GH5_gene9_E326A	6.99
GH5_gene9_E326G	5.05
GH5_gene9_E326S	5.05

Name	Conc (mg/ml)
GH5_gene10_wt	1.93
GH5_gene10_E313A	3.78
GH5_gene10_E313G	2.80
GH5_gene10_E313S	2.26
GH5_gene11_wt	3.14
GH5_gene11_E333A	5.01
GH5_gene11_E333G	4.69
GH5_gene11_E333S	3.67
GH5_gene12_wt	0.17
GH5_gene12_E303A	0.78
GH5_gene12_E303G	0.67
GH5_gene12_E303S	0.82
GH5_gene13_wt	0.78
GH5_gene13_E323A	0.03
GH5_gene13_E323G	4.19
GH5_gene13_E323S	3.10
GH5_gene14_wt	3.96
GH5_gene14_D263A	2.57
GH5_gene14_D263G	3.16
GH5_gene14_D263S	2.71
GH5_gene15_wt	0.39
GH5_gene15_E274A	0.16
GH5_gene15_E274G	0.38
GH5_gene15_E274S	1.23

Table 4: Protein Concentrations for the GH5 genes

For the proteins with the CBM3a domain, from the purification of the protein using the magnetic beads, we still observed significant cleaving. Thus, for the proteins which showed cleaving, a second step purification was performed using hydrophobic interaction

chromatography. In this technique, the proteins are separated based on hydrophobic properties. The HIC was performed using the NGS bio rad equipment. In this method, the HIC column has resin that binds based on the hydrophobic interaction of the two domains present from the elution of the first step purification. First, the proteins were eluted into to HIC starter buffer using buffer exchange (desalting) columns. Using the FPLC, the column was first saturated with HIC starter buffer, and then the protein was loaded. Once the entire protein was loaded, the elution buffer was passed through the column. The FPLC run showed us two peaks. The fractions for the two peaks were collected, and a protein gel was run to determine which peak was our protein of interest and the cleaved domain. From the protein gel image (figure 18), it was clear that the second peak seen in lanes 12 and 13 was our protein of interest, and thus those fractions were then consolidated and desalted into 10 mM MES buffer pH 6.5 and stored at -80°C.

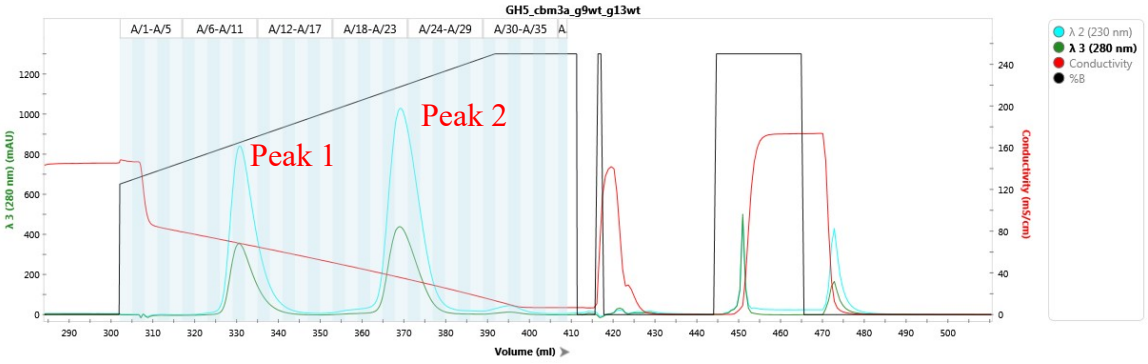

Figure 17: HIC run for the GH5-CBM3a gene13 wt

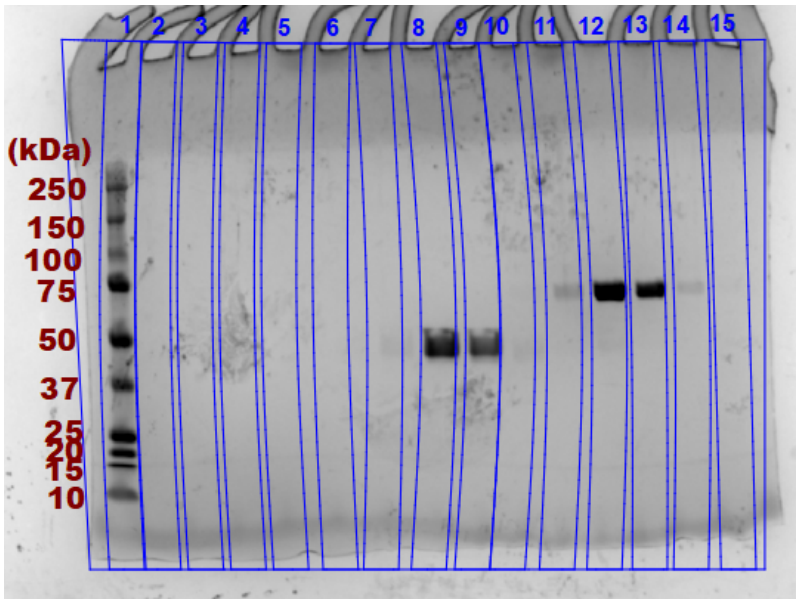

Figure 18: GH5-CBM3a gene13 wt HIC run

Protein Name	Conc mg/ml	IMAC Purification	HIC Purification
GH5_g1_CBM3a_wt	0.77	Yes	No
GH5_g1_CBM3a_E325A	0.33	Yes	No
GH5_g1_CBM3a_E325G	0.21	Yes	No
GH5_g1_CBM3a_E325S	0.26	Yes	No
GH5_g2_CBM3a_wt	0.31	Yes	No
GH5_g2_CBM3a_E324A	0.97	Yes	No
GH5_g2_CBM3a_E324G	1.53	Yes	No
GH5_g2_CBM3a_E324S	-	-	-
GH5_g3_CBM3a_wt	1.39	Yes	No
GH5_g3_CBM3a_E642A	0.54	Yes	No
GH5_g3_CBM3a_E642G	0.60	Yes	No
GH5_g3_CBM3a_E642S	-	-	-
GH5_g5_CBM3a_wt	1.44	Yes	No
GH5_g5_CBM3a_E310A	0.75	Yes	No
GH5_g5_CBM3a_E310G	0.68	Yes	No
GH5_g5_CBM3a_E310S	0.66	Yes	No
GH5_g6_CBM3a_wt	0.63	Yes	No
GH5_g6_CBM3a_E362A	0.61	Yes	Yes
GH5_g6_CBM3a_E362G	0.82	Yes	No

GH5_g6_CBM3a_E362S	1.13	Yes	No
GH5_g7_CBM3a_wt	1.50	Yes	No
GH5_g7_CBM3a_E643A	0.82	Yes	No
GH5_g7_CBM3a_E643G	1.01	Yes	No
GH5_g7_CBM3a_E643S	0.97	Yes	No
GH5_g9_CBM3a_wt	0.91	Yes	Yes
GH5_g9_CBM3a_E326A	0.81	Yes	No
GH5_g9_CBM3a_E326G	1.10	Yes	No
GH5_g9_CBM3a_E326S	0.75	Yes	No
GH5_g10_CBM3a_wt	0.12	Yes	Yes
GH5_g10_CBM3a_E313A	0.61	Yes	No
GH5_g10_CBM3a_E313G	0.55	Yes	No
GH5_g10_CBM3a_E313S	0.91	Yes	No
GH5_g11_CBM3a_wt	1.43	Yes	No
GH5_g11_CBM3a_E333A	1.58	Yes	No
GH5_g11_CBM3a_E333G	0.65	Yes	No
GH5_g11_CBM3a_E333S	1.52	Yes	No
GH5_g12_CBM3a_wt	0.63	Yes	No
GH5_g12_CBM3a_E303A	0.61	Yes	No
GH5_g12_CBM3a_E303G	0.82	Yes	No
GH5_g12_CBM3a_E303S	1.13	Yes	No
GH5_g13_CBM3a_wt	0.21	Yes	Yes
GH5_g13_CBM3a_E323A	3.16	Yes	Yes
GH5_g13_CBM3a_E323G	3.44	Yes	Yes
GH5_g13_CBM3a_E323S	1.82	Yes	No
GH5_g14_CBM3a_wt	1.94	Yes	No
GH5_g14_CBM3a_D263A	1.16	Yes	No
GH5_g14_CBM3a_D263G	1.26	Yes	No
GH5_g14_CBM3a_D263S	1.07	Yes	No
GH5_g15_CBM3a_wt	1.33	Yes	Yes
GH5_g15_CBM3a_E274A	1.81	Yes	Yes
GH5_g15_CBM3a_E274G	0.43	Yes	No
GH5_g15_CBM3a_E274S	0.22	Yes	Yes

Table 5: Protein Concentrations of GH5-CBM3a enzymes

3.5 pH optimization by Activity Assay Analysis

Once the protein was purified, an activity assay was performed used CMC and PASC as the substrates to ensure that the enzymes were active on cellulose. Since all the enzymes were not active on pNP-cellobiose, were tried changing the substrates to confirm that the enzymes were indeed cellulases. The activity assays with these substrates were utilized to optimize the pH conditions as well. Various pH for 10 mM MES was used to optimize the buffer conditions using the DNS assay. The DNS reagent (3,5-Dinitrosalicylic acid) reacts with reducing sugars to form 3-amino-5-nitrosalicyclic acid. This product absorbs light at 540 nm and can be detected using Molecular Devices Spectramax M5E microplate reader. This reaction is a colorimetric and changes the DNS reagent color from yellow to dark red in the presence of glucose sugars upon heating. The assay was set up in a 96 well microplate. For the setup of the assay, 60 µL of DNS stock reagent was added to each reaction well in a microplate along with 30 µL of sugar standard. The plate was sealed and heated using a thermocycler to 95°C for 5 minutes and then cooled to 10°C for 10 minutes. Next, 36 µL of the total reaction volume was transferred to a transparent bottom 96 well plate. Initially, the protein loading for each well was 25 pmoles, and the substrate concentration in each well was 5 g/L. MES buffer of pH 5.5, pH 6.5, and pH 7.5 were considered to determine the optimal pH for protein activity. The buffer well concentration was 50 mM. The temperature at which the activity assay was performed was 40°C. The reaction was diluted along with 160 µL of DI water. The absorbance of each well was read using a Molecular Devices Spectramax M5E microplate reader at 540 nm. A standard curve for varying concentrations of known concentration glucose standards was also generated to estimate the amount of glucose

released from the substrate during reaction for unknown samples. A pNP-cellobiose substrate bases assay was also set up to optimize the pH.

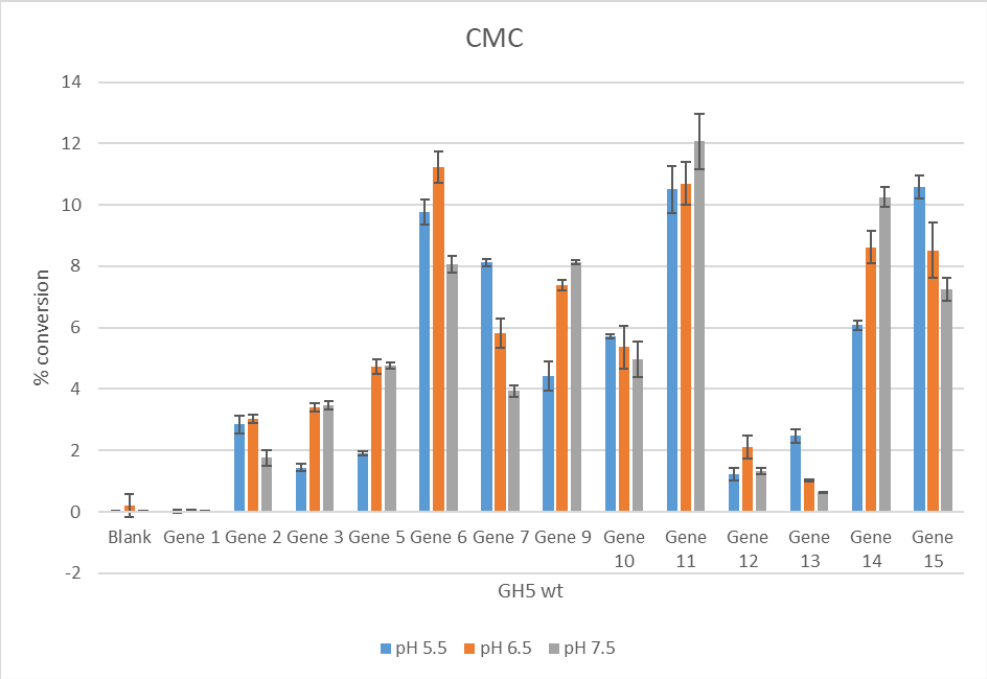

Figure 19: Activity Assay for GH5 wild type enzymes expressed at 25°C in RG2 cells and TB+g media with CMC Substrate

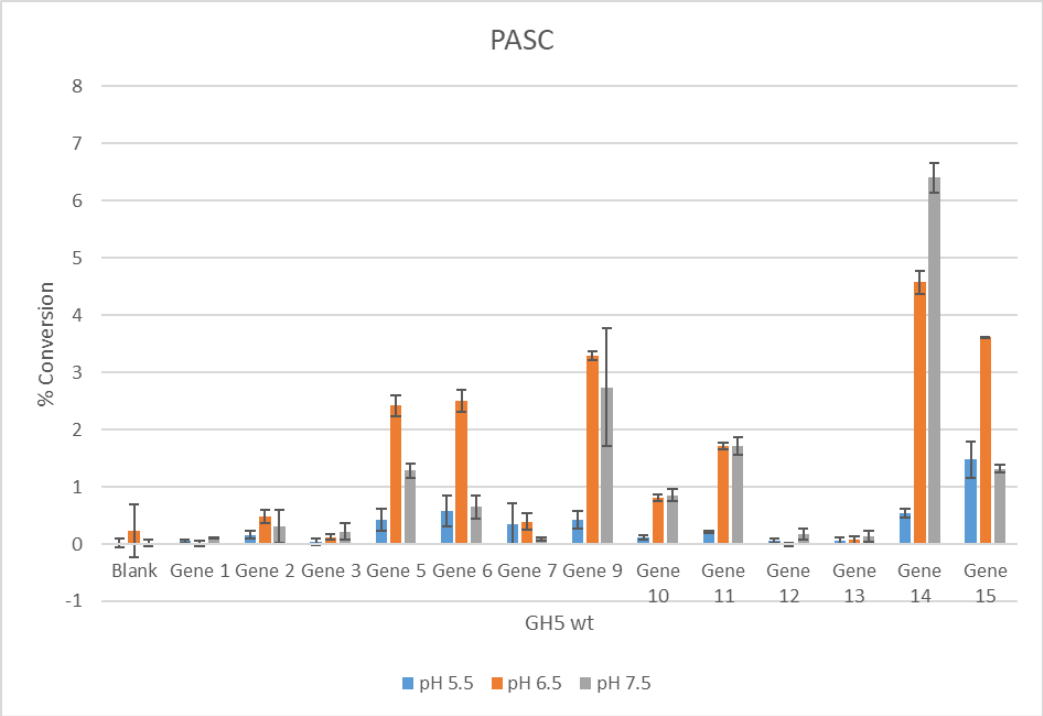

Figure 20: Activity Assay for GH5 wild type enzymes expressed at 25°C in RG2 cells and TB+g media with PASC Substrate

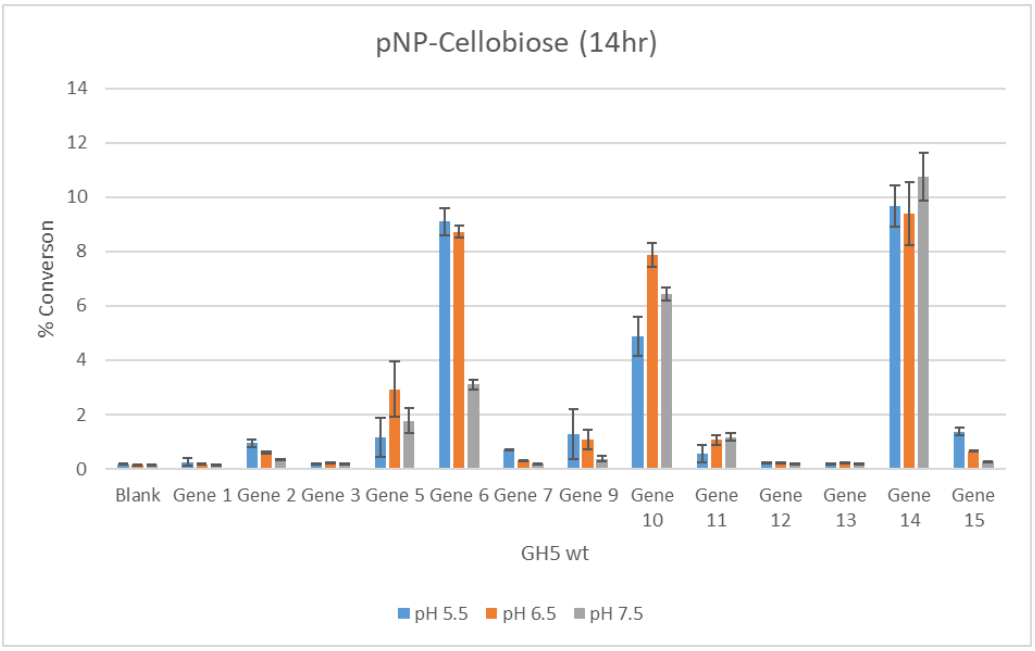

Figure 21: Activity Assay for GH5 wild type enzymes expressed at 25°C in RG2 cells and TB+g media with pNP-Cellobiose for pH optimization

From the graphs, that the proteins were active on the CMC and PASC substrates. This was important as these enzymes were cellulases, and the DNS assay confirmed so. Along with that, we can see that on an average pH 6.5 was a better pH for the activity of the enzymes.

3.6 Conclusion

From the small-scale expressions and the crude activity assay, the best expression conditions were when the proteins were grown in RG2 cells in TB+g media. The proteins for the GH5 without CBM3a constructs expressed well at 16°C. However, for experimental convenience, we considered culturing the proteins at 25 °C as they still expressed well at that temperature. The GH5-CBM3a constructs were initially expressed at 25°C. However, due to the lack of a substantial amount of pellets obtained from harvesting, we decided to grow the proteins at 16 °C. For most of the proteins, the magnetic rack purification using nickel-iron beds was used. For the GH5-CBM3a proteins that showed cleaving, hydrophobic interaction chromatography was used to purify the proteins. The proteins were stored and used for further assays to determine the optimal pH and substrate for the proteins. Since we wanted to study the transglycosylation activity and not hydrolysis, we focused only on the pNP based substrates. The crude activity assays determined that the substrate on which the activity assay showed predominant activity was that of pNP-Cellobiose.

From the DNS assay, the optimal pH for the activity assay was pH 6.5 for most of the proteins. The temperature was set to 40 °C. Thus, by fixing these conditions, further activity assays were conducted.

CHAPTER 4. ACTIVITY ASSAY

Microwell, plate-based activity assays, are set up to study the hydrolysis and potential transglycosylation activity in all the enzymes that have been purified. As mentioned previously, the substrate that enabled us to study the potential, transglycosylation activity was pNP-cellobiose. Thus, activity assays were set up, and this layer chromatography was performed to separate the products and study them. The activity assay was performed for all the generated constructs. For the proper understanding of the kinetics of the activities, we focus on one particular enzyme gene set.

4.1 Activity Assay Setup

The activity assay and data obtained was set up for all the enzymes in sets. Each set consisted of the parent enzyme from the table and its respective mutants. The following is an example of the gene7 set that was set up:

GH5 gene7 (without CBM3a)	**GH5 gene7 with CBM3a**
GH5_gene7_wt	GH5_gene7_CBM3a_wt
GH5_gene7_ala	GH5_gene7_CBM3a_ala
GH5_gene7_gly	GH5_gene7_CBM3a_gly
GH5_gene7_ser	GH5_gene7_CBM3a_ser

Table 6: Activity Assay Setup

From the previous crude trial assays, there was some evaporation detected from the wells that were on the edges. Loss of water could impact the results; thus, the assays were set in PCR tubes, and the total volume of the reaction was set to be 200µl. Thus, the 40µl of

20mM pNP-cellobiose was added to the reaction volume making the effective molar concentration as 4mM. The buffer concentration for the well was 1mM, and the plate well substrate concentration was 4mM. The rest of the reaction volume was made up by DI water. The PCR tubes were then completely sealed and placed in the thermomixer at 40°C. The reaction samples were extracted for the 1 hour, 4 hours, 24 hours, 72 hours, and 120 hours. The reaction was run for120 hours to ensure that equilibrium was achieved. For each of the time points, the pNP release was measured by the Spectramax instrument. A TLC was also performed for each of the time points. 1 µl of the same was spotted on AnalyticalTech Plates along with the standards. The mobile solvent used was ethyl acetate, 2-propanol, acetic acid, and water in the ratio of 3:2:1:1. The samples were placed in a pre-saturated chamber and then removed once the run was complete. The plates were then dried, and a UV image was taken. Once the image was captured, the plates were stained with orcinol and left to dry. The orcinol staining allows us to visualize the product by forming dark violet spots on applying heat to the plate. The plates were then heated, and the image of the plate was captured for analysis.

4.2 TLC Analysis

From the initial TLC plates images, there were apparent trends that were noticeable from each set of the enzymes. In each of the enzyme set, it observed that at least one of the CBM3a constructs with mutation showed either hydrolysis or transglycosylation activity as compared to the respective constructs without CBM3a. There were a few constructs that also showed activity with the without CBM3a. Overall, the products that were formed indicated that some of the enzymes were showing transglycosylation as there was

cellotriose sugar that was formed in the process that was retained even after 5 days. This meant that the formed oligosaccharide was not hydrolyzed future.

Analyzing the TLC plates, there was another trend of activities that was observed by the presence of the CBM3a. In a few of the mutant cases, the only visible product is that of pNP-glucose, glucose, and cellobiose. For some mutants, the products observed were pNP-glucose, cellobiose, and cellotriose. On the other hand, there were a few mutants that produced both glucose and cellotriose in varying amounts along with pNP-glucose and cellobiose. In the case of the constructs without CBM3a that showed activity, the counter construct with the CBM3a showed a different product profile. The following show examples of the trends that we see.

4.2.1 Activity type 1

From the following TLC images of the gene7, gene9, and gene11, we observe almost no activity in the enzymes without the appended CBM3a. However, the counter constructs for the mutants with the CBM3a, various products are formed, which suggests that there is some activity. Some of the CBM3a show transglycosylation like activity where we see the formation of oligosaccharides like cellotriose. Thus, from these results, we can conclude that the CBM3a does assist in enzyme in improving the efficiency of the activity of the enzyme.

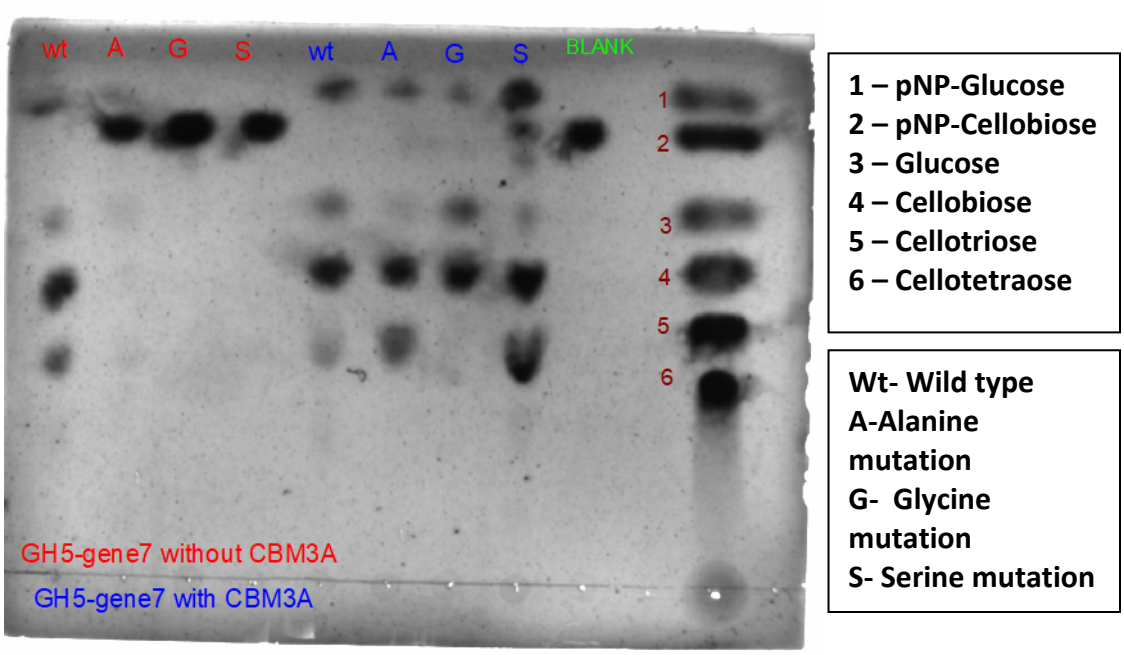

Figure 22: Thin layer chromatography image showing the formation of synthesized products for gene7 set in the presence of pNP-cellobiose

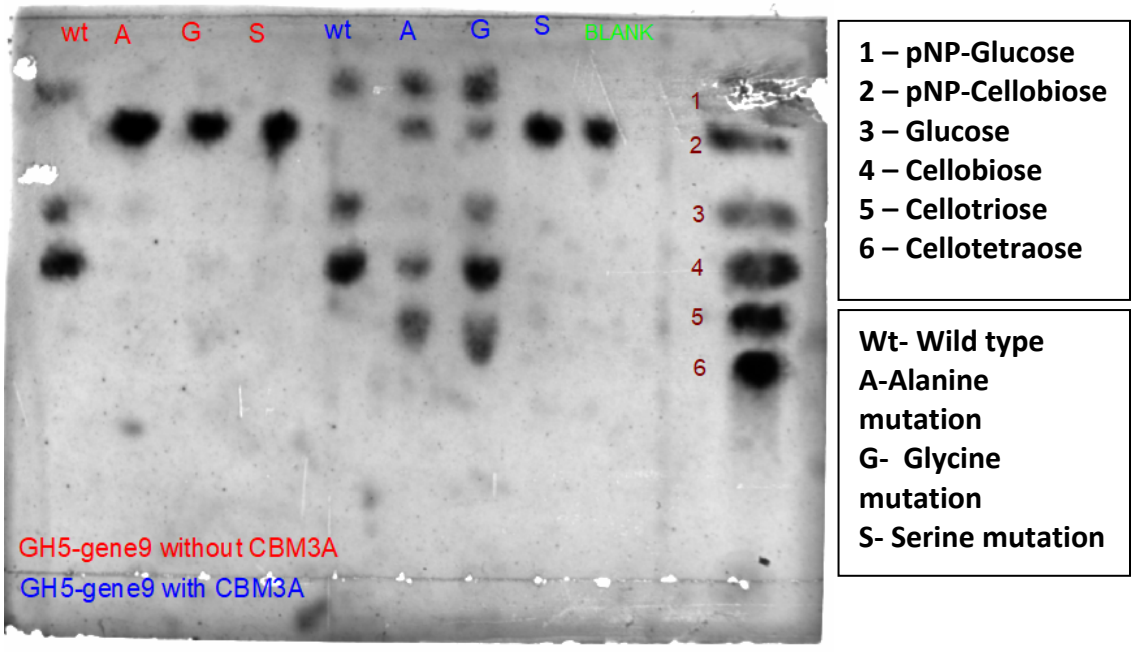

Figure 23: Thin layer chromatography image showing the formation of synthesized products for gene9 set in the presence of pNP-cellobiose

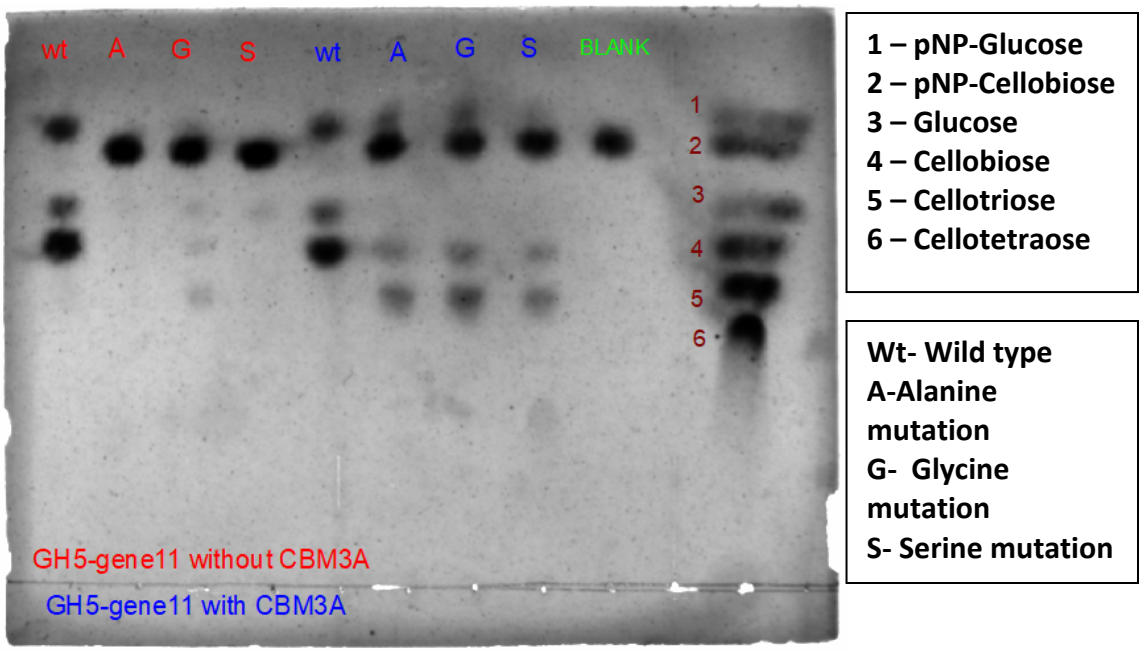

Figure 24: Thin layer chromatography image showing the formation of synthesized products for gene11 set in the presence of pNP-cellobiose

4.2.2 Activity type 2

The analysis of the product profiles for the following gene10, gene12, and gene 15 revealed a different trend that was observed among the following sets of enzymes. In this case, we do see activity from a few mutants without the CBM3a. However, the product profile of these mutants is varied from their counter mutants with CBM3a. For example, in the case of gene10, the GH5-gene10 glycine mutants showed activity, and the products formed were pNP-glucose, glucose, cellobiose, and cellotriose. Meanwhile, the product profile for the glycine mutant of GH5-gene10 with the CBM3a showed that the enzyme had consumed all of the pNP-cellobiose entirely as compared to the previously mention enzyme but did not produce cellotriose along with pNP-glucose, glucose, and cellobiose.

In the case of GH5-gene12 glycine mutants, the product profiles of the enzymes with and without CBM3a are different as there is no pNP-glucose formation for the GH5-gene12 glycine without CBM3a. For the GH5-gene15, the alanine and glycine mutants produce cellotriose but do not form pNP-glucose. While as the correspondingCBM3a mutants had a different product profile. Overall, it appeared like the enzyme activity was altered in the presence of a CBM.

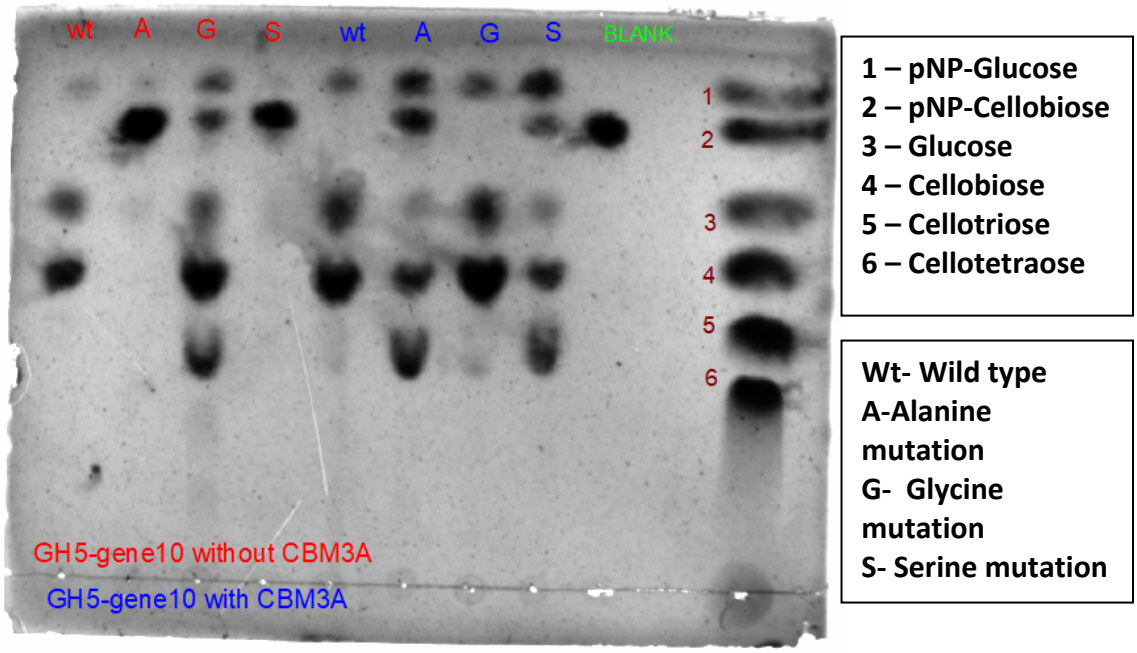

Figure 25: Thin layer chromatography image showing the formation of synthesized products for gene10 set in the presence of pNP-cellobiose

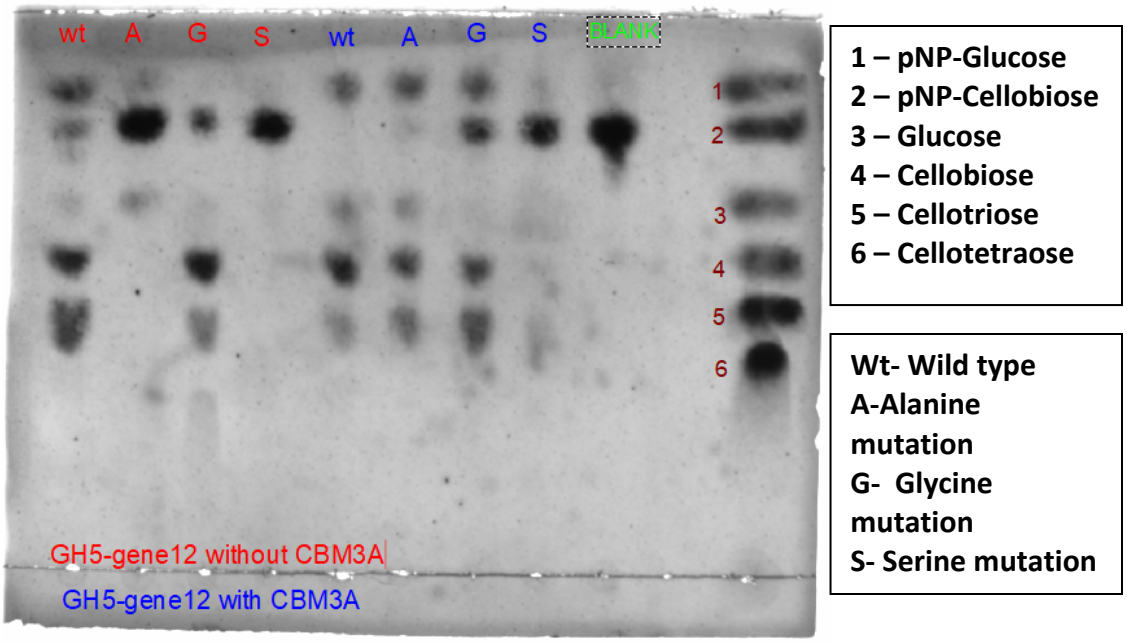

Figure 26: Thin layer chromatography image showing the formation of synthesized products for gene12 set in the presence of pNP-cellobiose

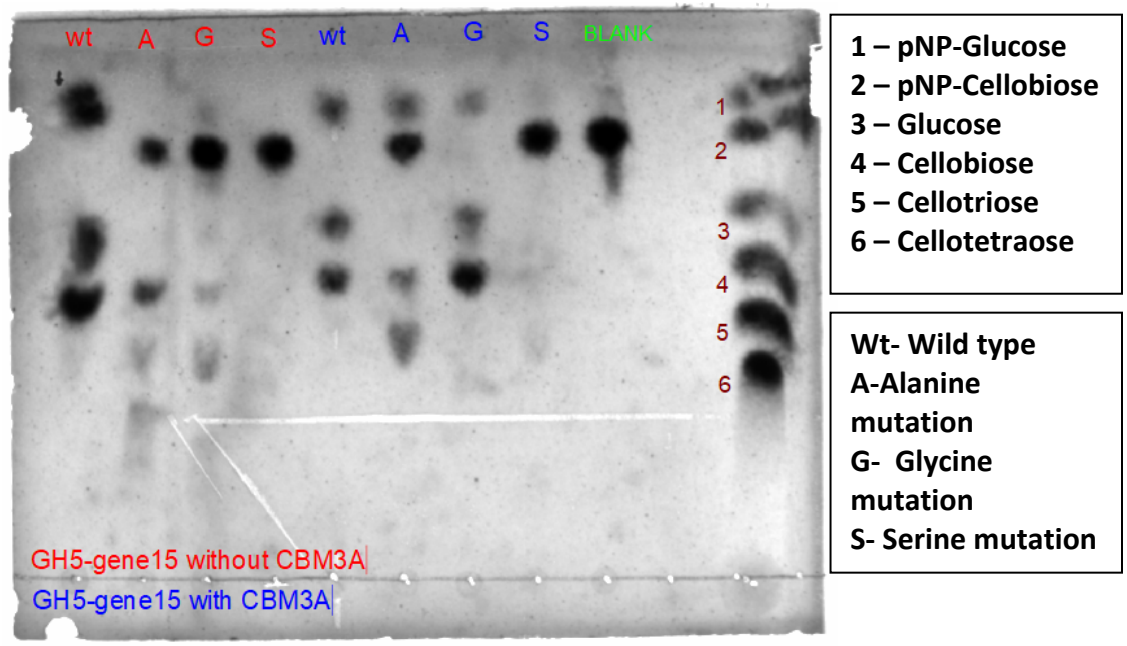

Figure 27: Thin layer chromatography image showing the formation of synthesized products for gene15 set in the presence of pNP-cellobiose

4.2.3 Activity type 3

In the case of GH5-gene2 and gene3, we see that the product profile for alanine and glycine mutants for the enzymes with and without CBM3a are almost similar. Nevertheless, the amount of product formed was slightly varied between the two constructs with the same mutation. For instance, the GH5-gene2 alanine enzyme without the CBM3a mutant profile is imitated by the alanine mutant with CBM3a, but the intensities of the TLC spots are different for the two. Thus, even in these experiments, we see that the occurrence of the CBM3a affects the efficiency of the activity of the enzyme.

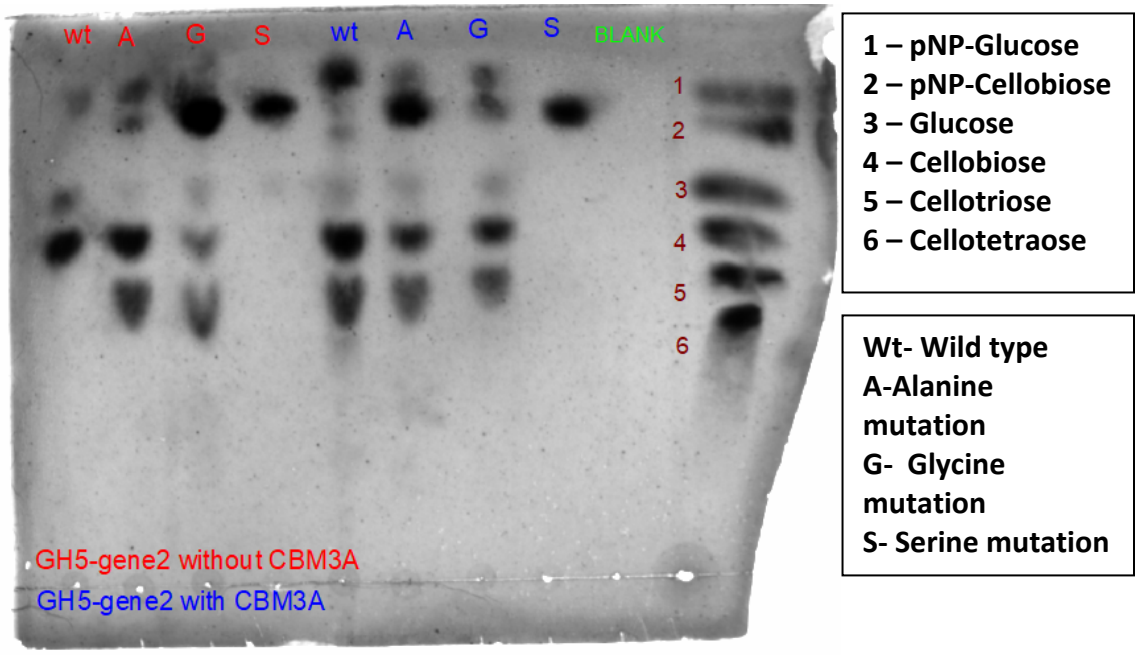

Figure 28: Thin layer chromatography image showing the formation of synthesized products for gene2 set in the presence of pNP-cellobiose

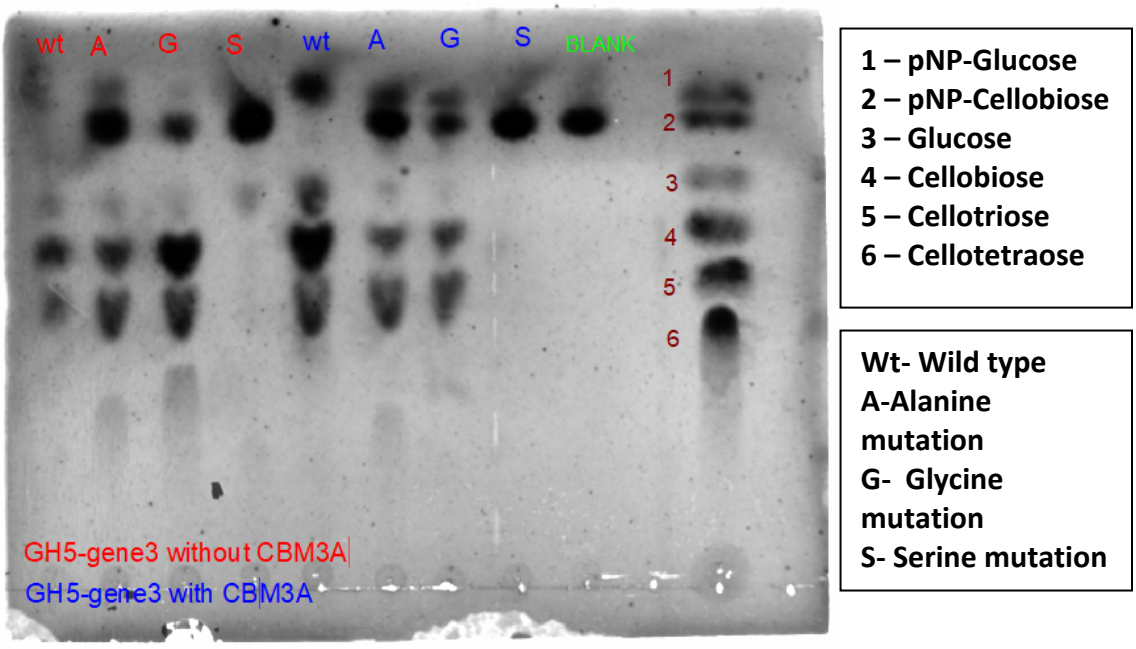

Figure 29: Thin layer chromatography image showing the formation of synthesized products for gene3 set in the presence of pNP-cellobiose

4.2.4 Activity type 4

Lastly, we also observed that the presence of the CBM3a might hinder the activity of the enzyme. In the case of GH5-gene1 mutations and GH5-gene2 alanine mutations. The mutants without the CBM3a consume all if not most of the substrate that is pNP-cellobiose. On the TLC plates, the mutants without CBM3a, we do not see intense spots corresponding to pNP-cellobiose. However, the mutants with the CBM3a, the spots on the TLC are visible and thus still retain pNP-cellobiose. This could imply that the presence of CBM3a could potentially affect the efficiency of the activity of the enzyme, contrary to the previously seen activities.

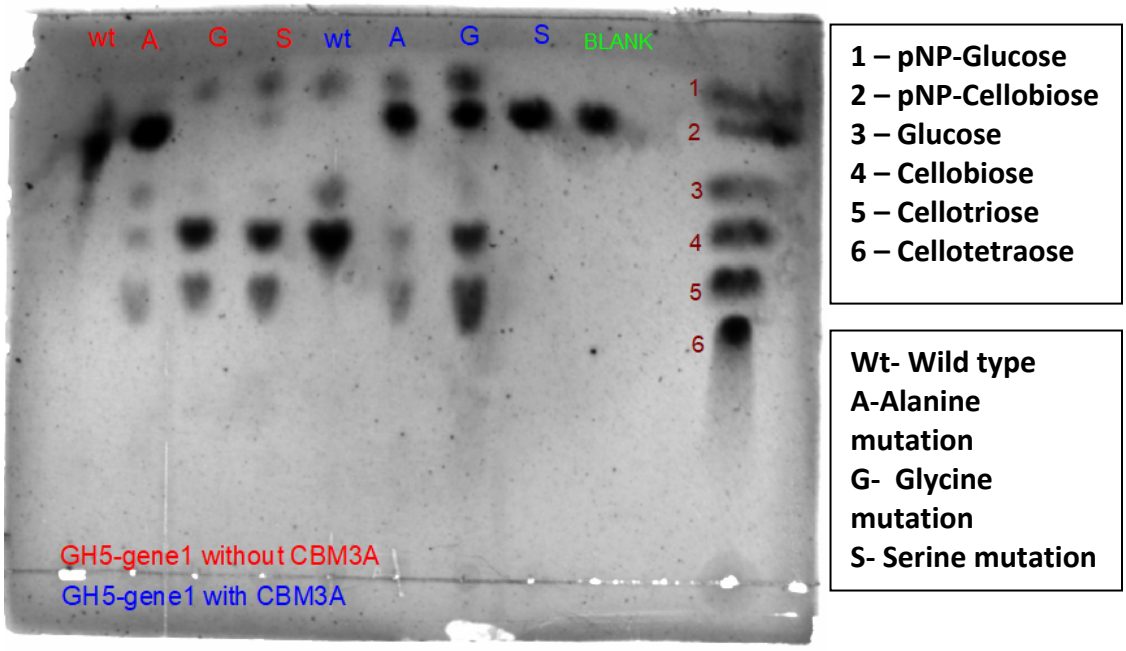

Figure 30: Thin layer chromatography image showing the formation of synthesized products for gene1 set in the presence of pNP-cellobiose

4.3 Kinetic Scheme

Once the TLC for all the plates was analyzed and the products formed were noted down, a general scheme was created to understand the kinetics of the enzyme in a better fashion. The following scheme shows possible pathways and product formations that the enzymes are potentially depicting.

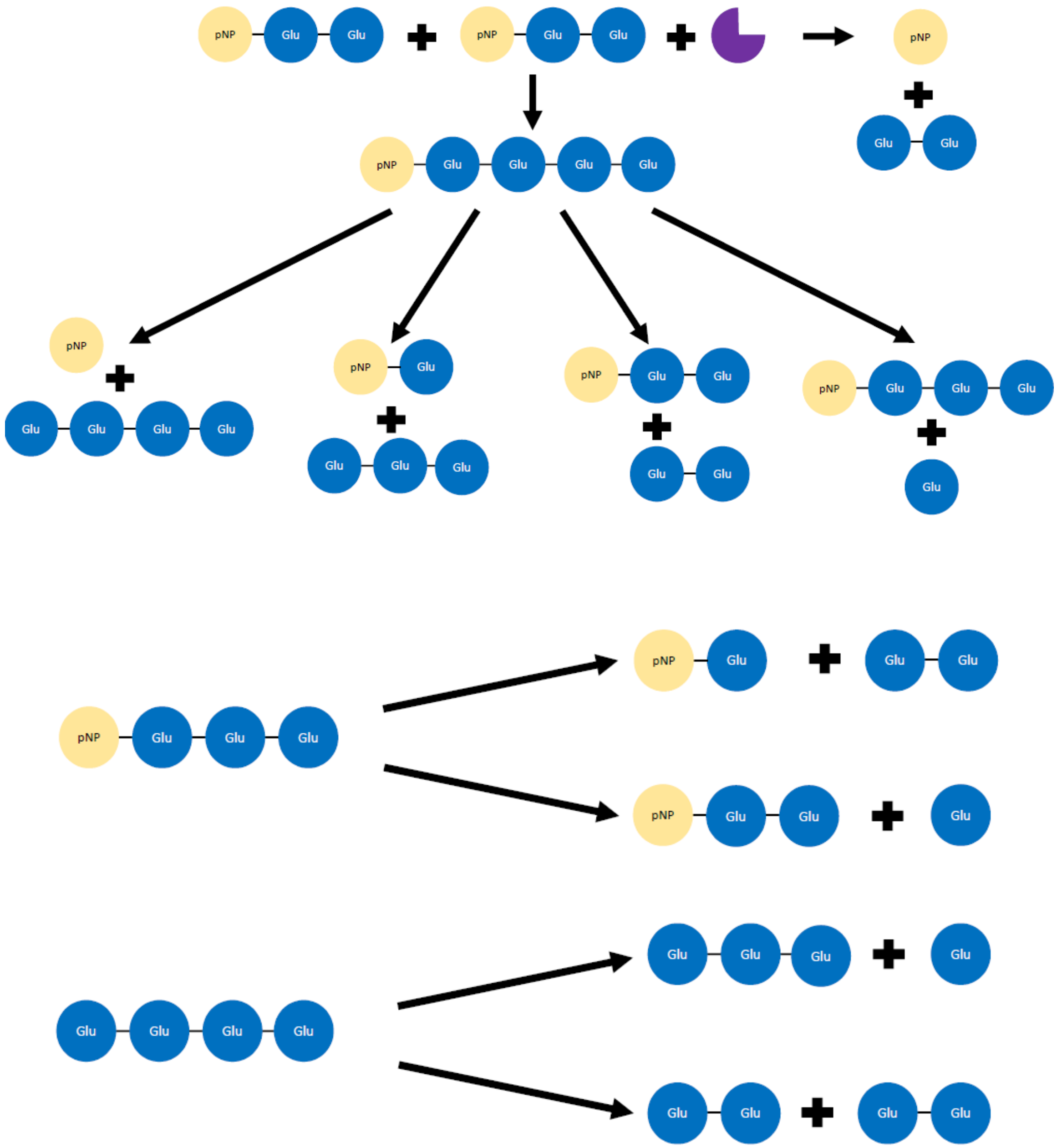

Figure 31: Proposed kinetic scheme depicting the formation of synthesized products

Form the scheme, we then find the apparent rates of the appearance or disappearance of the products. The first pathway is that of the hydrolysis of the substrate, pNP-cellobiose, to pNP and cellobiose. The second pathway that was presented is of the

transglycosylation in which 2 of the substrate molecules present in the active site form a pNP-cellotetraose. However, from our UV images and TLC images, we do not observe the prolonged formation of this product. This potentially means that the product formed in being hydrolyzed immediately. Form the TLC plates, we see the formation of various products such as glucose, cellobiose, cellotriose, etc. Thus, for the generic model for all the enzyme, we consider all possible pathways. Thus, the first is the formation of pNP and cellotetraose. The cellotetraose is then hydrolyzed to pNP-glucose and cellobiose. The cellotetraose could also be hydrolyzed to 2 cellobiose molecules or cellotriose and glucose. The second set of products that are hydrolyzed to pNP-glucose and cellotriose. Other products that are formed are pNP-cellobiose and cellobiose. the final possible set of products that are pNP-cellotriose and glucose. As we do not observe pNP-cellotriose and the reason for this could be as pNP-cellotriose is hydrolyzed to glucose or cellobiose and pNP-glucose. To summarize, there are nine products that are being either formed or consumed that follow:

Product Formed	
P1	pNP
P2	pNP-Glucose
P3	pNP-Cellobiose
P4	pNP-Cellotriose
P5	pNP-Cellotetraose
P6	Glucose
P7	Cellobiose
P8	Cellotriose
P9	Cellotetraose

Table 7: Abbreviations for Products formed

For these nine products, the corresponding kinetic equations that are as follows:

$$E + P3 \xrightarrow{k1} P1 + P7 + E$$

$$E + 2P3 \xrightarrow{k2} P5 + E$$

$$E + P5 \xrightarrow{k3} P2 + P8 + E$$

$$E + P5 \xrightarrow{k4} P3 + P7 + E$$

$$E + P5 \xrightarrow{k5} P4 + P6 + E$$

$$E + P5 \xrightarrow{k6} P1 + P9 + E$$

$$E + P9 \xrightarrow{k7} P6 + P8 + E$$

$$E + P9 \xrightarrow{k8} 2P7 + E$$

$$E + P4 \xrightarrow{k9} P2 + P7 + E$$

$$E + P4 \xrightarrow{k10} P3 + P6 + E$$

$$E + P8 \xrightarrow{k11} P6 + P7 + E$$

Where k is the rate constant for each equation.

The above equations were made dimensionless by dividing the equations by the substrate.

4.4 Parameter Estimation Using Curve Fitting

To obtain the data for numerical analysis, from the TLC plates, we acquire volumetric intensity data that helped us quantify the product concentrations. The TLC plates were analyzed using the Image Lab. The volumetric intensity values were then normalized using by the blank volumetric intensity values for each plate in that is 4mM pNP-cellobiose. The pNP release values were normalized to blank as well. Thus, the products

were normalized by the initial substrate concentrations. For each protein, the products were noted for the five time points, and the tables were generated.

GH5_gene7_wt									
Time (hr)	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**
1	0.053486	0.0000	0.8296	0.0000	0.0000	0.0000	0.0795	0.0000	0.0000
4	0.176651	0.0000	0.6766	0.0000	0.0000	0.1327	0.2213	0.0000	0.0000
24	0.51367	0.3706	0.3805	0.0000	0.0000	0.0000	0.3214	0.3514	0.0000
72	0.670505	0.1626	0.0279	0.0000	0.0000	0.0365	0.4538	0.3694	0.0000
120	0.849928	0.1797	0.0000	0.0000	0.0000	0.0903	0.8203	0.2725	0.0000

GH5_g7_CBM3a_wt									
Time (hr)	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**
1	0.108558	0.0000	0.7825	0.0000	0.0000	0.0000	0.1910	0.0000	0.0000
4	0.303766	0.0945	0.5064	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
24	0.655625	0.4487	0.0995	0.0000	0.0000	0.0000	0.4695	0.3173	0.0000
72	0.730946	0.1447	0.0000	0.0000	0.0000	0.0247	0.4981	0.2287	0.0000
120	0.846418	0.2361	0.0000	0.0000	0.0000	0.0898	0.7693	0.1091	0.0000

GH5_g7_CBM3a_ala									
Time (hr)	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**
1	0.043462	0.0000	0.9513	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.108766	0.0000	0.6145	0.0000	0.0000	0.1969	0.3466	0.0000	0.0000
24	0.455978	0.1542	0.7016	0.0000	0.0000	0.0000	0.2313	0.2401	0.0000
72	0.881851	0.1015	0.0408	0.0000	0.0000	0.0000	0.4572	0.2914	0.0000
120	0.990168	0.0867	0.0000	0.0000	0.0000	0.0000	0.3728	0.2936	0.0000

GH5_g7_CBM3a_gly									
Time (hr)	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**
1	0.072516	0.0000	0.9916	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.218462	0.0000	0.8041	0.0000	0.0000	0.0137	0.0098	0.0000	0.0000
24	0.785272	0.0000	0.4126	0.0000	0.0000	0.1669	0.5174	0.2775	0.0000
72	0.987212	0.0000	0.0821	0.0000	0.0000	0.1261	0.6835	0.0000	0.0000
120	1	0.0474	0.0000	0.0000	0.0000	0.1371	0.6789	0.0000	0.0000

GH5_g7_CBM3a_ser									
Time (hr)	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**
1	0.029583	0.0000	0.9852	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.053558	0.0000	0.9145	0.0000	0.0000	0.0333	0.0597	0.0000	0.0000
24	0.202179	0.2385	0.8861	0.0000	0.0000	0.0000	0.1482	0.1734	0.0000
72	0.457644	0.0943	0.2735	0.0000	0.0000	0.0000	0.1935	0.3013	0.0000
120	0.908053	0.2798	0.1828	0.0000	0.0000	0.0527	0.6282	0.9239	0.0000

Table 8: Fractional Concentration for Products for Gene7 set

For the estimation of the rate constants for each of the reactions, an ODE was established for each of the products that were mentioned in the kinetic scheme.

$$\frac{dP1}{dt} = k_1P3 + k_6P5$$

$$\frac{dP2}{dt} = k_3P5 + k_9P4$$

$$\frac{dP3}{dt} = -k_1P3 - k_2P3^2 + k_4P5 + k_{10}P4$$

$$\frac{dP4}{dt} = k_5P5 - k_9P4 - k_{10}P4$$

$$\frac{dP5}{dt} = k_2P3^2 - k_3P5 - k_4P5 - k_6P5$$

$$\frac{dP6}{dt} = k_5P5 + k_7P9 + k_{10}P4 + k_{11}P8$$

$$\frac{dP7}{dt} = k_1P3 + k_4P5 + \frac{1}{2}k_8P9 + k_9P4 + k_{11}P8$$

$$\frac{dP7}{dt} = k_1P3 + k_4P5 + \frac{1}{2}k_8P9 + k_9P4 + k_{11}P8$$

$$\frac{dP9}{dt} = k_6P5 - k_7P9 - k_8P9$$

From the ODEs that are mentioned above, we had 9 odes and 11 kinetic rate constants. The ODEs formed were made dimensionless by dividing the ode by the initial substrate concentration. Using MATLAB, the parameters that required to be assessed were done by curve fitting the assimilated fractional molar concentration data for the five time points. The function lscqcurvefit was used to sit the data points to the ODEs to get the values of the parameters. Once the values of the rate constants for each reaction were generated for the given set of data, for a given time span, the ODEs were run using the values of the constants, and the plots were generated. The graphs obtained show that the product concentration fraction vs. time (hour). The goodness of the fit was calculated using mean square error and compared between each the for each data point.

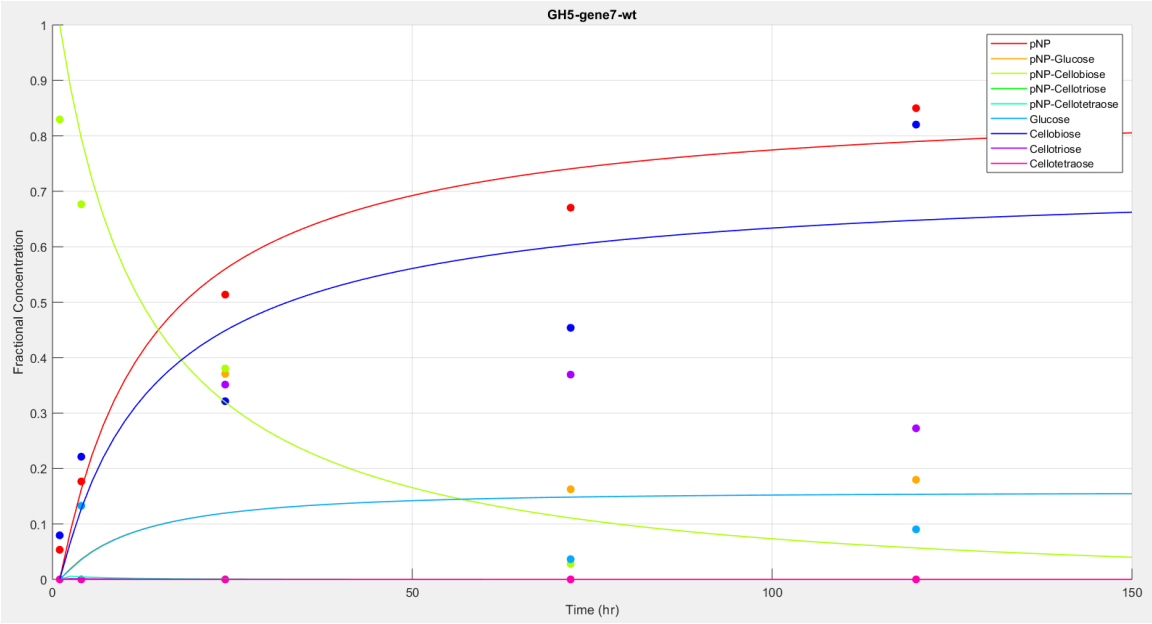

Figure 32: Fractional Concentration over time for GH5-gene7-wt

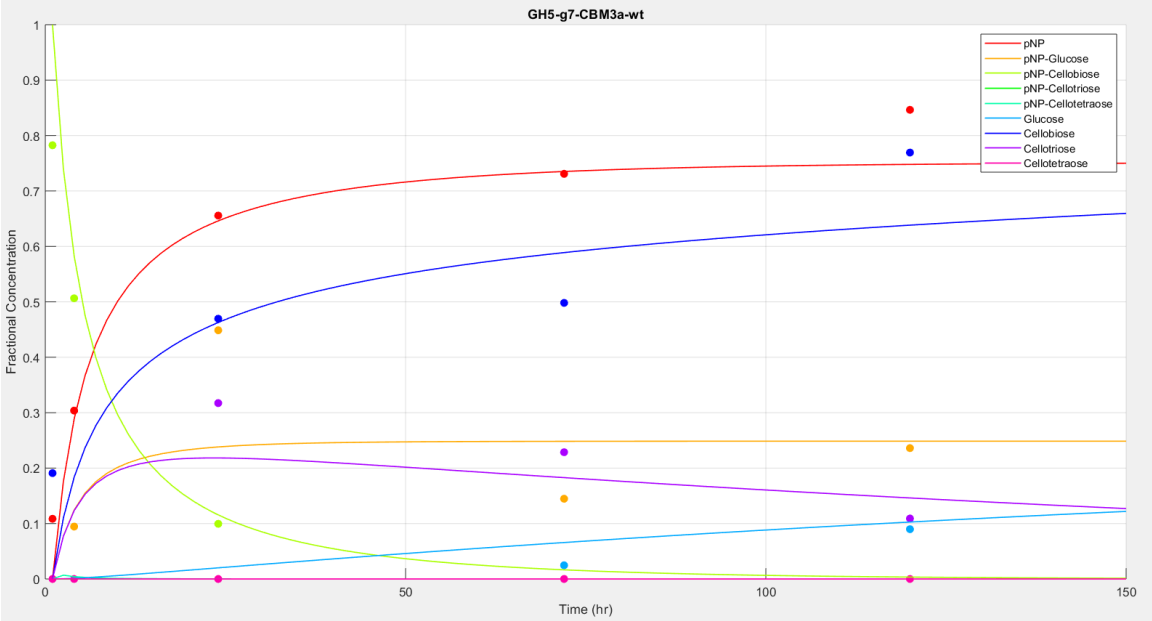

Figure 33: Fractional Concentration over time for GH5-CBM3a gene7-wt

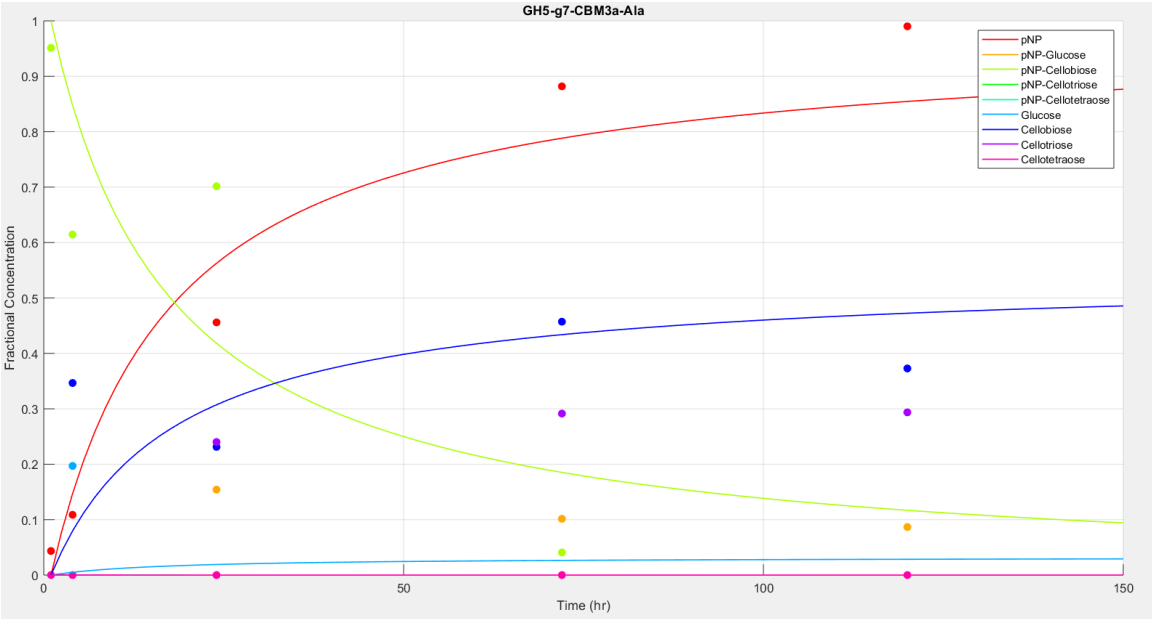

Figure 34: Fractional Concentration over time for GH5-CBM3a gene7 with Alanine mutation

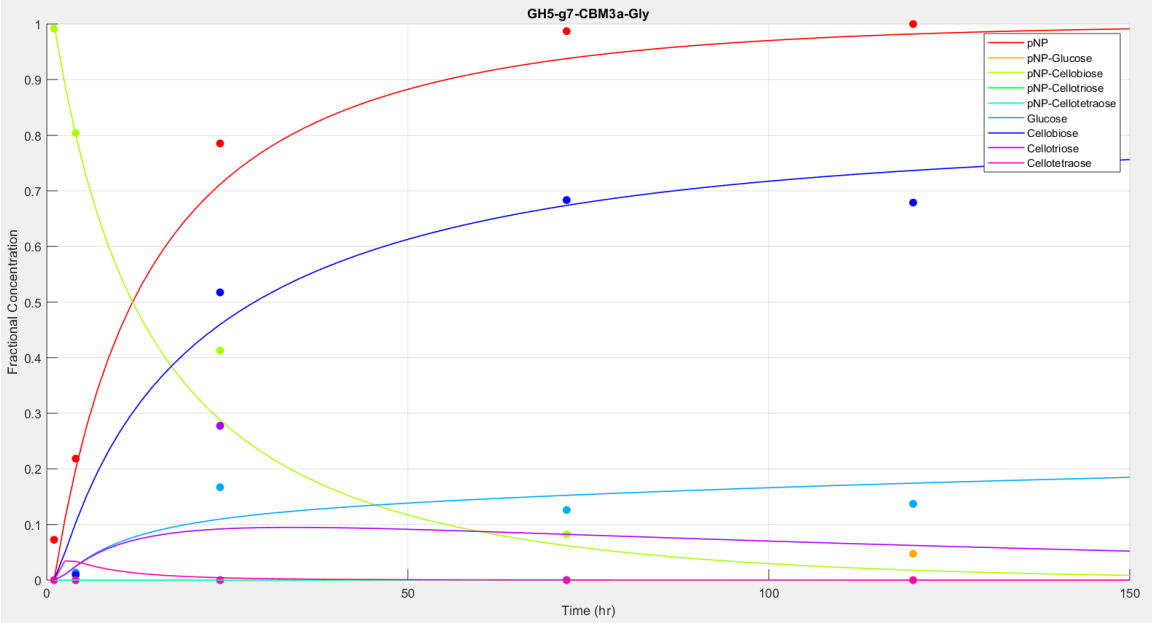

Figure 35: Fractional Concentration over time for GH5-CBM3a gene7 with Glycine mutation

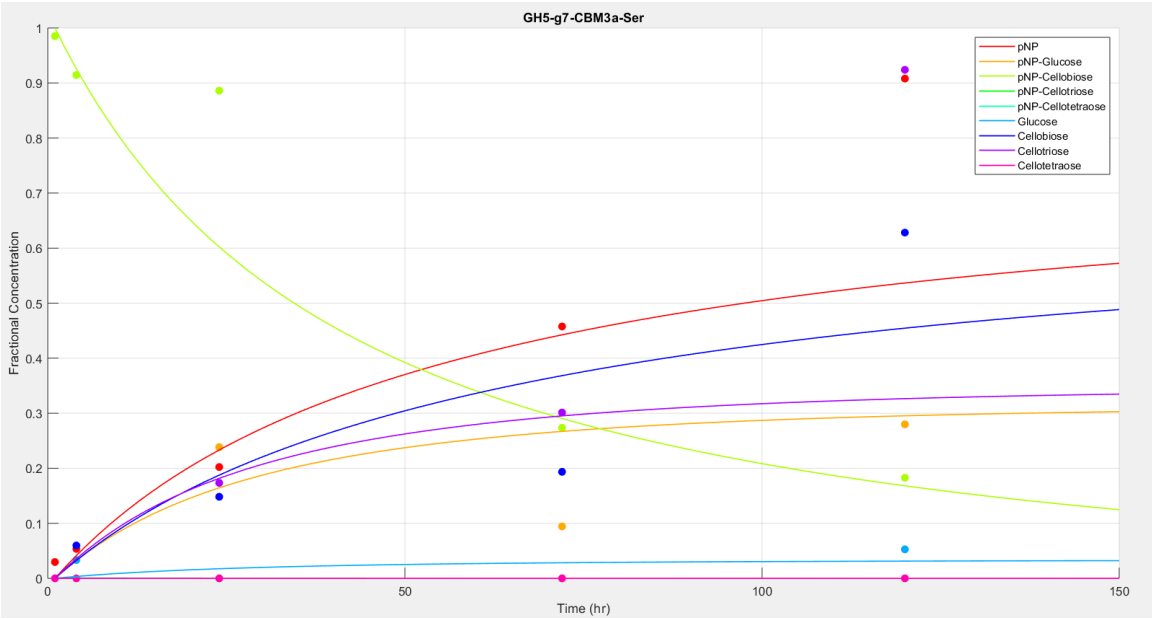

Figure 36: Fractional Concentration over time for GH5-CBM3a gene7 with Serine mutation

From the rate constant values that are obtained, we can understand the product formation or consumption reactions in a quantitative way.

	GH5_gene7_wt (kw)	GH5_gene7_ CBM3a_wt (kcw)	GH5_gene7_ CBM3a_Ala (kca)	GH5_gene7_ CBM3a_Gly (kcg)	GH5_gene7_ CBM3a_Ser (kcs)
k1	0.01	0.03	0.00	0.02	0.01
k2	0.08	0.20	0.06	0.06	0.02
k3	2.17	5.83	3.16	0.64	15.50
k4	0.00	0.12	0.00	0.33	1.53
k5	0.00	0.01	0.00	0.79	0.00
k6	8.37	10.21	91.24	4693.95	10.04
k7	0.61	0.06	0.00	0.28	2280.04
k8	10526.86	242.90	163.77	0.88	11705.11
k9	3.33	872.33	16.10	68915.49	11913.67
k10	0.81	4.82	98.67	99671.55	23371.43
k11	9.76	0.00	1528.54	0.01	0.00

Table 9: Rate constants from the parameter fitting

To begin, we first compare the values for each of the k values to the products that were formed from the reactions that were stated above. Holistically, the rate constants for specific product formations seem to correspond with the products that are being formed. The products formed in the wild type are pNP-glucose, glucose, cellobiose, and cellotriose. From the rate constant that has been calculated, the k values for the formation of pNP-glucose and cellotriose was relatively higher as we see there from the TLC images. The k6 represents the hydrolysis of the pNP-cellotetraose to pNP and cellotetraose. We do see the pNP from the UV images, but cellotetraose is not visible.

The k8 could explain this occurrence as this represents the formation of cellobiose by the hydrolysis of the cellotetraose that is formed. As we do see cellobiose as a prominent product, the high value of k8 is comprehensible. Lastly, k11 is also relatively higher as it represents the formation of glucose and cellobiose, and the TLC plates characterize both these products. In the case of the GH5 gene7 with the CBM3a, k9 is comparatively higher as compared to k8 as k9 depicts the formation of pNP-glucose. The pNP-glucose formed was more in comparison to the wild type, while the creation of cellotriose was not significant.

	kw/kcw	**kcw/kca**	**kcw/kcg**	**kcw/kcs**
k1	0.25	0.03	0.77	0.23
k2	0.39	0.30	0.29	0.10
k3	0.37	0.54	0.11	2.66
k4	0.01	0.00	2.72	12.67
k5	0.03	0.07	121.81	0.55
k6	0.82	8.94	459.76	0.98
k7	10.05	0.01	4.71	37705.06
k8	43.34	0.67	0.00	48.19
k9	0.00	0.02	79.00	13.66
k10	0.17	20.46	20664.61	4845.53
k11	2078.88	325584.40	1.29	0.00

Table 10: Rate Constant Comparisons

The k constants for the wild type enzyme was compared to that of the wild type enzyme with an appended CBM3a. The activity of the mutants was also compared with respect to the wild type with the CBM3a. The k6 comparisons between the wild types show that the

formation of cellotetraose was increased with the presence of CBM3a. In the case of the mutants with CBM3a, the k6 was either higher or about the same as that of the GH5-gene7-CBM3a enzyme. This suggests that the formation of the cellotetraose was improved with a mutation and the CBM present. We see a similar trend with a few more of the enzymes sets.

CHAPTER 5. CONCLUSION AND FUTURE STUDIES

In this project, we studied the effect of the CBM on the efficiency of the activity of the GH5 proteins and then attempted to develop a generic kinetic model that reflects the activity of the enzymes. From the activity assays, we can conclude that the activity of the enzyme was altered with an appended CBM3a. The way the activity changes probably depends on the structure of the enzyme and the residues that interact with the substrate. A kinetic model was proposed for the observed products that were formed. Since some of the products were not observed, future assays need to be performed to gather data points for the curve fitting. Overall, the kinetic model gives us a representation of the rate constant of activity of the enzyme.

However, some additional experiments will need to be performed in the near future. Primarily, the observed k constant values and the TLC analysis, it is noticeable that the rate constant does not necessarily match up with the amounts of products that are being formed in comparison. For a few data sets, the model fails, so calculate the values for the rate constants. This failure could occur due to the few reasons mentioned as follows. Products like pNP-cellotriose, pNP-cellotetraose, and cellotetraose are not observed on the TLC plates or the UV images. This could be since these products formed are generated and consumed quickly, and the time points for which we take the readings do not capture the product generations. The lack of these data points during curve fitting probably affects the values of the predicted rate constants. Another possible reason we do not see the formation of these products under UV or the TLC is lower substrate concentration during the activity assays. In the case of the observed activity with CelE-

CBM3a, at higher substrate concentrations, the oligosaccharides were detectable under UV and on the orcinol stained TLC plates. We tried to reproduce the same by setting an activity assay for the gene7 set with a higher concentration of pNP-cellobiose. The UV images and the TLC showed us the formation of oligosaccharides.

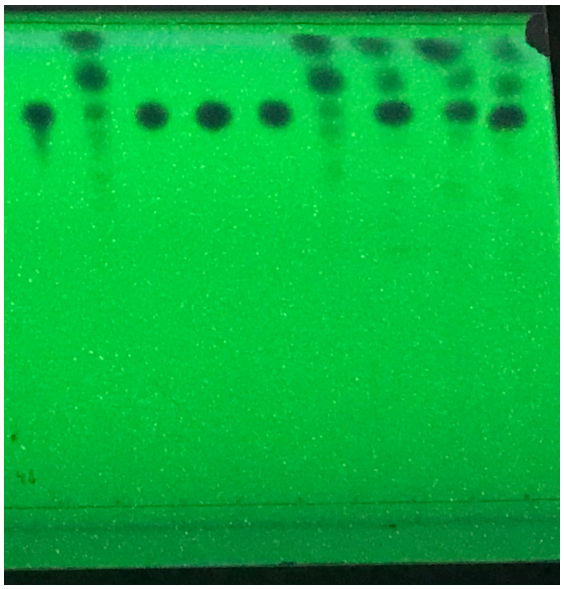

Figure 37: UV images for gene7 set with increasing substrate concentration

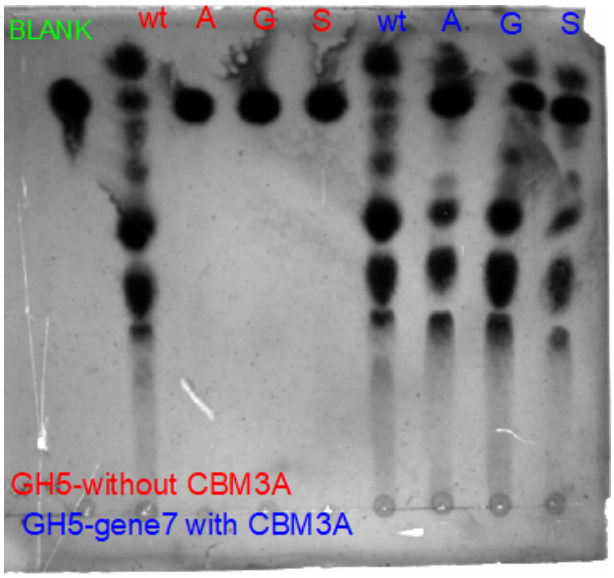

Figure 38: TLC image for gene7 set with increasing substrate concentration

This indicated that the kinetic scheme that was put forward was plausible. But from the TLC plates, we do see the formation of higher oligosaccharides. Thus, appropriate reaction equations need to be added to the kinetics scheme. For the models that the curve fitting failed for, either reading in earlier time points is required, or the increased concentration of the substrate is necessary to update the missing data points for the curve fitting. Future activity assays are required to gather data to optimize the model.

Furthermore, for the structural analysis, since the structures for the enzymes were not readily available, the PDB for the homology models were created using I-TASSER website. The homology models were compared to the CelE-CBM3a model by superimposing the 2 PDB files. The superimposition indicated that the enzyme structures closely resembled that of the CelE protein.

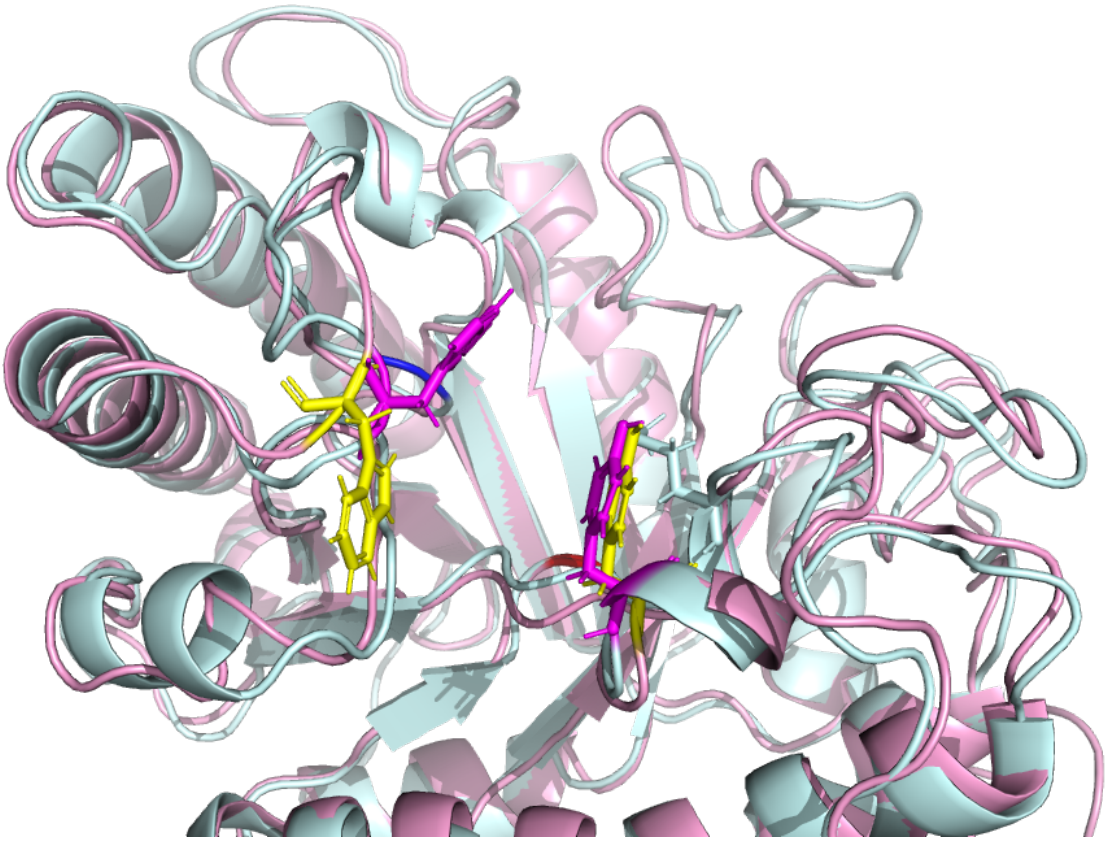

Figure 39: Superimpositions of the GH5 gene7 wild type on CelE

Comparing the residues at the active site, we discovered that there were a few aromatic residues that there present like the CelE protein. From the figure above, we observed that the aromatic residues of the gene7 enzymes shown in yellow were at similar positions of that of the CelE shown in pink. These aromatic proteins are essential for the stacking and enabling stable substrate docking. From experiments previously done by on the CelE protein mentioned in the introduction, the presence of these residues is vital for these enzymes, as mutating these residues resulted in killing the enzyme. A MAFFT alignment was done and observed that the set of these proteins had an aromatic residue around the similar location. In some proteins, we see the active site forming a canal-like structure due to the presence of these residues. The hypothesis was that the mechanism that

enables this kind of nucleophilic substitution is the S_Ni synthase-like reaction [12]. This mechanism does not require a catalytic nucleophile as the reaction takes place. The aromatic residues from around the catalytic site and the residues on the CBM probably stabilize the substrate in the catalytic pocket. The aromatic residues on the CBM assist with leaving group pNP by creating a strain on the bonds in the substrate. While the pNP departs, another substrate molecule attacks the anomeric carbon and bonds to create a higher oligosaccharide product. This could indicate that the CBM3a influence the substrate and enzyme activity.

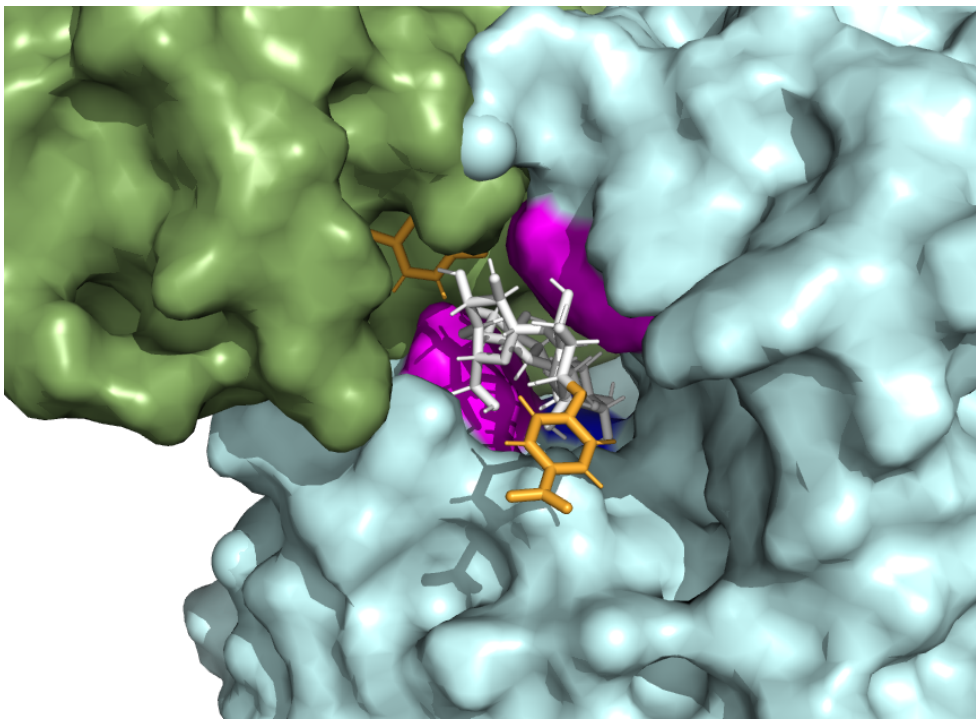

Figure 40: Docking of substrate pNP-cellobiose with CBM3a near the catalytic domain of GH5 gene7

Thus, to potentially see a structure and functional relationship in all the GH family 5 enzymes, appropriate mutations at these residues are required to see the effect of the CBMs on the transglycosylation activity of enzymes. Ultimately, we will be able to

characterize these proteins and achieve a deeper understanding of the mechanism taking place.

APPENDIX A

FASTA sequences for the enzymes

>Gene1

MFINSKTVKRFVSISQQRYVCSHSSARRYADAADNIMTAFQITENMKVGWNLGN
TLDRYAQKANPKDPSKPIPLDSAGLETETCWGCPEASQELFDAIKAKGFNTVRIP
TTWFQHLDENDNIDPAWMARVHQVVDYAYNIGLYVIINLHHEQNWINRADLAT
AYDDINPRLMKLWTQIATEFKDYDQHLIFECMNEPRAMDTPWEWWSATPVEER
DVINRLEANFVELIRGMDGPYAKTRLLMLPGYVASSDKTFLNQIVLPENDDFIAV
SIHAYTPYNFTMNTKTEEGAYHDTFTKEFSNDLAYNLQNFRDMFINKDIPVVIGE
MGTSDFGNTQARVEWTTQYFETTKKYGIPCVLWDNGQKKDPSNPNQNPGEVHG
YIDRKTGEWYENNLPIINKMMEIMNDASIVWGSEGKMPTYDHQDLSAGTLLLEN
SDIGRSSRRNIRRLNSRQGDLM

>Gene2
MKYKVFRSCAAMGLFFSMFSCAFSAEGGSGTAGGAKKFGSSVSSVNAAFTLAG
WNLGNTLDAFVDGHAGLETETSWGMPKTTKEMIEAIKKAGFSTIRLPVSWHNH
VDSNYKIDEKWMARVKEITDWALDSGLYVILNIHHDNLTEAQMKKGIAGFCLSL
DSELQAKSIQYLTSVWTQVSETFKKYDSRLVFEVMNEPRRIGESNEWGFSQPSDA
KEWNDLITVYEQKCLDTIRASGSENANRFVMCPEYAASPHFLDYYTLPKDSAND
KLILSTHAYDPYNFCMNKGSDNTFTSEVEGSINYLFNMLSEKYTSKGIGVVMGET
SASDKKNTAERVKWTKCYYKKASDAKIAVILWDNNVTVAGGGDIDSGECHGYF
NRKALSWYFPEINAAIKKSYGK

>Gene3
MNKRWRQRIFSLCTTAAMVSSLMPMVQATEASSTYKIGDSSPIFCQFYAQNQSG
NWDWMSAGGQNLTVGETTTLEFSALDDDNNSKFAVATDDAAFGFQFGDSKIIA
GDKSSVDLTIDEIVIKADGYEDLVIDPKQKDYKVDYLAEEVSWGIQGNTTSVDLK
SYLGDDFLPYLKAITGFEAKVTLNGYTYKGQAPKEEEKVEGEAIYKKGDVSQLF
TQFYAQENGGNWSWYSPGSENIVYGESNTFKFNTLDEDGNSIFANVGADDAFGI
QIGDSKLQDGDRGIFKSNIKQIVMKAEGYDDVVIDLDETYDLDFIAQEVAWGIDG
NAKMISLKKYVPGAFTEYVKAITSLEVTMSCDDYQFIAKPAEEPEFPEDYQHPTE
MRGLTAKELVKDMKVGWNLGNTLDSTGGETAWGNPKTTRKMIDAIKAAGFNT
VRVPVTWDTGTDEDVNISQEYMDRVETVVNYALANDMYVILNVHHAPKGWFY
TTPELEDSAAAHYAKLWEQIGARFSDYGDKLIFETMNEPRSNEQDWTGDKTAYD
IVNHYNRVALSAIRESGGNNADRLVMIPPYAASSNYSGATALEIPDDDMIAVSIH
AYSPYSFAMDTKSDKVTWGSDEDKRELQNLFKSLDEMFLSKDIPVVIGEFASTN
KNNDEARATHAKYYAQLCKSYDIPCVWWDNGAITENTTDAMGIFDRKRLTWFK

PGIVKALIEGYNGETIDISESALIFDGTATSTEWGQAVSLTPNKDIMLKNLTEGMN
IAVKYESESKPELVLQSWSGGPSWVKVAPARVENGVAYFRYEDMVEAYAKELE
EPSEETFPSLDQIHIGDTGSDLTVTKVYLSETEDPTLIYKGECTCTAWGQALTFIPG
TDIMMNKLGKNVKIAVKYESEEVPEIILQSWSGGASWAKAQPSEVKNGVAYFRY
EDMVKAYAEGTENYESYGEVFPCLNKVYIGAQNTDLKVTKVNYVFPVRVEDAS
FNTLPTLYQGDTVDLNDYLIITPDNATDKEVSKWSSSNEKVATVDENGNLVVLK
PGYTTITAKLDDNDMEVTYDLEVAIKFDADCKLGSDWGTGVNGELVFTNTQTT
PINDWTVEFDYDGEILSIWYGKIVSHEGSHYVVKNLGWNKEIPAGGTITVGFIAN
YEGDVPAPTNYKFLNPEKKDVEPTYDVALDVKSAWNGNFSGEIKITNTSEETISN
WSLDAEITGEIAQIWGGAVASQEDSKYTFINADYNKEIAPGETVTIGISGTANGDV
ELVSYALNF

>Gene5
MTKRSRKLCALITCIALIATLLFPGNGITQSKGATNLTAAQIQRGMGLGFNIGNTF
DSSNNDMGCLVSNHELHWGNPAVTQAYVDAIYDKGFRSIRLPITWYEFITEDNG
TYSIKPEYLARVKEVVDYAYNKNMYVIINVHHENWINRSDLAASYNSISPKLKG
VWKVIAEYFSDYDQRLIFEGMNEPRLVGVEGVEWVGNAEAYNVVNKLDKDFIS
TVRSVASPYKSTRLLMVPSYAASVNPVAYEKMDMTMFNDPYVAASIHAYSPYN
FAMGNGDHSDFSPYKAELESIFAGLRTTFTSKKIPVILGEFSSSNFNNQSARVAWA
KCYMEQAKKLGIPCVLWDNDVIAMQDDGEAHGYLNRATNKWYSESEPVVNAL
LSTLNDSSIVWKSESVYPIYKHNATTSGTELNVSEYFVFDNPSSIITSGKEIAIKYTT
KTPKIALANGAWGNWTEIMPTDLDIDNKIAYFSCDTMLNAWTSADTIANMKILD
SSGNGIQLPAYLISYSGERAETSEEDISSEDNTTSETSQTTQTSQSSQSSQTTQTSQT
KPDNSSESASSQDKSESKPNATQTIPSNSTTTVTPSSEVESFEIGTALENVKGTAKY
KVMGKTTDGTPMVQFVASKSASSATVTVPATVNINGTACKVTSIAPKAFYKNKK
LKKIVIGANIVRIGKQAFYGCTNLKKVIIKKKPKQIGKQAFGKCNKLKKISIANKN
VSAKKYK

>Gene6
MKKKALIMTMTLALVTGLFAGCGSEAVSSSASVSSKTEISKEAKTQTSTSAKAST
ETKTSTETSTETKTSTETSTETSTDTSSNNQEAYNVDRSRVFDILSNINIGWNLGNT
LDATGGGNSVNAETSWGNPKTTQEIVDTVNDRGFNAIRIPVTFANHLGPAPEYPI
SADWLARVKEVVDYAVNDGMYIILDTHHETNYWLKTDPNNEAALCEELAAIW
KQLAEAFKDYDEKLMFEGMNEPRMAGSAKEWSGGTPAERKLINAMNKAFIDA
VRATGGNNADRVLIICTYGHNSDEPTLKDLEIPSDPNIAVALHTYTPYFFTYVAD
GSYSVWNGSKKNDITWQYNNIKKYLIDKGIPVVITETGAQFKENTEDIVRWIGDY
VGTLDQDGVKCFIWDNNIYHGNGEKFGLLNRSLLKWYNDDIVDAYVNHAK

>Gene7
MRNKKRITSLVTGLAMLFTCAVGNTSLKVHADAQSIYTTKGETTKIYASAFTQN
TDDWTWMSMGDTASLVYQDVTNFNAVDANSAFAKSNSTANFGINISDGNLAEG

DSSTLKFHVGTVTVKANGYDDLVINLDKDYSEAYTATKSSWGLTGNTTQILLND
YLPKDTSAKATYLQKITSVKADITLSDYQYTKATPAPATGGQSIYTTKGQLTKIY
ASAFTQNTDDWTWMDIGDTATLVYQDVTNLKGVDSTKNAFAKSNSTANFGINIS
DGNLAEGDSSTLKFHVGTVTVKADGYDDLVINLDKDYSEAYTATKSSWGLTGN
TTQVLLNDYLPKDATEKANYLHKITGVTADITLSDYQYTKAAPPTPEFPSDYHYP
TTMRGLSSMDLVKDMKVGWDLGNSLESAGGETGWGNPATTRKMIDQMKKAG
FNTVRIPVRWDEHYTDSNYTIDPAYMSRVETVVNYALANGMYAIINIHHNKIQG
EMNAENKDSVINEGSAIWKQVGDHFKDYGDQLIFETINEPRTGDDWTGNSSYYN
VVNEYNAKTLSVIRATGGNNGKRLVMMPTYCASADYAKVAAMVVPNDPNVA
VSIHAYIPYNFAMNTDTTKGSYSTFGDADKACIDKTFRLLDKTFIKKGIPVVVGEF
AATDKNNLQDRVNFAKYYLQVASSYGIPCCWWDNNAFSPNSTDSMGIFNRKTL
QFVYPEIVQAMLDGWNNPKDNSNYDENSLFSGTATSKNWGQAVSFNYGLDFVD
DDFSNNLAIAVEYTSDKTPQLILQGNLTGTNWVTVNPAITKSSGTQSIAYYTLKDI
VSAYKNALANYDTYGQVLPGIQTIYIGDQGTNLTVTKVYKVYAQKLEQKVSAA
TINEANVYFDSVTKNASSLSVTLPTDLQKIVDNGLSTCTVSYKSSDDKIAQVSQD
GKITATGVGTATITTTVTVGNSTKTLTTTVEVNKDSSAALLGDVDGNGVVDSLDI
TLIRSYLAGKAVTINKTSADMNNDGRVSGADLMALMKKIAKGA

>Gene8
HEVILIIITESFINENLTINLTQFYLKMKMMEFAPKELFNKDNLVIDNGSKPLKPTIT
SKDIINDMGLGWNLGNTLDASNGLKQDIDSETSWGNPITTKEMIEELYRKGFRTI
RLPTTWRNHLIDENYTIDPKWMKRVKEIVDWCLEAGLYVILNDHHDNADLIEGS
SIPYGKGYYPLLKDLEESERFLFNVWKQITIAFNNGYDHHLIFEGLNEPRLKGHK
NEWSLKRGSYLGGACSALNKYLELIVKPIRESGGNNSRRFIMVTPLSAGFESAIKS
DFQFPNDKSYNSINKLILSVHMYQPYDFTMSKDLTLNIFKEEYKTSLDENFNKLY
DKYIKEGYSIIVGEFGARDKNNLEERIKWGKYYIETVRKNEMSCVIWDNGVFKK
EVDALGSFYGLLHREELTWQTDELVDTYIKAAELY

>Gene9
MANYRERGEKMSTAKKTLSVITAVAIFAIAALMTMCVTQVASAAEMNGNTATE
LVSFIGTGWNLGNSLDAIGGGNSLSSETSWGNPKTTKAMIDAVKAQGFNTVRVP
VSWGNHTTGDNFTIDSKWLARVKEVVDYCIDNDMYVILNIHHDTSTQYYYPSST
YKTQSMKFVKSIWTQVAKYFKDYDQHLVFETLNEPRLVGTGDEWWFPVNNPNS
AVRDSISVINTLNQTAVDAIRAAGGKNTDRCIMVPGYGASIDGCTTPTFKLPDDS
TPNRLIVSVHAYTPYNFALNANGTAEFTNDLKNEVDYLYSTIKSHFIDEGIPAIIGE
TSASNKNNAAERVKWAQYYMGKSAEYGVPCMLWDNNAFNGSDRGECHGHLN
RSTLGWYDKAFVDAVIKYGKKEQIPEKLTPPASVTNCTARHTTTTEIFINWDKVE
DADGYITEQYKGGKWVQINDSEYPAVDLYNLKPDTVYTLRIRAYKTQKGIYAYS
DYVRFAARTSKLVVKNVTNFKVKSATTNSVTLQWDKCEGEVGYFVERYKNGS
WCNIAELPVDTTSYTEKGLINGTTYSFRIRAYRYGDTVLTGLYSNVAGTTTLANV
TGFAKQSSTDSSITVKWNRNSCATGYVLEQYTGGKWVQLTKTASNTNTSYTAK
NLKASTTYTFRIKTYKTVNGKTTETAYIRLAARTSAPSVTNVTGFTKESVTQTTA

TLKWNKNTTATGYIVEQYKGGKWTQIAKTTSNSTLTCTAKGLKANTTYTFRIRA
YKTSGSSTKYSNYVRAAVTTAK

>Gene10
MALLGVSLALTSLKPLLPAKTVYAAGTDKTATEVVSDMTVGWNLGNSLDSYGQ
SSNFPYTSSNETYWGNPKTTKALIDAVAEAGFNTIRIPVSWGQYTTGSDYQIPDFF
MSRVKEVVDYAIANDMYVILNTHHDINSDYCFYVPNNANKDRSEKYFNSIWTQ
VANEFKNYDYHLVFETMNEPRLVGHSEEWWFPRNNPSSDIKEAVACINDYNQV
ALDAIRATGGNNATRCVMVPGYDASIEGCMTDGFKMPKDSASNRLILSVHAYIP
YTFALASDTYTKTFSDNLKGDIDSFFNDLDSKFLSKNIPVVVGETSATNRNNKDE
RVKWADYYWGAAAKHSNVAMVLWDNNVYENNSAGSNGECHMYIDRNSLQW
KDPEIISAIMKHVDGTPATINGKQIPTTDPQPQPDPQPQPDPQPQPDPQPSDDPAPV
NNGDLNATYTINNWGSGYQVIIKVKNDSASAVSNWTVKVKKSQVKIDTSWCVN
IEESGDYYVITPMS

>Gene11
MKKERNFLWAGYSRRLYAMALIFVIGFAAAACNNDQETDSGSEPDVTDGQKPF
VDTAAGKIVADIKIGWNLGNTLDATMNTTSSTSVNQMETAWIRPATTKAMITTI
KEAGFNAIRIPVSWTKAAGEAPDYTIRADWMKRVKEIVDYAVANDMYIILNTHH
DEEVFTFLNSNAEAGKAAFQKIWEQIADAFKDYNEKLIFEGLNEPRTPGSKAEWS
GGTAEERANLNAYYAIFVDTVRKSGGNNDKRVLLINAYAASADAAAMNALTLP
EDSAENRLIVSFHSYSPYDFALNKDPTKNSWSKSNSGDTSAITGPIDRYYNKFVK
NDIPVLIGEFGAMNKDNEEARAQWAEYYVSYAAGKGIKCFWWDNGATTGNGE
KFGLLNRQNNTFTYPVLLQGLIRGAGSAATTTPTVPTEPTTPTPPAIKGNMDKYII
QADGEGNYTQAVWTLSGTNLTTAKTAGAKLVLELSDAPTNVMKLVWVGPDKG
LWWNEKDILDNKGSVKVETKVTWDSGTKTLTIPLSANSVADYNNFTDQPSLKIII
AYYGGSNINDLGIVSANLQ

>Gene12
MIKHLLSRGKLLLFVSVMATSSIIAGGNAYGSTAFTGVRDVPAQQIVNEMKVGW
NLGNTMDAIGGETNWGNPMTTHAMINKIKEAGFNTLRLPVTWDGHMGAAPEY
TIDQTWMKRVEEIANYAFDNDMYVIINLHHENEWLKPFYANEAQVKAQLTKVW
TQIANNFKKYGDHLIFETMNEPRPVGASNEWTGGSYENREVVNRYNLTAVNAIR
ATGGNNATRYIMVPTLAASAMSTTINDLVIPNNDSKVIVSLHMYSPYFFAMDING
TSSWGSDYDKSSLDSEFDAVYNKFVKNGRAVVIGEMGSINKNNTAARVTHAEY
YAKSAKARGLTPIWWDNGYSVAGKAETFGIFNRSNLTWDAPEVMKAFIKGIGGS
STTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPQSAVEVTYAITNSWGSGASVNVTIK
NNGTTPINGWTLKWTMPINQTITNMWSASFVASGTTLSVTNAGYNGTIAANGGT
QSFGFNINYSGVLSKPTGFTVNGTECTVK

>Gene13
MFKKWKKFGISSLALVLVAAVAFTGWSAKASAADASQIVSEMGAGWNLGNQL
EAAVNGTPNETAWGNPTVTPELIKKVKAAGFKSIRIPVSYLNNIGSAPNYTINAA
WLNRIQQVVDYAYNEGLYVIINIHGDGYNSVQGGWLLVNGGNQTAIKEKYKKV
WQQIATKFSNYNDRLIFESMNEVFDGNYGNPNSAYYTNLNAYNQIFVDTVRQTG
GNNNARWLLVPGWNTNIDYTVGNYGFTLPTDNYRSSAIPSSQKRIMISAHYYSP
WDFAGEENGNITQWGATSTNPAKKSTWGQEDYLESQFKSMYDKFVTQGYPVVI
GEFGSIDKTSYDSSNNVYRAAYAKAVTAKAKKYKMVPVYWDNGHNGQHGFAL
FNRSNNTVTQQNIINAIMQGMQ

>Gene14
MKKITTIFVVLLMTVALFSIGNTTAADNDSVVEEHGQLSISNGELVNERGEQVQL
KGMSSHGLQWYGQFVNYESMKWLRDDWGINVFRAAMYTSSGGYIDDPSVKEK
VKEAVEAAIDLDIYVIIDWHILSDNDPNIYKEEAKDFFDEMSELYGDYPNVIYEIA
NEPNGSDVTWGNQIKPYAEEVIPIIRNNDPNNIIIVGTGTWSQDVHHAADNQLAD
PNVMYAFHFYAGTHGQNLRDQVDYALDQGAAIFVSEWGTSAATGDGGVFLDE
AQVWIDFMDERNLSWANWSLTHKDESSAALMPGANPTGGWTEAELSPSGTFVR
EKIRESASIPPSDPTPPSDPGEPDPTPPSDPGEYPAWDPNQIYTNEIVYHNGQLWQ
AKWWTQNQEPGDPYGPWEPLN

>Gene15
MLAGIRTESVSAATMRDVTAQEVVNAIDFGWNLGNDLDCFEENAGGYKGSPQA
YERLWGNPPATKAMIDKVKETGINAIRIPVTWYHNMEGTNIRSNWLSRVKEVVD
YAIDNDTYVIINVHHDTGVHGWLKATTSNAYQNEQKFVKIWEQVASYFKNYDE
HLLFEGYNEIIDNSSNWNDTSYENYQECNKLNQLFVDTVRKSGGNNAKRVLITN
TYAAAVQTSELNNFKKPTDSVPNKLIAEVHCYTPYKFCFKEYSDTHYGSDYDVY
EFMDRTVSRFKSLGMPVIIGEFGAVDKGNTSTRVKWADIYTKNAAQRGIKCFLW
DDGGDFKMLNRNGLSWYYPDVLKAMLKNYGITWNGYGKGTKDHPTNTPRPTA
TPTPQPWRPTATPTPQPWRPTATPTPQPWRPTATPTPLPQDGHERNEDGTYVGAE
IDESKIYMIQNVNSGKYLDAAGQNVQSGINVQQWDASTANPGINQTWIVKKVN
WGYYYIYSRLGDGSSLVLNVNDGNNGSNINLDYKNGQSTQYFKFVPNQDGTYEI
VTRASRDRSCVEIINAYTDNGANVQQYQWNNHPCQRWRLIPVGDHQGMTPTPK
PTATPRPTATPRPTNTPQPTVTQRPTSTPRPTPTPAGAERRRLRYHRIDLDGEPFLP

APPENDIX B

Primers for PIPE/SLIC.

FP- Forward Primer

RP- Reverse Primer

Name	Sequence	Tm (C)
pEC_GH5_gene1_E325A_FP	GTTGTTATTGGTGCAATGGGAACATCG	63.6
pEC_GH5_gene1_E325A_RP	CGATGTTCCCATTGCACCAATAACAAC	63.6
pEC_GH5_gene1_E325G_FP	GTTGTTATTGGTGGAATGGGAACATCG	63
pEC_GH5_gene1_E325G_RP	CGATGTTCCCATTCCACCAATAACAAC	63
pEC_GH5_gene1_E325S_FP	GTTATTGGTTCAATGGGAACATCG	59
pEC_GH5_gene1_E325S_RP	CGATGTTCCCATTGAACCAATAAC	59
pEC_GH5_gene2_E324A_FP	GTAGTTATGGGCGCAACATCAGCGA	65.6
pEC_GH5_gene2_E324A_RP	TCGCTGATGTTGCGCCCATAACTAC	65.6
pEC_GH5_gene2_E324G_FP	GTAGTTATGGGCGGAACATCAGCGAG	65.4
pEC_GH5_gene2_E324G_RP	CTCGCTGATGTTCCGCCCATAACTAC	65.4
pEC_GH5_gene2_E324S_FP	GTTATTGGTTCAATGGGAACATCG	64.8
pEC_GH5_gene2_E324S_RP	CGATGTTCCCATTGAACCAATAAC	64.8

pEC_GH5_gene3_E642A_FP	GGTTGTTATCGGTGCATTCGCTTCTAC	64.5
pEC_GH5_gene3_E642A_RP	GTAGAAGCGAATGCACCGATAACAACC	64.5
pEC_GH5_gene3_E642G_FP	GTAGTTATGGGCGGAACATCAGCGAG	63.9
pEC_GH5_gene3_E642G_RP	CTCGCTGATGTTCCGCCCATAACTAC	63.9
pEC_GH5_gene3_E642S_FP	GTTATTGGTTCAATGGGAACATCG	62.5
pEC_GH5_gene3_E642S_RP	CGATGTTCCCATTGAACCAATAAC	62.5
pEC_GH5_gene4_E276A_FP	GTTGTTATCGGTGCAATGGGTACCATC	64.3
pEC_GH5_gene4_E276A_RP	GATGGTACCCATTGCACCGATAACAAC	64.3
pEC_GH5_gene4_E276G_FP	GTTGTTATCGGTGGAATGGGTACCATCAAC	65.8
pEC_GH5_gene4_E276G_RP	GTTGATGGTACCCATTCCACCGATAACAAC	65.8
pEC_GH5_gene4_E276S_FP	GTTGTTATCGGTTCAATGGGTACCATC	62.3
pEC_GH5_gene4_E276S_RP	GATGGTACCCATTGAACCGATAACAAC	62.3
pEC_GH5_gene5_E310A_FP	TTATCTTAGGCGCATTTAGCTCGA	60.3
pEC_GH5_gene5_E310A_RP	TCGAGCTAAATGCGCCTAAGATAA	60.3
pEC_GH5_gene5_E310G_FP	CAGTTATCTTAGGCGGATTTAGCTCGAG	63.2
pEC_GH5_gene5_E310G_RP	CTCGAGCTAAATCCGCCTAAGATAACTG	63.2

pEC_GH5_gene5_E310S_FP	GTTATCTTAGGCTCATTTAGCTCGAG	59.4
pEC_GH5_gene5_E310S_RP	CTCGAGCTAAATGAGCCTAAGATAAC	59.4
pEC_GH5_gene6_E362A_FP	GTTGTTATCACCGCAACCGGTGCTCAG	68.2
pEC_GH5_gene6_E362A_RP	CTGAGCACCGGTTGCGGTGATAACAAC	68.2
pEC_GH5_gene6_E362G_FP	GGTTGTTATCACCGGAACCGGTGCTCAG	69
pEC_GH5_gene6_E362G_RP	CTGAGCACCGGTTCCGGTGATAACAACC	69
pEC_GH5_gene6_E362S_FP	GGTTGTTATCACCTCAACCGGTGCTC	65.8
pEC_GH5_gene6_E362S_RP	GAGCACCGGTTGAGGTGATAACAACC	65.8
pEC_GH5_gene7_E643A_FP	GTTGTTGGTGCATTCGCTGCTAC	63.3
pEC_GH5_gene7_E643A_RP	GTAGCAGCGAATGCACCAACAAC	63.3
pEC_GH5_gene7_E643G_FP	GGTTGTTGTTGGTGGATTCGCTGCTAC	66.6
pEC_GH5_gene7_E643G_RP	GTAGCAGCGAATCCACCAACAACAACC	66.6
pEC_GH5_gene7_E643S_FP	GTTGTTGTTGGTTCATTCGCTGCTACCG	67
pEC_GH5_gene7_E643S_RP	CGGTAGCAGCGAATGAACCAACAACAAC	67
pEC_GH5_gene8_E347A_FP	CATCGTTGGTGCATTCGGTGCTC	65
pEC_GH5_gene8_E347A_RP	GAGCACCGAATGCACCAACGATG	65
pEC_GH5_gene8_E347G_FP	CATCGTTGGTGGATTCGGTGCTCG	66.2
pEC_GH5_gene8_E347G_RP	CGAGCACCGAATCCACCAACGATG	66.2

pEC_GH5_gene8_E347S_FP	CTATCATCGTTGGTTCATTCGGTGCTCGTG	67
pEC_GH5_gene8_E347S_RP	CACGAGCACCGAATGAACCAACGATGATAG	67
pEC_GH5_gene9_E326A_FP	CTATCATCGGTGCAACCTCTGCTTC	63.7
pEC_GH5_gene9_E326A_RP	GAAGCAGAGGTTGCACCGATGATAG	63.7
pEC_GH5_gene9_E326G_FP	CTATCATCGGTGGAACCTCTGCTTC	63
pEC_GH5_gene9_E326G_RP	GAAGCAGAGGTTCCACCGATGATAG	63
pEC_GH5_gene9_E326S_FP	GCTATCATCGGTTCAACCTCTGCTTCTAAC	65.4
pEC_GH5_gene9_E326S_RP	GTTAGAAGCAGAGGTTGAACCGATGATAGC	65.4
pEC_GH5_gene10_E313A_FP	GTTGTTGGTGCAACCTCTGCTAC	62.5
pEC_GH5_gene10_E313A_RP	GTAGCAGAGGTTGCACCAACAAC	62.5
pEC_GH5_gene10_E313G_FP	GGTTGTTGTTGGTGGAACCTCTGCTACC	67.5
pEC_GH5_gene10_E313G_RP	GGTAGCAGAGGTTCCACCAACAACAACC	67.5
pEC_GH5_gene10_E313S_FP	GTTGTTGTTGGTTCAACCTCTGCTACC	64.5
pEC_GH5_gene10_E313S_RP	GGTAGCAGAGGTTGAACCAACAACAAC	64.5
pEC_GH5_gene11_E333A_FP	GTTTTAATCGGTGCATTCGGCGCGATG	67
pEC_GH5_gene11_E333A_RP	CATCGCGCCGAATGCACCGATTAAAAC	67
pEC_GH5_gene11_E333G_FP	GGTTTTAATCGGTGGATTCGGCGCG	67

pEC_GH5_gene11_E333G_RP	CGCGCCGAATCCACCGATTAAAACC	67
pEC_GH5_gene11_E333S_FP	GTTTTAATCGGTTCATTCGGCGCGATG	65.6
pEC_GH5_gene11_E333S_RP	CATCGCGCCGAATGAACCGATTAAAAC	65.6
pEC_GH5_gene12_E303A_FP	GTCATTGGAGCGATGGGCTCC	63
pEC_GH5_gene12_E303A_RP	GGAGCCCATCGCTCCAATGAC	63
pEC_GH5_gene12_E303G_FP	GTTGTCATTGGAGGGATGGGCTCCATTAAC	67
pEC_GH5_gene12_E303G_RP	GTTAATGGAGCCCATCCCTCCAATGACAAC	67
pEC_GH5_gene12_E303S_FP	GTTGTCATTGGATCGATGGGCTCC	63
pEC_GH5_gene12_E303S_RP	GGAGCCCATCGATCCAATGACAAC	63
pEC_GH5_gene13_E323A_FP	GTGGTCATTGGTGCATTTGGCAGCATTG	68
pEC_GH5_gene13_E323A_RP	CAATGCTGCCAAATGCACCAATGACCAC	68
pEC_GH5_gene13_E323G_FP	GTGGTCATTGGTGGATTTGGCAGC	65
pEC_GH5_gene13_E323G_RP	GCTGCCAAATCCACCAATGACCAC	65
pEC_GH5_gene13_E323S_FP	GTGGTCATTGGTTCATTTGGCAGCATTG	65.8
pEC_GH5_gene13_E323S_RP	CAATGCTGCCAAATGAACCAATGACCAC	65.8
pEC_GH5_gene14_D263A_FP	CTGCTACCGGTGCCGGTGGTGTTTTC	70

pEC_GH5_gene14_D263A_RP	GAAAACACCACCGGCACCGGTAGCAG	70
pEC_GH5_gene14_D263G_FP	GCTACCGGTGGCGGTGGTGTTTTCC	65
pEC_GH5_gene14_D263G_RP	GGAAAACACCACCGCCACCGGTAGC	65
pEC_GH5_gene14_D263S_FP	CTGCTACCGGTTCCGGTGGTGTTTTCC	57
pEC_GH5_gene14_D263S_RP	GGAAAACACCACCGGAACCGGTAGCAG	57
pEC_GH5_gene15_E274A_FP	GTTATCATCGGTGCATTCGGTGCTGTTG	67
pEC_GH5_gene15_E274A_RP	CAACAGCACCGAATGCACCGATGATAAC	67
pEC_GH5_gene15_E274G_FP	GTTATCATCGGTGGATTCGGTGCTG	64
pEC_GH5_gene15_E274G_RP	CAGCACCGAATCCACCGATGATAAC	64
pEC_GH5_gene15_E274S_FP	GTTATCATCGGTTCATTCGGTGCTG	63
pEC_GH5_gene15_E274S_RP	CAGCACCGAATGAACCGATGATAAC	63

APPENDIX C

Primers for site directed mutagenesis

Name	Sequence
pEC_GH5_gene1_E325A_FP	GTTGTTATTGGTGCAATGGGAACATCG

pEC_GH5_gene1_E325A_RP	CGATGTTCCCATTGCACCAATAACAAC
pEC_GH5_gene1_E325G_FP	GTTGTTATTGGTGGAATGGGAACATCG
pEC_GH5_gene1_E325G_RP	CGATGTTCCCATTCCACCAATAACAAC
pEC_GH5_gene1_E325S_FP	GTTATTGGTTCAATGGGAACATCG
pEC_GH5_gene1_E325S_RP	CGATGTTCCCATTGAACCAATAAC
pEC_GH5_gene2_E324A_FP	GTAGTTATGGGCGCAACATCAGCGA
pEC_GH5_gene2_E324A_RP	TCGCTGATGTTGCGCCCATAACTAC
pEC_GH5_gene2_E324G_FP	GTAGTTATGGGCGGAACATCAGCGAG
pEC_GH5_gene2_E324G_RP	CTCGCTGATGTTCCGCCCATAACTAC
pEC_GH5_gene2_E324S_FP	GTTATTGGTTCAATGGGAACATCG
pEC_GH5_gene2_E324S_RP	CGATGTTCCCATTGAACCAATAAC
pEC_GH5_gene3_E642A_FP	GGTTGTTATCGGTGCATTCGCTTCTAC
pEC_GH5_gene3_E642A_RP	GTAGAAGCGAATGCACCGATAACAACC
pEC_GH5_gene3_E642G_FP	GTAGTTATGGGCGGAACATCAGCGAG
pEC_GH5_gene3_E642G_RP	CTCGCTGATGTTCCGCCCATAACTAC
pEC_GH5_gene3_E642S_FP	GTTATTGGTTCAATGGGAACATCG
pEC_GH5_gene3_E642S_RP	CGATGTTCCCATTGAACCAATAAC
pEC_GH5_gene4_E276A_FP	GTTGTTATCGGTGCAATGGGTACCATC
pEC_GH5_gene4_E276A_RP	GATGGTACCCATTGCACCGATAACAAC
pEC_GH5_gene4_E276G_FP	GTTGTTATCGGTGGAATGGGTACCATCAAC
pEC_GH5_gene4_E276G_RP	GTTGATGGTACCCATTCCACCGATAACAAC
pEC_GH5_gene4_E276S_FP	GTTGTTATCGGTTCAATGGGTACCATC
pEC_GH5_gene4_E276S_RP	GATGGTACCCATTGAACCGATAACAAC
pEC_GH5_gene5_E310A_FP	TTATCTTAGGCGCATTTAGCTCGA

pEC_GH5_gene5_E310A_RP	TCGAGCTAAATGCGCCTAAGATAA
pEC_GH5_gene5_E310G_FP	CAGTTATCTTAGGCGGATTTAGCTCGAG
pEC_GH5_gene5_E310G_RP	CTCGAGCTAAATCCGCCTAAGATAACTG
pEC_GH5_gene5_E310S_FP	GTTATCTTAGGCTCATTTAGCTCGAG
pEC_GH5_gene5_E310S_RP	CTCGAGCTAAATGAGCCTAAGATAAC
pEC_GH5_gene6_E362A_FP	GTTGTTATCACCGCAACCGGTGCTCAG
pEC_GH5_gene6_E362A_RP	CTGAGCACCGGTTGCGGTGATAACAAC
pEC_GH5_gene6_E362G_FP	GGTTGTTATCACCGGAACCGGTGCTCAG
pEC_GH5_gene6_E362G_RP	CTGAGCACCGGTTCCGGTGATAACAACC
pEC_GH5_gene6_E362S_FP	GGTTGTTATCACCTCAACCGGTGCTC
pEC_GH5_gene6_E362S_RP	GAGCACCGGTTGAGGTGATAACAACC
pEC_GH5_gene7_E643A_FP	GTTGTTGGTGCATTCGCTGCTAC
pEC_GH5_gene7_E643A_RP	GTAGCAGCGAATGCACCAACAAC
pEC_GH5_gene7_E643G_FP	GGTTGTTGTTGGTGGATTCGCTGCTAC
pEC_GH5_gene7_E643G_RP	GTAGCAGCGAATCCACCAACAACAACC
pEC_GH5_gene7_E643S_FP	GTTGTTGTTGGTTCATTCGCTGCTACCG
pEC_GH5_gene7_E643S_RP	CGGTAGCAGCGAATGAACCAACAACAAC
pEC_GH5_gene8_E347A_FP	CATCGTTGGTGCATTCGGTGCTC
pEC_GH5_gene8_E347A_RP	GAGCACCGAATGCACCAACGATG
pEC_GH5_gene8_E347G_FP	CATCGTTGGTGGATTCGGTGCTCG
pEC_GH5_gene8_E347G_RP	CGAGCACCGAATCCACCAACGATG
pEC_GH5_gene8_E347S_FP	CTATCATCGTTGGTTCATTCGGTGCTCGTG
pEC_GH5_gene8_E347S_RP	CACGAGCACCGAATGAACCAACGATGATAG
pEC_GH5_gene9_E326A_FP	CTATCATCGGTGCAACCTCTGCTTC

pEC_GH5_gene9_E326A_RP	GAAGCAGAGGTTGCACCGATGATAG
pEC_GH5_gene9_E326G_FP	CTATCATCGGTGGAACCTCTGCTTC
pEC_GH5_gene9_E326G_RP	GAAGCAGAGGTTCCACCGATGATAG
pEC_GH5_gene9_E326S_FP	GCTATCATCGGTTCAACCTCTGCTTCTAAC
pEC_GH5_gene9_E326S_RP	GTTAGAAGCAGAGGTTGAACCGATGATAGC
pEC_GH5_gene10_E313A_FP	GTTGTTGGTGCAACCTCTGCTAC
pEC_GH5_gene10_E313A_RP	GTAGCAGAGGTTGCACCAACAAC
pEC_GH5_gene10_E313G_FP	GGTTGTTGTTGGTGGAACCTCTGCTACC
pEC_GH5_gene10_E313G_RP	GGTAGCAGAGGTTCCACCAACAACAACC
pEC_GH5_gene10_E313S_FP	GTTGTTGTTGGTTCAACCTCTGCTACC
pEC_GH5_gene10_E313S_RP	GGTAGCAGAGGTTGAACCAACAACAAC
pEC_GH5_gene11_E333A_FP	GTTTTAATCGGTGCATTCGGCGCGATG
pEC_GH5_gene11_E333A_RP	CATCGCGCCGAATGCACCGATTAAAAC
pEC_GH5_gene11_E333G_FP	GGTTTTAATCGGTGGATTCGGCGCG
pEC_GH5_gene11_E333G_RP	CGCGCCGAATCCACCGATTAAAACC
pEC_GH5_gene11_E333S_FP	GTTTTAATCGGTTCATTCGGCGCGATG
pEC_GH5_gene11_E333S_RP	CATCGCGCCGAATGAACCGATTAAAAC
pEC_GH5_gene12_E303A_FP	GTCATTGGAGCGATGGGCTCC
pEC_GH5_gene12_E303A_RP	GGAGCCCATCGCTCCAATGAC
pEC_GH5_gene12_E303G_FP	GTTGTCATTGGAGGGATGGGCTCCATTAAC
pEC_GH5_gene12_E303G_RP	GTTAATGGAGCCCATCCCTCCAATGACAAC
pEC_GH5_gene12_E303S_FP	GTTGTCATTGGATCGATGGGCTCC
pEC_GH5_gene12_E303S_RP	GGAGCCCATCGATCCAATGACAAC
pEC_GH5_gene13_E323A_FP	GTGGTCATTGGTGCATTTGGCAGCATTG

pEC_GH5_gene13_E323A_RP	CAATGCTGCCAAATGCACCAATGACCAC
pEC_GH5_gene13_E323G_FP	GTGGTCATTGGTGGATTTGGCAGC
pEC_GH5_gene13_E323G_RP	GCTGCCAAATCCACCAATGACCAC
pEC_GH5_gene13_E323S_FP	GTGGTCATTGGTTCATTTGGCAGCATTG
pEC_GH5_gene13_E323S_RP	CAATGCTGCCAAATGAACCAATGACCAC
pEC_GH5_gene14_D263A_FP	CTGCTACCGGTGCCGGTGGTGTTTTC
pEC_GH5_gene14_D263A_RP	GAAAACACCACCGGCACCGGTAGCAG
pEC_GH5_gene14_D263G_FP	GCTACCGGTGGCGGTGGTGTTTTCC
pEC_GH5_gene14_D263G_RP	GGAAAACACCACCGCCACCGGTAGC
pEC_GH5_gene14_D263S_FP	CTGCTACCGGTTCCGGTGGTGTTTTCC
pEC_GH5_gene14_D263S_RP	GGAAAACACCACCGGAACCGGTAGCAG
pEC_GH5_gene15_E274A_FP	GTTATCATCGGTGCATTCGGTGCTGTTG
pEC_GH5_gene15_E274A_RP	CAACAGCACCGAATGCACCGATGATAAC
pEC_GH5_gene15_E274G_FP	GTTATCATCGGTGGATTCGGTGCTG
pEC_GH5_gene15_E274G_RP	CAGCACCGAATCCACCGATGATAAC
pEC_GH5_gene15_E274S_FP	GTTATCATCGGTTCATTCGGTGCTG
pEC_GH5_gene15_E274S_RP	CAGCACCGAATGAACCGATGATAAC

BIBLIOGRAPHY

1. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkt1178
2. Guillén, D., Sánchez, S., & Rodríguez-Sanoja, R. (2010). Carbohydrate-binding domains: Multiplicity of biological roles. *Applied Microbiology and Biotechnology*. https://doi.org/10.1007/s00253-009-2331-y
3. de Souza, P. M., & e Magalhães, P. de O. (2010). Application of microbial α-amylase in industry - a review. *Brazilian Journal of Microbiology*. https://doi.org/10.1590/s1517-83822010000400004
4. Harris, A. D., & Ramalingam, C. (2010). Xylanases and its Application in Food Industry: A Review. *Journal of Experimental Sciences*.
5. Sela, D. A., Garrido, D., Lerno, L., Wu, S., Tan, K., Eom, H. J., … Mills, D. A. (2012). Bifidobacterium longum subsp. infantis ATCC 15697 α-fucosidases are active on fucosylated human milk oligosaccharides. *Applied and Environmental Microbiology*. https://doi.org/10.1128/AEM.06762-11
6. Sukumaran, R. K., Singhania, R. R., Mathew, G. M., & Pandey, A. (2009). Cellulase production using biomass feed stock and its application in lignocellulose saccharification for bio-ethanol production. *Renewable Energy*. https://doi.org/10.1016/j.renene.2008.05.008
7. Mackenzie, L. F., Wang, Q., Warren, R. A. J., & Withers, S. G. (1998). Glycosynthases: Mutant glycosidases for oligosaccharide synthesis. *Journal of the American Chemical Society*. https://doi.org/10.1021/ja980833d
8. Qiao, Y., Molina, H., Pandey, A., Zhang, J., & Cole, P. A. (2006). Chemical rescue of a mutant enzyme in living cells. *Science*. https://doi.org/10.1126/science.1122224
9. Matsubayashi, H., & Ueda, T. (2014). Cell-free protein synthesis system. *Kobunshi*. https://doi.org/10.1016/j.jbiotec.2004.02.014
10. Ii, T. (1989). Restriction enzyme digestion. *Electrophoresis*. https://doi.org/10.1371/journal.pone.0124783
11. Stevenson, J., Krycer, J. R., Phan, L., & Brown, A. J. (2013). A practical comparison of ligation-independent cloning techniques. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0083888
12. Iglesias-Fernández, J., Hancock, S. M., Lee, S. S., Khan, M., Kirkpatrick, J., Oldham, N. J., … Davis, B. G. (2017). A front-face “SNi synthase” engineered from a retaining “double-SN2” hydrolase. *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.2394
13. Sukharnikov, L. O., Cantwell, B. J., Podar, M., & Zhulin, I. B. (2011). Cellulases: Ambiguous nonhomologous enzymes in a genomic perspective. *Trends in Biotechnology*. https://doi.org/10.1016/j.tibtech.2011.04.008
14. Boraston, A. B., Bolam, D. N., Gilbert, H. J., & Davies, G. J. (2004). Carbohydrate-binding modules: Fine-tuning polysaccharide recognition. *Biochemical Journal*. https://doi.org/10.1042/BJ20040892

15. Abbott, D. W., & Boraston, A. B. (2012). Quantitative approaches to the analysis of carbohydrate-binding module function. In *Methods in Enzymology*. https://doi.org/10.1016/B978-0-12-415931-0.00011-2
16. 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*. https://doi.org/10.1016/j.pep.2014.02.013