2016

THE KINETIC CHARACTERIZATION OF TWO PUTATIVE GALACTOSIDASES, BT2857 AND SCO6594, AND THE PRELIMINARY ANALYSES OF SCO6595-97

Maria Ngo
Wilfrid Laurier University, ngox8930@mylaurier.ca

THE KINETIC CHARACTERIZATION OF TWO PUTATIVE GALACTOSIDASES, BT2857 AND SCO6594, AND THE PRELIMINARY ANALYSES OF SCO6595-97

By

Maria Ngo

Submitted to the Department of Chemistry and Biochemistry

Faculty of Science

In partial fulfillment of the requirements for the

Master of Science in Chemistry and Biochemistry

Wilfrid Laurier University

2016

ABSTRACT

Bacteroides thetaiotaomicron is a prolific bacterium found in the distal intestinal tract of humans that possesses the ability to breakdown complex polysaccharides through the release of carbohydrate-active enzymes (CAZymes). Overall, the saccharolytic prowess of *B. thetaiotaomicron* presents an intriguing model for understanding microbial-host nutrient exchange within the gut and to identify novel mechanisms for accessing carbohydrates. The successful expression, purification, and preliminary kinetic characterization of BT2857 are reported. BT2857 was found to be enzymatically active against the artificial substrate, 4-nitrophenyl β-D-galactopyranoside (pNP-Gal); suggesting its role as a putative β-galactosidase and allowed for the quantification of BT2857 activity by the determination of the Michaelis-Menten parameters k_{cat}, K_M, and V_{max}. A combinatorial molecular modeling approach of genomic context predicted a two-domain structure formed primarily of beta-sheets and loops. Polysaccharide Utilization Loci (PUL) analysis revealed BT2851-55 collectively belonging to GH36, GH42, GH3 and GH2 providing insight on potential substrates downstream from BT2857. The natural substrate remains unknown despite experiments demonstrating the upregulation of this region of the *B.thetaiotaomicron* genome in the presence of varying carbohydrates. Crystallization of the truncations of BT2857 will be a key finding for structure and functional characterization.

Streptomyces coelicolor is the representative species within its genus and predominantly found in soil with access to a diverse environment of carbohydrate sources. The main contribution of *S. coelicolor* is the metabolism of insoluble remains from plants

and animals indicating positive improvements on plant growth and the rhizosphere. To date, the kinetic characterization of a putative galactosidase, SCO6594 has been explored including the determination of Michaelis-Menten parameters. Solubility screening of SCO6595-67 also display potential expression and characterization of this gene cluster will enable an understanding of protein regulation and carbohydrate interaction.

ACKNOWLEDGEMENTS

I would like to acknowledge my supervisor, Dr. Michael Suits for giving me the opportunity to work in his lab, as well as the support and encouragement that I greatly appreciated. I would also like to acknowledge the collaboration with Dr. Alicia Lammerts van Bueren on working with BT2857. I would like to thank the members of my committee, Dr. Geoff Horsman and Dr. Allison McDonald, for their support and advice throughout my research. Big thanks to Joe Meissner for always lending help and space in the undergraduate teaching labs. Thank you to Dr. Lilian DeBruin and Dr. Masoud Jelokhani-Niaraki for all of your guidance and for making my educational experience at Laurier memorable. I would also like to thank everyone I've met from the Suits Lab, Horsman Lab, Smith Lab, Weadge Lab, and Jelokhani Lab who kept me sane through the past two years. Most of all, I'd like to thank Brigitte Hudy-Yuffa for being there for me and making my lab experience unforgettable; I could not have asked for a better friend who is always bursting with kindness to work with in the lab. Lastly, I would like to thank my parents for believing and supporting all of my endeavours, and my brother for putting up with all my antics.

DECLARATION OF WORK PERFORMED

All data presented and described in this thesis are the results of my own work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AA	auxiliary activity
APS	ammonium persulfate
BT	*Bacteroides thetaiotaomicron*
CAZyme	carbohydrate-active enzyme
CBM	carbohydrate-binding module
CE	carbohydrate esterase
FPLC	fast protein liquid chromatography
GH	glycoside hydrolase
GT	glycosyltransferase
IMAC	immobilized-metal affinity chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
LB	Luria-Bertani
MQ	MilliQ
NCBI	National Center for Biotechnology Information
NTA	nitrilotriacetic acid
oNP-Gal	(ortho) 2-nitrophenyl β-D-galactopyranoside
pI	isoelectric point
PL	polysaccharide lyase
pNP-Gal	(para) 4-nitrophenyl β-D-galactopyranoside
pNP-Man	(para) 4-nitrophenyl α-D-mannopyranoside

PCR	polymerase chain reaction
PUL	polysaccharide-utilization loci
rRNA	ribosomal ribonucleic acid
SDS-PAGE	sodium dodecyl sulfate - polyacrylamide gel electrophoresis
SCO	*Streptomyces coelicolor*
TEMED	tetramethylethylenediamine

CHAPTER 1 – INTRODUCTION

1.1 Glycobiology and Carbohydrate Variance

Carbohydrates also referred to as saccharides or sugars, are molecules that commonly feature in biosynthetic processes or as structural components for the cell [9]. They are arranged in over a million different isomers with variation in stereochemistry and are a crucial source of energy [10,11]. Carbohydrates are found throughout the cellular surface, within the extracellular matrix, and covalently bound to form glycoproteins; thus offering roles that span from the growth and function of cellular processes to adhesion and recognition for cell-cell interactions [12]. Oligosaccharides and polysaccharides are used in microbial interactions in living organisms including the storage of energy, as signal molecules in specific processes, the production of the cell wall, and much more [10,13]. The utilization of carbohydrates by bacteria have a crucial role as a source of energy and as a key constituent of the peptidoglycan which is made up of alternating units of N-acetyl glucosamine and N-acetyl muramic acid that forms the cell wall [14,15]. Oligosaccharides of collagen also have a role in the stability and physical condition of human tissue serving as a scaffold [12]. Overall, there are many processes and interactions (cell-cell, cell-carbohydrate, etc.) that are modulated by the presence of carbohydrates [16,17]. An example demonstrating the dependence of carbohydrate-protein interaction is the recognition of mannose-6-phosphate to initiate the exportation of lysosomal enzymes from the Golgi apparatus [18]. Moreover, carbohydrates are not only valuable as free-standing compounds, but also when they are covalently attached to secreted proteins or part of the cell surface [19]. The carbohydrate moiety of glycosylated proteins typically contribute to recognition and modification in

biological systems [12,20]. Over the years, the biotechnology industry has evolved towards the engineering of therapeutic glycoproteins such as enzymes, hormones and antibodies [21,22]. Although there are many advantages in the glycosylation of proteins, not all modifications are beneficial. In a study with bone and platelet osteonectin, it was revealed that the collagen-binding properties of the protein was altered after glycosylation [23]. Furthermore, granted that carbohydrates and glycosylation present many benefits, they are often found in complex forms or as bulky polysaccharides which make them unfavorable for cellular transportation [24]. Hence, the breakdown of complex carbohydrates by carbohydrate-active enzymes help release simpler forms of sugars (i.e. glucose, fructose, galactose) beneficial for situations like absorption by the host organism in symbiotic situations later discussed in the human gut microbiome [25,26].

1.2 Carbohydrate-Active Enzymes

Carbohydrate-active enzymes (CAZymes) utilize carbohydrates as substrate for metabolic processes involving the breakdown and synthesis of glycosidic bonds to form simple or complex sugars [27]. CAZymes are organized into classes and families based on sequence similarity with established groups of characterized proteins, therein providing inferences to structure and function by using a combination of predictive software [28]. The CAZy database (http://www.cazy.org/) serves as a guide by organizing sequences that encode for putative CAZymes or proteins that possess carbohydrate-binding modules (CBMs). Although this is a useful starting point, common research questions arise from the variance in substrate specificity even among enzymes of the same family. Therefore, a combination of glycogenomics with experimental

procedures is required for complete protein characterization. The classes of CAZymes that can be found in the CAZy database are glycoside hydrolases (GHs), glycosyltransferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), and auxiliary activities (AAs) [28]. Glycoside hydrolases (also referred to as glycosidases or glycosyl hydrolases) catalyze glycosidic bonds with the involvement of activated water molecules [29]. Glycosyltransferases catalyse the transfer of a sugar moiety to an acceptor molecule through the formation of a glycosidic bond [30]. Glycosyltransferases are commonly used in drug discovery and development for targeted drug delivery [20]. In general, glycosylation is also found to contribute to protein folding and transportation across the cell membrane [31]. Polysaccharide lyases perform similar reactions to glycoside hydrolases but differ in the absence of water to break glycosidic bonds [32]. The reaction by carbohydrate esters is unique from the other enzyme classes because the products result in a carboxylic acid and unlike the rest of the CAZymes, cleaves an ester bond rather than a glycosidic bond [27,33]. Finally, auxiliary activities are a family classified in connection with CAZymes involved in redox reactions [28]. An example of an auxiliary activity is the characterization of a protein as a mono-oxygenase belonging to GH61 [34]. Additionally, carbohydrate-binding modules (CBMs) are frequently associated with the catalytic domains of CAZymes but are not enzymatically-active themselves [34,35]. CBMs increase enzymatic activity by anchoring long-chain carbohydrates to increase the proximity of the substrate from the enzymatic domain [36]. Overall, CAZymes have demonstrated their usefulness in the metabolism of carbohydrates. In particular, bacteria are known to code for a large number of CAZymes that enable them to adapt to different environments and selection of carbohydrate sources

available [27,37]. Harnessing the activity from CAZymes have value in therapeutic applications of disease research as well as large-scale production of alternative biomass resources [38]. For instance, ethanol and butanol can be produced through the fermentation of monosaccharides released from GH activity of degraded cellulose providing potential for the growing market of flexible fuel vehicles [38]. Furthermore, understanding the mechanisms of such CAZymes can either be improved for increased efficiency (i.e.biofuel production) or restricted for inhibitory effects. The latter situation is advantageous when studying CAZymes involved in disease such as the involvement of glycosidases in the virulence of *Streptococcus pneumoniae* [39]. Site-directed mutagenesis of conserved catalytic residues is a useful technique for manipulating the kinetic activity of such enzymes [40]. In general, CAZymes have demonstrated expression in a widespread number of bacteria and allow for an understanding between the effect of carbohydrates and their local environments as well as the microbial processes of the species.

1.2.1 Focus on Glycoside Hydrolases

Of the CAZyme classes mentioned, glycoside hydrolases arises as the most common domain among the putative proteins addressed throughout this project. Glycoside hydrolases (GHs) are a group of enzymes that cleave the glycosidic bond via general acid/base catalysis [41]. There are two described catalytic mechanism of GHs, retaining or inverting GHs, and both involve a Bronsted acid and Lewis base (Figure 1) [10]. In retaining GHs, the Lewis base (B^-) directly attacks the O-glycosidic bond and deprotonates the acidic (AH) residue to release the product. A water molecule is then deprotonated by the free acid (A^-) from above to hydrolyze the bond between the Lewis

base and the remaining glycosyl compound resulting in retention of stereochemistry [10]. In inverting GHs, the Lewis base (B^-) activates a water molecule for hydrolysis of the O-glycosidic bond from below. The catalytic acid (AH) simultaneously protonates the glycosidic oxygen of the leaving group which results in the inversion of the glycoside stereochemistry [10].

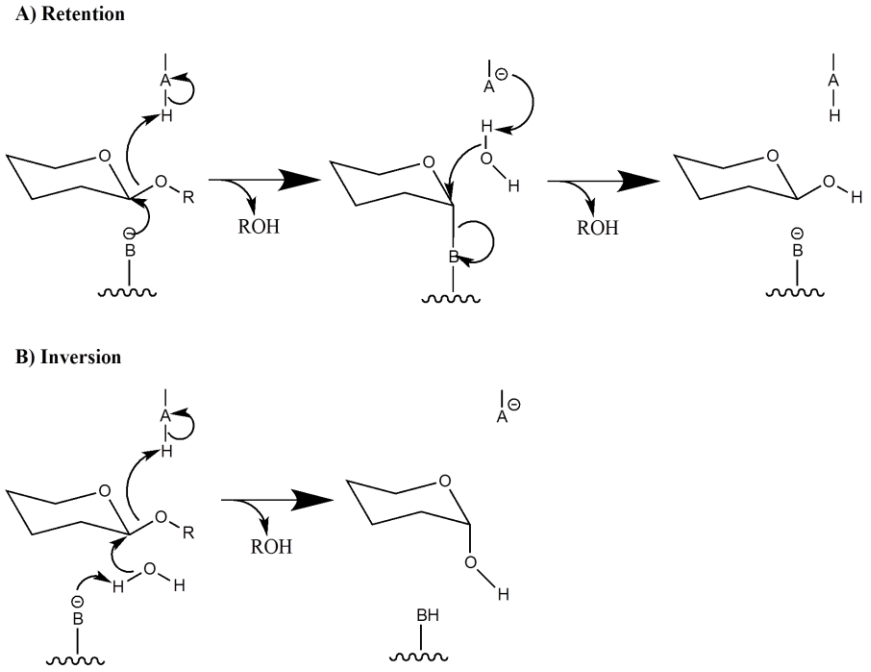


Figure 1. Mechanism of (A) retaining and (B) inverting glycoside hydrolases: (A) The Lewis base (B^-) directly attacks the anomeric carbon and the Bronsted acid (AH) protonates the leaving group. The Bronsted acid (A^-) activates a water molecule which hydrolyzes the bond between the base and substrate also reversing the stereochemistry to the original direction. **(B)** The Lewis base (B^-) activates a water molecule for hydrolysis thus inverting the substrate stereochemistry. The leaving group is protonated by the Bronsted acid (AH).

Other than the step-wise differences in the two mechanisms, there is also a 5.5Å and 10Å spatial difference between catalytic residues for retaining and inverting GHs, respectively

[42]. Up to date, there are 131 families of GHs in which evolutionary divergence can be observed through differences in substrate specificity within the same family [10]. Studies have also shown mechanistic development in GHs by manipulating the substrate into a "half-chair" or "sofa" conformation thus lowering the activation energy for catalysis [43]. Additionally, aspartate and glutamate residues are commonly known as the acid/base pair for catalysis, however other residues such as tyrosine was also found to play a role in stabilization of the transitional state of the substrate in retaining mechanisms [44,45].

Aside all the mechanistic developments amongst GH families, the active-site topology of this enzyme class categorizes into three types of general folds (Figure 2):

1. The pocket or crater proteins possess an active site that is buried within the core of the protein and envelopes the substrate. The pocket/crater shape is most ideal for exo-acting enzymes and monosaccharidases because of the limiting space of the enclosed active site. A model example of a pocket/crater protein is glucoamylasae from *Aspergillus awamori* (PDB ID: 1GLM) [46]. Polysaccharides with branched composition such as starch and amylopectin are common substrates for pocket/crater GHs because of their abundance in non-reducing free ends.
2. The cleft or groove proteins are endo-acting enzymes that more optimal for less fibrous substrates such as cellulose and chitin due to the nature of the U-shaped fold exposing the active site to the surroundings. An example of a cleft/groove protein is endoglucanase Cel6A from *Thermobifida fusca* (PDB ID: 2BOD) [47].
3. The tunnel proteins are enzymatically efficient for releasing cellobiose from cellulose as the substrate remains bound while the free end is fed through (as the

name suggests) the active site. Thus far, the only characterized GH with a tunnel fold is cellobiohydrolase Cel6A and Cel7A from *Trichoderma reesei* (PDB ID: 3CBH) [48,49].

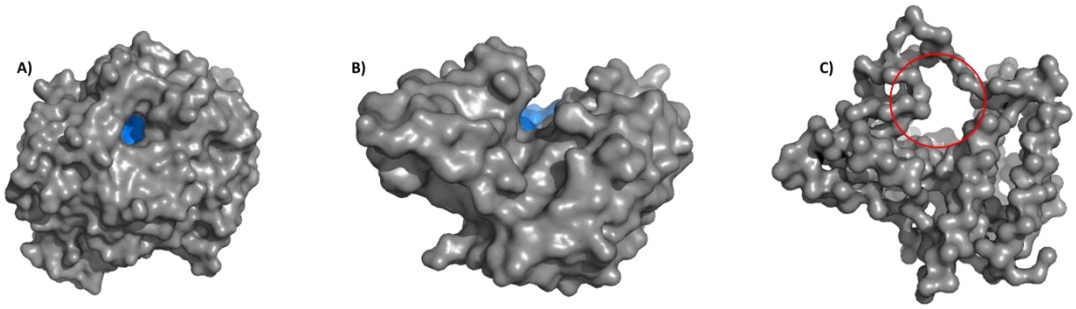


Figure 2. Surface models of GH active-site topologies for (A) pocket or crater, (B) cleft or groove, and (C) tunnel proteins: (A) The active site (blue) of pocket/crater proteins are buried within the protein and is typical for exo-acting enzymes. (B) The active site (blue) of cleft/groove proteins are exposed to the solvent and are most optimal for rigid substrates with less free ends such as cellulose and chitin. (C) Cellobiohydrolyase Cel6A and Cel7A from *T. reesei* (PDB ID: 3CBH) is the only GH observed with a tunnel fold (circled in red) and is useful for catalyzing long polysaccharides.

1.3 The Human Gut Microbiome and Carbohydrate-Processing

The human gut microbiome consists of tens of trillions of microorganisms with anaerobic species outnumbering aerobic species by nearly one thousand to one, and demonstrate mutualistic and commensialistic relationships [50–52]. Because of the beneficial health implications of probiotic supplements, dietary additives containing specific carbohydrates have been developed to attempt to manipulate the gut metabolism [53]. The gut microbiota consists of approximately 400 different species of bacteria;

however 70% of the population is accounted by species belonging to the genera *Cytophaga, Flavobacterium, Firmicutes, and Bacteroides*. Of the dominating genera, *Bacteroides* makes up 25% of that population [54]. The presence of such bacteria is key to the metabolism of dietary carbohydrates, because the human genome only codes for 17 CAZymes that are involved in digestion of foods containing starch, sucrose and lactose [25,50,55]. Therefore, the gut microbiota aid in metabolic activities with their wide production of enzymes that otherwise are not encoded by the human genome [50]. The evolution of the gut microbiota has demonstrated many benefits including the protection against pathogenic micro-organisms by triggering the human immune system, and most importantly the degradation of complex carbohydrates which is beneficial for intestinal absorption [50].

1.3.1 *Bacteroides thetaiotaomicron*

Bacteroides thetaiotaomicron is localized in the descending (or distal) large intestine and currently codes for approximately 300 CAZymes with a total of 55 GH and PL families, and 172 characterized GHs [50,53,56]. Its genome encodes for the most number of genes involved in carbohydrate metabolism including two notable outer membrane proteins, SusC and SusD, responsible for the transportation and binding of starch [57]. *B. thetaiotaomicron* produces approximately 11% of extracellular targeted CAZymes that are involved in the breakdown of indigestible polysaccharides (i.e. cellulose) thus leaving monosaccharides and disaccharides to be absorbed in the small intestine [56]. As mentioned earlier, one benefit of carbohydrate use by *B. thetaiotaomicron* is the formation of an intestinal mucosal barrier that induces the release of angiogenin to support the human immune response against pathogenic microorganisms

[58]. Another intriguing characteristic of *B. thetaiotaomicron* is its ability to upregulate the production of certain enzymes in relation to the availability of resources and environmental cues [59]. In a study conducted by Kotarski et al., they discovered that *B. thetaiotaomicron* did not produce a certain glycosyl hydrolase unless it was exposed to the substrate specific for the enzyme demonstrating the dependence between protein expression and carbohydrate availability [59–61].

1.3.2 Polysaccharide Utilization Loci

Polysaccharide utilization loci (PULs) are clusters of genes that are involved in the breakdown of complex glycans [62]. PULs were initially described specifically for the starch utilization system (sus) in *B. thetaiotaomicron* however further experiments have shown upregulation of gene clusters accordingly to changes in the environment and carbohydrate variability other than starch [6]. PULs are made up of co-regulated proteins that interact with the carbohydrate substrate in a five-part systematic procedure. The general PUL system is comprised by: (1) a protein located on the surface for substrate binding, (2) cleavage of complex sugars into more manageable lengths of oligosaccharides (3) the transportation of the oligosaccharides into the periplasmic space for (4) further degradation into monosaccharides, and finally (5) regulation of the PUL expression in response to carbohydrate availability and negative feedback from product saturation [63].

1.4 *Streptomyces coelicolor* and Implications on the Rhizosphere

Streptomyces coelicolor is a free-living soil bacteria and representative model species for the genus. *S. coelicolor* undergoes a unique morphological growth cycle

initiated by spore germination that develops into branched hyphae for substrate feeding, followed by the septation of reproductive aerial hyphae to produce spores to repeat the cycle [64,65]. As a ubiquitous species in soil, it has a significant role in contributing to the carbon cycle through vegetative decay and forms a close symbiotic relationship with plants for nutrient exchange. Rhizobacteria such as *S. coelicolor* densely populate the surroundings of plant roots because of the abundance of nutrients [66]. In return, rhizobacteria metabolize a variety of substrates including insoluble remains into useable carbohydrate sources to promote plant growth [67]. Understanding the relationship within the rhizosphere has been used to improve crop yields for agricultural purposes [68,69]. *S. coelicolor* are also found to produce a diversity of secondary metabolites with applications in human medicine such as anti-tumor and immunosuppressant drugs [65,70,71]. Interestingly, the production of such natural antibiotics were observed in correlation with increased carbohydrate metabolism indicating the importance of available carbohydrate substrates [72,73]. Additionally, *S. coelicolor* codes for anti-microbial genes such as a type I ribosome-inactivating protein (SCO7092) with N-glycosidase activity that was found to cleave a specific adenine residue from rRNA from ribose to prevent fungal ribosomal activity [74]. In general, studies of *S. coelicolor* revealed that the degradation of biopolymers was down-regulated when in the presence of glucose or other monosaccharides as a form of negative feedback to control catabolic activity [75]. The precise regulation in protein expression indicates an evolved system of carbohydrate metabolism by the bacterium.

1.5 Research Objectives

The long term objectives of this project was to determine the structure and function of putative carbohydrate-active enzymes as provided by the CAZy database as open reading frames. Each potential open reading frame was analyzed and dissected into constructs of the catalytic domain and carbohydrate-binding domain. Characterized proteins will be added to the CAZy database as an access point for future studies in engineering biochemical uses and understanding the glycobiology of *B. thetaiotaomicron* and *S. coelicolor*.

The first short term objective achieved was selecting potential open reading frames and running the sequences through a workflow of bioinformatics to predict putative characteristics or present carbohydrate-active domains of interest. From the CAZy database, the targets SCO6594-97 were selected while BT2857 was sent from Dr. Alicia Lammerts van Bueren of Gron Institute of Biomolecular Sciences & Biotechnology (Groningen, The Netherlands). Using bioinformatics and referencing to characterized proteins with high sequence similarity, target constructs were produced (Appendix I) for the downstream process of isolating and purifying recombinant, putative CAZymes.

The second objective was to obtain precursor conditions that would lead to the determination of protein 3-D structure. The steps to achieving this objective required the screening of protein stability and crystal growth in a range of buffers, pH, additives, and temperatures. Once optimal crystal growth was obtained, diffraction would be performed for X-ray crystallography.

Finally, the last objective of the project was to determine the function of the protein through kinetic assays and substrate-binding. Experimental testing of potential substrates for the enzyme activity was drawn from extensive literature review. The bioinformatics workflow would also provide leads to active domains or catalytic residues that might be conserved and critical for site-directed mutagenesis.

CHAPTER 2 – EXPERIMENTAL TECHNIQUES FOR STUDYING THE STRUCTURE AND FUNCTION OF EXTRACELLULARLY-EXPRESSED PROTEINS

2.1 Bioinformatics for Construct Design and Functional Hypotheses

The carbohydrate-active enzyme (CAZy) database is an essential starting point that indicates the hypothetical open reading frames that may code for enzymes partaking in carbohydrate metabolism and binding [76]. The list of hypothetical open reading frames are further organized into enzyme classes and families based on predicted domains and sequence similarity to existing characterized proteins. Using the sequence provided by *National Centre for Biotechnology Information (*NCBI*),* a series of bioinformatics resources will help predict the function and structure of the protein [77,78].

STRING-10 provides a visual representation and interactive map of associated proteins with both structural and functional similarities surrounding the target protein thus leading to other prospective targets within the species' genome [79]. *InterPro* is useful for predicting domains and important sequence-based features that might be present in the query sequence [80]. Domain predictions not only depict hypothetical structural folds, but also suggest possible functional activity. Additionally, *Phyre*^{2} compiles a list of proteins with sequence similarity to the query which can help determine potential truncations sites when designing constructs for recombinant proteins [3]. This program also generates a predicted model based on high identity from similar proteins to the query indicating the confidence level of secondary structures. Open reading frames

that are predicted as extracellular proteins are prioritized because they are generally easier to express and isolate without the need to replicate particular *in vivo* conditions for protein folding and stability. *SignalP 4.1* is a useful bioinformatics tool used to cleave any present signal peptide sequence from the query to prevent protein localization within the cell (such as transmembrane regions) when it is expressed [7]. Lastly, the *PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC* is a visualization software used to display the predicted 3D structure using the protein database file (.pdb) provided by *Phyre*^{2} [3]. Other useful features of *PyMOL* include overlap of protein structures for comparison, modification and design elements to optimize presentation, and display of molecular interactions (i.e. electrostatic). *Snapgene* ® software (from GSL Biotech; available at www.snapgene.com) is used to view vector maps, restriction cute sites, and primer design.

2.2 Protein Solubility Screening

To isolate expressed protein from cultures, the standard protocol solubilizes samples in 50mM Tris-HCl pH 7.5. However, if the protein is present in the pellet after chemical lysis, a quick alternative method to screen for protein solubility is sonicating cell pellets in different buffer conditions ranging from pH 4-10. Other variables for protein solubility include salt concentration, buffering agent, presence of urea, detergents (i.e. triton X-100, Tween-20, etc.) and/or glycerol. Sonication is a process of homogenization at ultrasonic frequencies (20kHz) by inserting a probe into the sample thus disrupting the cell membrane. The probe is connected to a generator which allows for control of frequency amplitude and pulse mode. Pulse mode is an important feature when homogenizing cell samples to prevent heat build-up and protein denaturation.

2.3 Methods of Protein Purification

Chromatographic techniques used throughout this study were immobilized-metal (nickel) affinity chromatography (IMAC), anion-exchange chromatography, and size exclusion (or gel filtration) chromatography.

Nickel-IMAC is a method that captures recombinant proteins with a poly-histidine tag (usually of six or more) attached to the N- or C-terminal. Resin beads are covalently attached to linker chains bound to nitrilotriacetic acid (NTA) which acts as a chelator to immobilize nickel ions. The free electrons from the nitrogen atoms of the histidine side group interact with the nickel ion by coordination hence the capture of recombinantly-tagged proteins. In theory, all remaining contaminant proteins without a histidine tag will flow through the column, past the nickel resin. However due to the large content of non-specific protein within the crude sample, protein-protein interactions are bound to occur therefore the resin is typically washed with a large volume of buffer containing 5mM of imidazole. To release the target protein from the resin, increasing concentration of imidazole is added because of its similar structure to histidine, imidazole will compete for the nickel sites.

Ion-exchange chromatography is a method that separates proteins based on their net charge. In anion-exchange, the column resin is positively charged and attracts proteins that are negatively-charged. The net charge of a protein can be manipulated to become negative-charged by suspending it in a buffer with a pH above its isoelectric point (pI). The release of protein from the column is achieved by introducing chloride ions (salt; NaCl) at increasing concentrations to compete for the positive sites.

Finally, gel filtration is the process that separates proteins based on molecular mass. The packed resin is a porous matrix with varying molecular weight cut-offs. In theory, larger proteins or proteins with larger hydrodynamic radii will flow through the column faster and elute first because they do not travel through the pores of the resin. Conversely, smaller proteins below the molecular weight cut-off will elute slower as they travel a longer distance through the pores of the resin.

2.4 Electrophoretic Separation of Proteins with SDS-PAGE

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is an electrophoretic technique used to separate a sample of proteins based on molecular weight. A typical SDS-PAGE gel will consist of a 4% stacking gel over-top a 12% resolving gel. The stacking gel is usually made at pH 6.8, while the resolving gel is made at pH 8.8. Both gels consist of a mixture of acrylamide and bis-acrylamide solutions that polymerize and form crosslinks with the addition of TEMED and APS. The purpose of the stacking gel is to initially line up the protein samples so that migration across the resolving gel occurs in uniform. Migration of the protein should run from the negatively charged cathode to the positively charged anode in the buffer system. Tertiary and quaternary structures of the protein is denatured and linearized with heat and the reducing agent, 2-mercaptoethanol, is useful for cleaving disulfide bonds. Proteins are then coated with SDS, giving them an overall negative charge to allow for migration across the gel using an electric current; higher molecular weight proteins will move slowly, while lower molecular weight proteins will be found close to the bottom because they move faster. Visualization of the proteins involves staining the gel with Coomassie stain and then de-staining either with water or destaining solution. Heating up the stain and de-stain

solutions help decrease the time, otherwise the process may take from a few hours to overnight before proteins are visible on the gel.

2.5 Spectral Properties of Proteins for Determining Concentration

Fluorescence spectroscopy can be applied to study intrinsic and extrinsic fluorescent properties of a protein. The three residues used to measure intrinsic fluorescence are tryptophan, tyrosine, and phenylalanine with optimal absorption wavelengths at 280, 274, and 257nm, respectively. It is very common to measure fluorescence via tryptophan because it typically produces higher quantum yields due to overlapping absorption spectra with tyrosine and phenylalanine. To determine protein concentration, Beer's Law is applied, where the predicted extinction coefficient is retrieved from inputting the protein sequence into the online bioinformatics tool, ExPASy ProtParam. The concentration units in M (mol/L) can be further converted into mg/mL given the calculated molecular weight of the protein (also from ExPASy ProtParam).

2.6 Enzyme Kinetic Activity Measurement Using Spectroscopic Characteristics of Nitrophenol Release from the Substrate

The kinetic rate and Michaelis-Menten variables can be determined for specific enzymatic activity through the cleavage of an artificial carbohydrate substrate with an attached nitrophenol moiety. This kinetic assay is particularly useful for enzymatic reactions such as glycoside hydrolases with mechanisms known to cleave apart sugars between the O-glycosidic bonds. The cleavage of a simple sugar (i.e. galactose, glucose, etc.) covalently bound to a nitrophenol group allows for colorimetric determination of activity as well as quantitative measurement by measuring the absorption at a wavelength

of 405nm. The artificial substrate is initially clear in color, but when nitrophenol is released it turns a yellow color due to the conjugated system from the resonant structure of the nitrophenolate ion. Consequently, the reaction is stopped with the addition of sodium hydroxide to alkaline pH to convert any remaining nitrophenol to its ionic form. The method not only allows for kinetic measurement, but also the determination of enzyme specificity to sugar linkages (i.e. ortho- or para- isomers). An assay can be easily set-up in a cuvette or microplate which allows for efficiency of producing replicate reactions for testing in a range of buffers with different pH values or varying concentrations of substrate over time. The kinetic rate of an enzyme can be calculated using the known extinction coefficient of nitrophenolate (18000 M^{-1} cm^{-1}) and Beer's law ($A = \varepsilon cl$, where 'A' is absorption, 'Ɛ' is extinction coefficient of nitrophenolate, 'c' is concentration, and 'l' is path length).

2.7 Site-Directed Mutagenesis of Potential Active Residues

To pinpoint the potential active residues that may reveal significant changes in kinetic activity of an enzyme, site-directed mutagenesis is used to produce such mutants with substitution of select amino acids. Complimentary primers should encompass the desired amino acid for substitution with approximately 15 base pairs flanking either ends of the codon for the selected amino acid. High-fidelity polymerases are typically used for these PCR reactions to allow full extension of the plasmid. The PCR product should also be treated with DpnI digestion to remove any parental methylated DNA that is leftover which also decreases the number of false positives when screening. Amino acids should be substituted with residues of similar R-groups (Ex: aspartate substituted with

asparagine) to prevent significant structural changes in the overall protein in order to investigate only kinetic changes.

2.8 Examination of Protein Sample Distribution and Stability with Dynamic Light Scattering

Dynamic light scattering is a non-invasive technique used to measure size distribution of particles in a sample including proteins. It can be used to distinguish particles with large differences in molecular mass in cases with protein that may form as monomers or dimers, or the presence of aggregation. A typical set up for analysis via DLS involves a laser or monochromatic light source that directs a beam towards the liquid sample contained in a microplate with a clear bottom and opaque sides to prevent reflective lighting. Upon contact of the incidental light beam with the sample, the wavelength of the refractive light will change as theorized using Brownian motion, which is the random motion caused by the collision of atoms or molecules. The change in frequency from the wavelength of the incoming light changes from the particle's movement. Using the collected change in wavelength, the translational diffusion coefficient (a variable in the Stokes-Einstein relation) is used to solve for the diameter of the particle (assuming that it is spherical). The Stokes-Einstein relation is as shown below:

$$D = \frac{kT}{3\pi\eta D_t}$$ where,

D = diameter of particle,

k = Boltzmann's constant,

T = temperature,

η = dynamic viscosity of the liquid/sample and,

D_t = diffusion coefficient.

Data collection from dynamic light scattering is also capable of computing molecular weight as well. The main prospect of this technique is population distribution to determine the purity within the sample after anion chromatography followed by crystallization to prevent formation of contaminant crystals or undesired precipitation if the protein happens to be unstable in particular mediums.

2.9 Crystallization and Optimization for Diffraction

The structure of a purified protein in crystal form can be determined by analyzing the diffraction patterns using X-ray crystallography. To first obtain such quality crystals for this method, a series of crystal trays must be setup to determine initial conditions to allow for crystallization; this typically requires a large selection of buffers and additives with varying conditions including pH, temperature, and concentration. Generally, the screening solutions of buffer and precipitants (collectively referred to as the 'mother liquor') is formulated in a reservoir, while the protein droplet is situated apart from the reservoir either on an inverted cover slip above the reservoir known as the 'hanging drop' method, or in a well raised above the reservoir known as the 'sitting drop' method. Over time, an exchange of vapor diffusion will occur between the protein drop and the reservoir to reach equilibrium and possibly reach crystallization. This method allows for a gentle and gradual process of dehydrating the protein droplet. Additionally, the protein droplet is usually mixed with varying ratios of mother liquor to aid with the eventual

equilibration. Successful crystallization outlines the packing of molecules in crystals with defined edges and three-dimensionality. Further crystal optimization from the initial screening involves tweaking the concentrations from the buffer conditions. Addition of carbohydrates that resemble potential endogenous substrates for the protein may also help to stabilize the structure for better crystal formation.

CHAPTER 3 – MATERIALS, EXPERIMENTAL DESIGN AND METHODS

3.1 Materials

The cloning of target gene, BT2857 in vector p15TV-LIC, was sent to us in collaboration with Dr. Alicia Lammerts van Bueren (Faculty of Mathematics and Natural Sciences, Microbial Physiology) from Gron Institute of Biomolecular Sciences & Biotechnology (Groningen, The Netherlands). Custom gene orders for SCO6594-97 in vector pET21a(+) were ordered from GenScript (Piscataway, NJ, USA). Competent NEB 5-alpha (catalogue #C2987) and BL21(DE3) (catalogue #C2527) *E. coli* cells were purchased from New England BioLabs Inc. (Ipswich, MA, USA). Novagen Tuner™(DE3)pLyss competent cells were purchased from EMB Millipore Corporation (Darmstadt, Germany). Restriction enzyme, Dpn I, CutSmart buffer, and Phusion® High-Fidelity PCR Kit were also ordered from New England Biolabs Inc. All primers were ordered through Integrated DNA Technologies (Coralville, IA, USA).

Tryptone powder, yeast extract, agar A, ammonium sulfate ($(NH_4)_2SO_4$), tris, potassium phosphate (KH_2PO_4), sodium (dibasic) phosphate heptahydrate (Na_2HPO_4), sodium (monobasic) phosphate dehydrate (NaH_2PO_4), glucose, D-lactose, hydrochloric acid, ferric chloride ($FeCl_3$), calcium chloride ($CaCl_2$), manganese chloride ($MnCl_2$), zinc sulfate ($ZnSO_4$), nickel (II) chloride ($NiCl_2$), deoxycholic acid sodium salt, Triton X-100, DNase I, sodium hydroxide (NaOH), sodium acetate trihydrate, urea, bromophenol blue, glycine, acetic acid (glacial), methanol, and D(+)-galactose were all purchased from BioBasic Canada Inc. (Markham, ON, Canada). Sodium chloride (NaCl), ampicillin sodium salt, glycerol, sucrose, lysozyme, magnesium chloride ($MgCl_2$), imidazole, 2-(N-

morpholino)ethanesulfonic acid (MES), ethylenediaminetetraacetic acid (EDTA), acrylamide 40% solution, tetramethylethylenediamine (TEMED), sodium dodecyl sulfate (SDS), Coomassie Brilliant Blue G-250, bis-tris methane, ethylene glycol were all purchased from BioShop Canada Inc. (Burlington, ON, Canada). Cobalt chloride ($CoCl_2$), Tween-20, n-Dodecyl-β-D-maltoside (DDM), sodium citrate dihydrate, ammonium acetate, 4-Nitrophenyl β-D-galactopyranoside, 2-Nitrophenyl β-D-galactopyranoside, 4-Nitrophenyl α-D-mannopyranoside, 4-Nitrophenyl α-D-glucopyranoside were all purchased from Sigma-Aldrich (St. Louis, MO, USA). JumpStart™ Taq ReadyMix™ and GenElute™ PCR Clean-Up Kit were also from Sigma-Aldrich.

Cupric chloride ($CuCl_2$) was from The British Drug Houses (BDH) and borrowed from Dr. DeBruin's lab. Ethanol was purchased from Commercial Alcohols (ON, Canada). 2% Bis-2™ Solution (bisacrylamide) and polyethylene glycol 8000 were purchased from Amresco (Solon, OH, USA). HisPur™ Ni-NTA resin and SnakeSkin® dialysis tubing clips (green; 50mm) were purchased from ThermoScientific (Waltham, MA, USA). Dialysis tubing closures (white; 110mm) were also ordered from Sigma-Aldrich. Ammonium persulfate, Fisherbrand® filter paper P4, and Fisherbrand® regenerated cellulose dialysis tubing were purchased from Fisher Scientific (Fair Lawn, NJ, USA). 2-Methyl-2,4-pentanediol was purchased from ACROS Organics (Geel, Belgium).

Precision Plus Protein™ Dual Color Standards was purchased from Bio-Rad Laboratories Inc. (Hercules, CA, USA). The ENrich™ SEC 650 10 x 300 column for gel filtration was also purchased from Bio-Rad. Resource™ Q columns for anion-exchange chromatography was ordered from GE Healthcare (United Kingdom). Amicon®

Ultracel®-10K regenerated cellulose centrifugal filters were purchased from Merck Millipore Ltd. (Germany). Microlytics MCSG1-4 crystal screen HT (1.7mL) were purchased from Anatrace (Maumee, OH, USA). Index Solution Set 1 reagents 1-48 (catalogue #HR2-144) and Solution Set 2 reagents 49-96 (catalogue #HR-144) were purchased from Hampton Research (Aliso Viejo, CA, USA). 3-Inch wide crystal clear sealing tape (catalogue #HR4-506) and polyethylene glycol 3,350 flakes (catalogue #HR2-591) were also from Hampton Research. High vacuum grease was ordered from Dow Corning (Midland, MI, USA) and micro cover glass for crystal trays were ordered from VWR International (Radnor, PA, USA). Multiwell™ 24-well polystyrene non-tissue culture treated flat bottom plates with low evaporation lids were ordered from Becton Dickinson (Franklin Lakes, NJ, USA) and INTELLI-PLATE® 96-well plates (catalogue #102-0001-03) were ordered from Art Robbins Instruments (Sunnyvale, CA, USA). Test (round-bottom) 96-well microplates were from Sarstedt (Montreal, QC, Canada).

All the water used in this study was filtered with the Millipak®20 MQ Millipore purification system. All centrifugation steps were done using Beckman-Coulter (California,USA) rotors, ThermoScientific Sorvall™ Legend™ Micro 21centrifuge, and/or Fisher Scientific standard mini centrifuge. The Amsco® Lab 250 LV 250 Laboratory Steam Sterilizer from Steris®(Mentor, OH, USA) was used for any autoclave treatment.

3.2 Experimental Design and Methods

3.2.1 Protein Expression and Isolation

Plasmids (BT2857) or custom genes (SCO6594-97) were transformed into BL21(DE3) *E. coli* competent cells. Transformations involved incubation of 1µL of plasmid with 50µL of competent cells in micro-centrifuge tubes on ice for half-hour, followed by heat shock at 42°C for 45 seconds, and placing them back on ice for 3 minutes. 200µL of Luria-Bertani (LB) broth was added to the tubes and incubated at 37°C for one hour at 240rpm. Transformations were then spread on ampicillin-resistant agar plates and incubated at 37°C overnight to allow for growth. The next day, a colony was picked from the plate and grown in 10mL of LB containing 100mM ampicillin at 37°C overnight at 240rpm. The 10mL cultures were used to inoculate 2L of LB with 100mM ampicillin and grown at the same conditions until they reached an OD_{600} of ~0.7. Overexpression of protein proceeded with the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) and turning the temperature down to 16°C for overnight at the reduced speeds of 160rpm. Cells were harvested by centrifugation at 8000xg for 12 minutes. Cell pellets were either stored at -20°C for later use or immediately treated with chemical lysis for protein isolation.

For chemical lysis, cell pellets were re-suspended in one volume of sucrose solution for 10 minutes or until completely dissolved, followed by the addition of 1mg/mL of lysozyme to mix for another 10 minutes. Two volumes of deoxycholate solution (50mM Tris-HCl pH 7.5) were added to stir for 10 minutes or until the mixture became a viscous consistency. 1M $MgCl_2$ and DNase I solution was added and stirred for another 10 minutes. If the final mixture still appeared viscous, then a second chemical lysis was performed by repeating the above steps until the consistency became watery. Cell debris was separated from protein solution by centrifugation at 15000xg for 35

minutes. The pellet was assumed to consist of cell debris, while the supernatant was kept and assumed to be crude protein extract.

In situations where the protein was insoluble and appeared in the pellet, sonication was performed to screen for solubility in different buffers instead of the standard 50mM Tris-HCl pH 7.5 in the deoxycholate solution. To do so, small samples of the cell pellet were divided into separate micro-centrifuge tubes and 1mL of buffer solution (with pH ranging from 3-10) was added. Samples were pulsed for 30 second intervals with amplitude of 40% and paused for 30 seconds inbetween to allow for cooling. All tubes were placed on ice during the sonication process to prevent protein denaturation. The sonicated solutions were centrifuged at 12000rpm for 10 minutes at room temperature, then the pellet and supernatant was run on SDS-PAGE to determine the condition/buffer in which the protein was soluble. The successful buffer was then replaced in the deoxycholate solution of the chemical lysis, or a large scale sonication protocol was used.

Immobilized nickel affinity chromatography was preceded by incubating the crude protein extract with nickel-NTA resin in 50mL falcon tubes for 1 hour on ice or at 4°C on a VWR rocking platform. The protein and nickel-NTA resin mixture was then poured into a glass Econo-column (Bio-Rad) and allowed to drain into a 250mL flask (labelled as ‘flow-through’). Once the resin settled evenly at the bottom of the column, 50mL of wash buffer (50mM Tris-HCl pH 8, 5mM imidazole, 300mM NaCl) was added with care not to disrupt the resin and collected in a 50mL falcon tube (labelled as ‘wash’). Elution buffers contained increasing concentrations of imidazole (10, 25, 50, 75, 100, 200, 300, 400, 500mM) and were each collected as 15mL fractions. The resin was cleaned

with one column volume of water, one column volume of NaOH, another column volume of water, and then stored in 20% (v/v) ethanol.

3.2.2 SDS-PAGE

To evaluate protein expression and isolation, the pellet (after lysis), crude protein extract, 'flow-through', 'wash' and all elution fractions were kept to run on SDS-PAGE. 12µL of each fraction was mixed with 3µL of 5X SDS-PAGE sample buffer in micro-centrifuge tubes. Each tube was heated at 95°C for 10 minutes, and then loaded into the SDS-PAGE gel alongside the Precision Plus Protein™ Dual Color Standards ladder. SDS-PAGE was run at 180V for 42 minutes or until the dye-front was 1cm from the bottom of the gel.

Coomassie stain was added to the gel in a Tupperware container and heated for 30 seconds in the microwave. The container was agitated for 30 seconds and the stain was poured back into its bottle through filter paper. Typically, water was used to de-stain the gel as visualization usually occurred almost immediately if a good yield of protein was present. Otherwise, de-staining solution was used and heated for 30 seconds in the microwave and agitated on the rocking platform until protein bands were present. The gel was imaged using Bio-Rad's Molecular Imager VersaDoc™ MP 4000 System and Quantity One® software to manipulate and export images.

3.2.3 Dialysis and Protein Concentration

The SDS-PAGE gel was used to determine which elution fractions contained the least amount of contaminating bands (protein) while still containing the band for the target recombinant protein. The process of dialysis was to decrease the concentration of

imidazole in the sample to prepare it for other methods of protein purification. Dialysis is a gradual technique for removing imidazole through exchange of molecules between the sample and surrounding buffer by diffusion across the concentration gradient to reach equilibrium. All of the elution fractions that met this standard was pooled into one dialysis tubing bag and clipped at both ends to prevent leaking. The dialysis bag was submerged into a 2L beaker containing MQ water with a stir bar and was allowed to mix in the cold room (4°C) for 1 hour. Afterwards, the dialysis bag was transferred to another 2L beaker containing 50mM Tris-HCl pH 8 and again was mixed in the cold room overnight.

Protein concentration was performed using the Amicon® Ultracel®-10K regenerated cellulose centrifugal filters from Millipore. Before use, the filter was filled with MQ water and centrifuged at 4000xg for 3 minutes to wet the membrane. Subsequently, the filter was filled with protein sample and centrifuged at 4000xg for 10 minutes using the Allegra X-14R centrifuge from Beckman Coulter, and was refilled with more protein sample between every interval. If the bottom of the sample in the filter appeared a brown colour, then a disposable pipette would be used to gently agitate the solution to re-suspend the concentration of protein. Typically, protein samples were centrifuged in the filters until they reached a concentration of at least 10mg/mL for crystal tray screening. For anion-exchange chromatography, the protein was not concentrated, and for gel filtration, the protein sample would need to be concentrated to ~3mL.

3.2.4 Purification Using Anion-Exchange Chromatography and Gel Filtration

Purification by anion-exchange chromatography was performed using protocols from the UNICORN 6.4 Workstation program to control the ÄKTA pure fast protein liquid chromatography (FPLC) system purchased from GE Healthcare (Missisauga, ON, Canada). All lines were primed with their respective buffers and solutions at a flow rate of 3ml/min until the UV signal at 280nm flat-lined to ensure no previous protein was left on the Resource™ Q column. The standard protocol for anion-exchange chromatography was selected and run with the sample protein. The wash buffer used for the system was 50mM Tris-HCl pH 8, and the elution buffer used was 50mM Tris-HCl pH 8 with 1M NaCl. The elution gradient was set to increase from 0-100% of salt, leading to the gradual release of protein as the chloride ions competed for interaction with the column. The elutions were collected as 2mL fractions. The elution fractions that encompassed the peak with the greatest UV_{280} signal was run through SDS-PAGE to confirm the presence of the target protein.

Purification by gel filtration depended on the quality of the protein sample after it was purified through nickel-IMAC and anion-exchange. If the sample appeared to contain contamination with a large enough difference in molecular weight to the target protein, then gel filtration was an appropriate choice for further purification. To prepare, the sample was concentrated to a volume of ~3mL and injected into the 1mL sample loop on the Bio-Rad NGC™ Chromatography System. The ChromLab™ software was used to control the load of sample onto the ENrich™ SEC 650 column, followed by the flow of wash buffer (50mM Tris-HCl pH 8) at a flow rate of 1mL/min. All elutions were collected as 1mL fractions. The elution fractions that encompassed the peaks with a UV_{280} signal was run through SDS-PAGE to determine the presence of the target protein.

40µL of each elution fraction was also loaded into a sterile 384-well black polystyrene plate with clear flat-bottom for DLS using the DynaPro PlateReader-II from Wyatt Technology(Santa Barbara, CA, USA). The DYNAMICS® software was used to analyze protein samples and indicate the purity of the particle population. Typical results would reveal a single peak of size distribution with low percentage of polydispersity indicating that the protein sample exists in a unimodal state. The calculated molecular weight might also be close to the predicted value depending on the shape of the protein (see section 2.8).

3.2.5 Determining Protein Concentration

The concentration of protein in a sample was determined by measuring the fluorescence of tryptophan residues at a wavelength of 280nm. First, the Genesys 10S UV-Vis spectrophotometer from ThermoScientific was blanked with MQ water contained in a quartz cuvette by Hellman® Analytics. To the same cuvette, 5µL of protein sample was added to the MQ water and mixed by inversion while covering the top with parafilm. The fluorescence spectra from 260-300nm was collected, but only the value at 280nm was used to calculate protein concentration. Beer's law was used to calculate the protein concentration:

$$A = \varepsilon \cdot c \cdot l$$

Where 'A' = absorbance, 'Ɛ' = extinction coefficient ($M^{-1}cm^{-1}$), 'c' = concentration (M), and 'l' = path length (cm). The path length of the cuvette was 1cm.

By re-arranging the formula, the concentration in M (mol/L) can be isolated:

$$c = \frac{A}{\varepsilon \cdot l}$$

To convert the units of concentration (M) to mg/mL, the molecular weight of the protein is required:

$$Concentration\left(\frac{mg}{mL}\right) = concentration(\frac{mol}{L}) \times MW(\frac{g}{mol}) \times \frac{1L}{1000mL} \times \frac{1000mg}{1g}$$

Also, the dilution factor must be taken into consideration:

$$Actual\ concentration = concentration\ \times \frac{1005uL}{5uL}$$

3.2.6 Crystal Screening and Expansion Plate Set-up

96-well crystal trays for preliminary crystal screening were set-up using the automated Gryphon system to draw from a series of crystal screen conditions. The protein to reservoir ratios used for this sitting drop method was 1:1 and 1.5:1. All trays were sealed with clear tape and stored at 18°C. They were observed under a microscope after a day of incubation. Successful hits were noted on the scoring sheet for further optimization of crystal growth.

Expansion trays were set up in 24-well plates using the hanging drop method. Each expansion tray was a compilation of the successful hits with slight changes in concentrations or additives to test for better crystal growth. Expansion trays were prepared by greasing the rims of each well with vacuum grease, followed by formulating the reservoir solution in each well. The lab bench was sterilized with ethanol before placing six glass cover slips on it. To each cover slip, the same ratio of protein to

reservoir droplet was produced but at larger volumes (e.g. 2µL of protein to 2µL of reservoir compared to the 0.2µL used in the 96-well plates). The top droplet was usually assigned as the 1:1 ratio, while the bottom droplet was 1.5:1. The placement of the droplets was also located closer to the center of the cover slip to prevent disruption from the vacuum grease or the well edges. The cover slip was then inverted and placed over the respective well with slight pressure applied to ensure the seal. The next row of six cover slips was set-up and steps were repeated until the entire tray was completed. The crystal tray was stored at 18°C and observed for crystals after an appropriate time period.

3.2.7 Enzyme Kinetic Assays

The kinetic activity of the enzyme was tested using para-nitrophenyl-type substrates for easy detection and quantification of reactions. All reactions were prepared in 96-well microplates and absorbance values collected with the Synergy HT Multi-Mode microplate reader from BioTek (Winooski, VT, USA). First, the pH profile of the enzyme was determined by incubating 1µL of 50nM enzyme with 10µL of 50mM para-nitrophenyl β-D-galactopyranoside (PNP-Gal) in a range of buffers at different pH to a final volume of 100µL.

Table 5. List of buffers used for pH profile in enzyme kinetics.

Buffer (50mM)	pH
Sodium citrate	3.0
Sodium citrate	4.0
Sodium citrate	5.0
Sodium citrate	5.5
Bis-Tris	6.0
Bis-Tris	6.5
Bis-Tris	7.0
Bis-Tris	7.5
Tris-HCl	8.0
Tris-HCl	8.5
Tris-HCl	9.0
Glycine-NaOH	10.0

All reactions were performed in triplicate at room temperature (~25°C) and at 37°C for 10 minutes. The optimal buffer, pH, and temperature for the enzyme reaction were used to continue with kinetic characterization. To produce the Michaelis-Menten curve, 1µL of 50nM enzyme was incubated with the substrate at increasing concentrations. Reactions were incubated for 15 minutes with absorbance collected every 45 seconds.

Table 6. Typical microplate set-up of enzyme kinetics for producing a Michaelis-Menten curve.

Well	Volume of 50nM Enzyme (µL)	Volume of 50mM Substrate (µL)	Volume of Buffer (µL)
A1→A3	1	1 (0.5mM)	98
A4→A6	1	2 (1mM)	97
A7→A9	1	3 (1.5mM)	96
A10→A12	1	4 (2mM)	95
B1→B3	1	5 (2.5mM)	94
B4→B6	1	6 (3mM)	93
B7→B9	1	12 (6mM)	87
B10→B12	1	18 (9mM)	81
C1→C3	1	24 (12mM)	75
C4→C6	1	30 (15mM)	69
D1	0	1	99
D2	0	2	98
D3	0	3	97
D4	0	4	96
D5	0	5	95
D6	0	6	94
D7	0	12	88
D8	0	18	82
D9	0	24	76
D10	0	30	70

*All reactions were performed in triplicate. Final concentrations of substrates are indicated in brackets beside the volumes. Wells D1-D10 represents blanks.

The rate (Abs/min) was calculated for each curve collected from each concentration of substrate. The unit of each rate was converted to M/min using Beer's

Law and the Michaelis-Menten curve was plotted using MyCurveFit (https://mycurvefit.com/) (unless otherwise stated). The Michaelis-Menten variables, V_{max} and K_M, were provided by the equation of the curve. k_{cat} was calculated using the following formula:

$$k_{cat} = \frac{V_{max}}{E_t}$$

Where 'E_t' = total enzyme concentration.

3.2.8 Galactosidase Activity Confirmation

Enzymes that were tested positive for activity with PNP-Gal were transformed into the Tuner(DE3)pLyss cells to re-confirm β-galactosidase activity. Tuner cells lack the presence of lacZ which codes for endogenous β-galactosidase. A notable disadvantage of Tuner cell is the significant decrease in protein expression, therefore a larger volume of cells must be cultured and harvested to reach the same yield as that produced by BL21(DE3) cells.

3.2.9 Site-Directed Mutagenesis

Two mutants of BT2857 were selected in attempt to elucidate active site residues. Primers were created using QuikChange Primer Design for the mutants, D208N and E311Q. Upon arrival, primers were re-suspended in nuclease-free water for a final concentration of 1mM. Working stocks of primers were prepared by further dilutions to final concentrations of 100uM.

Table 7. PCR reaction set-up for D208N mutant.

PCR Component	Volume Added (µL)
BT2857 plasmid (50-100ng/µL)	5
5X HF Buffer	10
10mM dNTPs	1
Forward Primer (100uM)	1
Reverse Primer (100uM)	1
$MgCl_2$	0.5
DMSO	0.5
Phusion polymerase	0.5
Nuclease-free water	30.5
Total reaction volume	**50µL**

Table 8. Thermocycler conditions used for PCR reactions.

Step	Temperature (°C)	Time (s)	
Initial Denaturation	98	30	
Denaturation	98	10	Repeat cycles 34X
Annealing	66.8-0.2°C/cycle	30	
Extension	72	210	
Final Extension	72	600	

PCR products were run on a 0.7% (w/v) agarose gel at 100V for 40 minutes with a 1kb NEB ladder for determining DNA size. Gels were imaged using the Molecular Imager VersaDoc™ MP 4000 System. PCR products that were observed around the correct size were treated with restriction enzyme, DpnI, to cleave any remaining parental DNA. The reaction was then cleaned using GenElute™ PCR Clean-Up Kit and transformed into NEB 5-alpha *E.coli* competent cells. Colonies were picked and grown in 10-mL cultures

for plasmid isolation to be sent for sequencing at Robarts Research Institute (London, ON, Canada).

CHAPTER 4 – BIOINFORMATIC REVIEW

4.1 Model prediction and Functional Hypothesis for BT2857

The protein sequence of BT2857 was run through a series of bioinformatics tools in order to a build a hypothesis on the structural and functional aspects of the protein. The sequence of BT2857 was inputted in TMHMM Server v. 2.0 to predict protein topology and whether the structure possessed any membrane-spanning domains (Figure 2) [1]. This was an important source of information for the expression of BT2857 and stability in solution.

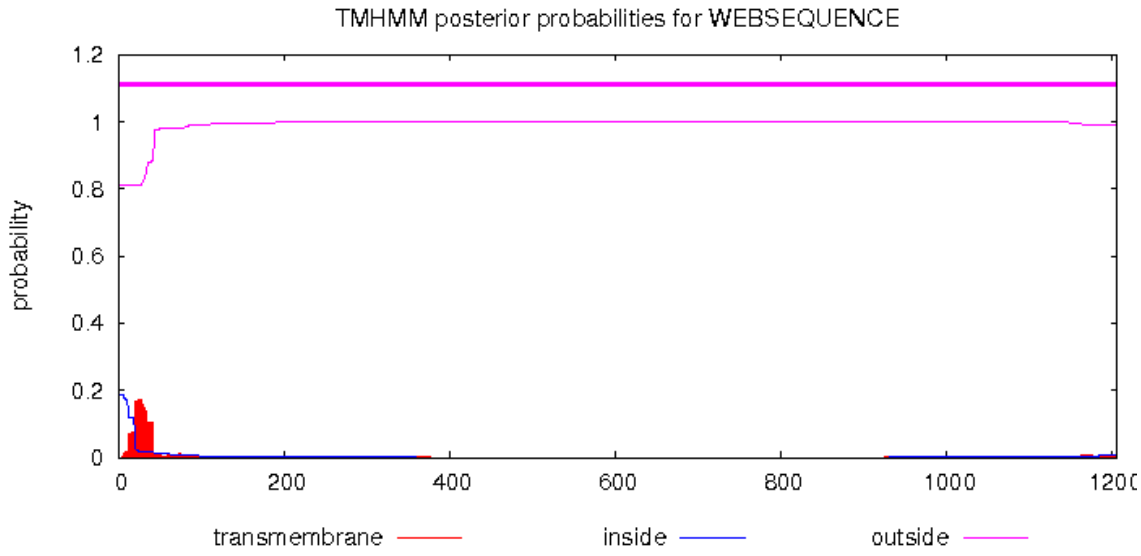


Figure 3. TMHMM Server v. 2.0 output for BT2857 [1]**:** A short peptide (red) is predicted to be located in the transmembrane, while the remaining sequence is found on the outside with a probability of 1 (pink) indicating that it is not a transmembrane protein.

The output from TMHMM depicted that a small portion of the N-terminus of BT2857 was predicted to possess transmembrane components, however the probability is

considerably low (~0.2). The remainder of the protein was found to be located outside with a probability above 1 giving confidence that BT2857 was suggested to be targeted extracellularly. In the results, the expression and isolation of BT2857 reflected the output by TMHMM. InterPro was used to predict the presence of domains or repeats based on sequence similarity [80]. The only output from InterPro suggested a detailed signature match to a cysteine knot. Cysteine knots are made up of six cysteine residues that link to form active forms of protein dimers in growth factors. This prediction suggested that the structure of BT2857 might exist in a dimer form or possessed more than one domain. To validate the structural predictions, RaptorX and Phyre^2 were used to produce models with characterized proteins of sequence similarity (Figure 3 and 4). Although both predicted models possessed low confidence in similarity with the listed output, the common structural aspect was observed in the beta-folds that may suggest a catalytic or substrate-binding domain. IntFOLD and SWISS-MODEL were two other servers used to predict the structure of BT857 and were able to cover a larger range of the sequence to produce a more complete model (Figure 5 and 6).

Figure 4. RaptorX predicted model of BT2857 based on template PDB ID:1EUT Chain A [2]**:** 100% of the input sequence was modelled based on the best template PDB ID: 1EUT. The structure is predicted as three domains forming barrel-like shapes and a long tail of residues towards the C-terminal.

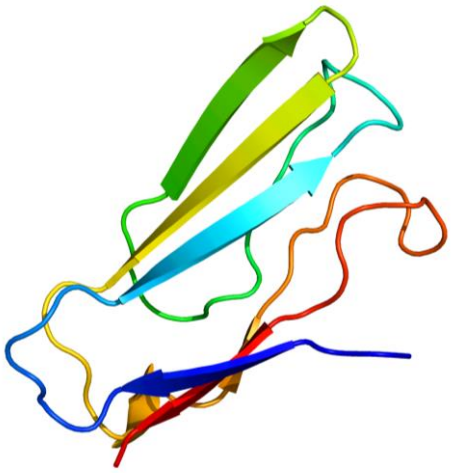


Figure 5. Phyre^2 predicted model of BT2857 based on template PDB ID:1RHF, an immunoglobulin superfamily fold: The model aligns with the N-terminal and covers 21% of the query sequence with 14% confidence in similarity. Residues G36-V120 are modelled [3].

The predicted models from IntFOLD and SWISS-MODEL provided better insight on the possible domains and position of active site residues within BT2857. Similar to the model from RaptorX, the output from IntFOLD predicted three distinct domains with beta-barrel like folds, while SWISS-MODEL only predicted two domains. The inconsistencies among models were most likely a result of low confidence and similarity between template and query sequences.

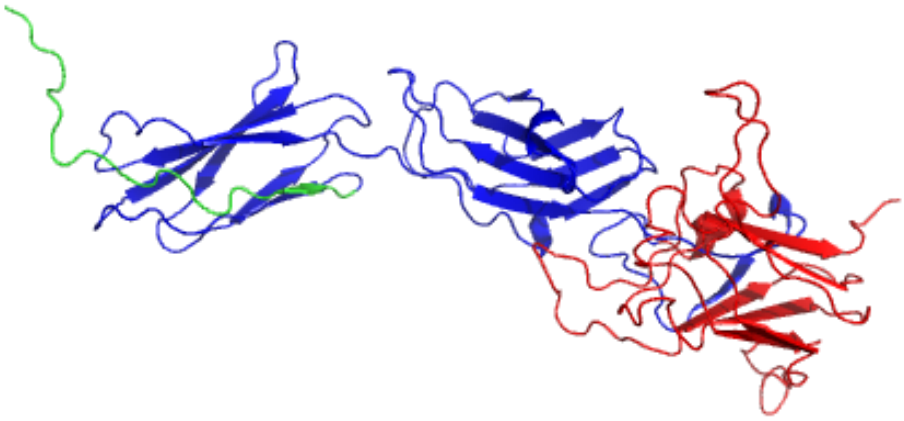


Figure 6. IntFOLD predicted model of BT2857 based on template PDB IDs: 1W8N and 4LQR [4]**:** The predicted model had medium confidence level with a p-value of 0.042. The input sequence displayed high values of residue error for the N-terminus (~1-200) compared to the templates. Residue error is based on the distance (in Angstroms) of an atom to the alpha-carbon (Cα). The C-terminus also displayed peaks of high residue error indicating that the overall predicted model is not entirely accurate. The overall structure is predicted to be made of beta-strands that form three barrel-like domains. The N- (blue) and C- (red) terminal truncations are highlighted on the model.

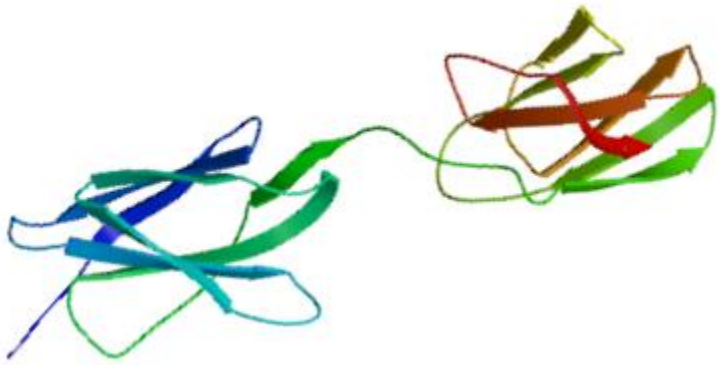


Figure 7. SWISS-MODEL predicted model of BT2857 based on template PDB ID: 4BQC [5]**:** The alignment covers residues 33-212 with a sequence similarity of 0.26 highlighting only the predicted structure of the N-terminus is made of beta-strands.

To investigate functionality, BT2857 was discovered to be experimentally validated as part of the polysaccharide-utilization loci (PUL) that could be found on the CAZy database (Figure 7) [6,55]. The PUL distinguishes genes organized around the susC/D gene pair for the transport and breakdown of complex glycans in the *Bacteroidetes* species. Upon expression in various conditions, studies demonstrated that aside from BT2857, several other proteins were upregulated and predicted to behave as an operon for carbohydrate uptake and regulation.

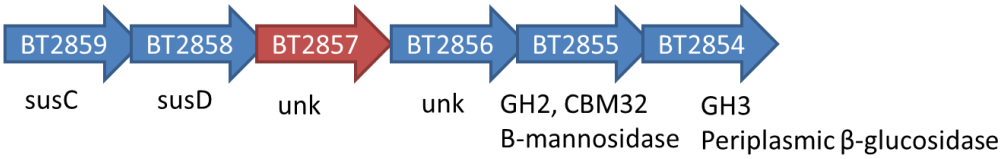


Figure 8. Experimentally validated upregulation of BT2854-59 as part of PUL labelled with putative functions: BT2858-59 are predicted to comprise of SusC and SusD domains used for carbohydrate transport into the cell. BT2856-57 remain unknown. BT2854 and BT2855 are predicted to belong to family of GH2 and 3, respectively. The known activity of the GH families are labelled below providing insight on potential substrates of the PUL [6].

Using the information from PUL, two characterized enzymes belonging to the gene cluster were revealed to have mannosidase and glucosidase activity, alluding to potential substrates to be tested with BT2857. Also, because *B. thetaiotaomicron* resides in the human gut, a series of endogenous substrates possessing mannose and glucose units were also possible substrates to be tested.

4.2 Predicted Functional Roles of SCO6594-97

Based on the cluster of *S. coelicolor* targets chosen, it was believed that SCO6594-97 may behave as a unit in carbohydrate metabolism much like a PUL for *Bacteroidetes* (Figure 8). Therefore the functional roles of SCO6594-97 were further investigated using a series of bioinformatics.

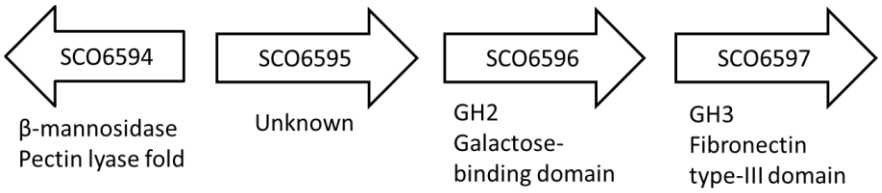


Figure 9. Organization of predicted functional roles of SCO6594-97: The activities of SCO6594-97 are predicted using InterPro and indicate that SCO6596 and SCO6597 comprise of domains belonging to GH family 2 and 3, respectively. The sequence prediction of SCO6594 resembles a pectin lyase fold where a majority of the structure is made of parallel beta-sheets.

The predicted function and fold of SCO6594-97 was collected from InterPro suggesting possible mannosidase activity as well as carbohydrate recognition for either substrate binding or cell adhesion. Open reading frames were explored upstream and downstream of this "operon" however no other predictions were revealed. Due to the lack of information available on SCO6595 for conclusive predictions on the structure and function, the section was omitted below. InterPro output for SCO6595 predicted an unknown domain, however further analyses did not uncover and possible functional roles.

4.2.1 Functional Predictions of SCO6594

SignalP 4.1 server was one of the primary sources used to screen the target sequence for presence of a signal peptide when ordering custom genes [7]. Signal

peptides are undesirable for protein localization within the cell because isolating protein may pose a challenge for solubility. Output for SCO6594 indicated a cut-off signal around Asp47 shown by the spike in C- and Y-score above 0.5 (Figure 9). The construct for SCO6594 was designed with a cut-off at Tyr72 after combining the output from all the tools. Tyrosine was reasoned as the better starting residue for protein solubility rather than a charged aspartate residue.

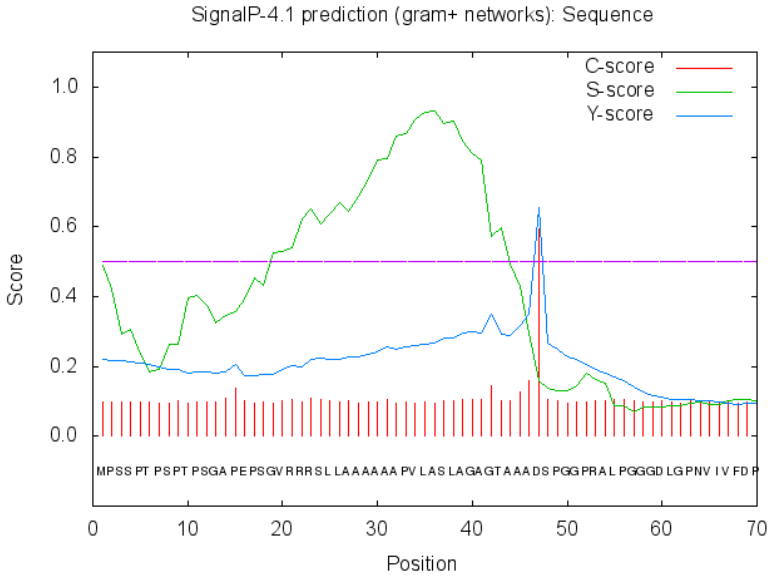


Figure 10. SignalP 4.1 output for SCO6594: The graph indicates a cut-off signal at Asp47 for removal of signal peptide indicated by the peak from the Y-score [7].

InterPro results (not shown) indicated that SCO6594 possessed an N-terminal domain with beta-mannosidase activity and C-terminal pectin-lyase fold suggesting its role in the degradation of plant carbohydrate sources [80]. Lastly, Phyre^2 predicted sequence similarity to a glucan 1,3-β-glucosidase and fructotransferase (Figure 10) [3]. The predicted model depicted a bulky protein with an abundance of beta-fold connected through loops. There was possibility that SCO6594 is made up of two domains, hence

another construct cleaving the sequence nearly in half from T72-V238 was designed to isolate for beta-mannosidase activity as predicted from InterPro.

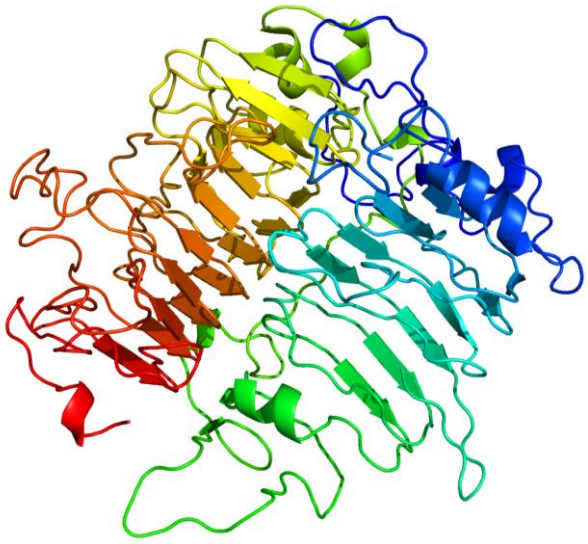


Figure 11. Phyre[2] predicted model of SCO6594 based on sequence homology with a beta-1,3-glucanase from *Phanerochaete chrysosporium*: The template sequence demonstrated 100.0% confidence and 95% coverage to the input sequence. The predicted model comprises of parallel beta-strands. Residues S12-V594 are modelled [3].

4.2.2 Functional Predictions of SCO6596

SignalP 4.1 output indicated the suggested cut-off signal at Ala42 which was taken into consideration when ordering custom genes [7]. BLAST was resourceful for indicating predicted domains based on sequence similarity to a database of unknown and characterized protein sequences [8].

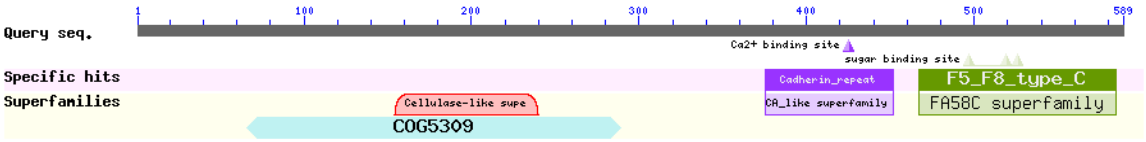


Figure 12. Basic local alignment search tool (BLAST) search of SCO6596: Results indicate predicted domains of a cellulase-like superfamily on the N-terminus, and cadherin repeat and F5/8 type C domain towards the C-terminus [8].

The BLAST search suggested a cellulase-like superfamily providing insight on cellulose and other similar carbohydrates as potential substrates (Figure 10) [8]. The cadherin repeat also suggested the requirement for calcium ions for protein activation or stabilization of interactions. Finally, the F5/8 type C domain did not reveal any predictions in overall SCO6596 function, but suggested the involvement in surface interaction or cell adhesion. Alternatively, InterPro was able to predict a glycoside hydrolase belonging to family 2 with galactose-binding domain uncovering more specific functions of SCO6596 in the hydrolysis of O-glycosyl compounds [80]. Phyre^2 also revealed sequence similarity to various beta-galactosidases including from *Arthrobacter sp. c2-2* and lacZ expressed as a protein from *Escherichia coli* (Figure 11). Truncations of SCO6596 were designed based on InterPro domain output and cross-referenced with Phyre^2 output to prevent cleavage within predicted secondary structures.

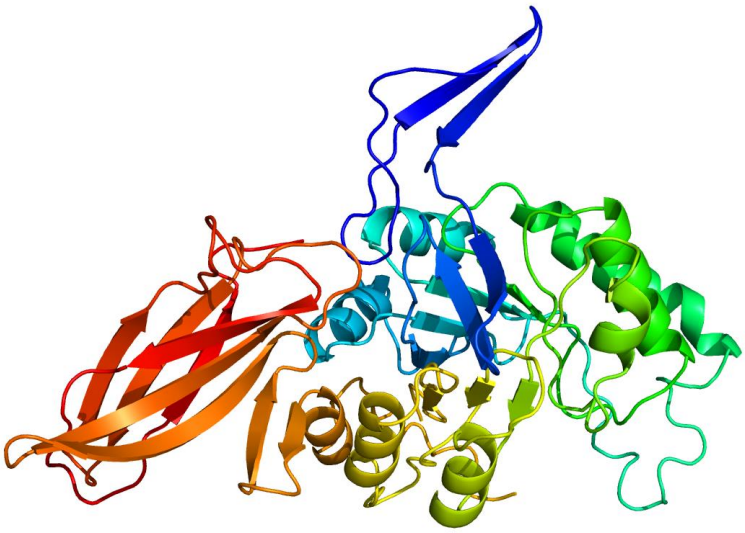


Figure 13. Phyre[2] predicted model of SCO6596 based on sequence homology with a beta-galactosidase from *Arthrobacter sp. c2-2*: The template sequence demonstrated 100% confidence and 63% coverage of the query sequence from the N-terminus. Residues R10-V443 are modelled [3].

4.2.3 Functional Predictions of SCO6597

A BLAST search of the protein sequence of SCO6597 revealed predicted glycosyl hydrolase family 3 (GH3) on both N- and C-terminus [8]. The C-terminal GH3 domain was also predicted to be involved in beta-glycan binding and catalysis. A fibronectin type III-like domain was also predicted at the C-terminal suggesting substrate recognition and binding. Output from InterPro also confirmed the prediction of a GH3 domain on either terminus providing insight on SCO6597 possible function in the hydrolysis of O-glycosyl substrates [80]. Finally, Phyre[2] revealed sequence similarity with a β-glucosidase from *Kluyveromyces marxianus* and *Streptomyces venezuelae* alluding to potential substrates made of glucose units [3].

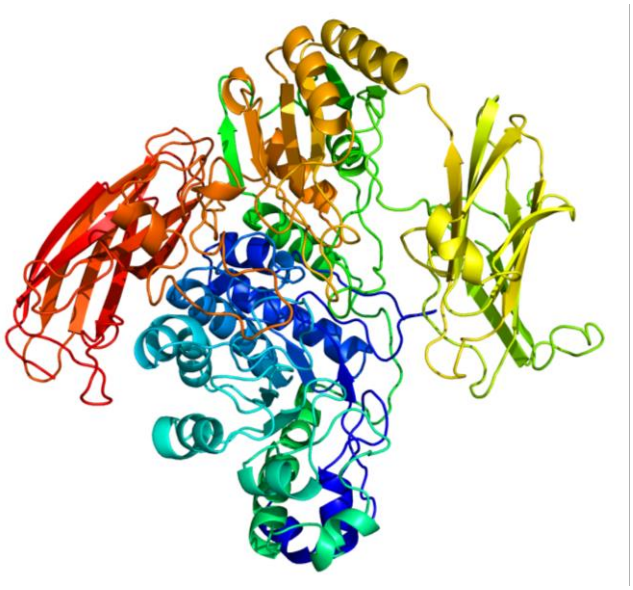


Figure 14. Phyre[2] predicted model of SCO6597 based on sequence homology with a beta-glucosidase from *Kluyveromyces marxianus*: The template sequence demonstrated 100.0% confidence and 93% coverage of the query sequence. Residues I12-A857 are modelled [3].

CHAPTER 5 – RESULTS

5.1.1 Expression and Purification of BT2857

The first objective was to express the recombinant protein, BT2857, which was achieved in 2L cultures of LB supplemented with ampicillin. The cells were harvested and underwent chemical lysis to isolate the protein. Following centrifugation to separate the cell debris (pellet) and supernatant (crude protein extract), the supernatant was incubated with nickel beads to run through nickel-IMAC (Figure 13).

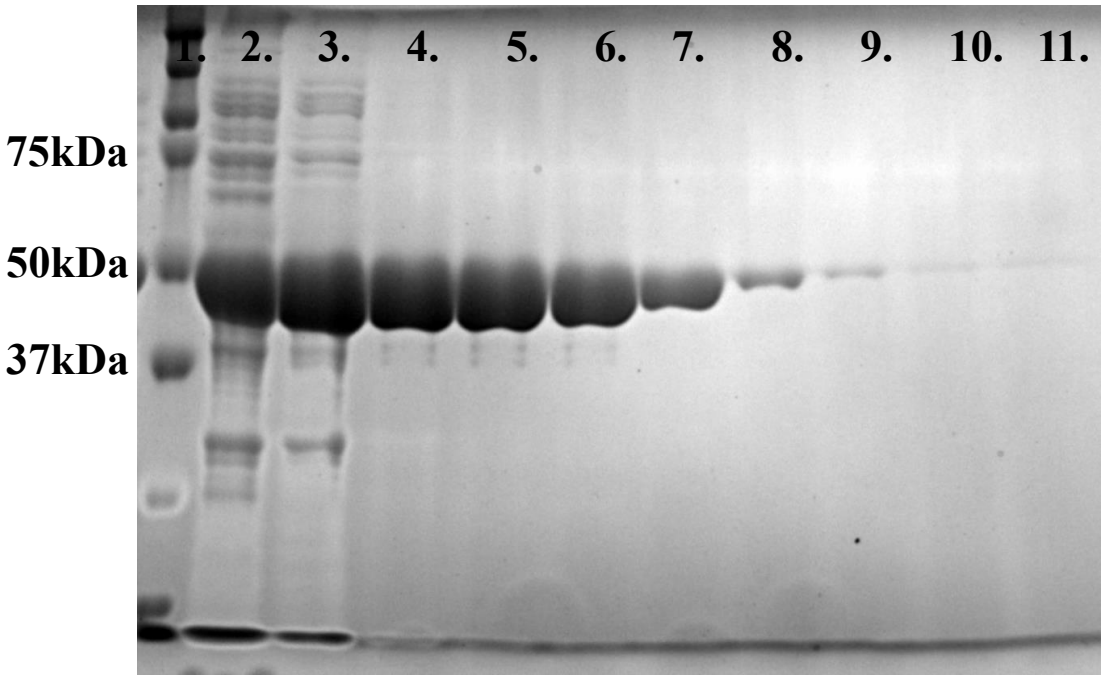


Figure 15. SDS-PAGE gel depicting the isolation of BT2857 (~46kDa) from crude protein extract via nickel-IMAC: 12% SDS-PAGE gel loaded with the elution fractions of increasing imidazole concentrations of 5, 10, 25, 50, 75, 100, 200, 300, 400, and 500mM from lanes 2-11. Lane 1 was loaded with Precision Plus Protein™ Kaleidoscope ladder.

An initial attempt to use gel filtration on BT2857 was conducted by pooling together fractions 4-7 from the nickel-IMAC (see Figure 13) and concentrating to a volume of 3mL. Unfortunately, gel filtration did not yield clean purification as the next step because of the contaminating bands that can be found directly below, as well as a few contaminating bands found above the target protein bands (Figure 14).

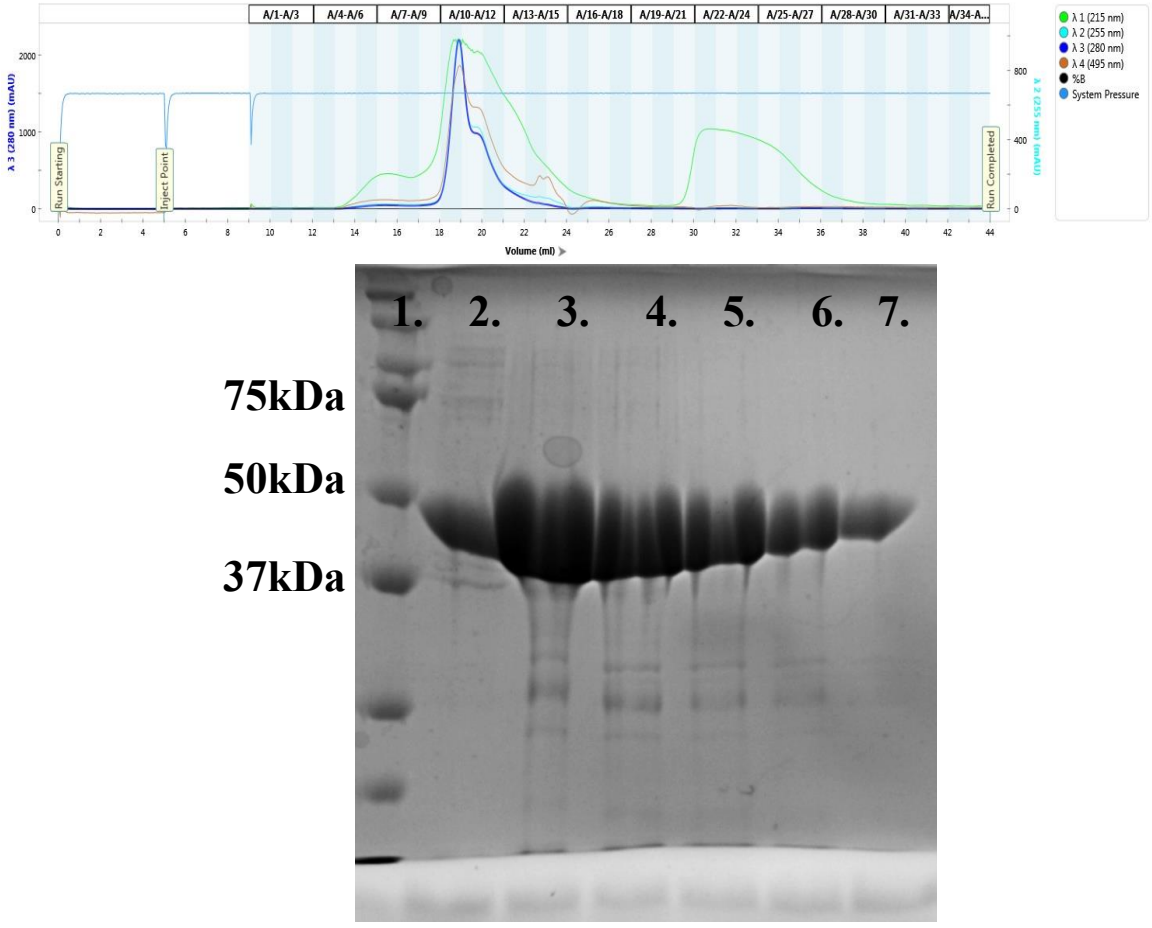


Figure 16. (Above) Chromatogram from gel filtration run, (Below) SDS-PAGE gel depicting the attempted purification of BT2857 (~46kDa) using gel filtration: 12% SDS-PAGE gel loaded with elution fractions from gel filtration. Lanes 2-7 were elution fractions A/9 to A/14 found under the peak of the chromatogram.. Lane 1 was loaded with Precision Plus Protein™ Kaleidoscope ladder.

Instead of using gel filtration, anion-exchange chromatography was attempted after nickel-IMAC instead which yielded better results (Figure 15). Correction to purification method can be observed as a sharper peak and less trailing in the chromatogram.

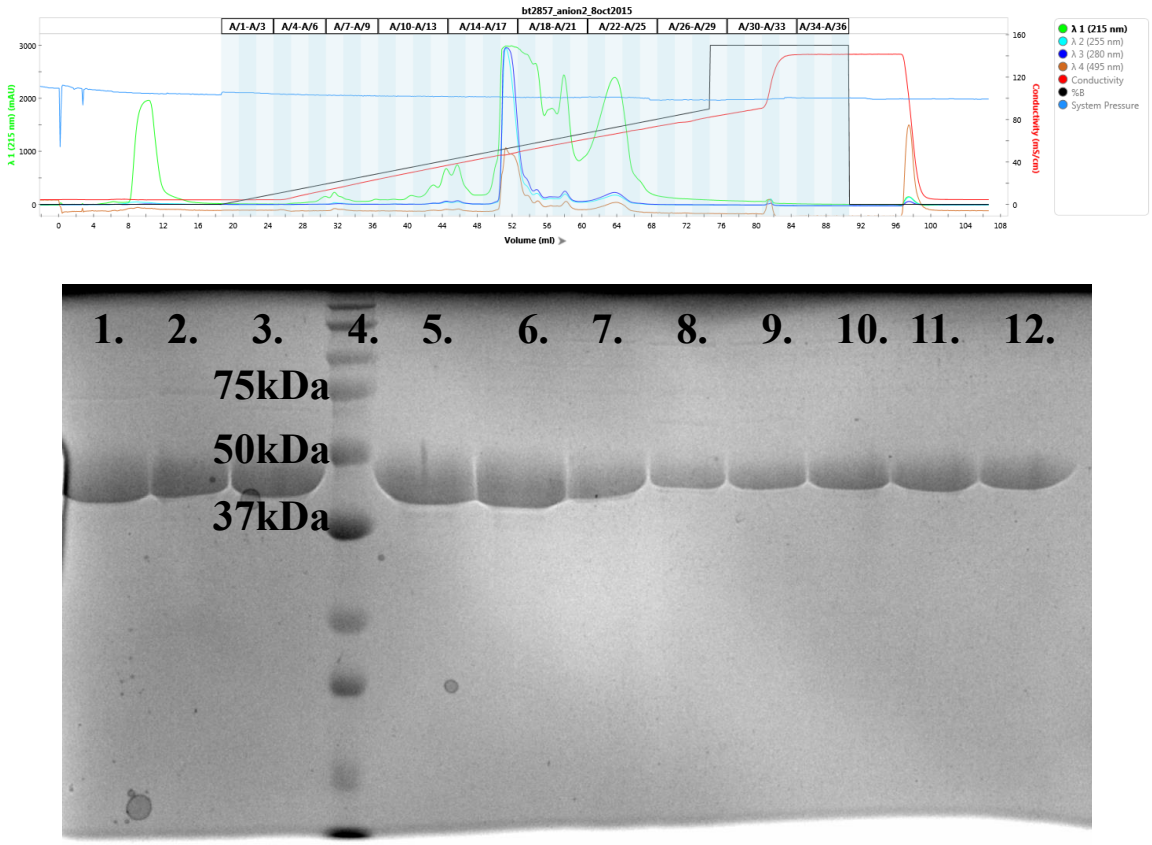


Figure 17. (Above) Example of chromatogram from anion-exchange chromatography, (Below) SDS-PAGE gel depicting the purification of BT2857 (~46kDa) using anion-exchange chromatography: (Above) The chromatogram is an example of BT2857 run through anion-exchange and how defined the peak appears with little tailing indicating the mono-dispersity of the sample. The chromatogram is an example from another run and does not correlate with the SDS-PAGE gel below. (Below) 12% SDS-PAGE gel loaded with elution fractions collected from an anion-exchange chromatography. Lane 4 was loaded with Precision Plus Protein™ Kaleidoscope ladder.

After successful purification of BT2857, the protein sample was concentrated in the centrifugal filters to a concentration of ~10mg/mL. A portion of the concentrated sample was used to prepare preliminary crystal trays for screening, while the leftover volume was used to test for kinetic activity. After a few days of incubation, the crystal trays were observed to show precipitation in many of the conditions indicating that the protein concentration might be too high or was unstable to crystallize as a full protein. After trial and error with decreased concentration, it was concluded that the protein was unable to crystallize due to the latter situation. Thus, the protein was truncated into N-terminal and C-terminal domains based off predicted models of BT2857 from bioinformatics sources (Figure 5). As a result, both domains have been successfully expressed and purified to some extent (Figure 16 and 17).

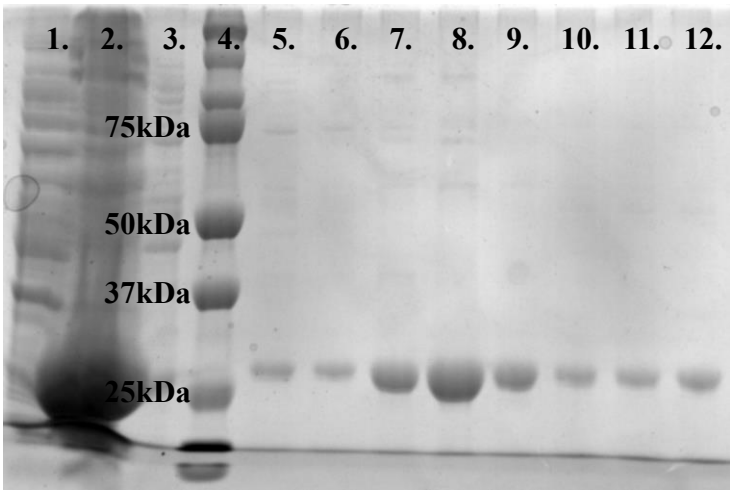


Figure 18. SDS-PAGE gel depicting the isolation of BT2857-N (~23kDa) using nickel-IMAC: 12% SDS-PAGE gel loaded with elution fractions of increasing imidazole concentrations (5, 5, 5, 10, 25, 50, 75, 100, and 500mM) in lanes 3, 5-12. The flow through and pellet were also run in lanes 1 and 2, respectively. The pellet indicates a large amount of protein that was not soluble perhaps dues to inadequate chemical lysis of the harvested cells or protein aggregation. Lane 4 was loaded with Precision Plus Protein™ Kaleidoscope ladder.

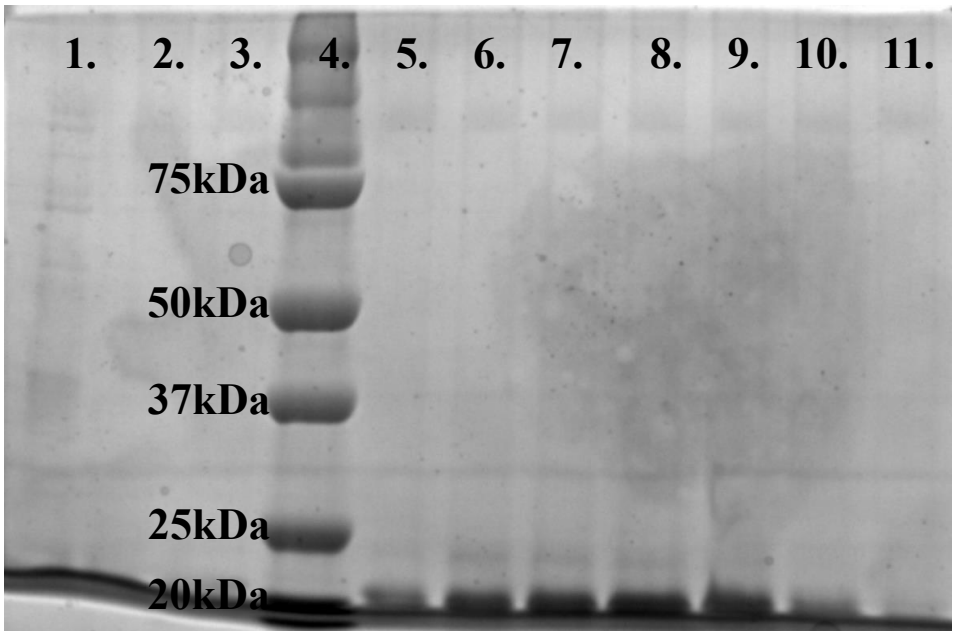


Figure 19. SDS-PAGE gel depicting the isolation of BT2857-C (~20kDa) using nickel-IMAC: 12% SDS-PAGE gel loaded with elution fractions of increasing imidazole concentrations (5, 5, 5, 10, 25, 50, 75, 100, and 500mM) in lanes 2,3, 5-11. The flow through was loaded into lane 1. Lane 4 was loaded with Precision Plus Protein™ Kaleidoscope ladder.

BT2857-N has been isolated through nickel-IMAC, while BT2857-C has been isolated with nickel-IMAC and purified using gel filtration (not shown). BT2857-C has been successfully crystallized at various conditions and is in the process of diffraction for X-ray crystallography. Crystal formation for BT2857-C initially started as thin platelets at a point of nucleation (can be observed in Figure 18, Middle image, towards top right corner of droplet), and later developed into more three-dimensional shapes (Figure 18, Bottom image, top of droplet).

Figure 20. Crystal growth for BT2857-C from optimization trays: The crystallization of BT2857-C was achieved in the following conditions; (Above) 33mg/mL protein in 1% (v/v) glycerol, 0.1M sodium acetate pH 4.0, 17.5% (w/v) 3.35K PEG, 0.1M NaCl, and 0.1M ammonium acetate, (Middle) 33mg/mL protein in 1% glycerol, 0.1M sodium acetate pH 4.0, 17.6% 3.35K PEG, 0.1M NaCl, and 0.1M ammonium acetate, and (Bottom) 33mg/mL protein in 2.5% glycerol, 0.1M sodium acetate pH 4.0, 18% 3.35K PEG, 0.1M NaCl, and 0.1M ammonium acetate. In the above condition, the formation of plate crystals are observed towards the center, however it is important to note the 3-dimensionality of the crystal (seen in blue) that is angled on its side showing the thickness. The middle and bottom conditions also display crystals large enough for diffraction with 3-dimensionality.

5.1.2 Kinetic Characterization of BT2857

A series of kinetic activity trials were run with the remaining volume of purified BT2857. The first parameters to test for were the temperature and pH at which BT2857 achieved optimal activity (Figure 19).

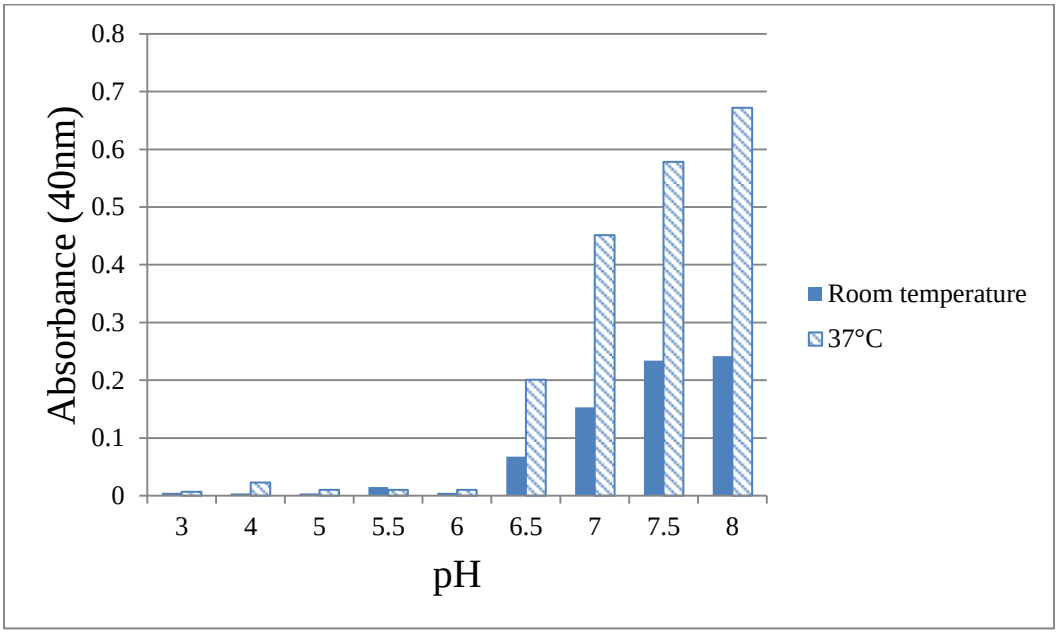


Figure 21. pH Profile and temperature-dependence of BT2857 enzymatic activity on PNP-Gal: Optimal activity is observed with 50mM Tris-HCl pH 8.0 at 37°C. Kinetic activity for pH above 8 resulted in protein precipitation revealing protein instability in alkaline pH values.

Once optimal conditions were determined for BT2857 kinetic activity, further reactions were prepared in triplicate for a Michaelis-Menten plot. Reactions with increasing concentrations of substrate were run in the microplate reader at 37°C for 15 minutes and readings taken every 45 seconds.

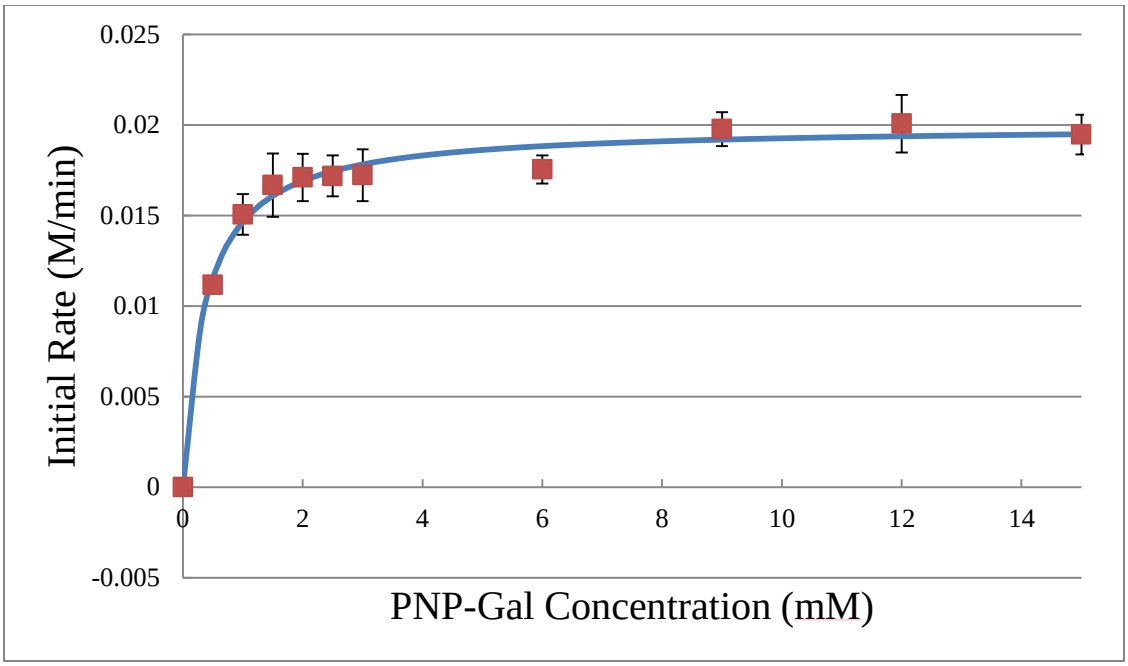


Figure 22. Michaelis-Menten plot of BT2857 kinetic activity with PNP-Gal: From the equation of the curve, the Michaelis-Menten variable were calculated where V_{max} = 0.01994M/min, K_M = 0.3660mM, and k_{cat} = 6645 s^{-1}. Data points were produced from a set of triplicates.

BT2857 presented positive activity as a putative β-galactosidase on the substrate, PNP-Gal, meaning that the transformation into Tuner(DE3)pLyss competent cells was required. Protein expression by the Tuner strain displayed significant decrease in yield, however was still deemed a sufficient amount to test for kinetic activity (Figure 21). Protein was reportedly concentrated to 14.6 mg/mL (or ~320 µM), however the actual concentration of pure BT2857 was suspected to be much lower. The reported concentration accounts for BT2857 as well as contaminating protein that was still present in the sample as seen after one run of anion-exchange chromatography (Figure 22). Kinetic activity testing was proceeded because theoretically the endogenous

galactosidase activity expressed by lacZ should be removed by expression in Tuner cells thus extreme protein purity did not seem necessary.

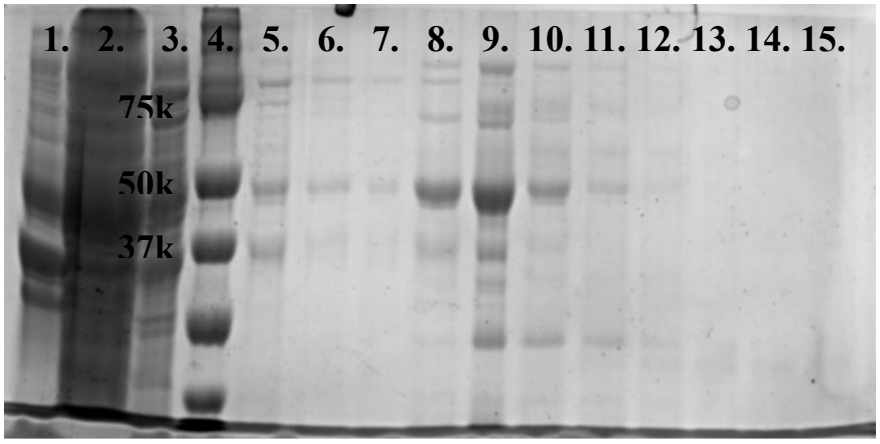


Figure 23. SDS-PAGE gel depicting the isolation of BT2857 expressed in Tuner(DE3)pLyss cells using nickel-IMAC: 12% SDS-PAGE gel loaded with the elution fractions of increasing imidazole concentrations of 5, 5, 5, 10, 25, 50, 75, 100, 200, 300, 400, and 500mM from lanes 3, 5-15. The flow through and pellet were also loaded into lanes 1 and 2, respectively. Lane 4 was loaded with Precision Plus Protein™ Kaleidoscope ladder.

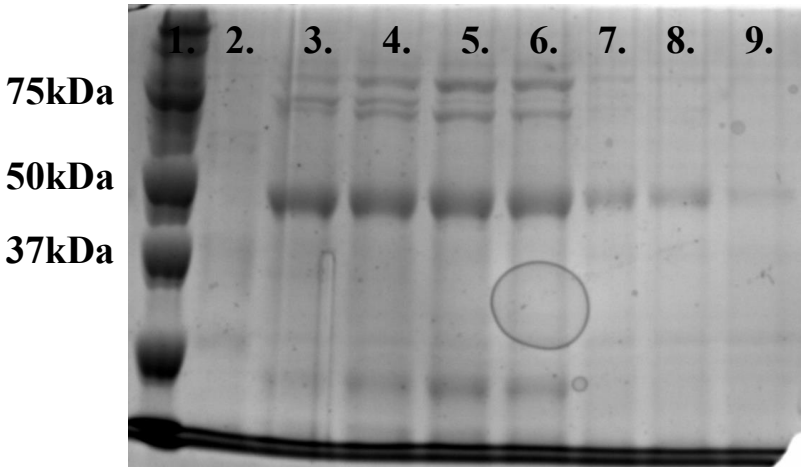


Figure 24. SDS-PAGE gel depicting the attempted purification of BT2857 (expressed in Tuner(DE3)pLyss cells)) using anion-exchange chromatography: 12% SDS-PAGE gel loaded with elution fractions under the peak (chromatogram not shown)

in lanes 2-9. The elution fractions demonstrate the presence of contaminating bands therefore the purification was deemed unsuccessful. Lane 1 was loaded with Precision Plus Protein™ Kaleidoscope ladder.

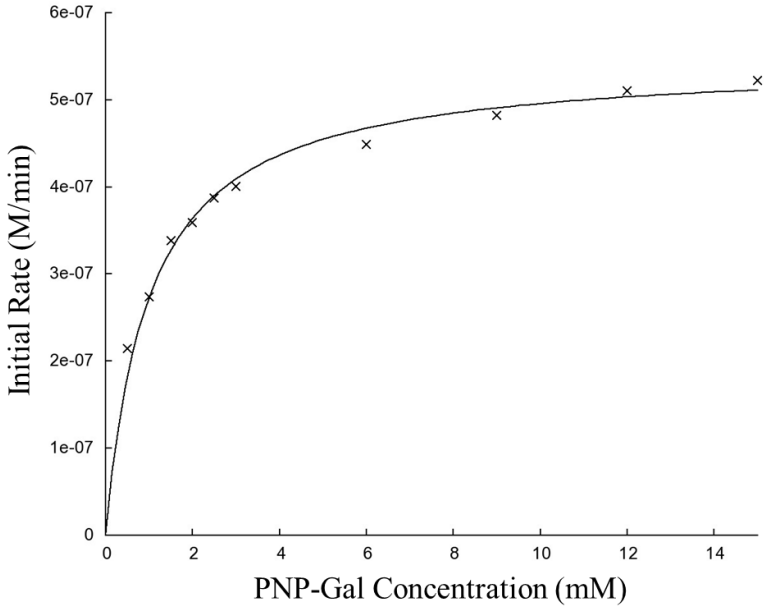


Figure 25. Michaelis-Menten plot of BT2857 expressed in Tuner(DE3)pLyss cells to confirm galactosidase activity with PNP-Gal: From the equation of the curve, the Michaelis-Menten variables were calculated where V_{max} = 5.453e-7 ± 1.105e-8 M/min, K_M = 1 ± 0.0832 mM, and k_{cat} = 1.704e-3 s^{-1}. Curve produced using the online curve-fitting program 'K_M V_{max} Tool Kit' found on http://ic50.tk/kmvmax.html.

Despite various efforts to increase protein concentration of BT2857 expressed in Tuner cells, the same yield as expressed in BL21 was unachievable. However, low concentrations of BT2857 in Tuner cells still allowed for kinetic quantification as it did test positive for activity on PNP-Gal (after long periods of incubation compared to BL21) (Figure 23). The Michaelis-Menten variables obtained were much lower than when conducted in BL21 cells due to challenges in concentration, however the typical kinetic

trend could be observed where the rate of reaction leveled off at increased substrate concentrations (V_{max}).

Primers were ordered for two mutants, D208N and E311Q, of BT2857 to determine potential active site residues. The neutral analogous residues (asparagine and glutamine) were chosen to replace the respective charged residues (aspartate and glutamate) in hopes of revealing a decrease in kinetic activity. Therefore PCR reactions were set-up for site-directed mutagenesis of BT2857 (Figure 24).

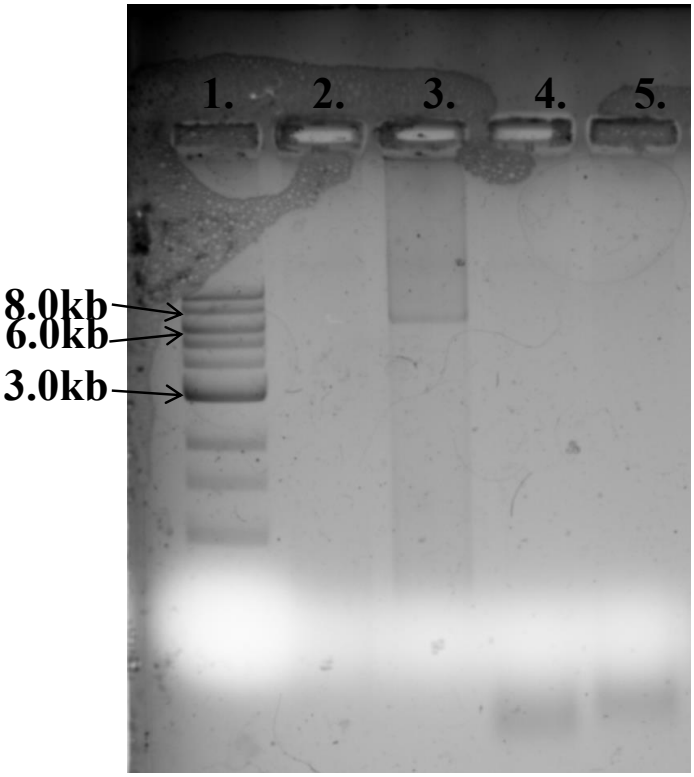


Figure 26. Agarose gel depicting PCR products for attempted trials of site-directed mutagenesis of BT2857: 0.7% (w/v) Agarose gel loaded with PCR products for site-directed mutagenesis of D208N (lanes 1 and 2), and E311Q (lanes 3 and 4). PCR reactions and thermocycler conditions are listed in Table 3 and 4 (section 3.2.9 Site-Directed Mutagenesis). Lane 1 was loaded with NEB 1kb DNA ladder.

The possible PCR hit for D208N was immediately treated with the restriction enzyme, DpnI, to remove any parental DNA. The plasmid was then transformed into NEB 5-alpha competent cells to ensure high efficiency uptake. Colonies were picked and cultured in 10mL of LB with ampicillin (final conc. 100mM). Cultures were mini-prepped to isolate plasmid and sent to Robarts Research Institute (London, ON) for sequencing and confirmation of successful mutation (Appendix 4). Because BT2857 confirmed β-galactosidase activity in the Tuner(DE3)pLyss cells, the D208N mutant was expressed in BL21 cells for better yields of protein and concentrated to approximately 20 mg/mL. Due to time constraints, D208N activity was only roughly measured by incubating it with pNP-Gal at 37°C for 30 minutes. Some remaining BT2857 previously expressed in Tuner cells was incubated under the same conditions as the mutant as a positive control.

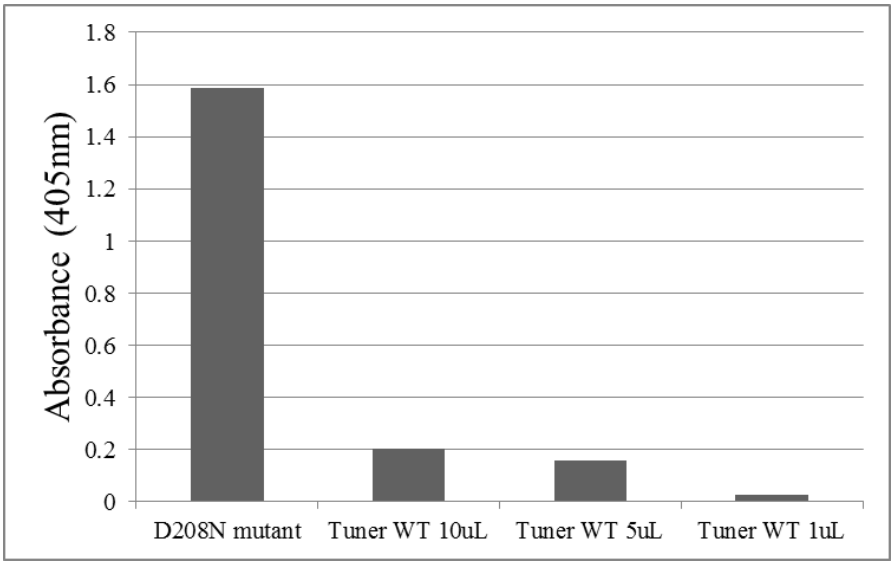


Figure 27. Comparison in activity of BT2857 variants with PNP-Gal: The D208N mutant possesses the greatest activity with an absorbance reading of ~1.6, while BT2857 expressed in Tuner(DE3)pLyss cells demonstrated little to no activity with varying

volumes of enzyme (due to extremely low concentration and possible denaturation of protein).

5.2 SCO6594-97 Solubility Screen

Upon arrival of custom genes for SCO6594-97, each gene was immediately transformed into BL21(DE3) competent *E. coli* cells and inoculated into 1L LB and ampicillin cultures for growth. Cells were harvested and a small portion was used for solubility screening by sonication to test which genes were expressed and readily soluble in the standard buffer condition (50mM Tris-HCl pH 7.5) used in chemical lysis. Sonications were performed on ice at 40% amplitude for 30 seconds of pulsing on and 30 seconds of rest for a total of 3 minutes. Sonicated samples were centrifuged at 12000xg for 10 minutes and run on SDS-PAGE to assess expression and solubility (Figure 26 and 27). It was difficult to differentiate target bands to the high levels of noise of contaminating bands on the gel, however SCO6584 showed potential indicated by an arrow. The pellet of SCO6597 (Figure 26 Lane 7) demonstrated a large expressed protein however further solubility analyses would need to be conducted.

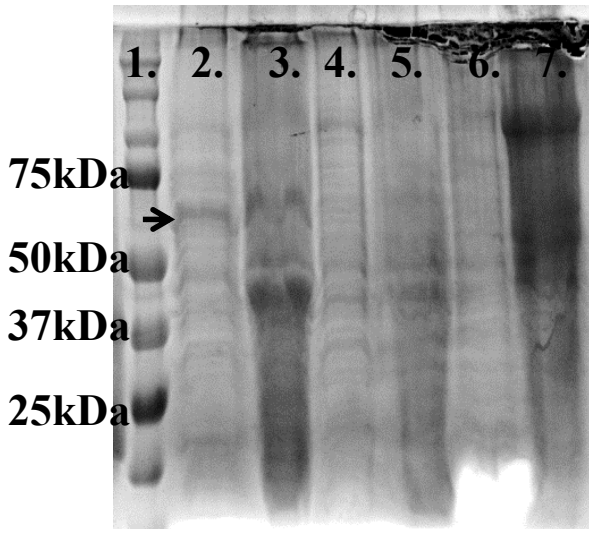


Figure 28. SDS-PAGE gel depicting the solubility screening of SCO6594 (~64kDa), SCO6596 (~63kDa), and SCO6597 (~89kDa) in 50mM Tris-HCl pH 7.5 via sonication: 12% SDS-PAGE gel loaded with the supernatant and pellet of sonicated samples with 50mM Tris-HCl pH 7.5. Lanes 2 and 3 were loaded with SCO6594 supernatant and pellet, respectively. The black arrow indicates protein expression in the supernatant for SCO6594. Lanes 4 and 5 were loaded with SCO6596 supernatant and pellet, respectively. Lanes 6 and 7 were loaded with SCO6597 supernatant and pellet, respectively. Lane 1 was loaded with Precision Plus Protein™ Kaleidoscope ladder.

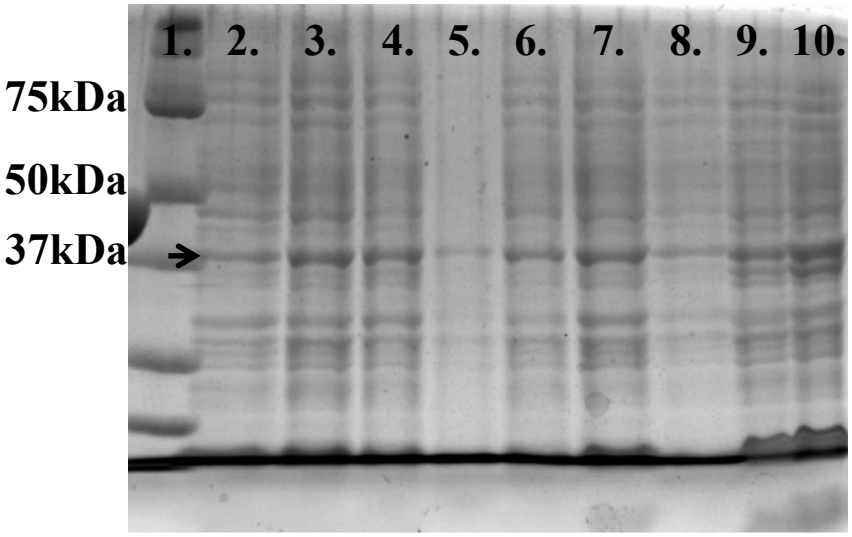


Figure 29. SDS-PAGE gel depicting the solubility screening of SCO6595 (~39kDa) in various buffer conditions (see Table 5) via sonication: 12% SDS-PAGE gel loaded

with supernatant of SCO6595 after sonication in various buffer conditions and salt concentrations. The black arrow indicates the row at which SCO6595 is expressed and soluble in each condition. The best yield was observed in lane 3, 7, and 10 which corresponds to 50mM Tris-HCl pH 8.0, 50mM Sodium acetate pH 7.0 (300mM NaCl), and 50mM MES pH 7.0 (300mM NaCl), respectively. Lane 1 was loaded with Precision Plus Protein™ Kaleidoscope ladder.

Table 9. List of Buffer Conditions Used in SCO6595 Solubility Screening (Table corresponds with Figure 28).

Lane	Buffer Conditions
2	50mM Tris-HCl pH 7.0, 300mM NaCl
3	50mM Tris-HCl pH 8.0
4	50mM Tris-HCl pH 8.0, 300mM NaCl
5	50mM Sodium phosphate pH 7.0, 300mM NaCl
6	50mM Sodium phosphate pH 8.0, 300mM NaCl
7	50mM Sodium acetate pH 7.0, 300mM NaCl
8	50mM Sodium acetate pH 8.0, 300mM NaCl
9	50mM MES pH 7.0, 300mM NaCl
10	50mM MES pH 7.0, 300mM NaCl

5.3 Expression, Purification, and Kinetic Characterization of SCO6594

SCO6594 demonstrated protein expression and solubility with the standard chemical lysis buffer, therefore it was selected as the first of the custom genes to be isolated for characterization. Isolation of SCO6594 proceeded with nickel-IMAC as previously used for BT2857 (Figure 28). Elution fractions were relatively clean off nickel-IMAC with some contamination seen right underneath the target bands.

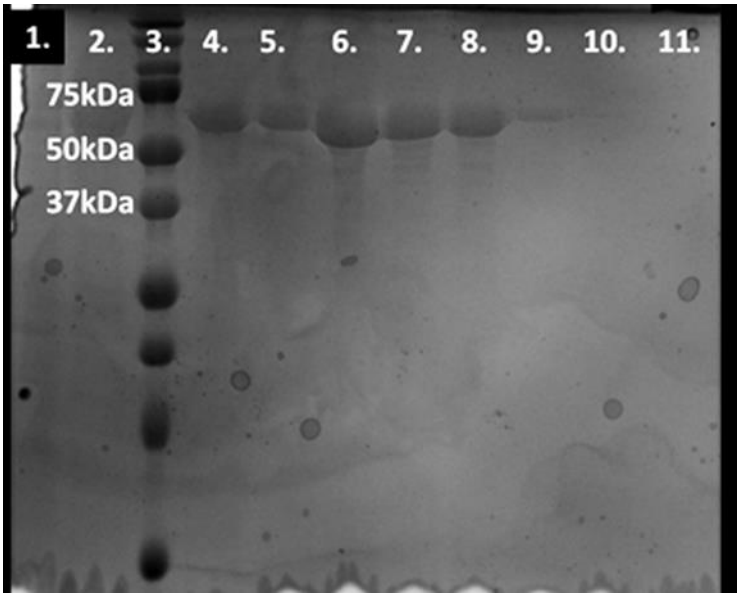


Figure 30. SDS-PAGE gel depicting the isolation of SCO6594 (~64kDa) via nickel-IMAC: 12% SDS-PAGE gel loaded with elution fractions of increasing imidazole concentrations (5, 5, 5, 10, 25, 50, 75, 100, and 500mM) in lanes 2, 4-11. Lane 1 was loaded with flow through, and lane 3 was loaded with Precision Plus Protein™ Kaleidoscope ladder.

The fractions in lanes 4-8 of Figure 28 were pooled together for dialysis to decrease imidazole concentration in the pooled fractions. The collective protein sample was then run through anion-exchange chromatography for further purification (Figure 29).

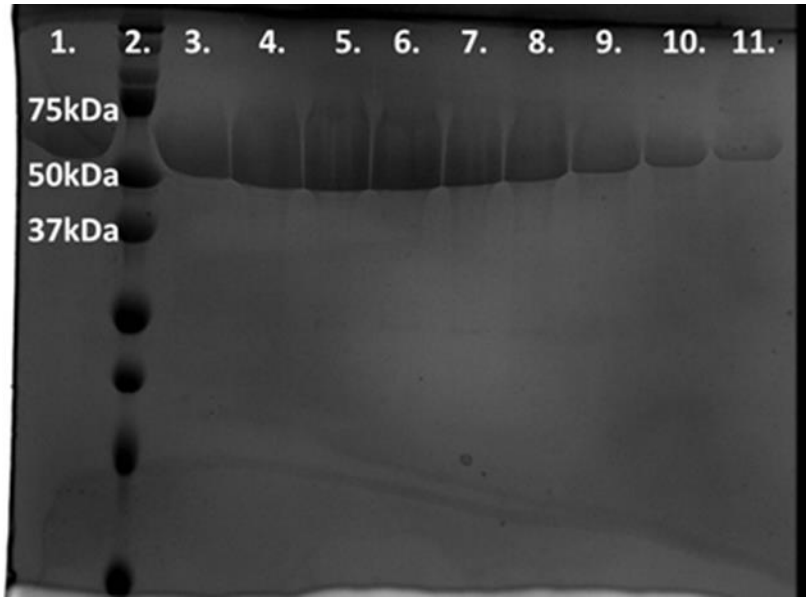


Figure 31. SDS-PAGE gel depicting the purification of SCO6594 (~64kDa) through anion-exchange chromatography: 12% SDS-PAGE gel loaded with elution fractions that appeared under the peak (chromatogram not shown) in lanes 1, 3-11. Lane 2 was loaded with Precision Plus Protein™ Kaleidoscope ladder.

The successful purification of SCO6594 and subsequent concentration of the protein sample to ~11mg/mL lead to preliminary crystal tray set-up using the Gryphon against MCSG-2 (Microlytics), Index HR2-144 (Hampton Research) and Crystal Screen 2 HR2-112 (Hampton Research) crystal screens. The remaining volume of concentration protein was used for kinetic testing, where it was crudely tested with PNP-Gal, ONP-Gal, and PNP-Man in a centrifuge tube to incubate at room temperature for 10 minutes. It was observed that the reaction with ONP-Gal exhibited greater activity (reaction turned dark yellow) than PNP-Gal (reaction turned light yellow), while the reaction with PNP-Man displayed zero activity (reaction remained colorless).

Similar to the initial kinetic testing with the pH profile that was conducted for BT2857, the same protocol was used for SCO6594 with the only difference being the substrate used was ONP-Gal instead of PNP-Gal (Figure 30).

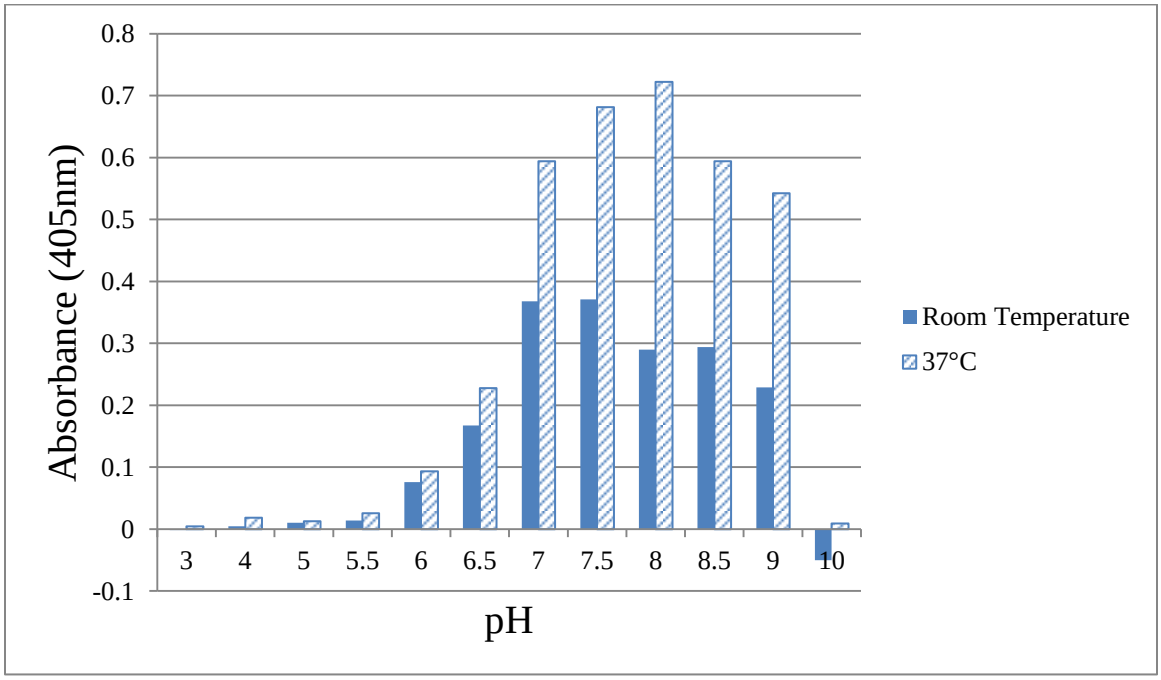


Figure 32. pH Profile and temperature-dependence of SCO6594 kinetic activity with ONP-Gal: The kinetic activity begins to increase from pH 6.0, where to reaches optimal activity in 50mM Tris-HCl pH 8.0 at 37°C for SCO6594. The kinetic activity decreases thereafter and is completely absent at pH 10. Incubation at 37°C also yields greater levels of activity.

Following the same protocol as previous, the kinetic characterization of SCO6594 was performed in triplicate at the optimal buffer and temperature conditions of 50mM Tris-HCl pH 8 at 37°C. Reactions were set up in a microplate and incubated for 15 minutes while taking absorbance readings every 45 seconds (Figure 31).

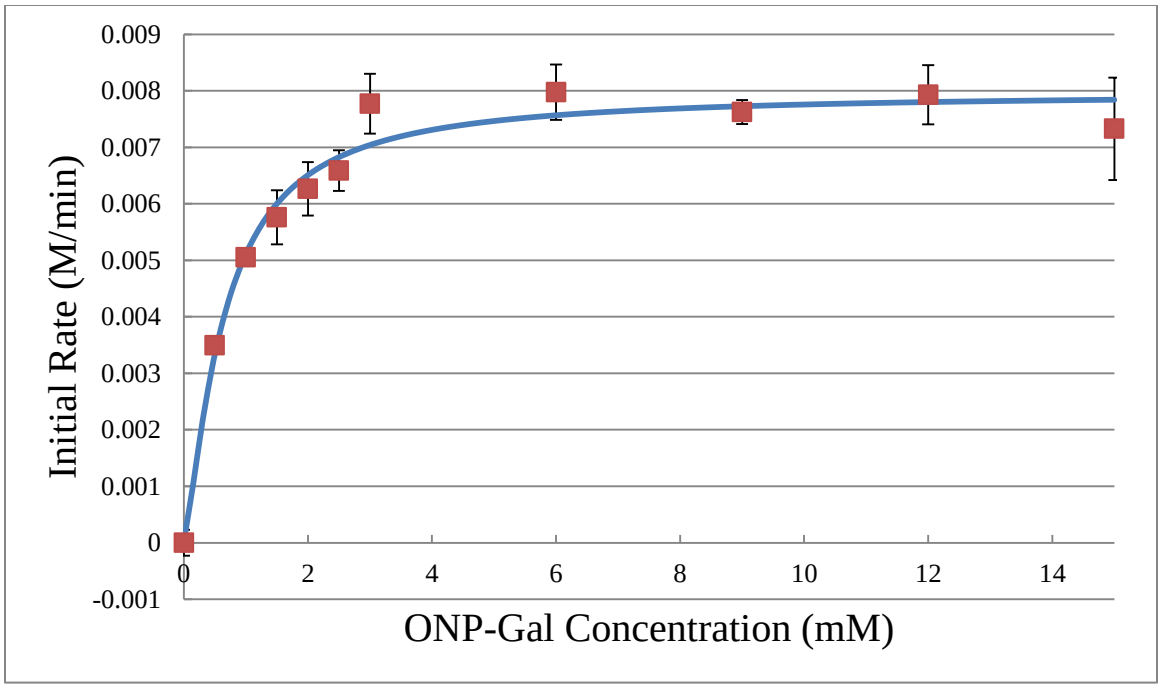


Figure 33. Michaelis-Menten plot of SCO6594 kinetic activity with ONP-Gal: From the equation of the curve, the Michaelis-Menten variables were calculated where V_{max} = 0.007966M/min, K_M = 0.6451mM, and k_{cat} = 2655.3 s^{-1}.

CHAPTER 6 – DISCUSSION

6.1 Interpretation of BT2857 Kinetics

The objective of the research project was the characterization of carbohydrate-active enzymes. The target proteins that were taken on in this project were BT2857, SCO6594-97. Each protein required expression, isolation and purification before functional assays or crystal trays could be conducted to reveal functional or structural characteristics. BT2857 was successfully expressed and purified for kinetic characterization with the artificial substrate, 4-nitrophenyl β-D-galactopyranoside. From the Michaelis-Menten curve that was plotted, the V_{max} = 0.019936M/min, K_M = 0.36598mM, and k_{cat} = 6645.3 s^{-1} were calculated. The values calculated were slightly greater compared to literature values, where the reported k_{cat} of endogenous β-galactosidase in *E. coli* was determined to be 750 ± 30s^{-1} [81]. The reported K_M of the endogenous β-galactosidase was also determined to be 0.12 ± 0.02mM [81], revealing a three-fold difference with BT2857. Even with the most active mutant produced in studies with endogenous β-galactosidase, the k_{cat} and K_M values were not comparable to that of BT2857. The difference in BT2857 activity with endogenous *E. coli* β-galactosidase activity might be that ONP-Gal was conducted in the study rather than PNP-Gal that was used with BT2857. Substrate variability can affect the kinetic rates of enzymatic activity due to favorable substrate-binding depending on substrate orientation and spacial geometry. The kinetic rates of BT2857 might also be slightly inflated because they were expressed in BL21(DE3) competent *E.coli* cells which possessed lacZ coding for endogenous β-galactosidase. Although isolation and purification steps performed on

BT2857 should somewhat prevent background activity of endogenous *E. coli* galactosidase, it was not completely guaranteed unless BT2857 is expressed in Tuner(DE3)pLyss cells. The kinetic rates obtained from BT2857 expressed in Tuner cells were much lower at what seems like negligible values mainly due to the suspected low concentration of protein obtained. The reported protein concentration was ~14.6 mg/mL however observation of the sample appeared a brown/yellow color which is atypical of a pure sample which was expected to be clear and colorless. However, when the sample was incubated with PNP-Gal, the color change to yellow was observed indicating that there was indeed catalytic activity. A negative control was also monitored alongside Tuner BT2857 to validate the color change was not from auto-hydrolysis of the substrate alone.

6.2 Truncations of BT2857 for Crystal Growth and Site-Directed Mutagenesis

Furthermore, the structural aspect of BT2857 was unattainable due to precipitation of protein even at concentrations below 10mg/mL when initial crystal trays were set-up. To overcome this challenge, BT2857 was truncated into N- and C-terminal domains. Analysis of the bioinformatics conducted on the sequence of BT2857 revealed little predictions in function and structure. Therefore, it was imperative to explore all general aspects of BT2857 to compile enough background for the hypothesis that BT2857 was a putative β-galactosidase that existed as two domains. InterPro, RaptorX, SWISS-MODEL, and IntFOLD were all consulted and suggested that BT2857 was composed of either two or three domains. Although evenly matched on output and confidence levels on the structure of BT2857, it was the two-domain hypothesis that held a greater percentage of similarity in sequence alignments. Two of the three domains that

were predicted from IntFOLD also demonstrated similarity with the two-domain hypothesis, while the third domain had low percentage in confidence. The active residues of each characterized protein with similarity to BT2857 from the query were compiled to predict appropriate truncation sites. The active residues within each characterized protein showed conservation of Asp92 and Trp290 [82,83]. A BLAST alignment was also conducted to predict any possible conserved residues, which is where the mutants D208N and E311Q for the study came to fruition. Although Asp92 and Trp290 came up as active residues, the placement of the residues on the predicted model of BT2857 was located in unusual areas outside of the two-domain model. Therefore, using the leads of each bioinformatics tool, the domain truncations were M24-T225 and E221-D398, which encompassed each active residue in each domain. The crystallization of the C-terminal domain of BT2857 was successful as a result of producing truncations of the protein (Figure 20). In hopes of splitting BT2857 into two domains, the active site could also be narrowed down, however in crude incubations of C-terminal and N-terminal isolated proteins there was activity present in both indicated by the color change to yellow with PNP-Gal. The D208N mutant was designed to explore potential catalytic residues, where the removal of the charged side group of aspartate would remove or lower activity. At a protein concentration of ~20 mg/mL, D208N activity was quantified with an absorbance reading of ~1.6 which is comparable to wild-type BT2857 readings of ~0.7 at a protein concentration of ~10 mg/mL. Therefore, it was concluded that Asp208 did not have a role in catalysis.

6.3 Interpretation of SCO6594 Kinetics

The solubility screen performed on SCO6594-97 through sonication revealed that SCO6594 was readily soluble in the supernatant of the standard chemical lysis buffer. Even when SCO6594 was found to be soluble in 50mM Tris-HCl pH 7.5, it was purified in 50mM Tris-HCl pH 8.0 which displayed no changes in protein stability. Following up with kinetic characterization of SCO6594 revealed that the enzyme's optimal condition was at 50mM Tris-HCl pH 8.0 at 37°C with the artificial substrate, 2-nitrophenyl β-D-galactopyranoside. From the Michaelis-Menten curve that was plotted, the Vmax = 0.007966M/min, K_M = 0.6451mM, and k_{cat} = 2655.3 s^{-1} were calculated. These values compared to the literature values of endogenous *E. coli* β-galactosidase with the same artificial substrate were also considerably greater, with an almost four-fold difference in k_{cat} values [81]. As discussed, the kinetic rates of SCO6594 may have appeared inflated due to expression in BL21(DE3) competent *E. coli* cells coding for endogenous β-galactosidase. No structural characteristics of SCO6594 have been elucidated, however the notable difference in enzyme activity between 2-nitrophenyl β-D-galactopyranoside and 4-nitrophenyl β-D-galactopyranoside of SCO6594 suggest a favorable interaction with the ortho-isomer and perhaps geometry of substrate fit to the active site.

6.4 Solubility of SCO6595-97

The solubility screen also revealed that SCO6595 was soluble in various buffer conditions. The best yield was observed in 50mM Tris-HCl pH 8.0, 50mM Sodium acetate pH 7.0, 300mM NaCl , and 50mM MES pH 7.0, 300mM NaCl. Although there was no common component among the three conditions, the pH does not differ greatly indicating the stability of SCO6594 in pH 7-8. The presence of sodium chloride also does

not interfere with protein solubility; however it is a preferable that the protein is soluble in sodium chloride for increased stability [84].

Expression of SCO6596 was observed in neither pellet nor supernatant; therefore expression trials should be conducted to track protein expression after induction with IPTG. SCO6597 might possibly be expressed in the pellet where a prominent band appears around 100kDa. Although the observed band is considerably larger than the expected molecular weight of SCO6597 (~89kDa), the SDS-PAGE gel was run inconsistently which may impact the alignment of sample to the ladder. A second solubility screen with the presence of various metals could be run with SCO6596 and SCO6597 to identify the buffer condition with adequate yield for protein characterization. Protein expression may be improved by testing in different growth mediums such as Terrific Broth (TB) or Super Optimal Broth with Catabolite repression (SOC). TB differs from LB for containing glycerol, and SOC contains traces of KCl, $MgCl_2$, and glucose.

CHAPTER 7 – CONCLUSIONS AND FUTURE STUDIES

Bioinformatics tools had predicted that BT2857 possessed a galactose-binding domain, however it was unclear whether it was simply a carbohydrate-binding domain or if it was a domain as part of an enzyme. Through the kinetic characterization, it has been discovered that the enzyme is active with the artificial substrate 4-nitrophenyl β-D-galactopyranoside thus partially elucidating the functionality of BT2857 as a putative β-galactosidase. The V_{max} = 0.019936M/min, K_M = 0.36598mM, and k_{cat} = 6645.3 s^{-1} was calculated using the Michaelis-Menten equation. Future research in the potential active residues of BT2857 is needed to determine the mechanism of kinetic activity. Piecing together the crystal structures of both N- and C- terminals will also aid in visualizing enzyme-substrate interactions.

The preliminary steps have been conducted in revealing the expression and solubility of SCO6595-97, while kinetic characterization has been established for SCO6594. The V_{max} = 0.007966M/min, K_M = 0.6451mM, and k_{cat} = 2655.3 s^{-1} was calculated using the Michaelis-Menten equation. Similar to the complete structural and functional protocols in the characterization of BT2857, the same process for future studies should be implemented with SCO6594-97.

Future studies for the collaborative project of protein structure and function characterization also include thin-layer chromatography for predicted substrates. UV-difference is also another procedure that could be conducted to investigate carbohydrate-protein interactions.

REFERENCES

[1] S. Möller, M.D. Croning, R. Apweiler, Evaluation of methods for the prediction of membrane spanning regions, Bioinformatics, 17 (2001) 646–653.

[2] H. Wang, S. Wang, J. Peng, Z. Wang, H. Lu, J. Xu, Template-based protein structure modeling using the RaptorX web server, Nat Protoc., 7 (2016) 1511–1522.

[3] L. Kelley, M. Sternberg, Protein structure prediction on the web : a case study using the Phyre server, Nat. Protoc., 4 (2009) 363–371.

[4] L.J. McGuffin, J.D. Atkins, B.R. Salehe, A.N. Shuid, D.B. Roche, IntFOLD: an integrated server for modelling protein structures and functions from amino acid sequences, Nucleic Acids Res., 43 (2015) W169–73.

[5] T. Schwede, J. Kopp, N. Guex, M.C. Peitsch, SWISS-MODEL: An automated protein homology-modeling server, Nucleic Acids Res., 31 (2003) 3381–3385.

[6] E.C. Martens, H.C. Chiang, J.I. Gordon, Mucosal Glycan Foraging Enhances Fitness and Transmission of a Saccharolytic Human Gut Bacterial Symbiont, Cell Host Microbe, 4 (2008) 447–457.

[7] T.N. Peterson, S. Brunak, G. von Heijne, H. Nielson, SignalP 40 : discriminating signal peptides from transmembrane regions, Nat. Methods, 8 (2011) 785–786.

[8] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol., 215 (1990) 403–10.

[9] C.K. Mathews, K.E. Van Holde, K. Ahern, Biochemistry, 3rd Editio, Prentice Hall, 1999.

[10] G. Davies, B. Henrissat, Structures and mechanisms of glycosyl hydrolases, Structure, 3 (1995) 853–859.

[11] R.A. Laine, Invited Commentary: A calculation of all possible oligosaccharide isomers both branched and linear yields 105 × 1012 structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems, Glycobiology, 4 (1994) 759–767.

[12] A. Varki, Biological roles of oligosaccharides: all of the theories are correct, Glycobiology, 3 (1993) 97–130.

[13] L.V. Hooper, J.I. Gordon, Glycans as legislators of host-microbial interactions: spanning the spectrum from symbiosis to pathogenicity, Glycobiology, 11 (2001) 1R–10R.

[14] J.F. Turner, D.H. Turner, The Regulation of Carbohydrate Metabolism, Annu. Rev. Plant Physiol., 26 (1975) 159–186.

[15] S.O. Meroueh, K.Z. Bencze, D. Hesek, M. Lee, J.F. Fisher, T.L. Stemmler, S. Mobashery, Three-dimensional structure of the bacterial cell wall peptidoglycan, Proc. Natl. Acad. Sci. U. S. A., 103 (2006) 4404–4409.

[16] I. Bucior, M.M. Burger, Carbohydrate-carbohydrate interaction as a major force initiating cell-cell recognition, Glycoconj. J., 21 (2004) 111–123.

[17] S. Hakomori, Carbohydrate-to-carbohydrate interaction, through glycosynapse, as a basis of cell recognition and membrane organization, Glycoconj. J., 21 (2004) 125–137.

[18] S. Kornfeld, Trafficking of lysosomal enzymes, FASEB J., 1 (1987) 462–468.

[19] K. Drickamer, M.E. Taylor, Biology of Animal Lectins, Annu. Rev. Cell Biol., 9 (1993) 237–264.

[20] T.W. Rademacher, R.B. Parekh, R.A. Dwek, Glycobiology, Ann. Rev. Biochem., 57 (1988) 785–838.

[21] N. Jenkins, R.B. Parekh, D.C. James, Getting the glycosylation right: implications for the biotechnology industry, Nat. Biotechnol., 14 (1996) 975–981.

[22] D. Ghaderi, M. Zhang, N. Hurtado-Ziola, A. Varki, Production platforms for biotherapeutic glycoproteins Occurrence, impact, and challenges of non-human sialylation, Biotechnol. Genet. Eng. Rev., 28 (2012) 147–175.

[23] R.J. Kelm, K.G. Mann, The collagen binding specificity of bone and platelet osteonectin is related to differences in glycosylation, J. Biol. Chem., 266 (1991) 9632–9639.

[24] P. Kotrba, M. Inui, H. Yukawa, Bacterial phosphotransferase system (PTS) in carbohydrate uptake and control of carbon metabolism, J. Biosci. Bioeng., 92 (2001) 502–517.

[25] T. Bhattacharya, T.S. Ghosh, S.S. Mande, Global profiling of carbohydrate active enzymes in human gut microbiome, PLoS One, 10 (2015) 1–20.

[26] A. Hölemann, P.H. Seeberger, Carbohydrate diversity: Synthesis of glycoconjugates and complex carbohydrates, Curr. Opin. Biotechnol., 15 (2004) 615–622.

[27] B.I. Cantarel, P.M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, B. Henrissat, The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics, Nucleic Acids Res., 37 (2009) 233–238.

[28] V. Lombard, H. Golaconda Ramulu, E. Drula, P.M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013, Nucleic Acids Res., 42 (2014) 490–495.

[29] B. Henrissat, M. Vegetales, F. Grenoble, A classification of glycosyl hydrolases

based sequence similarities amino acid, Biochem. J., 280 (1991) 309–316.

[30] U.M. Ünligil, J.M. Rini, Glycosyltransferase structure and mechanism, Curr. Opin. Struct. Biol., 10 (2000) 510–517.

[31] G.J. Williams, J.S. Thorson, Natural Product Glycosyltransferases: Properties and Applications, in: Adv. Enzymol. Relat. Areas Mol. Biol., 2009: pp. 55–120.

[32] V. Lombard, T. Bernard, C. Rancurel, H. Brumer, P.M. Coutinho, B. Henrissat, A hierarchical classfication of polysaccharide lyases for glycogenomics, Biochem. J., 432 (2010) 437–444.

[33] R.J. Linhardt, P.M. Galliher, C.L. Cooney, Polysaccharide Lyases, Appl. Biochem. Biotechnol., 12 (1986) 135–176.

[34] A. Levasseur, E. Drula, V. Lombard, P.M. Coutinho, B. Henrissat, Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes, Biotechnol. Biofuels, 6 (2013) 41.

[35] H. Hashimoto, Recent structural studies of carbohydrate-binding modules, Cell. Mol. Life Sci., 63 (2006) 2954–2967.

[36] O. Shoseyov, Z. Shani, I. Levy, Carbohydrate binding modules: biochemical properties and novel applications, Microbiol. Mol. Biol. Rev., 70 (2006) 283–295.

[37] V. Tremaroli, F. Bäckhed, Functional interactions between the gut microbiota and host metabolism, Nature, 489 (2012) 242–249.

[38] D.B. Wilson, Cellulases and biofuels, Curr. Opin. Biotechnol., 20 (2009) 295–299.

[39] D.W. Abbott, M.S. Macauley, D.J. Vocadlo, A.B. Boraston, Streptococcus pneumoniae endohexosaminidase D, structural and mechanistic insight into substrate-assisted catalysis in family 85 glycoside hydrolases, J. Biol. Chem., 284 (2009) 11676–11689.

[40] T. V. Vuong, D.B. Wilson, Glycoside hydrolases: Catalytic base/nucleophile diversity, Biotechnol. Bioeng., 107 (2010) 195–205.

[41] D.E. Koshland, Stereochemistry and the mechanism of enzymatic reactions, Biol. Rev., 28 (1953) 416–436.

[42] J.D. McCarter, S.G. Withers, Mechanisms of enzymatic glycoside hydrolysis, Curr. Opin. Struct. Biol., 4 (1994) 885–892.

[43] R. Kuroki, Y. Ito, Y. Kato, T. Imoto, A covalent enzyme-substrate adduct in a mutant hen egg white lysozyme (D52E), J. Biol. Chem., 272 (1997) 19976–19981.

[44] W.P. Burmeister, B. Henrissat, C. Bosso, S. Cusack, R.W.H. Ruigrok, Influenza B virus neuraminidase can synthesize its own inhibitor, Structure, 1 (1993) 19–26.

[45] S.J. Crennell, E.F. Garman, W.G. Laver, E.R. Vimr, G.L. Taylor, Crystal structure

of a bacterial sialidase (from Salmonella typhimurium LT2) shows the same fold as an influenza virus neuraminidase, Proc. Natl. Acad. Sci. U. S. A., 90 (1993) 9852–9856.

[46] A.E. Aleshin, C. Hoffman, L.M. Firsov, R.B. Honzatko, Refined crystal structures of glucoamylase from Aspergillus awamori var X100, J. Mol. Biol., 238 (1994) 575–591.

[47] A.M. Larsson, T. Bergfors, E. Dultz, D.C. Irwin, A. Roos, H. Driguez, D.B. Wilson, T.A. Jones, Crystal structure of Thermobifida fusca endoglucanase Ce16A in complex with substrate and inhibitor: The role of tyrosine Y73 in substrate ring distortion, Biochemistry, 44 (2005) 12915–12922.

[48] J. Rouvinen, T. Bergfors, T. Teeri, J.K.C. Knowles, T.A. Jones, Three-dimensional structure of cellobiohydrolase II from Trichoderma reesei, Science (80-.)., 249 (1990) 380–386.

[49] C. Divne, J. Stahlberg, T. Reinikainen, L. Ruohonen, G. Pettersson, J.K.C. Knowles, T. Teeri, T.A. Jones, The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from Trichoderma reesei, Science (80-.)., 265 (1994) 524–528.

[50] A. El Kaoutari, F. Armougom, J.I. Gordon, D. Raoult, B. Henrissat, The abundance and variety of carbohydrate-active enzymes in the human gut microbiota, Nat. Rev. Microbiol., 11 (2013) 497–504.

[51] C.L. Sears, A dynamic partnership: Celebrating our gut flora, Anaerobe, 11 (2005) 247–251.

[52] F. Bäckhed, R.E. Ley, J.L. Sonnenburg, D. a Peterson, J.I. Gordon, Host-bacterial mutualism in the human intestine, Science, 307 (2005) 1915–1920.

[53] H.J. Flint, E. a Bayer, M.T. Rincon, R. Lamed, B. a White, Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis, Nat. Rev. Microbiol., 6 (2008) 121–131.

[54] H.M. Wexler, Bacteroides: The good, the bad, and the nitty-gritty, Clin. Microbiol. Rev., 20 (2007) 593–621.

[55] N. Terrapon, V. Lombard, H.J. Gilbert, B. Henrissat, Automatic prediction of polysaccharide utilization loci in Bacteroidetes species, Bioinformatics, 31 (2015) 647–655.

[56] L.E. Comstock, M.J. Coyne, Bacteroides thetaiotaomicron: A dynamic, niche-adapted human symbiont, BioEssays, 25 (2003) 926–929.

[57] J.A. Shipman, J.E. Berleman, A.A. Salyers, Characterization of four outer membrane proteins involved in binding starch to the cell surface of Bacteroides thetaiotaomicron, J. Bacteriol., 182 (2000) 5365–5372.

[58] J. Xu, J.I. Gordon, Honor thy symbionts, Proc. Natl. Acad. Sci. United States Am., 100 (2003) 10452–10459.

[59] A.A. Salyers, J.K. Palmer, T.D. Wilkins, Laminarinase (beta-glucanase) activity in Bacteroides from the human colon, Appl. Environ. Microbiol., 33 (1977) 1118–24.

[60] A.A. Salyers, S.F. Kotarski, Induction of chondroitin sulfate lyase activity in Bacteroides thetaiotaomicron, J. Bacteriol., 143 (1980) 781–788.

[61] S.F. Kotarski, a a Salyers, Isolation and characterization of outer grown on different carbohydrates Isolation and Characterization of Outer Membranes of Bacteroides thetaiotaomicron Grown on Different Carbohydrates, 158 (1984) 102–109.

[62] N. Terrapon, B. Henrissat, How do gut microbes break down dietary fiber?, Trends Biochem. Sci., 39 (2014) 156–158.

[63] M.A. Mahowald, F.E. Rey, H. Seedorf, P.J. Turnbaugh, R.S. Fulton, A. Wollam, N. Shah, C. Wang, V. Magrini, R.K. Wilson, B.L. Cantarel, P.M. Coutinho, B. Henrissat, L.W. Crock, A. Russell, N.C. Verberkmoes, R.L. Hettich, J.I. Gordon, Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla, Proc. Natl. Acad. Sci., 106 (2009) 5859–5864.

[64] R.F. Seipke, M. Kaltenpoth, M.I. Hutchings, Streptomyces as symbionts: An emerging and widespread theme?, FEMS Microbiol. Rev., 36 (2012) 862–876.

[65] A.R. Hesketh, G. Chandra, A.D. Shaw, J.J. Rowland, D.B. Kell, M.J. Bibb, K.F. Chater, Primary and secondary metabolism , and post-translational protein modifications , as portrayed by proteomic analysis of Streptomyces coelicolor, Mol. Microbiol., 46 (2002) 917–932.

[66] J.M. Lynch, J.M. Whipps, Substrate flow in the rhizosphere, Plant Soil, 129 (1990) 1–10.

[67] B.R. Glick, The enhancement of plant growth by free-living bacteria, Can. J. Microbiol., 41 (1995) 109–117.

[68] C.P. Vance, Rhizobium Infection and Nodulation: A Beneficial Plant Disease?, Annu. Rev. Microbiol., 37 (1983) 399–424.

[69] P.K. Sharma, B.S. Kundu, R.C. Dogra, Molecular mechanism of host specificity in legume-rhizobium symbiosis, Biotechnol. Adv., 11 (1993) 741–779.

[70] N.R. Thomson, K.D. James, D.E. Harris, M.A. Quail, S.D. Bentley, D. Harper, A. Bateman, S. Brown, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, L. Larke, L. Murphy, K. Oliver, E. Rabbinowitsch, K. Rutherford, et al., Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2), Nature, 417 (2002) 141–147.

[71] G.L. Challis, D. a Hopwood, Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by Streptomyces species, Proc. Natl. Acad. Sci. U. S. A., 100 Suppl (2003) 14555–14561.

[72] L.F. Wright, D. a Hopwood, Actinorhodin is a chromosomally-determined antibiotic in Streptomyces coelicolar A3(2), J. Gen. Microbiol., 96 (1976) 289–297.

[73] Y. Ryu, M.J. Butler, K.F. Chater, K.J. Lee, Engineering of Primary Carbohydrate Metabolism for Increased Production of Actinorhodin in Streptomyces coelicolor Engineering of Primary Carbohydrate Metabolism for Increased Production of Actinorhodin in Streptomyces coelicolor, Appl. Environ. Microbiol., 72 (2006) 7132–7139.

[74] A.G. Reyes, N. Geukens, P. Gutschoven, D. Mot, A. Mejı, S. De Graeve, J. Anne, The Streptomyces coelicolor genome encodes a type I ribosome-inactivating protein, Microbiology, 156 (2010) 3021–3030.

[75] H. Nothaft, S. Parche, A. Kamionka, F. Titgemeyer, In Vivo Analysis of HPr Reveals a Fructose-Specific Phosphotransferase System That Confers High-Affinity Uptake in Streptomyces coelicolor, J. Bacteriol., 185 (2003) 929–937.

[76] V. Lombard, H. Golaconda Ramulu, E. Drula, P.M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013, Nucleic Acids Res., 42 (2014) 490–495.

[77] E.W. Sayers, T. Barrett, D.A. Benson, S.H. Bryant, K. Canese, V. Chetvernin., D.M. Church, M. Dicuccio, R. Edgar, S. Federhen, M. Feolo, L.Y. Geer, W. Helmberg, Y. Kapustin, D. Landsman, D.J. Lipman, T.L. Madden, D.R. Maglott, V. Miller, I. Mizrachi, et al., Database resources of the National Center for Biotechnology Information, Nucleic Acids Res., 37 (2009) 5–15.

[78] D.A. Benson, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, E.W. Sayers, GenBank, Nucleic Acids Res., 37 (2009) D26–D31.

[79] L.J. Jensen, M. Kuhn, M. Stark, S. Chaffron, C. Creevey, J. Muller, T. Doerks, P. Julien, A. Roth, M. Simonovic, P. Bork, C. von Mering, STRING 8 - A global view on proteins and their functional interactions in 630 organisms, Nucleic Acids Res., 37 (2009) 412–416.

[80] A. Mitchell, H.-Y. Chang, L. Daugherty, M. Fraser, S. Hunter, R. Lopez, C. McAnulla, C. McMenamin, G. Nuka, S. Pesseat, A. Sangrador-Vegas, M. Scheremetjew, C. Rato, S.-Y. Yong, A. Bateman, M. Punta, T.K. Attwood, C.J. a. Sigrist, N. Redaschi, C. Rivoire, et al., The InterPro protein families database: the classification resource after 15 years, Nucleic Acids Res., 43 (2014) D213–D221.

[81] M. Martinez-Bilbao, R.E. Holdsworth, L.A. Edwards, R.E. Huber, A highly reactive beta-galactosidase (Escherichia coli) resulting from a substitution of an aspartic acid for Gly-794, J. Biol. Chem., 266 (1991) 4979–86.

[82] J.N. Watson, S. Newstead, V. Dookhun, G. Taylor, A.J. Bennet, Contribution of the active site aspartic acid to catalysis in the bacterial neuraminidase from Micromonospora viridifaciens, FEBS Lett., 577 (2004) 265–269.

[83] M.S. Rogers, E.M. Tyler, N. Akyumani, C.R. Kurtis, R. Kate, S.E. Deacon, S. Tamber, S.J. Firbank, K. Mahmoud, F. Knowles, S.E. V Phillips, M.J. Mcpherson, D.M. Dooley, The stacking tryptophan of galactose oxidase: a second coordination sphere residue that has profound effects on tyrosyl radical behavior and enzyme catalysis, Biochemistry, 46 (2008) 4606–4618.

[84] S. Damodaran, J.E. Kinsella, The effects of neutral salts on the stability of macromolecules A new approach using a protein-ligand binding system, J. Biol. Chem., 256 (1981) 3394–3398.

APPENDICES

Appendix 1: Full List of Target Proteins with Truncation Sites and Predicted Domains

	Construct(s) of Predicted Domains
Streptomyces coelicolor	
SCO6594	T72 – V594 (Signal peptide removed) T72 – V238 (β-mannosidase)
SCO6595	M1 – P360 (Full ORF)
SCO6596	R10 – V443 (Signal peptide removed) V45 – G345 (GH3/42 family domain) V45 – V443 (GH3/42 family and galactose-binding domain)
SCO6597	L15 – A857 (Signal peptide removed) L364 – A857 (GH3 family domain)
Bacteroides thetaiotaomicron	
BT2857	M1 – E401 (Full ORF) M24 – T225 (N-terminus) E221 – D398 (C-terminus)

Appendix 2: List of Bioinformatics Tools

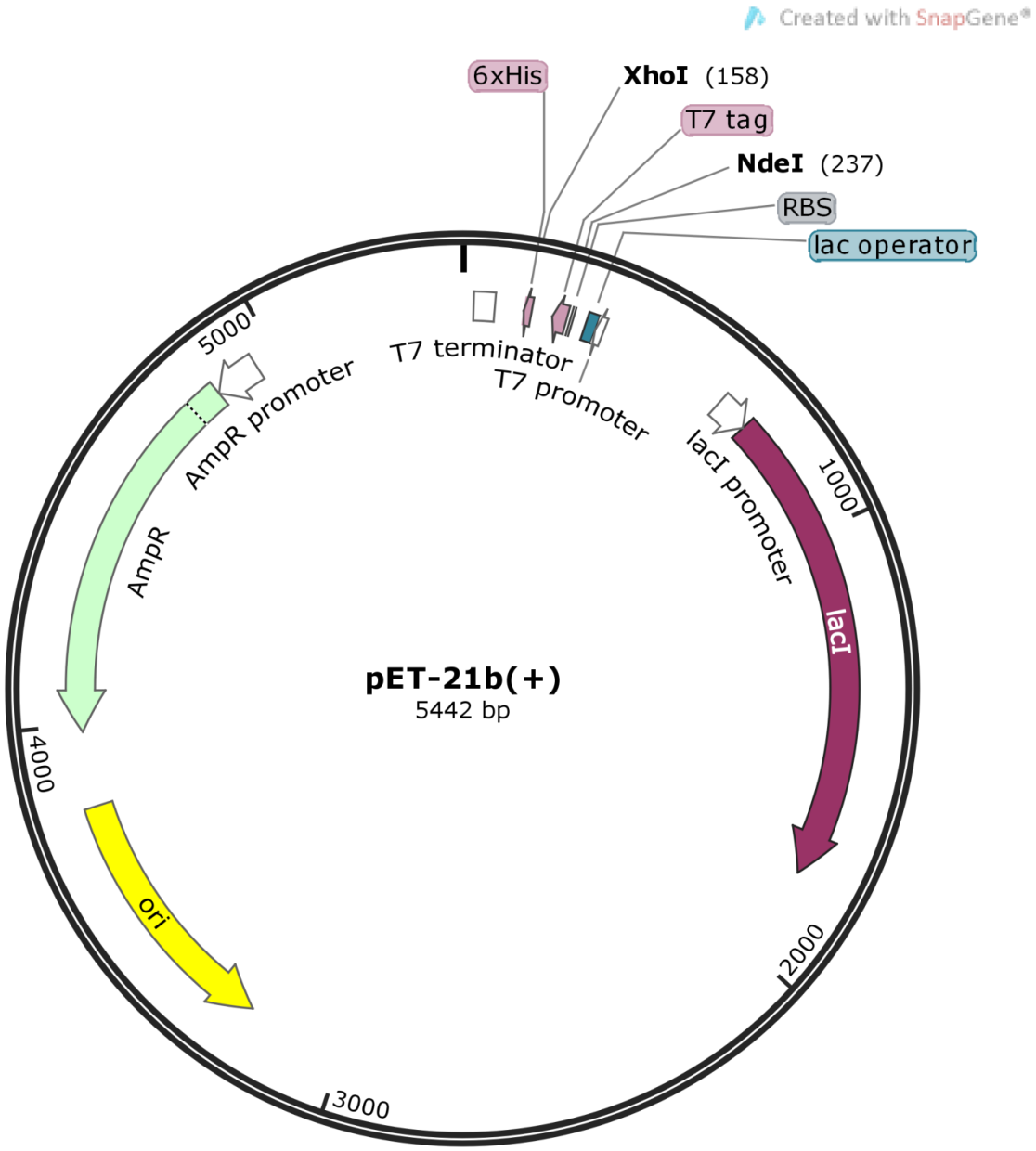
Program	Information Available	Web Address
RCSB Protein Data Bank	Characterized proteins 3D structure and related articles	http://www.rcsb.org/pdb/home/home.do
CAZy Database	List of hypothetical CAZymes Links to characterized protein under enzyme classes and families	http://www.cazy.org/
BLAST	Runs query against sequence databases to find local sequence similarity	http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastHome
ExPASy ProtParam	Computes molecular weight, theoretical pI, amino acid composition, extinction coefficient, and other chemical parameters	http://web.expasy.org/protparam/
Phyre^2	Runs query to find protein homology and predicted secondary and tertiary structures	http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index
STRING	Interactive map of predicted functional partners with target, and associated interactions	http://string-db.org/newstring_cgi/show_input_page.pl?UserId=f_EFjMiUT_IB&sessionId=A5dzsaoT96zY
SignalP	Distinguished predicted localization of protein with presence of signal peptide	http://www.cbs.dtu.dk/services/SignalP/
Interpro	Predicts domains, families, or motifs by comparison of signature sequences between query and sequence databases	http://www.ebi.ac.uk/interpro/

Appendix 3: List of Primers for Site-Directed Mutagenesis of BT2857

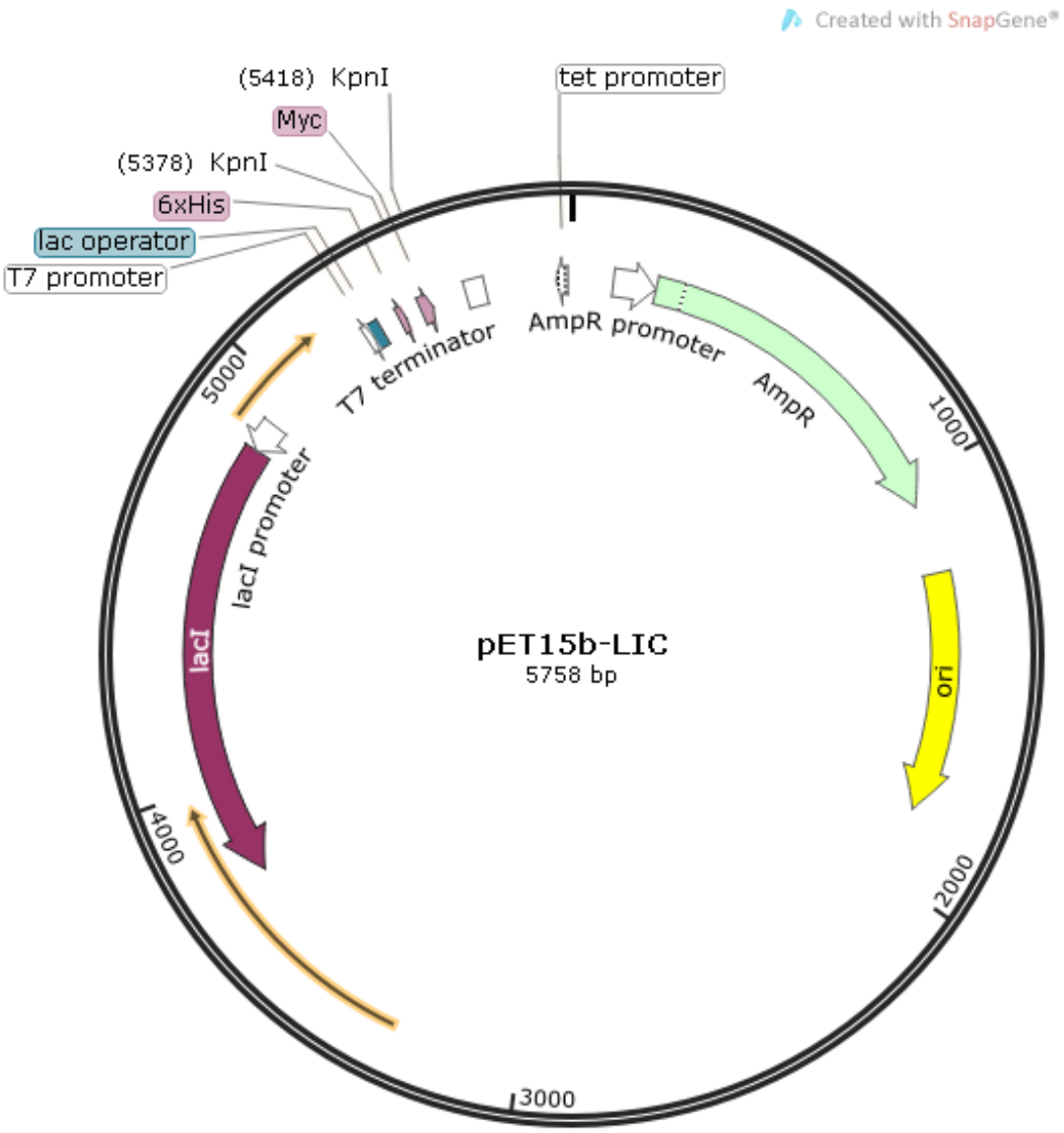
D208N	
Forward Primer	5’- AAATATGGGAATGAGACAAATACGCTGTTCACCGAGC -3’
Reverse Primer	5’- GCTCGGTGAACAGCGTATTTGTCTCATTCCCATATTT -3’
E311Q	
Forward Primer	5’- GCCCGAGACGCTACCAGGTATGGGGATGCGC -3’
Reverse Primer	5’- GCGCATCCCCATACCTGGTAGCGTCTCGGGC -3’

Appendix 4: Sequencing Results for BT2857 D208N Mutant

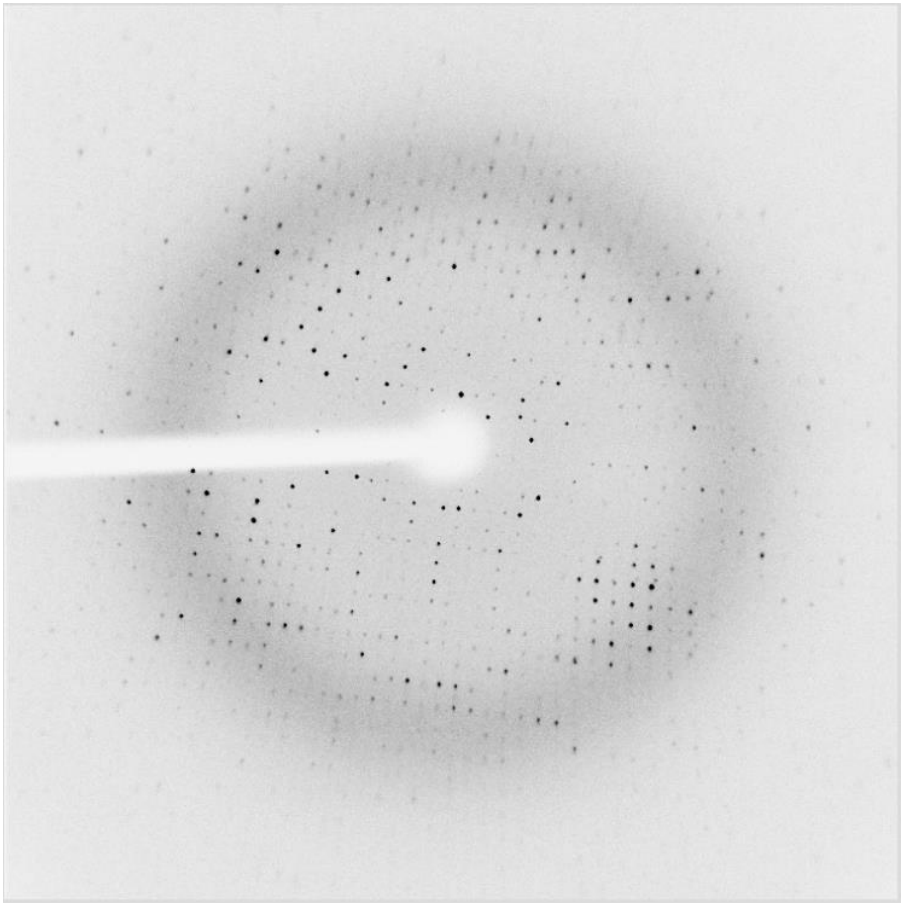
Appendix 5: pET21b(+) Vector with XhoI and NdeI Cut Sites Labelled for Insertion of Custom Genes for SCO6594-97



Appendix 6: pET15b-LIC Vector with KpnI Cut Sites Labelled for Insertion of BT2857



Appendix 7: Crystal Diffraction Pattern and Index for BT2857-C



Lattice 1

Solution	Lat.	Pen.	a	b	c	α	β	γ	σ(x,y)	σ(φ)	δ beam
1 (ref)	aP	0	49.8	62.0	68.0	117.1	90.0	90.1	0.31	0.94	0.28 (0.1)
2 (ref)	mC	0	62.1	121.1	49.8	90.0	90.1	90.0	0.42	0.94	0.27 (0.1)
3 (ref)	aP	0	49.8	62.0	68.0	62.9	90.0	89.9	0.31	0.94	0.28 (0.1)
4 (ref)	mP	1	62.1	49.8	68.0	90.0	117.1	90.0	0.30	0.95	0.28 (0.1)
5 (ref)	oC	1	62.1	121.1	49.8	90.0	90.0	90.0	0.42	0.94	0.28 (0.1)
6 (ref)	mC	1	121.1	62.1	49.8	90.0	90.0	90.0	0.42	0.93	0.29 (0.1)
7 (reg)	mC	71	62.0	121.1	49.8	90.0	90.1	90.0	-	-	-
8 (reg)	oC	71	68.0	111.0	49.8	90.0	90.0	90.0	-	-	-
9 (reg)	mC	71	111.0	68.0	49.8	90.0	90.0	90.0	-	-	-
10 (reg)	hP	72	65.0	65.0	49.8	90.0	90.0	120.0	-	-	-