Comparative functional analysis of two α-glucosidases, Family 31 Glycoside Hydrolases from the human gut symbiont *Bacteroides thetaiotaomicron*

by

Marcia M. Chaudet

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Science

in

Biology

Waterloo, Ontario, Canada, 2012

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The human gut is home to a significant number of microorganisms including the dominant symbiont *Bacteroides thetaiotaomicron*. This microbe is predicted to possess an array of glycoside hydrolases, majority of which are involved in starch utilization. Presented here is a comparative functional analysis of two α-glucosidases, Family 31 Glycoside Hydrolases from *Bacteroides thetaiotaomicron*. Enzymatic kinetics revealed these enzymes both preferentially cleave α-(1,6) linkage in comparison to the predicted α-(1,4) and favour maltose derived substrates of longer length.

Acknowledgments

I would like to voice a big thank you to my supervisor Dr. David Rose. I am very grateful that he gave me a chance and took me on as student and welcomed me into his lab. I am also very thankful that he opened my eyes to the world of crystallography and has spent time teaching me about this fascinating area of science.

I am also very thankful for the support Kyra Jones has provided me during my time in David's lab. Her guidance and constant encouragement has been very helpful. I also would like to thank present and past members of our lab for their support with my project. Additionally, I would like to also recognize the help that the Holyoak lab has provided me with my project and appreciate their support.

I would like to thank my committee members, Dr. Trevor Charles and Dr. Bernie Glick. Their recommendations and guidance with my project was very valuable and much appreciated.

I would like to thank my mother for her support through my undergraduate degree and further more with my Masters degree. Her encouragement was very important and I am very grateful for it.

My success in graduate school would not of been obtained it were not for the love and support of my close friends and family.

Lastly, I would like to thank Matt McGuire. His constant love and support was the keystone in my determination in school. I want to thank him for always encouraging me to follow my heart and my dreams and always there for me when I stumble along the way.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction and Background Information

1.1 Human-Microbial Interactions

1.1.1 The Human Microbiome

For centuries the complex interaction between humans and microorganisms has sparked interest in defining this intriguing relationship. The specifics of this relationship remain unclear; however, accumulating evidence demonstrates that these organisms are key components in our survival. From the moment we are born, we are exposed to an array of microorganisms and maintain this interaction throughout our life. This indigenous microbiota can have weighty effects on our physiology and have been predicted to exceed the number of human cells ten to one[1]. When assessing the physiological processes of humans it is not enough to only include our own specific functions, rather it would be beneficial to include those of our microbiota.

The human body is a reservoir to a multitude of microorganisms and results in a highly diverse, specialized ecosystem. These organisms can be found in various areas of the body and primarily populating regions that are exposed to the exterior environment. Microbial colonization of these regions is specific for that niche and the interaction with the host is highly specialized. The relationship that exists between the human and its microbiota is one that generates numerous effects on both partners and is considered symbiotic[2]. Symbiosis is when two organisms reside in close contact with one another and is categorized as mutualistic, commensal or pathogenic. During the host's lifetime the relationship with its microbiota can cycle between these three variants and is greatly influenced by diet, age and health[3].

1.1.2 Gastrointestinal Tract Microbiota and the Highly Populated Colon

The gastrointestinal tract is specifically of interest because of the high number of microbial cells that are present with an estimated population to be upwards of 10^{14} [4]. Physiologically, the gastrointestinal tract is quite diverse with each region demonstrating characteristically distinct

environments, which reflects the resident microbiota (Figure 1.1). The pH varies significantly through the tract, ranging from 2.0 to 7.7 with gradual transitions from one region into another. The physical nature of the epithelium also varies along the tract with an evolution of smooth epithelium in the stomach, to significant crypts and microvilli in the small intestine and slightly decreased ridges in the walls of the colon.

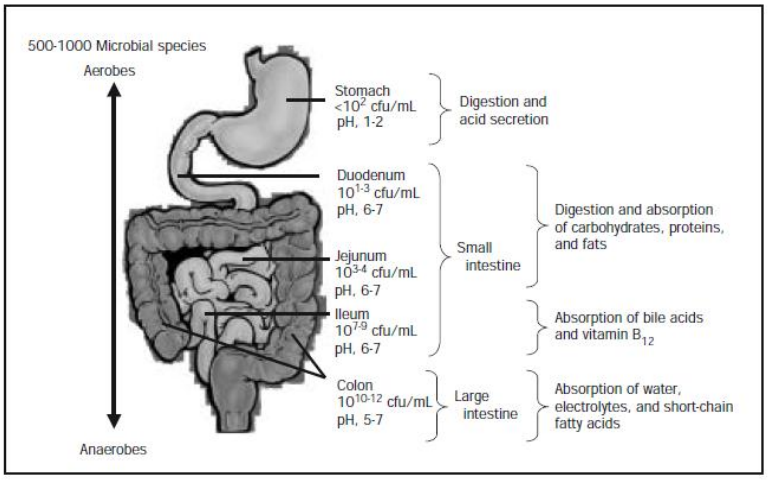


Figure 1.1 Physiology of the gastrointestinal tract. The various regions of the gastrointestinal tract demonstrate differing physiological characteristics that can regulate the specific populations of microorganisms that reside in these regions. (Figure was reproduced with permission from Elsevier)[5].

In general, majority of the indigenous population resides in the small intestine and the colon, as a result of the physical attributes of these regions. More specifically, the total population observed in the colon represents more bacterial cells than the combined cells of all other regions of the human body, highlighting it as an ideal site to study and understand the interaction between microorganisms and the human host. It has been estimated that the colon has upwards of 10^{11} microbial cells and characterized as one of the most densely populated regions in nature [1]. A major influence on the microbial population of the colon is their ability to utilize substrates passing

through this region. The composition of the partly digested matter found in the colon is highly composed of indigestible polysaccharides as a result of upstream human digestive mechanisms. The majority of the ingested polysaccharides are degraded into simple sugars in the small intestine by action of the human luminal glycoside hydrolases, resulting in the low availability of simple sugars in the colon. Therefore, the substrates entering the colon are more apt to be complex sugars that the host is unable to degrade. In order for an organism to reside in the colon, it must be able to utilize these polysaccharides that pass through or alternatively, have the ability to take up simple sugars that are released by surrounding microorganisms[6]. These undigested polysaccharides move very slowly at this point in digestion due to the limited fluidity in the colon. The combination of this with the physiological aspects of the colon enables microorganisms to flourish and function highly in this region[3].

The colon microbial population is not only quite diverse but also the composition has been shown to be highly individualistic. It has been observed that there is a significant variation between individuals which is strongly reflective in the individual's diet. This population is established by the creation of a specific equilibrium of substrates in the colon, which affects the competition between microbial populations[7]. A large component of the human diet is dietary fibre, the edible parts or component of plants or similarly related carbohydrates, which are resistant to human degradation and therefore are available for microbial breakdown[8]. There are a multitude of physiological benefits from having a diet composed of these substrates. Benefits include an observed improvement in the functional mucosal layer of the epithelial, an environment that is conducive to the proliferation of intestinal microflora, an improvement in the balance between water and electrolytes and also a reduced glycemic response to hunger. There is still a degree of ambiguity in the quantity of polysaccharides as provided by the host that is available for microbial digestion in the colon. However, measurements of diet composition at various regions of digestion have

suggested that 10 to 20% of an average meal of white bread in estimated to enter the colon and becomes available for microbial use[9]. Also, an indirect method of calculating bacterial cells in fecal samples and correlating that to required replenishment of the population and the amount of substrate to sustain such a population, demonstrated that an average of 15% of dietary sugar and starch enters the colon. In general it can be assumed that a minimum of 10% of the human diet of starch and sugar enter the colon and are considerably broken down by the indigenous microbiota. This is a significant portion of the human diet that is unaccounted for and reveals that the large intestine is an important site to study the contribution the residing microbiota can have in nutrition[10].

A mutualistic relationship is highly evident in the colon between the microbes and the human host. The human host provides a habitat and a constant supply of nutrients to the microbes, while the microbes complete multiple functions that which the host cannot perform independently. Polysaccharides provided by the host through its diet are ideal sources for microbial energy acquisition by means of fermentation. The indigenous population ferments these substrates into functional small molecules which includes an excess of short-chain fatty acids. Butyrate is an example of a short chain fatty acid and has been shown as an energy source for the colonic mucosa; upwards of 95% of its production is absorbed across the epithelia of these cells. The conversion of these metabolites into short chain fatty acids is important for colonic health and has been shown to protect against colitis and colorectal cancer[11]. Specifically, it was shown that butyrate is capable of directly supressing colon cancer cell proliferation and subduing tumour growth in the large intestine[12]. Simply, the host provides polysaccharides through its diet to the microflora and in turn the microbial by-products create a healthy colon (Figure 1.2).

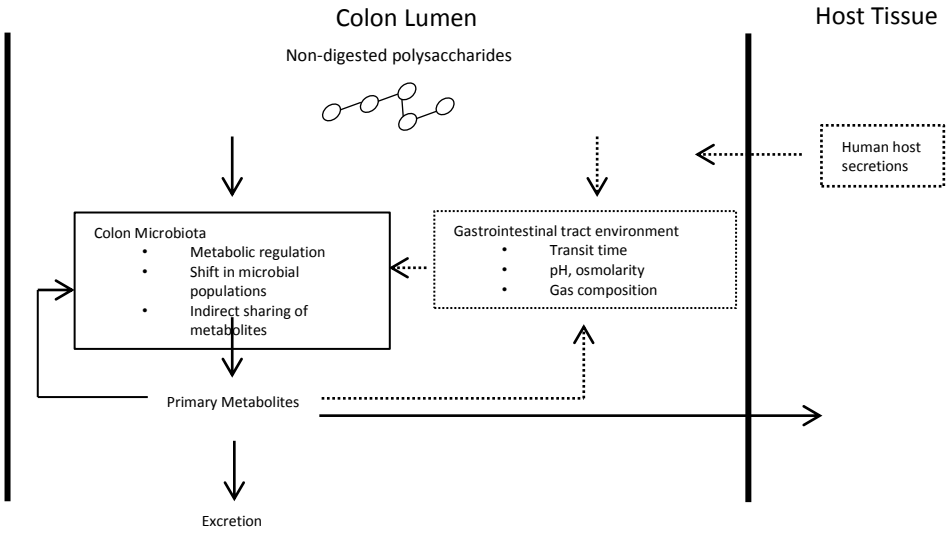


Figure 1.2 Balance in the colon. Various processes in the colon fluctuate the equilibrium in order to maintain a balance between with the host and the indigenous microbiota [13].

Utilization of dietary plant polysaccharides results in an additional relationship to develop in the colon between cohabitating microbes. This relationship occurs by means of one population utilizing the host provided substrate and in the process providing additional substrates to the environment. This process is known as cross-feeding and results from a microorganism utilizing the primary substrate and with its cleavage results in the release of a smaller form, which can then be utilized by other microorganisms within the same population (Figure 1.3). This relationship nurtures a mutualistic interaction between these populations and further nurtures the one with the host.

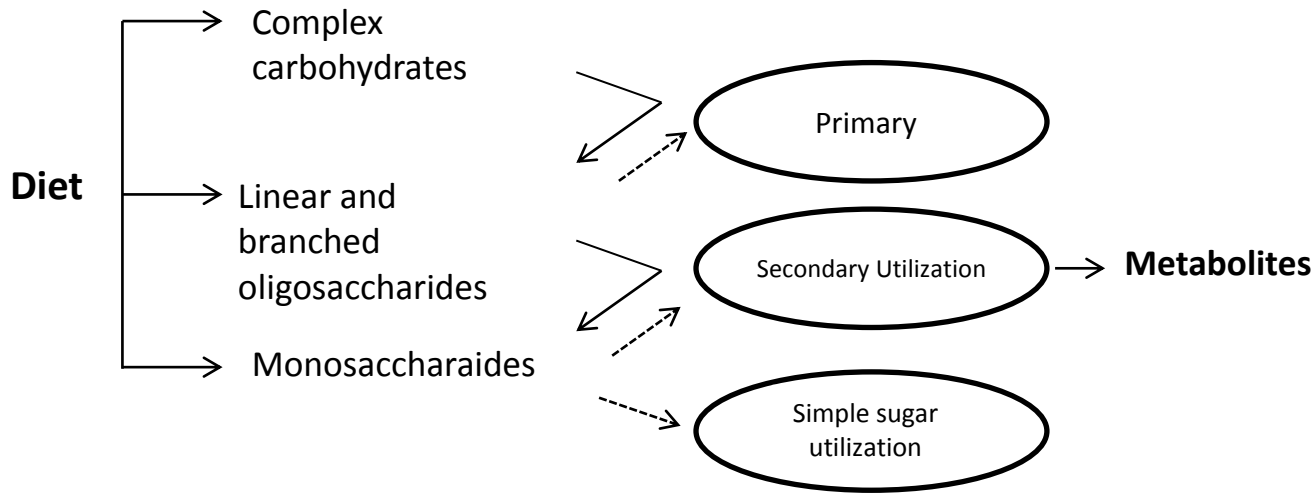


Figure 1.3 Cross-feeding. Multiple levels of breakdown of the host's diet allow for a relationship between microbial populations, resulting in sequestering substrates in order to obtain metabolites[7].

Composition of an individual's microbiota is also dependent on environmental influences and genetic predisposition. Classifications of the composition of the colon microflora was completed by 16S rRNA gene sequence-based enumeration and revealed significant diversity across the sampled pool. Over 17 phyla of Bacteria were detected with a surprising 90% of the phylogenetic types belonging to two main divisions, *Firmicutes* and *Bacteroidetes*[14].

1.2 *Bacteroides thetaiotaomicron* as a Dominant Symbiont of the Colon Microbiota

Bacteroides thetaiotaomicron is an anaerobic gram negative bacterium that is part of the natural flora of the gastrointestinal tract. It was originally isolated from fecal samples in the process of isolating pathogens from the tract[15]. This organism is of interest due to its dominant presence in the colon microbiota and has been identified as a significantly player in the biological activities of this region. *Bacteroides thetaiotaomicron* is a member of the *Bacteroidetes/Chlorobi* superphylum, which diverged prior to the appearance of gram positive and gram negative of microorganisms.

Bacteroides has a significant evolutionary distance from other phyla, such as *Proteobacteria* and *Firmicutes*. This distance is highly illustrated in its genetic distinctness from these phyla[6]. Sequencing of *Bacteroides thetaiotaomicron* genome revealed this organism has highly distinctive genetic traits with a significant portion of its genome dedicated to the acquisition and utilization of polysaccharides. These characteristics exemplify how this organism is able to thrive in the highly populated intestinal ecosystem and is able to out-compete the other hundreds of species[16].

In general, the relationship between the host and *Bacteroides thetaiotaomicron* is a blend of mutualism and commensalism. It can be suggested that it is a mutualistic relationship based on this microorganism's ability to utilize the dietary carbohydrates. This degradation ultimately results in a portion of these simple sugars being available in the lumen and is predicted to allow for the host epithelium to absorb these molecules, resulting in a contribution to the host energy intake[3]. Release of free sugars into the surrounding environment can also benefit other organisms in the intestinal population, which reveals a mutualism relationship between *Bacteroides thetaiotaomicron* and these surrounding microorganisms (Figure1.4).

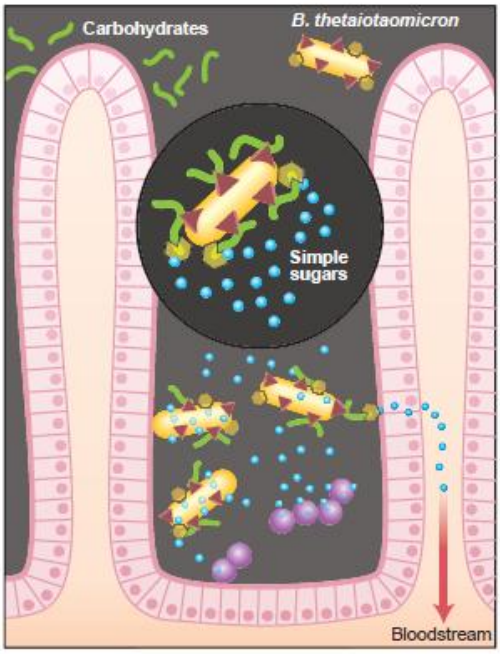


Figure 1.4 *Bacteroides thetaiotaomicron's* role in the colon. The predicted role of *Bacteroides thetaiotaomicron* in colon includes polysaccharide utilization that produces simple sugars which can be utilized by the organism, by surrounding organism and absorbed by the human host. (Figure was reprinted with permission from AAAS and illustrator Katharine Sutliff) [17].

However, it has not been empirically shown that this is the definitive role of this organism in the colon. It has also been hypothesized that the role of this organism is commensal and that its presence in this region of the body does not noticeably affect the human host.

The distal region of the colon provides an environment that is conducive to the proliferation of *Bacteroides thetaiotaomicron*. The ascending portion of the colon has an approximate pH of 5.6 due to the rapid fermentation of short chain fatty acids by the residing microbial population. The absorption of these degraded fatty acids results in a considerably shift in pH to 6.6 in the distal region. This shift generates an ideal environment with a stable pH for this organism to have a

dominant role[3]. This organism has been identified as occupying this region of the colon in significant numbers yet defining the specific region of the distal colon that this organism occupies can be quite complicated due to the complexity of this environment. The colon has multiple surfaces in which *Bacteroides* can adhere, such as the lumen region, the mucosal layer and the epithelial surface. Despite crypts in the colon being significantly shorter than those in the small intestine, they can still provide a site for this organism to proliferate. Also, *Bacteroides thetaiotaomicron* are also able to form a stable environment on the substrates that pass through the lumen, overall making it difficult to identify these organisms in their natural environment[3].

1.3 Negative Influences Associated with *Bacteroides thetaiotaomicron*

1.3.1 Colon Microbial Population as a Source of Antibiotic Resistance

There are many roles that *Bacteroides thetaiotaomicron* plays in connection to the human host. It has been shown through a number of studies that microbes derived from the gut have the ability to have unfavourable effects on its host. Recent research has shown a correlation between post-surgery infections and microorganisms derived from this region. Treatment of these infections revealed these gut microorganisms possess various degrees of antibiotic resistance with an observed difficulty associated with treatment. The specific contribution of *Bacteroides thetaiotaomicron* to such infections was investigated due to its dominance in the colon. Genetic analysis revealed that *Bacteroides thetaiotaomicron* does have a significant role in this pathology and investigation of conjugative elements illustrated the basis for the observed resistance in this organism. Analysis of various samples specified that two resistant genes were observed, *tetQ* and *ermF*. The microflora of the colon is able to obtain these resistance elements via two postulated methods. Species that reside in the colon that possesses such elements are able to share them among other organisms of the same

population (Figure 1.5). Also, they can acquire these elements from microorganisms that do not primarily reside in this environment, but rather are passing through the gut and subsequently are able to pass them on to the indigenous population[18].

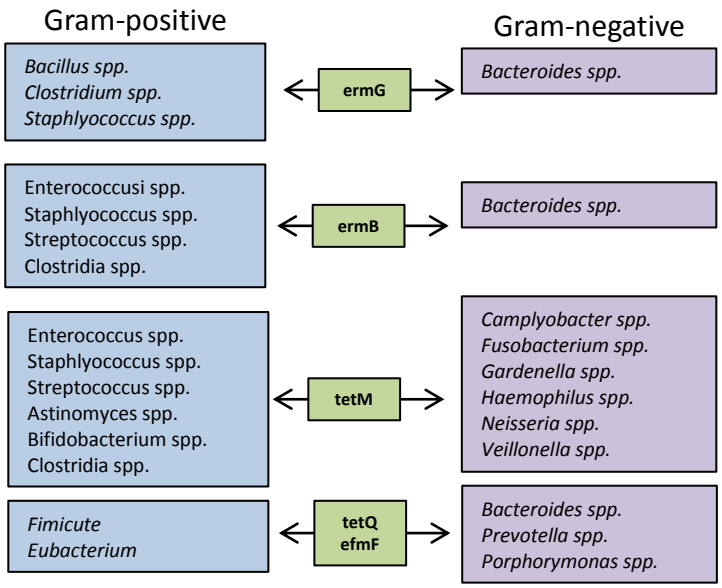


Figure 1.5 Antibiotic resistant genes derived from the colon. Predicted method of bacterial strains obtaining antibiotic resistance occurs by transfer of restraint genes between Gram-positive and Gram-negative bacteria that originate in the colon[18].

1.3.2 Implications of *Bacteroides thetaiotaomicron* in Non-Gut Pathology

Additional negative associations have been observed with *Bacteroides* species and their involvement in various human diseases. Observations of anaerobic infections and bacterial strains that reside in the infectious area revealed that the bacterium *Bacteroides* is one of the dominant participants. The most commonly associated infections with *Bacteroides* is intra-abdominal sepsis, which results from this organism disembarking from the gut due to intestinal rupture or intestinal

surgery leading to contamination of nearby tissues. This infection can arise from a surgical wound, appendicitis or malignancies with *Bacteroides* species colonizing in the later stages of the infection. *Bacteroides* species have also been shown to be implicated in bacteraemia, as the causative agent in the immuno-compromised and elderly. This infection stems from damage to mucosal barriers, which similarly, allows for this bacterium to remove itself from its common environment and colonize other regions. This organism has been classified as an opportunistic pathogen, yet the mechanism by which it relocates to other regions of the body demonstrates that it doesn't act in such a manner. This gives the impression that this organism doesn`t take advantage of the weakened state of the host, rather appears to seek out additional sources of nutrition outside the gut. *Bacteroides thetaiotaomicron* gives the impression that it is highly commensal with its host if the healthy boundaries of the intestinal mucosa are maintained[19].

1.4 Positive Associations of *Bacteroides thetaiotaomicron* in the Colon

1.4.1 *Bacteroides thetaiotaomicron* and the Maturation of the Immune System

Bacteroides thetaiotaomicron has also been highly implicated in positive associated interactions with the human host as demonstrated in the interaction with the colon immune system. The colon is an environment that is highly immune-competent, with direct interactions with the host's immune system. It remains unclear as to why this highly populated region of the gastrointestinal tract is able to exist closely without bringing harm to the human. It has been predicted that the microbiota of the colon is an essential component to the development of the human's immune system and these commensal organisms present specific factors that allow the host immune system to mature[20].

One speculation is that the indigenous population has specific factors on their cell surface that induce system maturation. Bacteria possess capsular and cell walls that are composed of

carbohydrates and have been shown to be recognized as antigens by the host's immune system. These carbohydrate containing structures stimulate T-cells which in turn induce an IgM response. Researchers assessed the effect of inoculating germ-free mice with *Bacteroides fragilis* and observed the effects on the host's immune system. Despite this study involving a single species that originates from a multispecies environment, it was relevant in showing that the presence of this bacterium had a positive effect on the mouse's immune system. The presence of *Bacteroides fragilis* was adequate to facilitate a correction in CD4+T-cell defection in the spleen of the mice. Furthermore, it was found that a mutated strain of *Bacteroides fragilis*, was unable to elicit the same observed response, further demonstrating an important interaction between the host and its residing microbiota[21]. Though these results were observed with *Bacteroides fragilis*, this specie is strongly related to *Bacteroides thetaiotaomicron*, therefore, a similar interaction can be assumed to exist.

Analysis of the effects of the microflora on the immune system revealed that *Bacteroides* species induced expression of antimicrobial molecules, which directly adhere and eliminate potential pathogens in the colon. Paneth cells are components of the lower gastrointestinal tract that produce granules containing antimicrobial proteins. These cells were evaluated by DNA microarray analysis in germ free mice and conventionally raised mice to observe changes in expression levels. It was observed that mice with an indigenous microbial population had an increased level of expression of a specific protein from the *reg* gene family. This secreted protein has a conserved carbohydrate recognition domain and previously was reported to be expressed during inflammatory responses during mucosal cells invasion. Purified Reg protein was demonstrated to bind significantly to peptidoglycan forms, commonly found on the surface of bacterial cells and indirectly affects the survival of pathogens. Evaluation of mice inoculated with *Bacteroides thetaiotaomicron* showed an increase in Reg expression; though it was not as high as conventionally raised mice that have a multispecies population. The presence of this indigenous population allows for the functioning of

the host immune system in the gut and specifically the cells found in the mucosal layer, and assist in immune functioning against pathogenic microorganisms[22].

Further evidence was demonstrated in the investigation of Ang4, a member of the human angiogenin protein family, which is commonly associated with the growth of the human tumour cells in mice. It was revealed that Ang4 is produced by Paneth cells secreted in to the gut lumen and was shown to have specific antimicrobial activity for intestinal microbes. Expression of this protein was found to be influenced by the presence of *Bacteroides thetaiotaomicron,* further supporting the notion that intestinal commensal microbes play an important role in the human host's innate immune system[23].

The multitude of microbes residing in the brush border and microvilli of the colon wall also creates a physical limitation for invading microbes. The indigenous population limits the space available for pathogens to invade and colonize. The microbes that naturally reside in the colon are able to retain their position by having an interaction between the microbial polysaccharides that lie on the cell surface and the epithelial mucins. This mechanism allows these microorganisms to have a resistance to being washed away due to the peristaltic movements of the colon and thus lowering the chances of pathogens adhering. The presence of the microorganisms also triggers the release of granules of mucins, which in turn reinforces their adherence to the mucosal surface[24].

These studies support the idea that the presence of *Bacteroides thetaiotaomicron* in the colon contributes to the development and the maintenance of the human immune system. Specifically, the colonization of this organism in this region of body limits the invasion of foreign pathogens that ultimately can have a detrimental effect on the physiological state of the host.

1.4.2 Probiotics and Prebiotics as Contributors to a Healthy Colon

The use of probiotics has been around for many years and recently has become an important tool in treating digestive illnesses. A probiotic is defined as a live microorganism that when consumed in significant quantities can have positive health benefits on the host[25]. Probiotics are commonly used to treat illnesses that include lactose intolerance, inflammatory bowel disease, and gastroenteritis, along with evidence to suggest that this treatment can influence the origin of colon cancer[26]. There are various mechanisms that probiotics can act in the gastrointestinal tract which influence the host and aide in restoring the colon to a healthy environment (Figure 1.6). Probiotic strains can colonize the epithelial surface and limit the physical available space for pathogens to adhere. They encourage goblet cells to produce mucus which limits the ability of pathogens to adhere to the epithelial surface. They allow for tight junction proteins to be unregulated, resulting in the limitation of unwanted pathogenic molecules to permeate the surface. Lastly, they are also capable of producing secreted molecules that hinder pathogenic strains from functioning normally[27].

It has also been observed that the presence of these probiotics in the colon can result in an increase in the amount of short chain fatty acids contributing to the maintenance of a healthy colon. Through animal studies it was also shown that probiotics can provide a protective barrier in the ridges between crypts and limit the colonization of pathogens. These are fundamental features of a healthy colon due to the presence of pathogens resulting in inflammation, which over time can lead to possible sites for digestive disorder development [27].

Microbes from the gastrointestinal tract have been used for a number of years as probiotics. The most commonly used strains are lactic acid bacteria. However, recent evidence has demonstrated that probiotics that contain additional strains have positive effects on patients suffering from viral diarrhea. Studies involving *Lactobacillus casei* and *Bacteroides thetaiotaomicron* strains demonstrated modifications to cell surface glycans. Rotaviruses are observed to enter the cell by disturbing cell surface glycans, resulting in cell rupture and leading to gastroenteritis. Exposure of

intestinal cells to these bacterial strains results in strengthening the cell wall through modifications. Ultimately, the exposure of epithelial cells to these specific strains indirectly is able to lower the possibility of infection from rotavirus[28]. Utilizing strains specifically identified to be members of the colon microbiota can provide a less invasive method of treatment for individuals suffering from various digestive disorders.

A developing area of nutrition is the use of prebiotics for the maintenance of a balanced digestive tract. Prebiotics involve the ingestion of medical fibres, which are purified from their natural source, and in essence provide a substrate that can encourage the proliferation of the microflora of the gastrointestinal tract. The effects of prebiotics have been shown to be very beneficial, demonstrating similar healthy effects as observed with consuming natural dietary fibre. Prebiotics are not to be used in replace of nutrition that provides natural dietary fibres, but rather should be used if an individual is limited by their diet, perhaps due to allergies or a gastrointestinal disorder[8]. There has been significant documentation reporting of the positive benefits prebiotics can have on individuals with type 2 diabetes and those suffering from heart disease. There is no specific evidence to show that prebiotics specifically promote the proliferation of *Bacteroides thetaiotaomicron*. None the less, it can be assumed that use of prebiotics can have a positive effect and provide a substrate that would encourage this bacterium's proliferation. These reports further illustrate that nurturing the microbiota in the gastrointestinal tract can have profound positive effects on the individuals overall health.

1.4.3 *Bacteroides thetaiotaomicron* and Inflammatory Bowel Diseases

Bacteroides thetaiotaomicron has been observed to be associated with various inflammatory disorders of the gastrointestinal tract, such as Irritable Bowel Diseases (IBD), Irritable Bowel Syndrome (IBS) and Ulcerative Colitis (UC). These disorders are characterized as having

symptoms of abdominal pain, bloating and altered bowl movements. Studies evaluating the gut microbiota and their interaction with these illnesses demonstrated a strong correlation in the composition of the flora and the observed symptoms[27]. Assessment of gene composition of the microbial population in the human gut revealed that individuals with IBD have 25% fewer genes in comparison with average individual[29]. A significant link was found to exist between the specific species that are present in the microflora of the gut and these illnesses, as shown through the evaluation of fecal bacterial composition in healthy individuals and those suffering from IBS and UC. There was an observed lower level of *Bacteroides* species in unhealthy individuals in comparison to the healthy controls[30]. This highlights the impact the microbiome can have on the human host and specifically the metabolic stability of the host's digestive system. As previously highlighted, a poorly established bacterial population can result in inflammation of the colon epithelium, ultimately leading to the development of these disorders[24].

1.5 *Bacteroides thetaiotaomicron's* Influence on the Human Host's Nutrition

1.5.1 The Effect of *Bacteroides thetaiotaomicron* on Obesity

The balance of energy in an individual is based on the intake of calories and the metabolic expenditure and storage. The main cause of obesity is an excess of caloric intake due to a highly functioning metabolism, but also generally due to an excess of provided nutrients. Obesity is of recent concern due to its increasing rate in developing countries. It has been characterized as an epidemic and has become a recent focus in nutritional research in reference to the functioning of the gastrointestinal tract[5]. Current research involving obese and lean mice revealed the composition of an individual's gut microbiota can have a strong influence on an individual's metabolic processes (Figure 1.6). It was observed that germ free mice inoculated with a distal gut microbiota community

from normally cultured mice demonstrated an increase in body fat. Over a period of 10-14 days it was found that the colonized mice had a significant increase in body fat during the period of limited food consumption. This study brings attention to a specific interaction between the gut microbiota and the host with an observed physiological consequence on the host. It was also revealed that the molecular level of this interaction involve the community of microbes fermenting the polysaccharides entering the distal gastrointestinal tract, leading to absorption of monosaccharaides and short chain fatty acids. These acids are then converted into complex lipids in the host's liver and further regulate the transcription of host genes responsible for lipid disposition. Overall, this study lends to the hypothesis of obese individuals with a microbiota result in being more efficient at extracting energy from nutrients in comparison with lean individuals[31].

Characterization of the gene content of the gut microbiota was completed to further support the postulated link between the gut microbial community and obesity. Investigation of the distal gut microbiome of genetically obese, leptin deficient mice and the associated lean littermates was completed by metagenomic techniques. Additionally, obese and lean humans were assessed and genetic information of the same region of the microbiome was taken into account in the analysis. This type of comparative analysis defines a consistency between mice and human gut microbiome. Specifically, it was observed that the proportion of *Bacteroidetes* and *Firmicutes* correlates with the body mass index of the individual and the observed obese organisms from both groups have a significant decrease in the percent of *Bacteroidetes*[4].

Analysis of microbial communities from adult female monozygotic and dizygotic twin pairs of obese and lean physiology, as measured by 16S rRNA sequencing, addressed various influences on the gut microbiota, that include the host, genotype, environmental exposure and host adiposity. It was revealed that the gut microbiome is shared among family members but additionally it was shown that there is a 'core' microbiome at the gene level for all individuals. Results here also

support the understanding that obesity is associated with phylum-level changes of the microbial population in the gut and reduced bacterial diversity. An interesting link was found between the level of functional diversity in the microbiome and the abundance of *Bacteoridetes*[32].

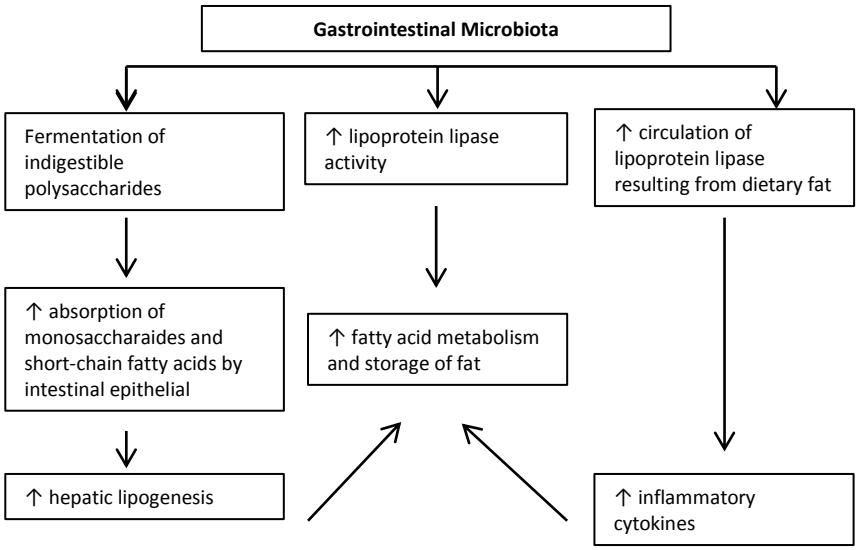


Figure 1.6 Connection of the gastrointestinal microbiota with obesity. Various contributions arise from the gut microbiota that can have a strong influence on the development of obesity in the human host[5].

Investigating the link between obesity and the host microbiota has revealed an interesting relationship, where in the case of obesity, a change in the microbiology of the gut can directly affects the host's pathology[33]. Specifying this link also highlights a unique opportunity to have an active role in the health of obese individuals. The specific content of this region of the microbiome can possibly act as a 'biomarker' to indicate individuals who are predisposed to the development of obesity and furthermore, a therapeutic target for people already suffering from this disorder. Focus of research on this link and contributing variables is an important step in understanding the

association of genetic and environmental factors and how they influence on the composition of the gut microbiota that is specific for each individuals with obesity[34].

1.5.2 Starch Utilization by *Bacteroides thetaiotaomicron*

Bacteroides thetaiotaomicron has also been highly implicated in having a strong effect on the metabolism of healthy individuals. It has been postulated that its presence in the colon directly impacts the utilization of various polysaccharides. Its contribution to nutrition has been investigated by the analysis of its growth requirements and it was revealed that it demonstrates the ability to make use of a variety of polysaccharides. *Bacteroides thetaiotaomicron* was observed to utilize polysaccharides that varied in length, as well as the composition of the sugar residues[35]. Specifically, it was found that this organism was able to grow on various profiles of starch[36]. Biochemical and genetic studies revealed that *Bacteroides thetaiotaomicron* possesses several genes that are fundamentally involved in the utilization of starch[37]. Investigation of the genes located near these identified starch genes revealed multiple outer membrane proteins involved in the binding and utilization of starch. Mutation studies further demonstrated that these proteins heavily influence the ability of this organism to utilize starch as an energy source[38]. This multiple protein system is collectively referred to as the Starch Utilization System (Sus) and has been stated to enable *Bacteroides thetaiotaomicron* to bind, degrade and utilize polymers of starch (Figure1.7).

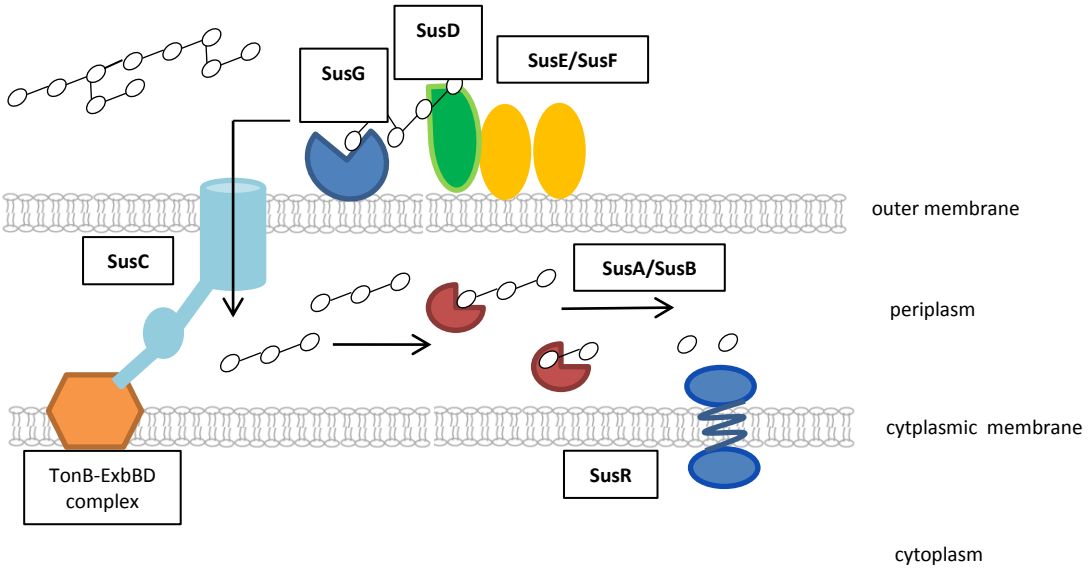


Figure 1.7 Starch Utilization System of *Bacteroides thetaiotaomicron*. This system has been characterized to consist of multiple proteins attached to the membrane that are able to bind and hydrolyze starch and export smaller oligosaccharides into the bacterial cell. These can then be further degraded into smaller units which are utilized for energy or involved in up-regulation of the system[39].

The Sus system consists of an outer-membrane binding starch protein (Sus-D), membrane associated glucosidases (Sus-G, possible Sus-E and F), an outer membrane transporter that transports smaller oligosaccharides into the periplasm (Sus-C), a multiple periplasm associated glucosidases (Sus-A and B) and a transcriptional regulator (Sus-R) regulated by the released monosacchraides[39]. Crystal structures of Sus-D illustrated direct contact with β-cylcodextrin, identifying its involvement in the binding of starch[40]. Crystal structure of Sus-G, an α-amylase, revealed that this protein is also involved in the utilization of starch by cleaving starch polymers into smaller units[41].

Understanding the details of *Bacteroides thetaiotaomicron's* ability to utilize starch was explored by expression studies utilizing microarray analysis. Growth on glycan mixtures and analysis of the resulted expressed genes, demonstrated multiple loci involved in the utilization of these glycans. These genes were profiled and catalogued in Polysaccharide Utilization Loci (PUL), which consisted of multiple proteins involved in starch utilization with multiple homologs to Sus proteins[39]. Over 80 PUL's were identified and characterized to contribute to *Bacteroides thetaiotaomicron* ability to utilize the provided polysaccharides and further supported the characterization of this organism as being highly capable of utilizing substrates derived from the colon[42].

1.5.3 Family 31 Glycoside Hydrolases

The ability of an organism to utilize carbohydrate molecules is defined by whether they possess enzymes that are capable of cleaving the bonds that make up the carbohydrate chain. These enzymes are known as glycoside hydrolases and are specifically characterized as catalyzing the hydrolysis of glycosidic linkages releasing simple sugars[43]. These glycoside hydrolases are classified into over 100 families by their specific enzymatic functionality, related substrate specificity and amino acid composition. These families encompass a wide range of capabilities along with various preferred substrates[44].

Family 31 is a glycoside hydrolase family that is relevant to nutrition as a result of it being one of two families that has been shown to demonstrate α-glucosidase activity and indicated to be involved in the degradation of starch polymers. Human Family 31 enzymes have been investigated due to their significant contribution to human digestion and have been shown to be specifically involved in starch digestion. These enzymes, maltase-glucoamylase (MGAM) and sucrase-isomaltase (SI) contain four α-glucosidases that are part of two small intestinal membrane bound enzymes with

multiple activities of maltase and sucrose isomaltase. These enzymes are composed of duplicate domains and have been characterized as contributing to metabolism from the human diet by the cleavage of starch into individual glucose units that are utilized for energy[45]. A variety of enzymes are members of Family 31, however, α-glucosidases are the primary enzymes of interest in relation to digestion. Family 31 α-glucosidases are endo-acting enzymes and are predicted to act on the terminal non-reducing end of starch polymers, resulting in the release of free glucose[44]. The specific mechanism of this family is the retaining mechanism, which was first proposed by Koshland[46]. The retaining mechanism involves two successive nucleophilic substitutions at the anomeric center (Figure 1.8). The first step is the nucleophilic residue acting on the anomeric center by the assistance of a second residue acting as an acid, creating the glycosyl-intermediate. In the second step, the deprotonated acidic residue, acts as a base and accepts a hydrogen ion from a water molecule, resulting in the cleavage of the intermediate, producing the liberated free glucose[47].

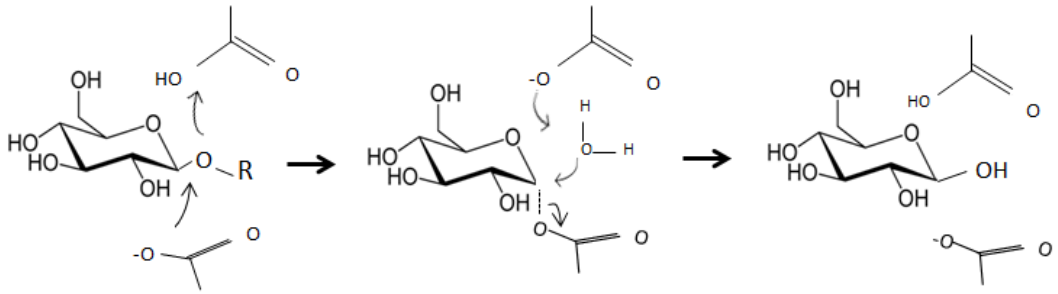


Figure 1.8 Retaining enzymatic mechanism. The predicted mechanism of Family 31 Glycoside Hydrolases is the retaining mechanism[48].

Multiple families utilize this specific mechanism in their hydrolysis of substrates, yet the defining characteristics of Family 31 Glycoside Hydrolases is the residues involved in the cleavage of the substrate, specifically characterized as aspartate residues.

Interestingly, both Bacteria and Eukaryotes possess enzymes that have been predicted to be members of Family 31. These organisms are distinctly different and overall homology is quite low, however, some of these enzymes have a common signature sequence (Figure 1.9). This brings attention to the idea that though there is limited similarity in the sequence of these enzymes, they may have a common functionality.

Figure 1.9 Signature sequence of Family 31 α-glucosidases. A comparison across kingdoms of Family 31 α-glucosidases reveals a fairly consistent signature sequence. NT-MGAM, NT-SI, CT-MGAM and CT-SI are human enzymes, MalA is a from *Sulfolobus solfataricus*, *R.obeum* is an enzyme from *Ruminococcus obeum*, YicI is from *Escherichia coli* and BT_3299, BT_3086 and BT_0339 are from *Bacteroides thetaiotaomicron.*

1.5.4 Family 31 Glycoside Hydrolases from *Bacteroides thetaiotaomicron*

Sequencing the genome of *Bacteroides thetaiotaomicron* revealed a significant percentage of its genome dedicated to glycoside hydrolases[1]. In comparison with other members of the colon microbiota, it is demonstrated that *Bacteroides thetaiotaomicron* possess an advantageous number

of these carbohydrate enzymes which possibly reflects the dominate role it has in this region of the gastrointestinal tract (Figure 1.10).

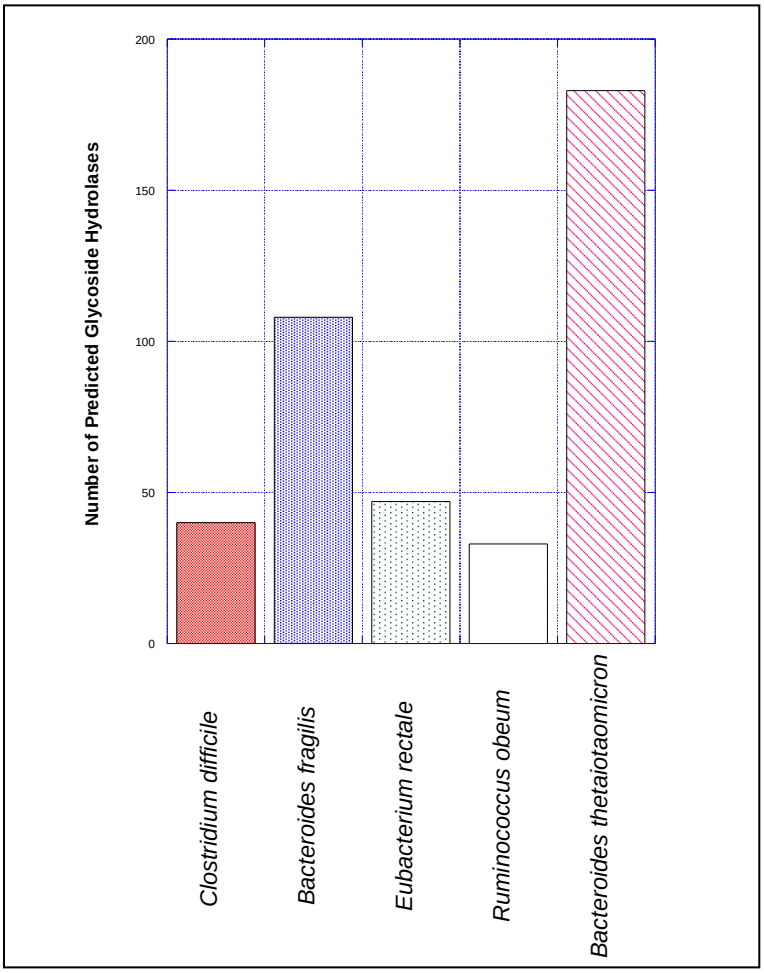


Figure 1.10 Comparison of predicted glycoside hydrolases. Comparison of predicted glycoside hydrolases with common members of the colon microbiota demonstrates *Bacteroides thetaiotaomicron* has an advantageous number over these other microorganisms[44].

In comparison to other microorganisms found in the gut, *Bacteroides thetaiotaomicron* has a greater number of Family 31 Glycoside Hydrolases, with three predicted α-glucosidases. These enzymes are predicted to utilize starch as part of its metabolism and release free glucose, which can be indirectly absorbed by the human host. Progressing the understanding of these enzymes can provide

information on the ability of these enzymes to contribute free glucose to the host at comparable levels of the human forms of these enzymes. Understanding the functionality of these enzymes can also provide some comprehension of a process that is predicted to readily occur in *Bacteroides thetaiotaomicron.*

1.6 Focus of Family 31 Glycoside Hydrolases from *Bacteroides thetaiotaomicron*

The objective of this research was to express a recombinant Family 31 Glycoside Hydrolase, an α-glucosidase, from *Bacteroides thetaiotaomicron.* This study incorporated the assessment of this enzyme and its ability to hydrolyze a range of substrates, which are derivatives of starch, a common component of the human diet. The goal was to provide an understanding of how this specific enzyme functions within the human gut and observe its predicted contribution to human nutrition. It was hypothesized that this enzyme is an α-glucosidase which is able to contribute to the digestion of starch-derived substrates by liberating free glucose that indirectly can be utilized by the host. This study highlights an area of research that can provide insight into the complex commensal environment that resides within the human intestines that can profoundly affect the host's diet and nutrition.

The research proposed here will build upon past research by investigating a probable glycoside hydrolase from Family 31 Glycoside Hydrolase derived from the organism *Bacteroides thetaiotaomicron*. BT_0339 is the specific α-glucosidase that is focused on in this study with the purpose of defining its functional capabilities in-vitro. This enzyme has an observed limited similarity to the other members of Family 31 from *Bacteroides thetaiotaomicron* (Figure 1.11).

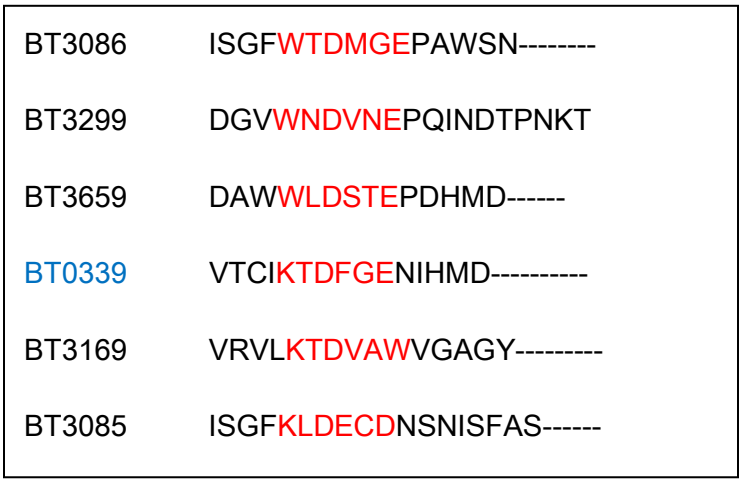


Figure 1.11 Comparison between Family 31 Glycoside Hydrolases from *Bacteroides thetaiotaomicron.* Signature sequence comparison across this family for *Bacteroides thetaiotaomicron* reveals that there are distinct differences amongst this family.

However, BT_0339 was compared across *Bacteroides* species and it was found that there was a significant sequence similarity and the signature sequence was consistent (Figure 1.12). This indicates that there is a consistent enzymatic function that has been retained across this genus and preserved during evolution and may contribute to the specific niche of the gut microbiota. Therefore, the proposed hypothesis in this study is relevant to other predicted family 31 α-glucosidases in other members of *Bacteroides*.

Bacteroides ovatus	VTCIKTDFGENIHMDALYKGMKPELLNN
Bacteroides xylanisolvens	VTCIKTDFGENIHMDALYKGMKPELLNN
Bacteroides thetaiotaomicron	VTCIKTDFGENIHMDAVYKGMKPELLNN
Bacteroides dorei	VTCIKTDFGENIHMDALYKGMKPELLNN
Bacteroides vulgatus	VTCIKTDFGENIHMDALYKGMKPELLNN

Figure 1.12 Comparison of Family 31 α-glucosidases from *Bacteroides* species. Signature sequence comparison between *Bacteroides* species found in the colon results in a highly conserved sequence.

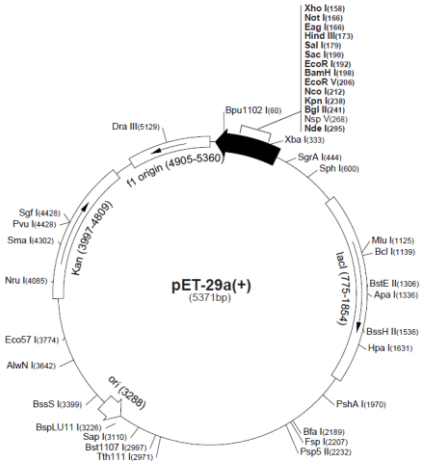
Additionally, a previously expressed and purified α-glucosidase, BT_3299[49], that is also a member of Family 31, was assessed in this study. The understanding of the functionality of this enzyme can contribute to the evaluation of BT_0339 and to the comprehension of how *Bacteroides thetaiotaomicron* is able to utilize starch and ultimately how it's able to contribute to the human diet.

Chapter 2

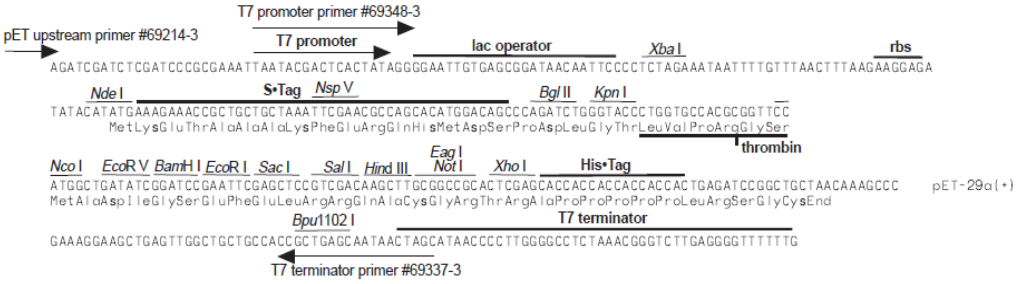
Methods

2.1 Expression System for *bt_0339*

The Novagen pET System was the selected expression system, a commonly used instrument in protein expression. This specific system allows for the expression of the gene of interest under the control of a highly effective bacteriophage T7 transcription and translation signals[50]. Two vectors were selected from this system, allowing for the expression of two forms of the protein, providing a degree of flexibility with the protein product. The pET-29a was the initial vector selected and contains a C-terminus purification His-tag, Kanamycin resistance as a selectable marker and a multiple cloning site (Figure 2.1).



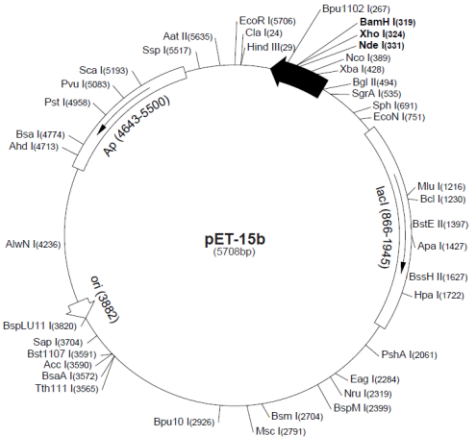
a)



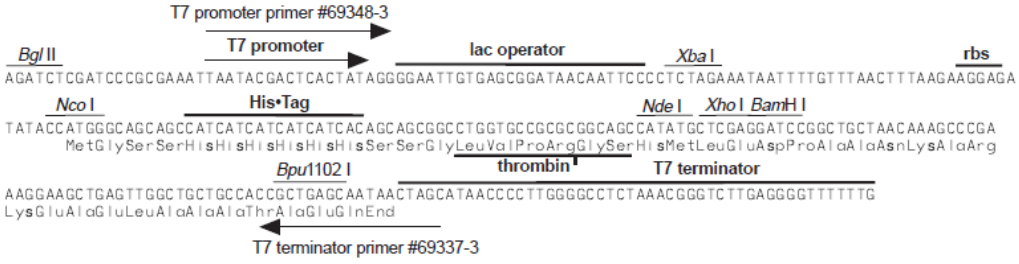
b)

Figure 2.1 pET-29a vector. a) Vector map b) cloning/expression region. (Picture was reproduced with permission from EMD) [50].

pET-15b was also selected as an expression vector and contains a N-terminus purification His-tag, Ampicillin resistance as a selectable marker and a multiple cloning site (Figure 2.2). Alteration in the location of the His-tag (N-terminus or C-terminus) will result in a slight variation in the produced protein and manipulation of such a variable could be beneficial in crystallization trials.



a)



b)

Figure 2.2 pET15b vector. a) Vector map b) cloning/expression region. . (Picture was reproduced with permission from EMD) [50].

2.2 Cloning of the gene *bt_0339*

Genomic DNA of *Bacteroides thetaiotaomicron*-VPI 5482 was purchased from ATCC. The predicted gene *bt_0339* was amplified from the genomic DNA by polymerase chain reaction (PCR) with primers designed with two restriction sites (Table 2.1). Primers contained an NdeI and XhoI cut site for the forward and reverse respectively and align with the same cut sites in the expression vector.

Table 2.1 Primers used for cloning of BT_0339. Underlined region indicates the restriction site.

Primer	Sequence (5'-3')	Restriction Site
Forward	CGAGACATATGATGAAACCTACTAATTACCATTTG	NdeI
Reverse	CGAGACTCGAGTTTAGTCCAAAATCCTTTGTAAT	XhoI

The PCR protocol consisted of denaturation at 98°C for 10 sec, annealing at 62°C for 30 sec and extension at 72°C for 1 min, for 35 cycles. PCR product was purified, digested and ligated into the two respective vectors. Constructs were transformed into *Escherichia coli* strain, DH5α. Screening of colonies for the vector with the desired insert was completed by purification of the vector, digestion with a single restriction enzyme and screening by electrophoresis for an increase in vector size proportional to the insert. Confirmation of the insert was completed by DNA sequencing with primers specific for the T7 site in the vector and was completed in both forward and reverse directions.

2.3 Expression of α-glucosidase gene *bt_0339*

The pET-29a plasmid, containing *bt_0339* (BT_0339-29a) was transformed into an *Escherichia coli* Rosetta (DE3) pLysS strain, a highly stringent expression strain containing a chromosomal copy of the gene T7 RNA polymerase and Chloramphenicol resistance as a selectable marker. In order to

obtain high transformation efficiency, transformation was completed with freshly prepared competent cells. Cultures were plated on LB-agar plates with both selective antibiotics. A single colony was used to inoculate LB broth with the appropriate antibiotics and grown overnight. Small scale expression was initiated by using this overnight culture to inoculate fresh LB broth with the appropriate antibiotics and grown to an OD_{590} reading of approximately 0.4 at 37°C. Expression of T7 RNA polymerase was induced through the addition of isopropyl-beta-D-thiogalactopyranoside (IPTG) for a final concentration of 1mM. Cell pellet was re-suspended in lysis buffer (50mM NaH_2PO_4 and 300mM NaCl) and cells were lysed by repeated rounds of sonication. The cell free extract and lysed cells were run on a SDS-PAGE for confirmation of the expressed protein, which was ~87,000 Daltons in size. Optimization of expression was completed by varying the expression time period, incubation temperature and the final concentration of IPTG.

The optimized conditions were then used in a large scale expression. Large scale expression was completed with incubation at 37°C until an OD_{595} of 0.4 was reached, induced expression by the addition of IPTG at a final concentration of 50μM and expression in an overnight time period at 30°C. Cell pellet was re-suspended in 200ml of lysis buffer and cells were emulsified at 30psi for three rounds.

Expression was also completed with the construct in the vector pET-15b (BT_0339-15b). Small scale optimization of expression was completed in the same manner as with BT_0339-29a. The optimized conditions for this expression were incubation at 37°C until an OD_{595} of 0.4 was reached. Expression was then induced by the addition of IPTG at a final concentration of 500μM and incubated overnight at 30°C. Cell pellet was re-suspended in 200ml of lysis buffer and cells were emulsified at 30psi for three rounds.

2.4 Purification of BT_0339

Affinity chromatography was the first step in the purification of BT_0339-29a. The 6xHis-tag that is adjacent to the recombinant protein was the tool utilized in this chromatography which has a high affinity for the nitrotriacetic acid chelating resin that is charged with nickel ions. Small scale purification was completed and fractions were run on an SDS-PAGE to determine the optimal elution that the protein of interest eluted off the column. Full purification was completed in a batch incubation of the lysed cells with NTA-Ni resin slurry in a ratio of one part resin slurry and ten parts lysed cells. Protein/resin slurry was incubated overnight with rotation at 4°C. Slurry was then packed into a column for gravity flow purification. The packed column was washed twice with wash buffer (50mM NaH2PO4, 300mM NaCl and 10mM imidazole, pH 8.0) and the protein of interest was eluted off in a concentration gradient of imidazole (25-250mM). Fractioned elutions were collected in 1ml volume for the optimized elution. The pooled elution was desalted into a stable buffer (20mM Tris-HCl, pH 7.5). This step prevents the protein from salting out of solution in the presence of high amounts of imidazole and allows the protein to be in a compatible buffer for the next step of purification. To ensure the protein solution is homogenous, an additional purification step was required. Fast Protein Liquid Chromatography (FPLC) purification was completed with the use of an anion exchange column, allowing for the separation of contaminates that are in close size to BT_0339 by utilization of the different isoelectric points of these proteins. The selected pH range for the buffer for the purification with this specific column and protein was Tris-HCl, pH 7.6-8.6, due to the pI of BT_0339 being 5.7. Protein solution was loaded onto the column and eluted off in a concentration gradient of 1M NaCl from 0-100%. Fractions with an indication of protein were assessed by SDS-PAGE to ensure the fractions to be pooled did not have residual contaminates.

Confirmed fractions from a slender peak on the FPLC absorbance of 280nm were pooled and a buffer exchange was completed in order to remove residual NaCl. The final working solution was then concentrated to reach the desired working concentration in the range of >5mg/ml. Other fractions that had remaining amounts of the protein of interest were then pooled and used for enzyme assays.

Confirmation of the isolated protein was completed to ensure it corresponds to the predicted α-glucosidase prior to initiation of the enzyme assays and crystallization trials. A sample was sent to the Aims Laboratory at the University of Toronto where they completed Mass Spectrometry-Electron Ionization. This specific method vaporizes the protein to produce ions and then the analyzer determines the mass to charge ratio. The detector then records the occurrence of these ions and reports the overall the mass[51]. This estimated mass can then be compared to the predicted mass that was calculated from the amino acid sequence of the enzyme.

Purification was completed for BT_0339-15b in the same manner as with BT_0339-29a

2.5 Substrate Specificity of BT_0339

Substrate specificity is an important characteristic to define for the enzyme of interest and can lead to an understanding of its function and structure. A variety of substrates were assessed using a standardized method, glucose oxidase assay. This assay evaluates the enzyme's ability to utilize the provided substrate and release free glucose in solution. This free glucose arises from its cleavage from the provided substrate and colour development occurs with the addition of the glucose oxidase/peroxidase reagent. The free glucose reacts with the glucose oxidase reagent, producing hydrogen peroxide which in turn, reacts with peroxidase. Reaction with peroxidase produces a colour change that can then be detected by a spectrometer.

Substrates evaluated were maltose, isomaltose, maltotriose, maltotetraose, maltopentaose, maltohexaose, maltoheptaose, sucrose and lactose. Maltose was the initial substrate assessed as a result of the predicted substrate specificity for BT_0339. Optimization of maltose allowed for the determination of the concentration range that allowed for a complete enzymatic curve to be calculated. All enzyme assays were completed with the enzyme BT_0339-29a and at a final concentration of 10nM and at a pH of 6.5. The assay reaction consisted of 40µl of maltose solution, 10µl of enzyme at 10nM and 0.1mg/ml of BSA. Assay was incubated at 37°C for 45 min and then the reaction was quenched by the addition of 3M Tris, pH 6.9. The reaction was developed by the addition of 125µl of Glucose Oxidase reagent and incubated at 37°C for 30 min. Absorbance of the reactions were assessed by a spectrometer at 450nm. A glucose standard curve was completed alongside each assay for quality control and to determine the amount of liberated free glucose for each reaction.

Reactions were completed in a minimum of triplicates and incubated in a range of substrate concentration from 2mM to 200mM. There were limitations with the quantitative availability of some of the substrates, resulting in these assays having a maximum substrate concentration of 30 to 50mM.

To ensure the location of the His-tag was not interfering with the functioning of the enzyme, an assay with maltose was completed with the enzyme BT_0339-15b. Concentration of maltose assessed was also 1-200mM. The pH of the assay was also assessed to determine if this is a significant variable affecting the enzymes functionality. The assay was completed in the same manner as previously described, yet the maltose solution was assessed in a pH range of 5.5 to 9.0 with six replicates for each buffer.

Kinetics of the enzyme BT_0339 was determined using the program KalediaGraph4.1. The program was used to fit the data to the Michaelis-Menten equation (Figure 2.3).

$$v = V_{\max} \frac{[S]}{K_m + [S]}$$

Figure 2.3 Michaelis-Menten Equation[52]

2.6 Additional α-glucosidase, BT_3299

An additional α-glucosidase, BT_3299, was also assessed in the same manner as BT_0339. The construct was generated in a previous experiment and expression and purification was previously optimized[49]. Transformed expression cells containing the construct for BT_3299 in the vector pET-29a were grown overnight. The overnight culture was used to inoculated LB broth and incubated at 37°C until an OD595nm of 0.6 was reached. Expression was induced by the addition of IPTG for a final concentration of 400μM and grown overnight at 30°C. Pelleted cells were re-suspended in lysis buffer and emulsified at 30psi for three rounds.

Purification of BT_3299 was initially completed in the same manner as BT_0339 by affinity chromatography. Elution fractions were pooled, desalted and purified additionally by FPLC on an anion exchange column. Fractions from this stage were pooled and buffer exchanged completed into 20mM Tris, pH 8.0. Solution was concentrated to a final working concentration of >5mg/ml, which was used in crystallization trials and enzyme assays.

Substrate specificity was assessed in the same approach as completed for BT_0339. Assays were completed at a final concentration at 10nM and data was fit to the Michaelis-Menten equation in KalediaGraph.

2.7 Operon Prediction for BT_0339 and BT_3299

The operon that the genes *bt_0339* and *bt_3299* are members of was investigated by use of the MicrobeOnline Database. This database functions by connecting functional genomic data with comparative genomic analysis[53]. The outputted data is based on a method where operons are predicted to be composed of genes based on comparative genomic measures and the distance that separates adjacent genes[54].

2.8 Crystallization Screening for BT_0339 and BT_3299

Once each enzyme was purified to sufficient homogeneity, the conditions under which it can be crystallized were investigated. Sitting Drop Vapor Diffusion method was used in the preliminary crystallization screening. Hampton Index Screen[55] was the initial screen used to narrow the conditions of crystallization. Results from this screen can be used to determine a set of conditions to optimize the crystallization conditions. BT_0339-29a was screened at 5mg/ml, BT_0339-15b was screened at 7mg/ml and BT_3299 was screened at 5mg/ml. Screen consisted of 100µl reservoir solution and a 2µl drop in a ratio of 1:1 protein and reservoir. These plates were incubated at room temperature and were observed daily for crystallization. Conditions that demonstrated conditions that allowed for crystallization were further optimized with custom made screens that include variations in pH, concentration of salt, buffer and precipitant, and protein concentration in drop.

Chapter 3
Results

3.1 Amplification of *bt_0339* gene by Polymerase Chain Reaction

The gene *bt_0339* from *Bacteroides thetaiotaomicron* was successfully amplified by Polymerase Chain Reaction (PCR). The amplified product was run on a 1% agarose gel for confirmation of the size of the amplified fragment (Figure 3.1).

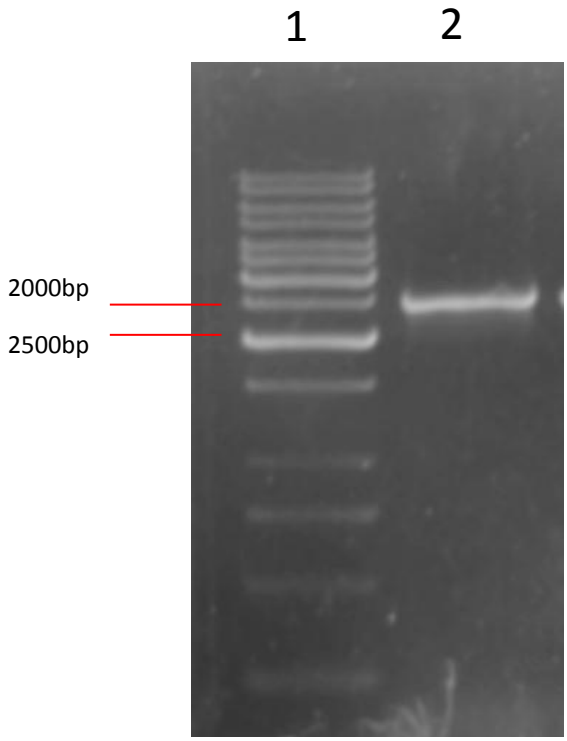


Figure 3.1 Amplification of *bt_0339* by Polymerase Chain Reaction. Lane 1 was the DNA ladder and lane 2 is the amplified product of *bt_0339.*

The expected size of the amplified product was ~2300 base pairs as confirmed by comparison with the DNA ladder. Cloning was completed successfully and constructs for both vectors were confirmed by DNA sequencing (Figure 3.2).

Figure 3.2 Alignment of protein sequence as confirmed by DNA sequence of *bt_0339* constructs. Alignment was completed for both constructs in comparison with the predicted amino acid sequence for BT_0339. Dark coloured region indicates the amino acid deduced from confirmed DNA sequence and the light region is the amino acids that are unknown. Clone Bt-29-2.2 contains the construct in vector pET-29a and clone Bt-15-20.3 contains the construct in vector pET-15b.

A large proportion of the gene was successfully sequenced with approximately 70% coverage of the sequence for the protein of interest. This was adequate amount of information regarding the constructs in order to move on to the next stage of this study.

3.2 Expression of BT_0339 in pET-29a vector

Initially the optimal temperature for expression of BT_0339-29a construct was assessed in a small scale format. Cultures were grown at 37°C until an OD_{595} of 0.4 was achieved and then the culture

was induced with IPTG for a final concentration of 1mM and grown overnight at 37°C, 30°C and 20°C. Soluble and insoluble fractions were compared on an SDS-PAGE (Figure 3.3).

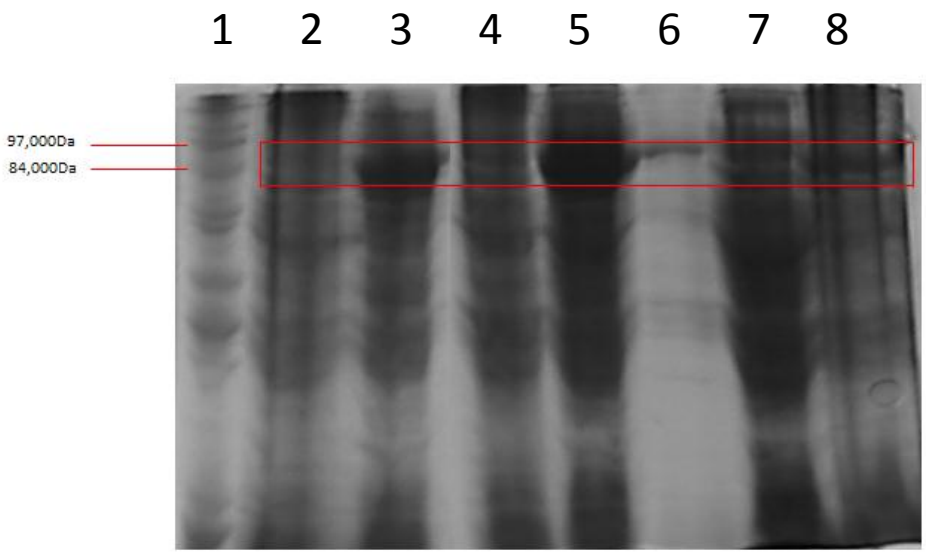


Figure 3.3 O**ptimization of temperature for the expression of BT_0339-29a**. Lane 1 is a protein ladder, lane 2 is the control (uninduced), lanes 3 and 4 are induction at 37°C, supernatant and pellet respectively, Lanes 5 and 6 are induction at 30°C, supernatant and pellet respectively, and lanes 7 and 8 are induction at 20°C, supernatant and pellet respectively.

It was demonstrated that the temperature that provided the highest degree of expression of the protein of interest was 30°C. At this given temperature there is minimal amount of insoluble protein and a high degree was soluble protein.

Additionally, temperature for expression was also optimized for BT_0339-15b construct. Expression was completed overnight at 37°C, 30°C and 20°C. Soluble and insoluble fractions were run on an SDS-PAGE for comparison (Figure 3.4)

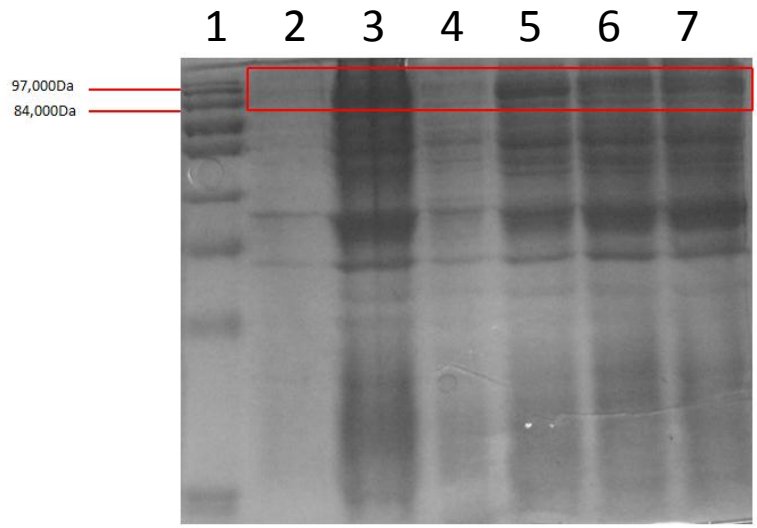


Figure 3.4 Optimization of temperature for the expression of BT_0339-15b. Protein of interest was expressed at 37, 30 and 20°C. Lane 1 is the protein ladder, lane 2 and 3 is 37°C, insoluble and soluble fractions respectively, lane 4 and 5 is 30°C, insoluble and soluble fractions respectively, and lanes 6 and 7 is 20°C, insoluble and soluble fractions.

It was found that 30°C was also the optimal temperature for expression of BT_0339-15b. At this given temperature there was a significantly higher degree of expressed protein which was soluble.

The length of time of expression is a variable that can have a powerful effect on the level of protein expressed and therefore was investigated for BT_0339. BT_0339-29a was expressed at various time intervals and soluble fractions were run on an SDS-PAGE for comparison (Figure 3.5).

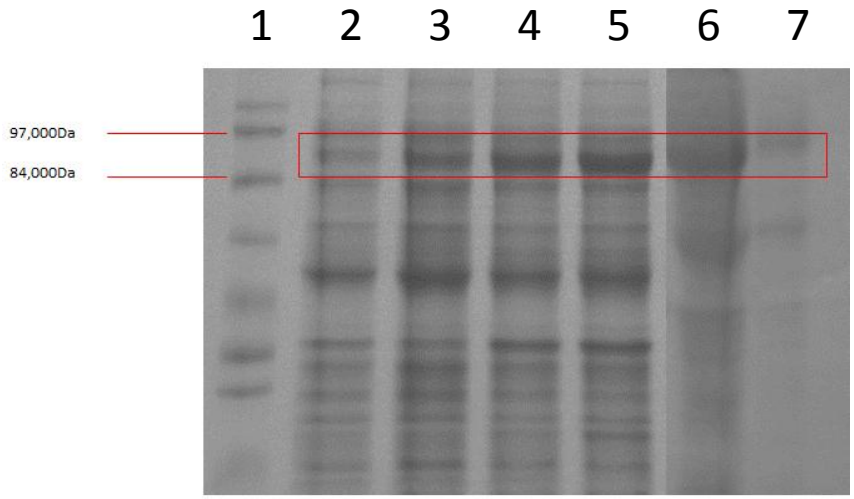


Figure 3.5 Optimization of the length of incubation for the expression of BT_0339-29a. Lane 1 is the protein ladder, lane 2 is 1hr, lane 3 is 2 hr, lane 4 is 3 hr, lane 5 is overnight and lane 6 is 5 hr.

It was revealed that the optimal length of incubation required for a high level of expression was overnight. This amount of time produced a good level of soluble protein in comparison with the other fractions.

The amount of time required for expression was also optimized for BT_0339-15b vector. The construct was expressed at various time intervals and the soluble fractions were run on an SDS-PAGE for comparison (Figure 3.6)

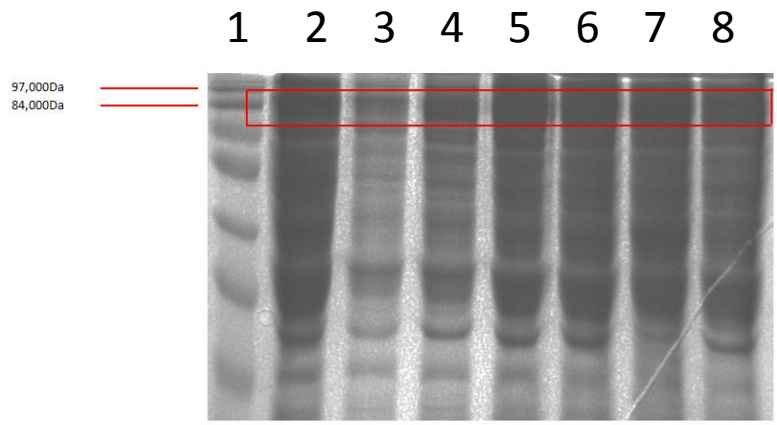


Figure 3.6 Optimization of the length of incubation for the expression of BT_0339-29a. Lane 1 is the protein ladder, lane 2 is a control (no induction), lane 3 is 30 min, lane 4 is 1hr, lane 5 is 2hr, lane 6 is 3hr, lane 7 is 5hr and lane 8 is overnight.

The length of time appeared to not have a significant impact on the level of expression of BT_0339-15b. Therefore, overnight expression was selected as the preferred expression time period due to technical aspects.

The amount of IPTG required for induction was also optimized. Highly levels of IPTG have been reported to be correlated with high levels of bacterial host proteases with ultimately lead to degradation of the protein of interest[56]. Lowering the level of IPTG without having a dramatic effect on the level of protein expressed can be beneficial. BT_0339-29a was grown at 37°C until an OD_{595} of 0.4, followed by induction with various levels of IPTG. These cultures were grown overnight and observed by running the cell lysates on an SDS-PAGE for comparison (Figure 3.7).

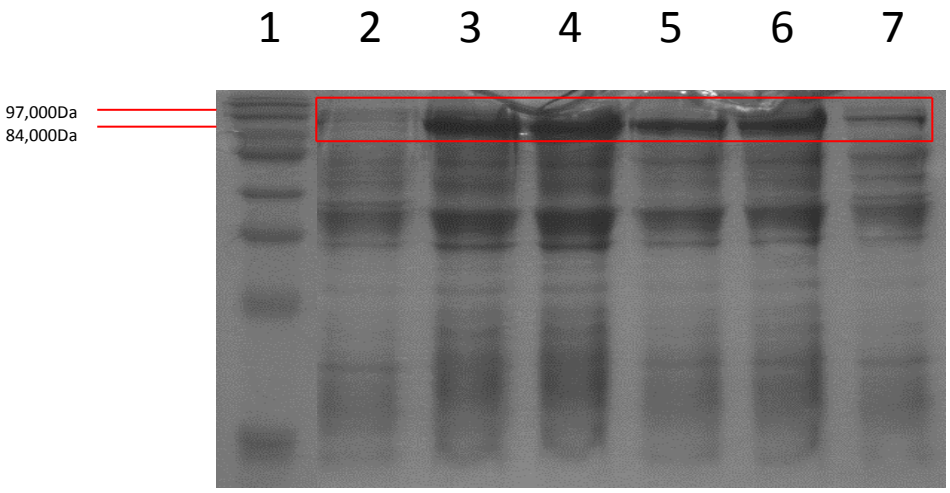


Figure 3.7 Optimization of IPTG for the expression of BT_0339-29a. Lane 1 is the protein ladder, lane 2 is a control (uninduced), lane 3 1mM IPTG, lane 4 is 500µM, lane 5 is 100µM, lane 6 is 50µM and lane 7 is 10µM.

It was observed that a final concentration of 100µM IPTG in the media had little effect on the level of expressed protein. At this concentration of IPTG, there were still significant levels of protein expressed in comparison to the control.

The amount of IPTG required to induce the expression of the desired protein was also tested with the construct in vector pET-15b. Analysis was completed by running cell lysates on an SDS-PAGE for comparison (Figure 3.8).

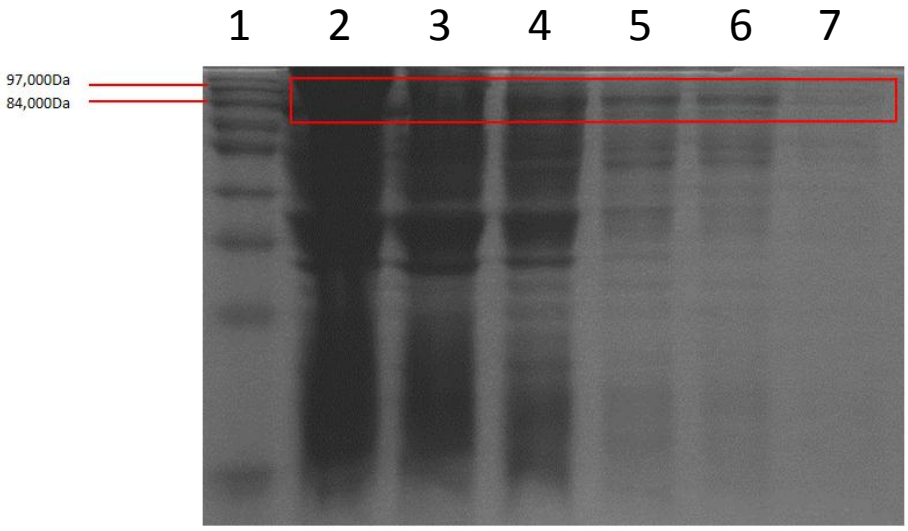


Figure 3.8 Optimization of ITPG for the expression of BT_0339-15b. Lane 1 is the protein ladder, lane 2 is 1mM, lane 3 is 500µM, lane 4 is 100µM, lane 5 is 50µM and lane 6 is 10µM of IPTG.

A final concentration of 100µMof IPTG provided was shown to induce a reasonable level of expression. In comparison to the additional samples, this level of IPTG allows for limited degradation an sufficient soluble product.

3.3 Affinity Purification of BT_0339

BT_0339 was purified by affinity chromatography by use of Ni-NTA resin. This resin has a strong affinity for the His-tag that resides on either end of BT_0339 constructs. Purification of BT_0339-29a was completed in a packed column at ambient pressure with elution by an imidazole gradient. Fractions were retained and collectively ran on an SDS-PAGE for analysis (Figure 3.9)

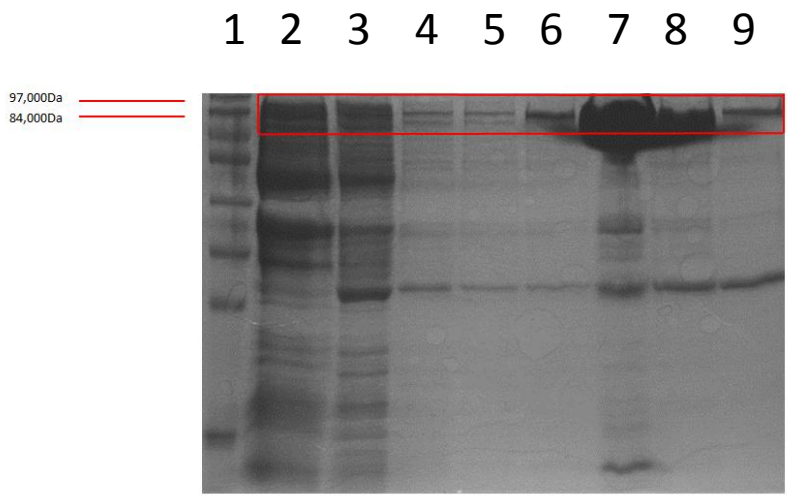


Figure 3.9 Affinity purification of BT_0339-29a with Ni-NTA resin. Lane 1 is a protein ladder, lane 2 is the flow through, lane 3 is wash 1, lane 4 is wash 2, lane 5 is 25mM imidazole, lane 6 is 50mM imidazole, lane 7 is 100mM imidazole, lane 8 is 200mM imidazole and lane 9 is 250mM imidazole elutions.

Majority of BT_0339-29a came off at the elution containing 100mM of imidazole. There was a degree of BT_0339 that came off in the additional elutions; however, these fractions may consist of BT_0339 in slightly different conformations, which is undesirable for crystallization. It can also be observed that there was some non-specific proteins that bound to the column, which were also eluted at the 100mM elution and as a result will need to be further purified to remove these containments.

Affinity purification was also completed with BT_0339-15b construct with the same method (Figure 3.10).

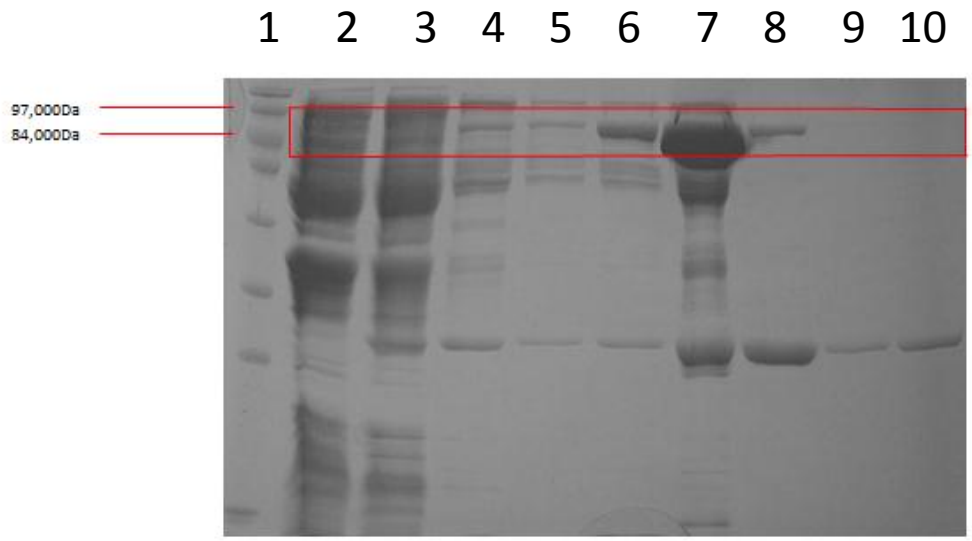


Figure 3.10 Affinity purification of BT_0339-15b with Ni-NTA resin. Lane 1 is the protein ladder, lane 2 is the flow through, lane 3 is wash1, lane 4 is wash 2, lane 5 is 25mM imidazole, lane 6 is 50mM imidazole, lane 7 is 100mM imidazole, lane 8 is 200mM imidazole and lane 9 is 250mM imidazole elutions.

It was observed that the elution of 100mM imidazole demonstrated the highest amount of desired protein coming off the column. Similarly, there were some undesirable products eluted off the column in the same fractions as BT_0339-15b, which revealed this elution required further purification prior to crystallization trials.

3.4 FPLC Purification with Anion Exchange Column

A buffer exchange was completed in order to remove the imidazole from the pooled elution and was followed by an additional purification step. An anion exchange column was utilized and protein elution was detected at an absorbance of 280nm (Figure 3.11).

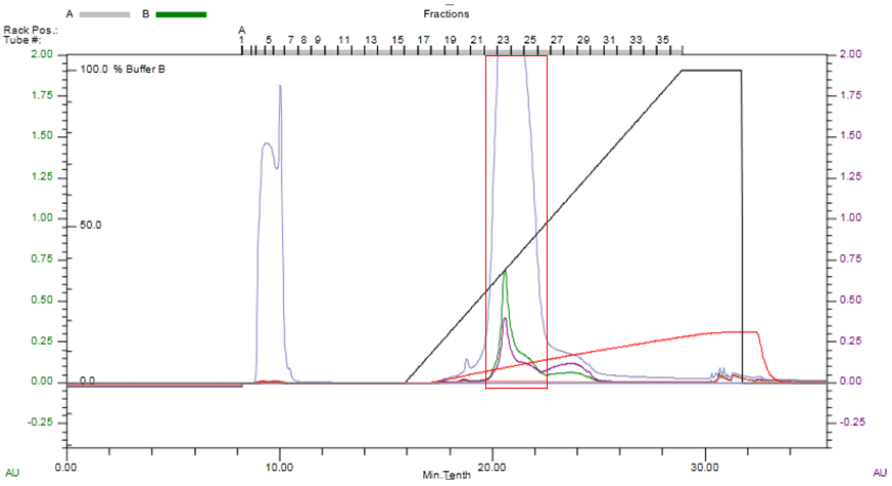


Figure 3.11 FPLC Purification of BT_0339-29a by anion exchange. Outlined region indicates the fractions that were pooled.

It was observed that there was a desirable amount of protein that was eluted off the anion column. Fractions 21 to 26 were pooled and are assumed to be the protein of interest. The pooled product and additional fractions with observed levels of protein were run on an SDS-PAGE for confirmation (Figure 3.12).

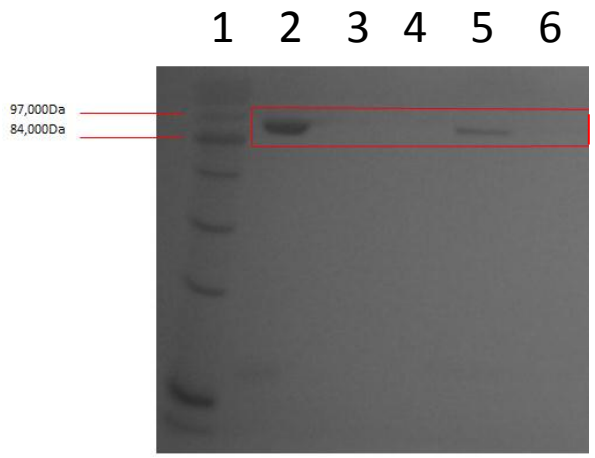


Figure 3.12 SDS-PAGE analysis of FPLC purification fractions for BT_0339-29a. Lane 1 is a protein ladder, lane 2 is fraction 21-26, lane 3 is fraction 27, lane 4 is fraction 28, lane 5 is the pooled fractions 29 and lane 6 is fraction 20.

SDS-PAGE analysis of the pooled fractions from the FPLC purification revealed the pooled sample did contain the protein of interest and an improvement in the sample homogeneity. There was a faint band below the protein of interest; however, it still appears that this protein is worth pursing for crystallization trials.

BT_0339-15b was also purified in an additional step by FPLC by use of an anion exchange column. This step of purification was completed in the same manner as for BT_0339-29a (Figure 3.13).

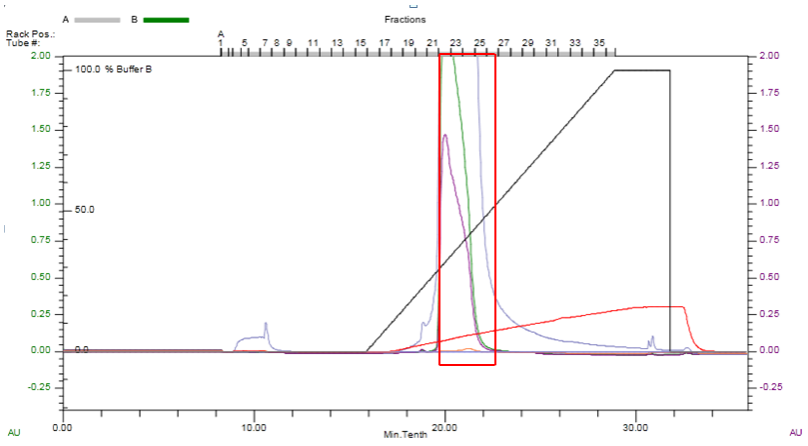


Figure 3.13 FPLC Purification of BT_0339-15b by anion exchange. Fractions that contained protein and that were subsequently pooled are highlighted.

A significant quantity of BT_0339-15b came off the anion exchange column in the FPLC purification. Fractions 21 to 26 were pooled and run on an SDS-PAGE for confirmation of purification, along with additional fractions with detected protein (Figure 3.14).

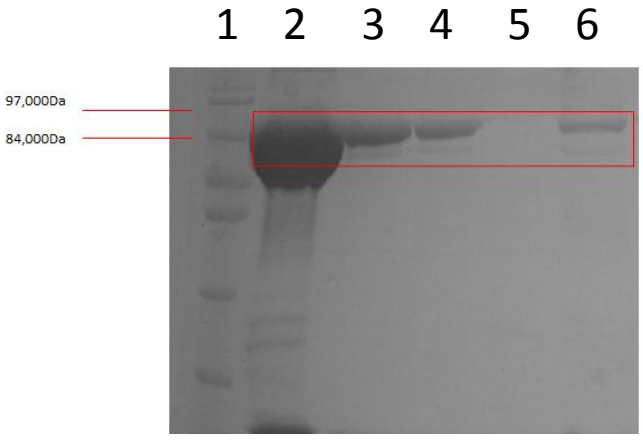


Figure 3.14 SDS-PAGE analysis of FPLC purification fractions for BT_0339-15b. Lane 1 is the protein ladder, lane 2 is pooled fractions 21 to 26, lane 3 is fraction 20, lane 4 is fraction 27 and lane 6 is fraction 28.

The pooled fractions on an SDS-PAGE demonstrated an improvement in purification, despite the observed faint low molecular weight containments. This protein was cleaned up sufficiently allowing the experiment to proceed with crystallization trials.

3.5 Mass Spectrometry Analysis of Purified BT_0339

Analysis of BT_0339 was completed prior to initiation of enzyme assays and crystallization trials in order to ensure the identity of the protein of interest is that of the prediction. The expected size of the purified protein was approximated to be 87968Da and mass spectrometry results confirmed the weight was 87970Da. Mass spectrometry generally can provide an error rate of ~1-3 Daltons[57], demonstrating that the purified protein was the predicted α-glucosidase, BT_0339 (Figure 3.15).

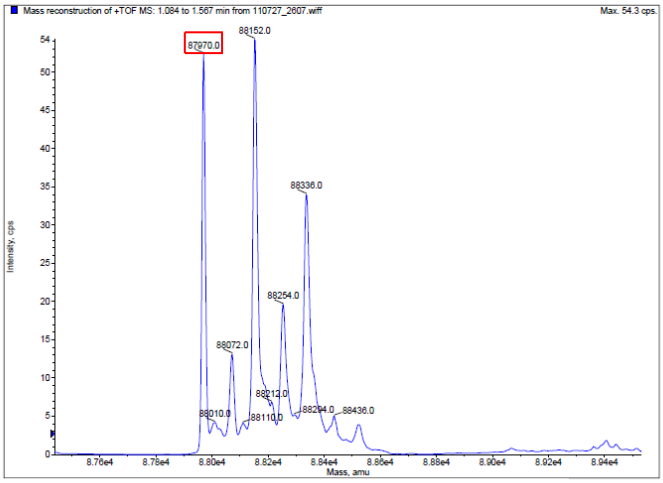


Figure 3.15 Mass Spectrometry analysis for BT_0339-29a. Highlighted mass was the observed peak for BT_0339.

3.6 Enzyme Kinetics of BT_0339

Glucose-oxidase assay was initially performed on BT_0339-29a with maltose as a substrate. BT_0339 was observed to successfully hydrolyze the α-(1,4) linkages of maltose and liberate free glucose (Figure 3.16). However, it was also observed that the rate of which this enzyme was able to hydrolyze maltose was quite slow. Due to limitations with the substrate solubility, Vmax could not be determined and therefore, the Km could not be reported by the Michaelis-Menten equation.

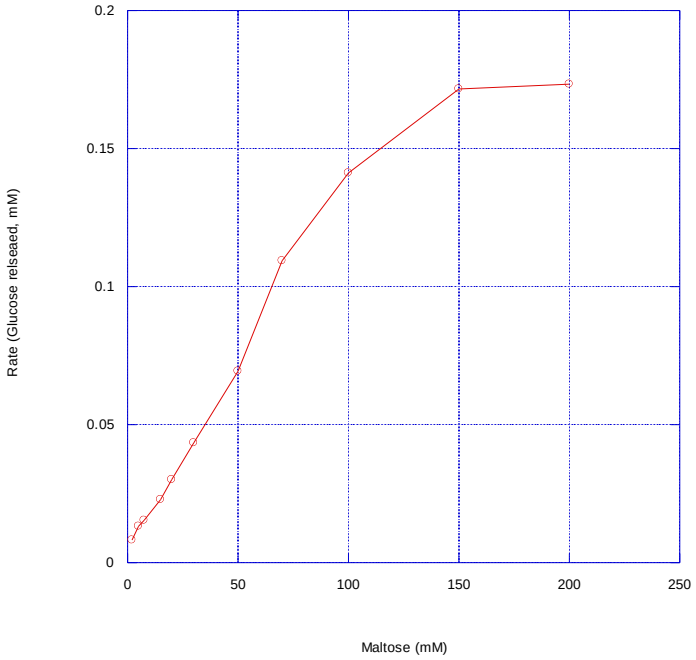


Figure 3.16 Hydrolytic activity of BT_0339 with maltose as a substrate. Assay was completed in a range of substrate concentrations of 2 to 200mM.

Hydrolytic ability on additional substrates was also investigated in order to determine the specifics of BT_0339's substrate specificity (Figure 3.17). It appears that this enzyme prefers isomaltose as a

substrate over maltose, demonstrating that is preferentially cleaves α-(1,6) over α-(1,4) linkage. Additionally, it was observed that BT_0339 prefers a substrate that is longer in length when comparing the levels of released glucose for the maltose derivative substrates. Also, BT_0339 was unable to utilize sucrose or lactose as a substrate. These results also indicate that BT_0339 does not preferentially cleave maltohexaose, demonstrating a level of cleavage that is slightly above the background level. The observed low activity of this enzyme resulted in a maximum substrate concentration of 30-50mM being assessed, due to restrictions with substrate availability. This limitation allowed for the calculation of the linear portion of the kinetic curve but the inability to report the Vmax and Km (refer to Supplementary Data).

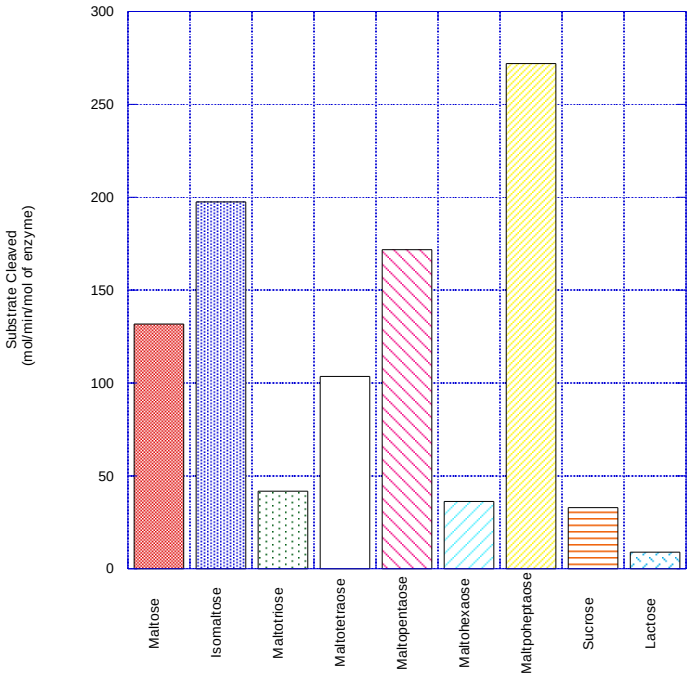


Figure 3.17 Glucose-oxidase assays for BT_0339 with isomaltose, maltooligosaccharides, lactose and sucrose. Values derived from kinetic plots for each individual substrate at a concentration of 30mM.

As mentioned, two constructs of BT_0339 were created in vectors pET-29a and 15b. Both of the enzymes expressed by these constructs were assessed for enzymatic ability on maltose as a substrate, with a range of concentrations, 2-200mM (Figure 3.18). It was found that the selection of vector had no significant impact on the functionality of the enzyme BT_0339 with observed an similar curves for both constructs. Deviations between the vectors in relation to the catalytic efficiency of these enzymes was comparable and the assessment of 30mM of maltose revealed that the catalytic efficiency for BT_0339-29a was 1.4uM and for BT_0339-15b was 1.1uM.

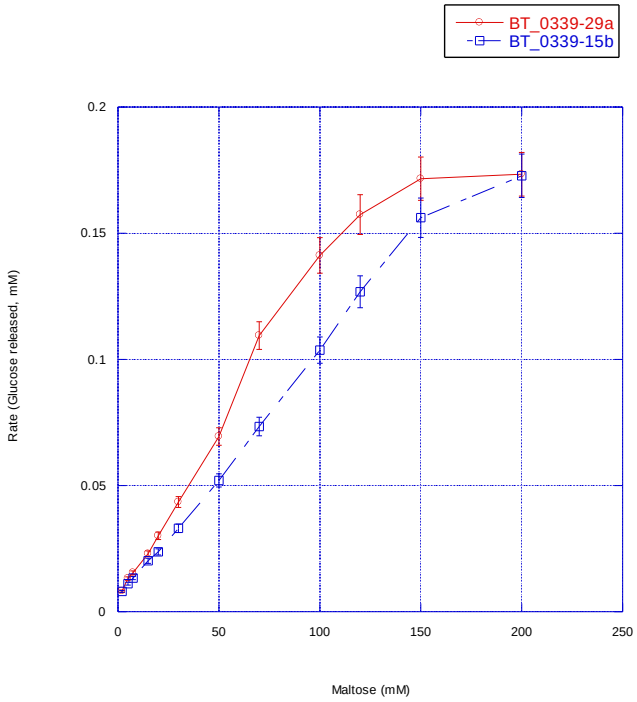
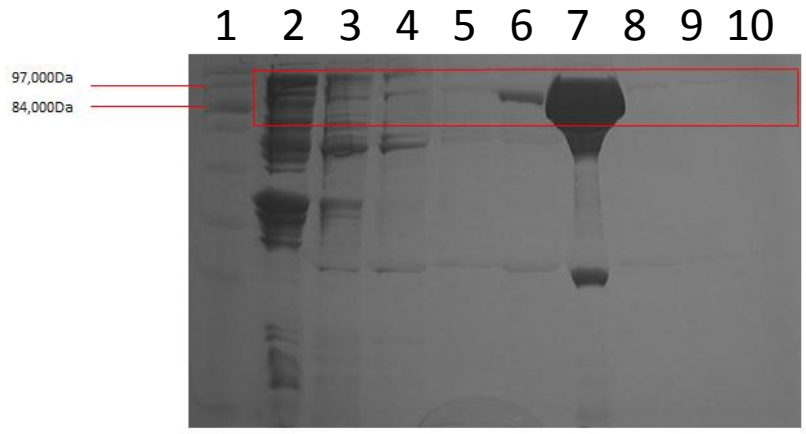


Figure 3.18 Analysis of vector selection for BT_0339 with maltose as a substrate. Comparison of vector selection was completed with vector pET-29a is shown in the above figure by the circled curve and vector pET-15b is indicated by the square curve.

Additionally, the pH of the assay for BT_0339 was assessed and demonstrated no significant difference between the range of 5.5 and 9.0.

3.7 Purification of BT_3299

The additional α-glucosidase, BT_3299, was purified by affinity chromatography, using NTA-Ni resin. Fractions were collected and run on an SDS-PAGE for comparison (Figure 3.19). BT_3299 was successfully eluted off at 100mM imidazole elution and fractions were pooled, yielding a significant amount of protein.



.

Figure 3.19 Affinity Purification of BT_3299 with Ni-NTA resin. Lane 1 is the protein ladder, lane 2 is the flow through, lane 3 and 4 is wash 1 and 2 respectively, lane 5 is 25mM elution, lane 6 is 50mM elution, lane 7 is 100mM elution, lane 8 is 200mM elution, lane 9 is 250mM elution and lane 10 is 500mM elution of imidazole.

The pooled sample for BT_3299 was further purified by FPLC with an anion exchange column. Fractions coming off the column were detected at 280nm absorbance (Figure 3.20).

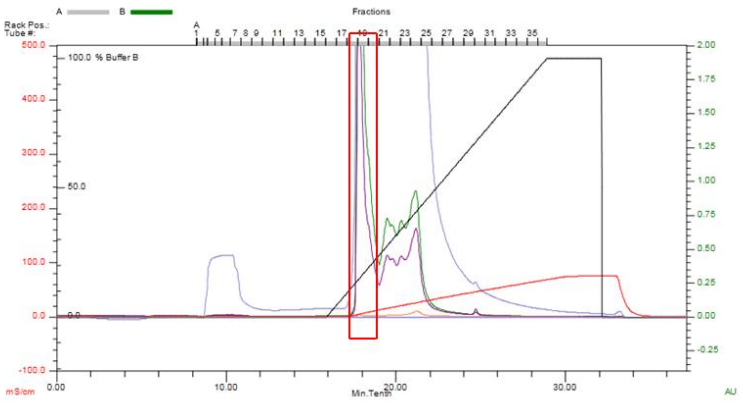


Figure 3.20 FPLC Purification of BT_3299 by anion exchange. Exchange was completed with an elution gradient of 0-100% 1M NaCl. Indicated region were the elution fractions that were pooled.

It was demonstrated that an initial slender peak came off the anion exchange column, which was presumed to be the protein of interest. Fractions 17 to 20 were pooled together and were run on an SDS-PAGE, along with the other fractions for confirmation purposes (Figure 3.21).

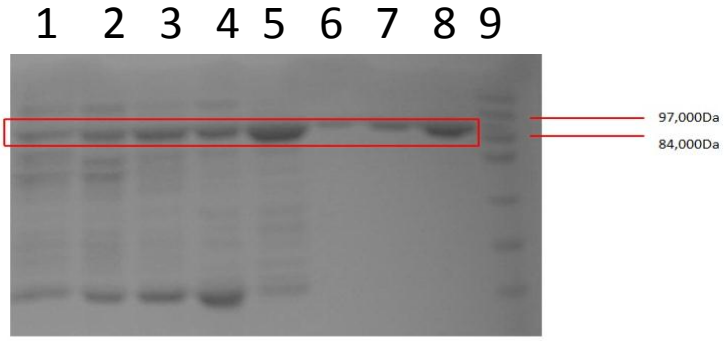


Figure 3.21 SDS-PAGE analysis of FPLC Purification fractions for BT_3299. Lane 1 is fraction 21, lane 2 is fraction 22, lane 3 is fraction 23, lane 4 is fraction 24, lane 5 is fraction 25, lanes 6 to 8 is dilution series of the pooled fractions 17-20 and lane 9 is a protein ladder.

Pooled fractions 17-20 revealed a significantly purified sample, with reduced level of contaminates. This pooled sample was of enough homogeneity to initiate crystallization trials. Previous studies of BT_3299 completed a Mass Spectrometry analysis on the purified enzyme and reported the purified product was the predicted protein, BT_3299[49].

3.8 Enzyme Kinetics of BT_3299

Substrate specificity of BT_3299 was assessed in the same manner as BT_0339. The initial substrate assessed was maltose in a concentration range of 2-200mM. Results were fit to a Michaelis-Menten curve in attempts to define the Vmax and Km (Figure 3.22).

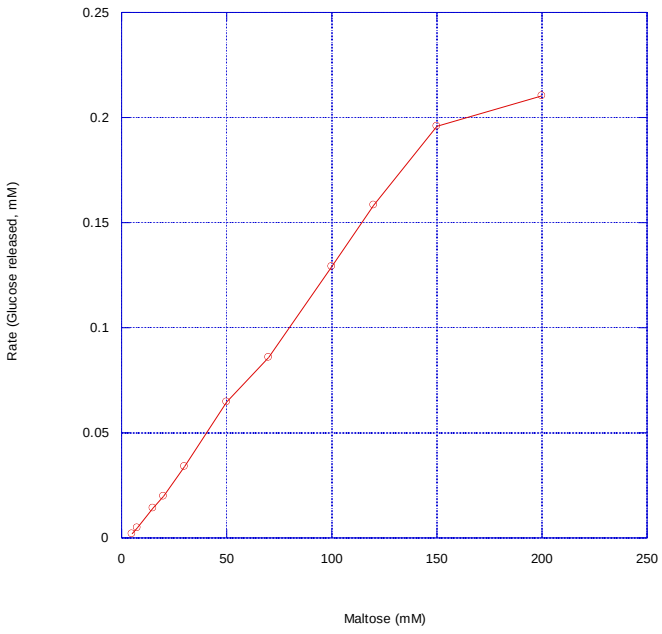


Figure 3.22 Hydrolytic activity of BT_3299 with maltose as a substrate. Assay was completed in a range of substrate concentrations of 2 to 200mM.

It was revealed that there was a similar mechanism for BT_3299 as observed for BT-0339 for maltose as a substrate. BT_3299 was successfully able to cleave the α-(1,4) bond between the residue, releasing free glucose. Similarly, BT_3299 hydrolytic capabilities on maltose was at a slow rate and due to the limitations of substrate solubility, it was not possible to determine an accurate Vmax or Km. Substrate specificity was further investigated with the assessment of isomaltose, maltose derivatives, lactose and sucrose as substrates (Figure 3.23).

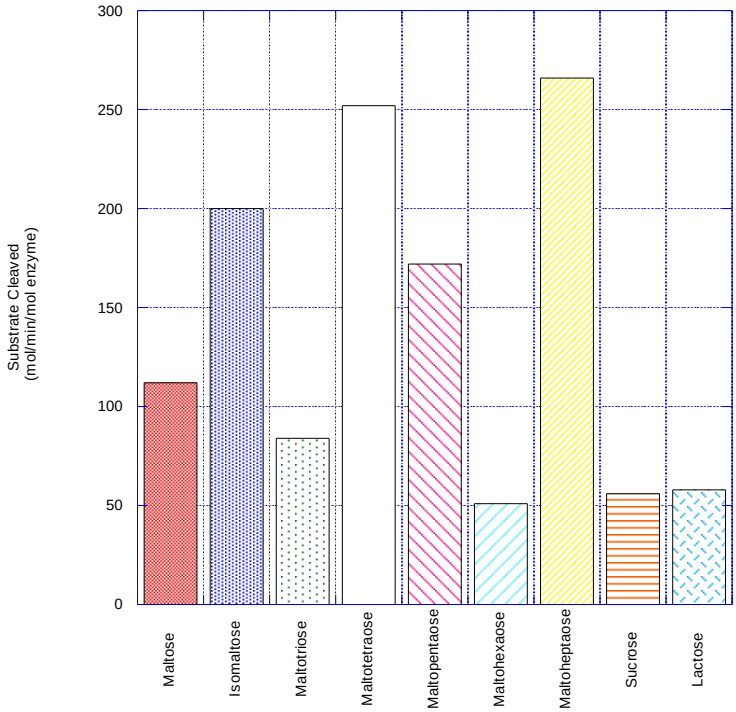


Figure 3.23 Glucose-oxidase assay for BT_3299 with isomaltose, maltooligosaccharides, lactose and sucrose. Values were obtained from kinetic plots for each substrate and plotted at a concentration of 30mM.

It was observed that the specificity of BT_3299 is similar to that of BT_0339, preferring isomaltose over that of maltose, revealing a specificity for the α-(1,6) linkage. Additionally, BT_3299 had an

observed higher activity on maltotetraose and maltoheptaose. BT_3299 was unable to utilize sucrose or lactose as substrate and release free glucose revealing its lack of specificity for these substrates. It was also shown that BT_3299 does not cleave glucose residue from maltohexaose polymer, demonstrating a similar specify observed for BT_0339.

The pH of the assay was previously assessed for BT_3299 and was found to demonstrate its optimal activity at a pH of 6.5[49].

3.9 Operon Prediction for BT_0339

Operon prediction was completed for the gene *bt_0339* using the program MicrobeOnline. The gene *bt_0339* was predicted to be part of an operon consisting of multiple genes (Figure 3.24).

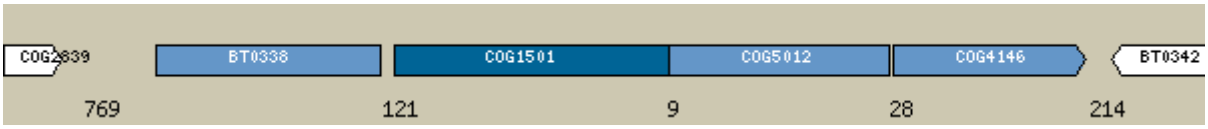


Figure 3.24 Predicted operon for the gene *bt_0339*. The query gene is indicated by the darker blue colour, nearby genes that are predicted to be part of the same operon are in light blue and genes upstream/downstream of operon are in white. The arrowhead indicates the direction of transcription and the numbers are the intragenic distance between the two adjacent genes[53].

A number of genes were predicted to be part of the operon of that *bt_0339* is predicted to be a member (Table 3.1). The predicted operon for BT_0339 consists of pectin lyases fold/virulence factor, a trimethylamine coronoid protein 2 and a Na+/glucose co-transporter.

Table 3.1 Genes predicted to be part of the operon for *bt_0339*.

Gene	Protein Name
bt_0338	pectin lyase fold/virulence factor
bt_0339	alpha-glucosidase
bt_0340	trimethylamine coronoid protein 2
bt_0341	Na+/glucose co-transporter

3.10 Operon Prediction for BT_3299

The operon of *bt_3299* was also predicted using the program MicrobeOnline[53] (Figure 3.25).

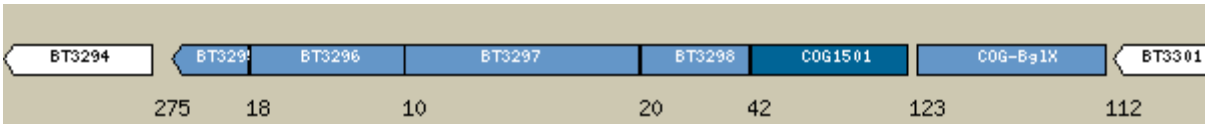


Figure 3.25 Predicted operon for the gene *bt_3299*. The query gene is indicated by the arker blue colour, nearby genes that are predicted to be part of the same operon are in light blue and genes upstream/downstream of operon are in white. The arrowhead indicates the direction of transcription and the numbers are the intragenic distance between the two adjacent genes[53].

There were also a number of genes predicted to be part of the operon of *bt_3299* in comparison with that of *bt_0339* (Table 3.2). A few of the genes were predicted to be glycoside hydrolases that

would be assumed to have their function independent of that of *bt_3299*. Yet, there are homologs of Sus proteins that have been shown to be related to the function of utilizing polymers of starch.

Table 3.2 Genes predicted to be part of the operon for *bt_3299*.

Gene	Protein
bt_3295	hypothetical protein
bt_3296	Outer membrane protein: SusD homolog
bt_3297	hypothetical protein: SusC homolog
bt_3298	hypothetical protein with a conserved domain-cell surface receptor
bt_3299	α-glucosidase II
bt_3300	β-glucosidase

3.11Crystallization Trials for BT_0339

Crystallization trials were completed on both constructs of BT_0339. Hampton Research Index[55] screen was the first screen used for both of these enzymes. The Index Screen for BT_0339-29a was observed to produce a few positive hits that were further investigated (Table 3.3).

Table 3.3 Initial crystallization hits for BT_0339-29a. Screen was completed with the Hampton Research Index Screen.

Reagent #	Salt	Buffer	Precipitant
72	0.2M Sodium chloride	0.1M Hepes, pH 7.5	25% Polyethylene glycol 3,350
80	0.2M Ammonium acetate	0.1M Hepes, pH 7.5	25% Polyethylene glycol 3,350

The Index Screen for BT_0339-15b was also observed to produce a number of positive hits that were further investigated for optimization (Table 3.4).

Table 3.4 Initial crystallization hits for BT_0339-15b. Screen was completed with the Hampton Research Index Screen.

Reagent #	Salt	Buffer	Precipitant
72	0.2M Sodium chloride	0.1M Hepes, pH 7.5	25% Polyethylene glycol 3,350
79	0.2M Ammonium acetate	0.1M Bis-Tris, pH 6.5	25% Polyethylene glycol 3,350
81	0.2M Ammonium acetate	0.1M Tris, pH 8.5	25% Polyethylene glycol 3,350

Custom optimization screens were set up for the positive hits observed in the Index screen. These optimized conditions varied in in pH, concentration of salt, precipitant and buffer components and concentration as related to the initial screen. These conditions resulted in an improved number of

conditions with crystal growth and the type of crystals that formed. However, it appeared that majority of the conditions resulted in poor lattice formation resulting in twinned crystals (Figure 3.26 and 3.27)

Figure 3.26 Crystallization of BT_0339-29a in an optimized condition.

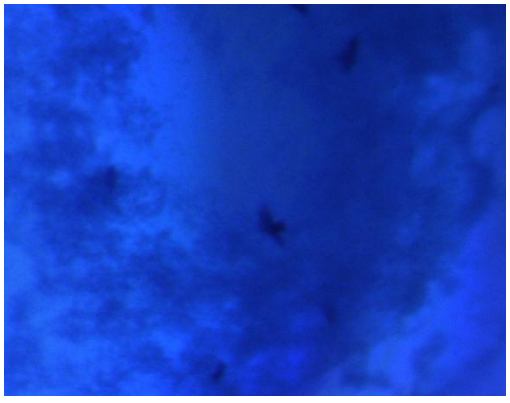


Figure 3.27 Crystallization of BT_0339-15b in an optimized condition.

Further optimized conditions consisted of the addition of additives to both BT_0339 constructs. These small molecules were added to some of the previously optimized conditions that were observed to have reasonably sized crystals. BT_0339-15b did not lead to much better conditions

with the addition of the additives, whereas BT_0339-29a resulted in a number of conditions yielding much better quality crystals that were attractive to screen for diffraction (Figure 3.28). These conditions consisted of 0.2 M ammonium acetate, 0.2M Hepes, pH 7.5 and 15% PEG 3350 (GH Screen-19) with the separate addition of the additives 0.1M calcium acetate hydrate, 0.1 M calcium chloride dehydrate and 0.1 M urea.

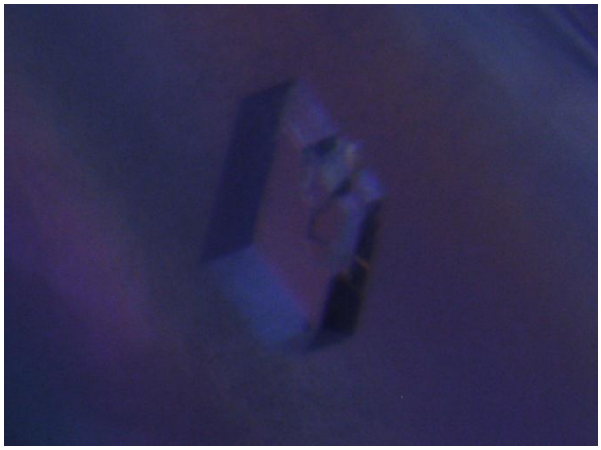


Figure 3.28 Crystallization of BT_0339-29a in GH Screen-19 with urea as an additive.

These conditions for BT_0339-29a were screened for quality of diffraction. All the crystals screened produced some degree of diffraction; however, they produced only low resolution reflections (Figure 3.29).

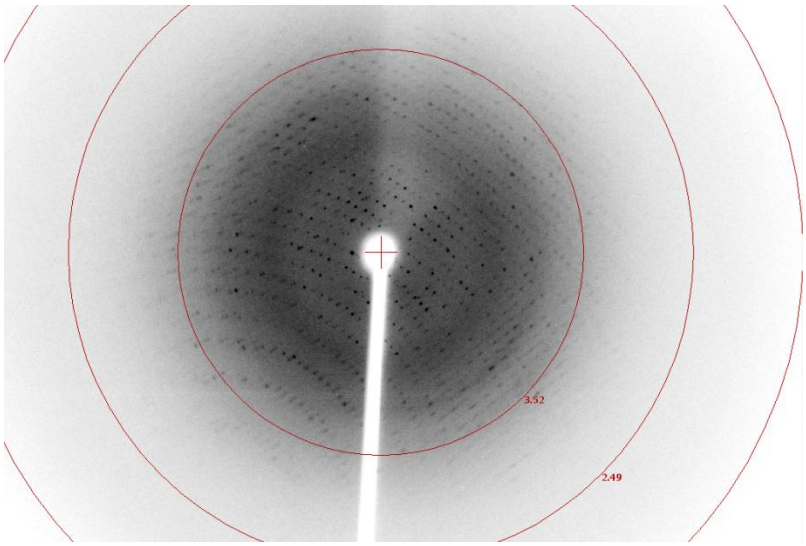


Figure 3.29 Diffraction image for BT_0339-29a. Crystal was grown in the optimized condition of GH Screen-19 with urea as an additive.

3.12 Crystallization trials of BT_3299

Crystallization of BT_3299 was also completed initially with the Hampton Research Index screen and a number of conditions produced successful conditions worth optimizing (Table 3.5).

Table 3.5 Initial crystallization hits for BT_3299. Screen was completed with the Hampton Research Index Screen

Reagent #	Salt	Buffer	Precipitant
61	0.2M L-Proline	0.1M Hepes, pH 7.5	10% Polyethylene glycol 3,350
68	0.2M Ammonium sulfate	0.1M Hepes, pH 7.5	25% Polyethylene glycol 3,350
69	0.2M Ammonium sulfate	0.1M Tris, pH 8.5	25% Polyethylene glycol 3,350

These conditions were then further optimized by custom optimization screens. These screens consisted of optimizing pH, concentration of salt, precipitant and buffer, and buffer composition in relation to the initial condition. The index screen produced crystals that were poor quality, mainly consisting of plate clusters. The optimized conditions allowed for the growth of better quality crystals that were diamond shape, but majority of these had some degree of twinning characteristic (Figure 3.30).

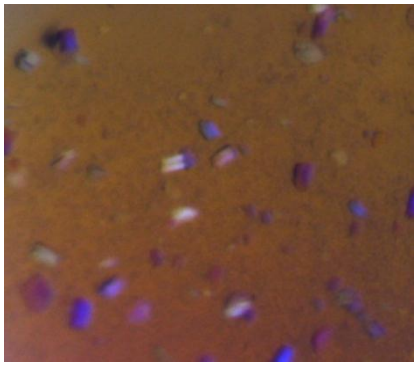


Figure 3.30 Crystallization of BT_3299 enzyme in an optimized condition.

Addition of additives to these conditions was then pursued, resulting in less twinned crystals. A condition consisting of 0.2 M sodium chloride, 0.2 M Tris-HCl, pH 7.5, 15% PEG 3350 (GH Screen-31) and the addition of 0.1 M calcium acetate hydrate was screened for diffraction.

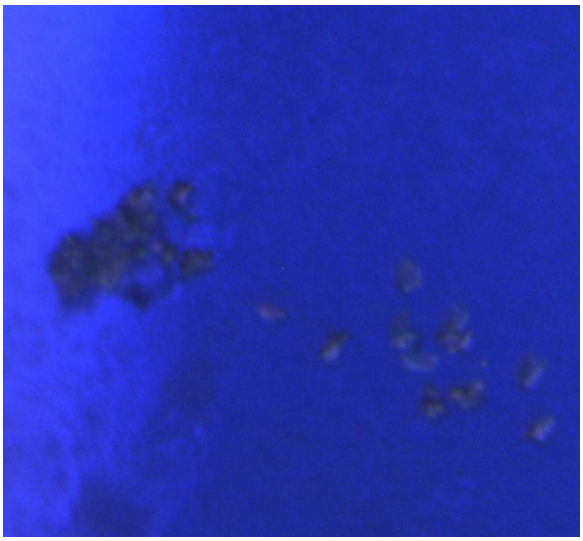


Figure 3.31 Crystallization of Bt_3299 in GH Screen-31 with the additive 0.1 M calcium acetate hydrate.

This condition resulted in successful protein diffraction but the pattern was fairly high in mosaicity and poor resolution (Figure 3.32).

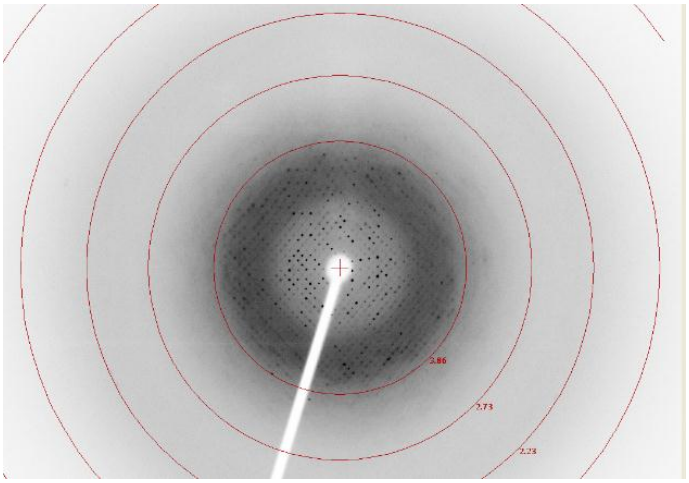


Figure 3.32 Diffraction image for BT_3299. Crystal was grown in the optimized condition of GH Screen-31 with calcium acetate hydrate as an additive.

Chapter 4

Discussion

4.1 Results and their Implications

Recent research has illustrated the significant impact microorganisms can have on humans and their ability to influence our wellbeing. These microbial cells populate the human body in numbers that exceed our own cells and have been shown to affect various human biological processes that range from nutritional, developmental and immunological. Microorganisms that reside in the gastrointestinal tract thrive in a specific niche that has direct interaction with the human host and are present in significantly high numbers. Specifically, the colon demonstrates an environment that is highly favourable for the establishment of these indigenous organisms, where they have been observed to contribute to nutrition by utilizing indigestible dietary plant polysaccharides[35]

As highlighted, a particular member of the colon microbiota that is of significant interest in relation to human digestion is *Bacteroides thetaiotaomicron.* This anaerobic bacterium is one of the dominant members of the colon region and has been shown to potentially perform biological activities that can contribute to the human host's nutrition. Predicted proteins from the sequenced genome revealed this organism is highly capable of utilizing various forms of polysaccharides and possesses a noteworthy number of glycoside hydrolases. Investigation of the function and structure of these hydrolases can reveal specific processes that this organism is capable of completing that can influence the physiology of the host.

This report investigated the functionality of two α-glucosidases from the Family 31 Glycoside Hydrolase from the bacterium *Bacteroides thetaiotaomicron*. These two enzymes were predicted to cleave glucose residues from the terminal end of starch polymers. Kinetic studies of the α-glucosidase, BT_0339, revealed it has a substrate specificity that is interestingly varied from the predicted function. It was observed that BT_0339 preferred isomaltose over maltose as a substrate, revealing this enzyme's preference to cleave an α-(1,6) linkage over an α-(1,4) linkage. When

assessing maltooligosaccharides it was observed that BT_0339 also prefers substrates that are longer in length. There was an observed correlation between the increase in moles of glucose provided in the substrate and the level of glucose liberated. It can be predicted that a longer substrate is better accommodated in the enzyme, resulting in a more effective mechanism of cleavage. However, one inconsistency in the data was observed with maltohexaose as a substrate, where BT_0339 was unable to cleave glucose at a level that is proportionate to its length. This result may reflect a possible secondary structure that maltohexaose is able to form in solution, which may slow down the hydrolytic activity of BT_0339 on this substrate. BT_0339 is also unable to utilize sucrose or lactose as a substrate, further demonstrating that this enzyme favours an α-(1,4) linkage with the polymer composed of repetitive units of glucose. Kinetic properties of BT_0339 showed low activity against the provided substrates resulting in the inability to accurately calculate the Vmax or Km for each substrate. Despite the limited catalytic activity of this enzyme it was demonstrated to hydrolyze various starch derived substrates.

Kinetic parameters of BT_3299 were investigated using the glucose-oxidase assay and it was observed to prefer isomaltose over maltose as a substrate, revealing a similar preference for an α-(1,6) linkage over an α-(1,4) linkage. Assessment of the hydrolytic activity of BT_3299 on various maltose derived substrates revealed this enzyme prefers maltoheptaose followed by maltotetraose as a substrate. These kinetic observations suggest that these specific substrates form a secondary structure or complex in solution that is optimal for BT_3299 to hydrolyze. Similarly, it was observed that BT_3299 hydrolyzed maltohexaose at an unexpectedly slow rate, which can also reflect the possible limitation this enzyme has at accessing this substrate in solution. Kinetic data for the Km and Vmax for the provided substrates could not be reported for BT_3299, as a result of the limited activity and the restricted availability of some of the maltose derived substrates. However,

the hydrolytic ability of this enzyme in the performed assays reveals its ability to utilize various starch derived substrates.

The kinetic patterns of the hydrolytic activity of both α-glucosidases as observed in these assays are quite comparable, yet are they a true representation of their activity in the environment of the human colon? A degree of understanding can come from comparing the kinetic properties of these enzymes with those obtained in a similar study involving a Family 31 α-glucosidase from *Ruminococcus obeum*, a common member of the gut microbiota. Substrate kinetics for this α-glucosidase (RUMOBE_03919) revealed a similar low level of hydrolytic activity on the provided starch derived substrates. Kinetic analysis of RUMOBE_03919 was assessed using the glucose-oxidase assay and it was reported that this enzyme had a substrate preference for isomaltose over maltose, revealing specificity for α-(1,6) over α-(1,4) glycosidic linkage. Comparison of various substrates including maltose derivatives resulted in a distinct profile of this enzyme's substrate preference. Kinetic data revealed the highest level of hydrolytic activity was on isomaltose at 40mols of glucose cleaved per min per mol of enzyme in the reaction. Maltose was observed to have 25mols of glucose cleaved per min per mol of enzyme in the reaction. In comparison, BT_0339 yielded the following values: for maltoheptaose at 270, isomaltose at 200 and maltose at 120mols of glucose cleaved per min per mol of enzyme in the reaction. BT_3299 was observed to have maltoheptaose at 270, maltotetraose at 250 and isomaltose at 200mols of glucose cleaved per min per mol of enzyme. In general, the catalytic efficiency of BT_0339 and BT_3299 is comparable to that of the *Ruminococcus obeum* enzyme. This comparison highlights that perhaps this level of cleavage by these microbial Family 31 Glycoside Hydrolases is a true representation of hydrolysis of starch substrates in the colon. Interestingly, the kinetic patterns that were observed suggest that though these enzymes have a similarity in how they function, it is evident that they have specific characteristics that define their activity. This specificity can be reflected in the amino acid

composition of these enzymes and the slight differences in their predicted signature catalytic sequences. Furthermore, it can be assumed that there is slight amino acid variation in the active site of these enzymes, which allows for the slight variations in the observed substrate specificity amongst these Family 31 enzymes.

When assessing the functionality of these enzymes, it is important to also consider the collective effect they may have on the host in addition to assessing them on an individual basis. The implications of the identified starch utilization 'like' systems in *Bacteroides thetaiotaomicron* demonstrates that though these enzymatic systems may function independently, they may also be simultaneously regulated. The identification of over 80 Polysaccharide Utilization Loci also provides insight into that a combination of these systems can lead to a significant multi-enzymatic system that is capable of degrading starch. If the products of these loci are a summed process, then it can be assumed that ultimately they can have a significant effect on the human host. In the process of liberating free glucose for its own energy requirements, *Bacteroides thetaiotaomicron* indirectly releases this simple sugar to the surrounding environment and undoubtedly can be taken up by surrounding bacteria in the lumen, as well as by the host intestinal epithelia. Accepting the connection between BT_0339 and BT_3299 with the human host suggests that though these enzymes have an observed low level of activity, they are also capable of exerting an influence on the host's metabolism and energy levels. Furthermore, the function of these enzymes can potentially influence the ability of other residing microbes to flourish in the colon, which likewise, affects the host.

Assessing BT_0339 and BT_3299 and its constitutive operons revealed interesting information and additional proteins that are relevant to the ability of these enzymes to utilize starch derived substrates. BT_0339 was found to have multiple proteins in its operon, including a pectin lyase fold/virulence factor, a trimethylamine corrinoid protein 2 and a Na+/glucose co-transporter. The

pectin lyase fold/virulence factor and trimethylamine corrinoid protein 2 appear to have no obvious function that is relatable to BT_0339. However, the Na+/glucose co-transporter is one component that appears to be highly relatable to the function of BT_0339. This transporter is predicted to function by an ion gradient that allows for an active transport. A generated high concentration of Na+ on the outside of the bacterial cell creates a favourable gradient for transporting Na+ into the cell, which in-turn pumps glucose into the bacterial cell. Interestingly, this type of protein is found in the intestinal mucosa of the small intestine and enables free glucose to be transported into the blood stream from the intestinal lumen[58]. Assuming this is the mechanism that this transporter performs, it is plausible that BT_0339 acts as a glycoside hydrolase in the lumen outside the bacterial cell. With this assumption, BT_0339 cleaves starch and releases free glucose, which is then transported into the bacterial cell by the transporter and is utilized for energy. Based on the functionality of this transporter, it can be suggested that at a certain level the transporter would become saturated with glucose and would be unable to transport these molecules into the bacterial cell, resulting in free glucose in the lumen. With that, it can also be assumed that some proportion of the free glucose is transported into the blood stream by the human host's intestinal transporters. Based on the results from the enzymatic kinetic analysis, this effect can be assumed to be quite minimal, but nonetheless, it is could be a contributing factor to glucose uptake in the colon by the human host. Additionally, the functionality of BT_0339 may be tied into a dual functioning with the transporter. It is possible that there is a requirement for these proteins to work in unison in order to demonstrate the full functioning capability of BT_0339. These proteins maybe tethered to the membrane and the presence of one enhances the abilities of the other. However, this type of analysis is beyond the scope of this project.

Predicted operon analysis of BT_3299 revealed an operon composed of multiple proteins. This predicted operon contains multiple glycoside hydrolases, Sus-like proteins and a cell surface

receptor. These results were compared to those reported in a previous study where PUL's were identified in *Bacteroides thetaiotaomicron* and interestingly it was found that BT_3299 was a member of PUL #55 (Table S2 in Supplementary Data)[42]. This predicted operon reveals multiple glycoside hydrolases and proteins that were characterized as Sus homologs. PUL #55 consists of a SusD and SusC homologs, as well as a two component system regulator. These predicted proteins illustrate that BT_3299 is a member of an organized system and functions in a manner related to these proteins. Based on the predicted function of these proteins in comparison with the previously characterized Starch Utilization System, it is assumed that the SusD-like protein binds a starch polymer and a degree of cleavage occurs by an uncharacterized glycoside hydrolase. The liberated short polymers of glucose or single glucose molecules are transported into the bacterial cell by the SusC-like protein and are further cleaved into simple sugars within the periplasm. These simple sugars are available to the cell for energy and also act as to induce expression of the entire system, as detected by the two-component system regulator. The predicted function of this operon follows the basis of the Sus-like paradigm. This operon does not highlight whether BT_3299 would be found residing in the lumen, tethered to the membrane or found in the periplasm. However, it does reveal that a starch polymer is predicted to be cleaved outside of the cell and then transported into the cell. Similarly to BT_0339, a proportion of the cleaved glucose molecules will be in the colon lumen and are assumed to be available to be utilized by the surrounding microbiota and the human host.

Crystallization of these enzymes was initiated and significant progress has been made towards data collection. BT_0339 was optimized in a condition that was altered by a slight decrease in the level of PEG and the addition of an additive. Similarly, the optimized condition of BT_3299 also consisted of a decreased amount of PEG and the addition of an additive. Both of these altered factors enabled a condition that allowed for the protein molecules to able to form a more organized

lattice. Initial, screening of diffraction patterns resulted in poor diffraction data with mainly low resolution and high mosaic images. It can be assumed that though the protein molecules are able to form a lattice there is still a degree of disorder in the crystal formation. In order to achieve better quality data there are multiple avenues to pursue to determine the influence causing the mosaicity. Flash cooling the crystals can have an effect on the mosaicity, as a result of the mother liquor freezing and cracking which can create disorder within the crystal. Optimizing the amount of liquid surrounding the crystal or dehydrating it can possibly provide an improvement in the diffraction data. Investigation of the type of cyro-protectant solution used in freezing these crystals could also provide better diffraction patterns. Sometime mosaicity cannot be avoided due to the fact it is a characteristic of how the crystal lattice forms therefore, further investigating additives that allow for three dimensional crystal formations can perhaps provide better diffraction. Simply, another solution to obtaining good quality diffraction data is to continually screen the optimized condition thus far, in hopes of eventually achieving a crystal with a well packed lattice[59].

4.2 Future Directions

It would be interesting to further investigate the substrate specificity of these enzymes and further define their abilities to utilize longer oligosaccharides derived from isomaltose. There are limitations in purchasing isomaltose-derived substrates, which reflects the reason for only using isomaltose in this study. However, an interesting substrate worth pursing is limit dextrins, which are derivatives of starch that enter the colon after small intestinal enzymatic degradation. Limit dextrins are mainly composed of α-(1,4) linked glucose with periodic branched α-(1,6) linkages that will vary due to the source of the originally ingested starch polymer. Investigating a variety of limit dextrins may provide additional information regarding substrate specificity of these enzymes and

particularly address their ability to utilize the maltose-derived substrates at differing rates. Furthermore, having knowledge of the original nutritional source of these limit dextrins can provide a degree of understanding of specifics components of the human diet that these enzymes are able to utilize. Limit dextrins from a variety of sources are available to us through our collaboration with Dr. Bruce Hamaker (Purdue).

Further understanding the predicted operons and how they function can increase the understanding of these enzymes. It is unknown at this point how the components of both operons function and behave structurally, yet their presence is highly related to how BT_0339 and BT_3299 perform. Additionally, it is unclear whether these proteins are under individual control or if a promoter controls them collectively. Completing mutation studies of these operons can reveal whether their functionality is reflected in the survival of this organism and if they are related to one another. This type of analysis can possibly render information regarding these enzymes functioning in a collective manner. Also, assessing the proteins from the operons on an individual level will provide an understanding of how these α-glucosidases function in-vivo and in relation to the human host.

Continued investigation of the optimal condition for diffraction for both enzymes can potentially lead to conditions that will allow for the collection of good quality data. The structure of these enzymes can divulge structural information about the active site and reveal how their structure is relatable to their function. Additionally, structures of these enzymes can possibly provide information regarding their functionally on the substrates assessed in this study and further illustrate the hydrolytic abilities of these enzymes on starch derived substrates. Soaking a substrate into the crystal can reveal specifics about the interaction between the substrate and the active site. It can visually demonstrate how these enzymes demonstrate the observed substrate specificity. This information can possibly be used to complement the kinetic properties observed in this study and further support the functionality of these enzymes in the relation to *Bacteroides thetaiotaomicron.*

It is quite apparent through many completed studies that *Bacteroides thetaiotaomicron* is an organism that possesses specific characteristics that have enabled it to establish its dominant role in the colon microflora. Accumulating evidence further demonstrates that this organism has a multitude of glycoside hydrolyses that enables it to utilize the various undigested polysaccharides passing through the colon. Specifically, it has been shown in this report that this organism has multiple functional glycoside hydrolyses that are members of the Family 31 Glycoside Hydrolase. This report has demonstrated through kinetic analysis that these enzymes are able to use various derivatives of starch and release free glucose. It is hypothesized that these released glucose molecules are able utilized by the organism, but also, are free to be used by surrounding microorganisms and as well as the human host. These results reveal a small but significant role that *Bacteroides thetaiotaomicron* has in contributing to the human host's nutrition and energy levels. Further investigation of these enzymes structurally and their functional operon counterparts can provide additional information regarding the role of this organism as a starch degrader. Ultimately, this knowledge can help us to understand the colon microbiota that has the ability to strongly influence human nutrition.

Chapter 5

Supplementary Data

The kinetics for the enzymes BT_0339 and BT_3299 were completed on various substrates in addition to maltose. These assays were completed in a range of substrates and the data was fitted to a Michaelis-Menten plot.

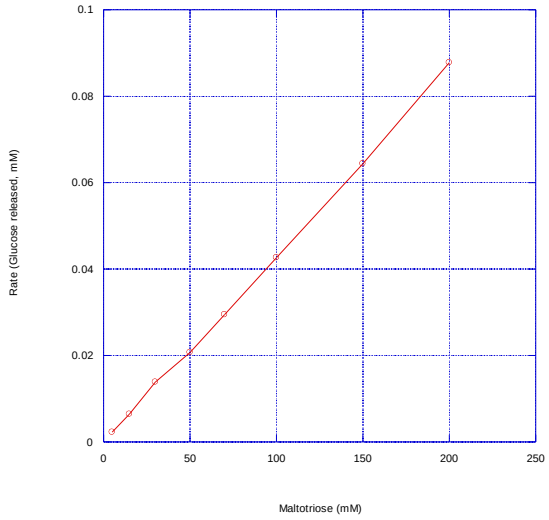
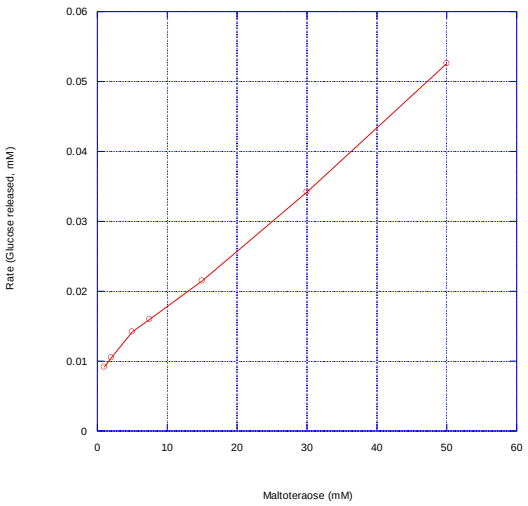
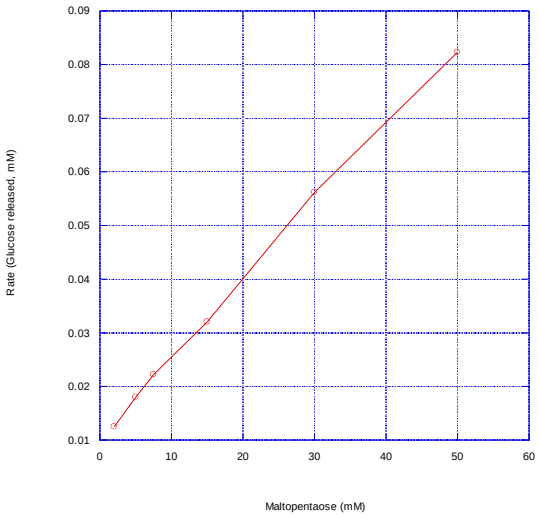


Figure S1 Michaelis-Menten plot for BT_0339 with maltotriose as the substrate. Assay was completed with substrate in the range of 5-200mM.



igure S2 Michaelis-Menten plot for BT_0339 with maltotetraose as a substrate. Assay was completed with substrate in the range of 1-50mM



.

Figure S3Michaelis-Menten plot BT_0339 with maltopentaose as a substrate. Assay was completed with substrate in the range of 2-50mM.

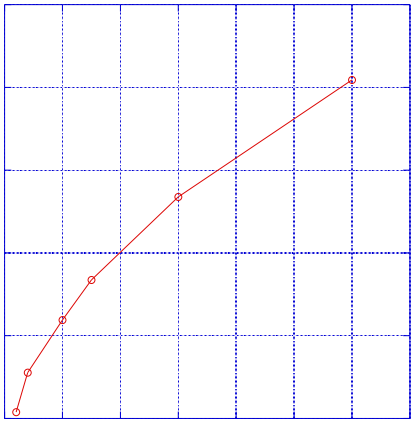


Figure S4 Michaelis-Menten plot for BT_0339 with maltohexaose as a substrate. Assay was completed with substrate in the range of 1-30mM.

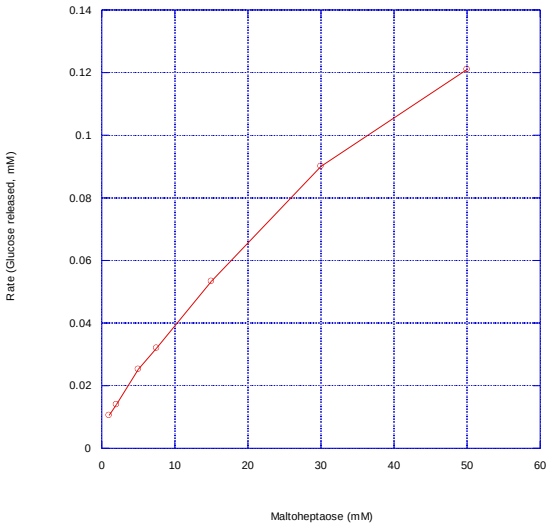


Figure S5 Michaelis-Menten plot for BT_0339 with maltoheptaose as a substrate. Assay was completed with substrate in the range of 1-50mM.

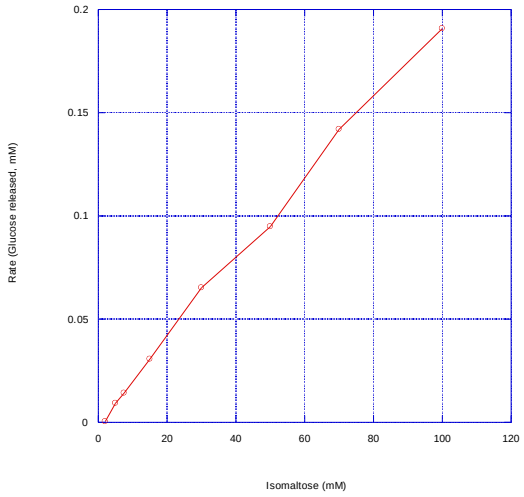


Figure S6 Michaelis-Menten plot for BT_0339 with isomaltose as a substrate. Assay was completed with substrate in the range of 2-100mM.

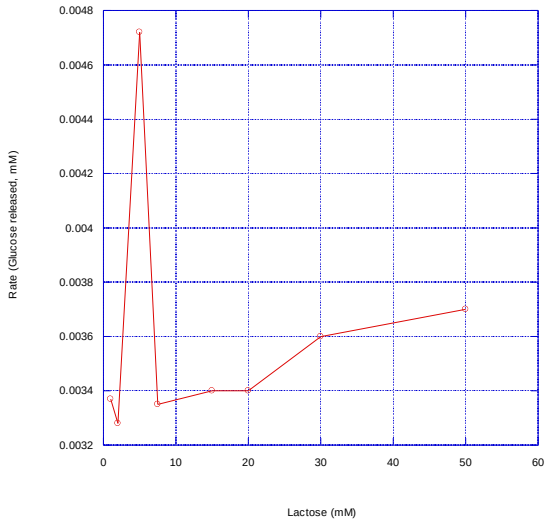


Figure S7 Michaelis-Menten plot for BT_0339 with lactose as a substrate. Assay was completed with substrate in the range of 1-50mM.

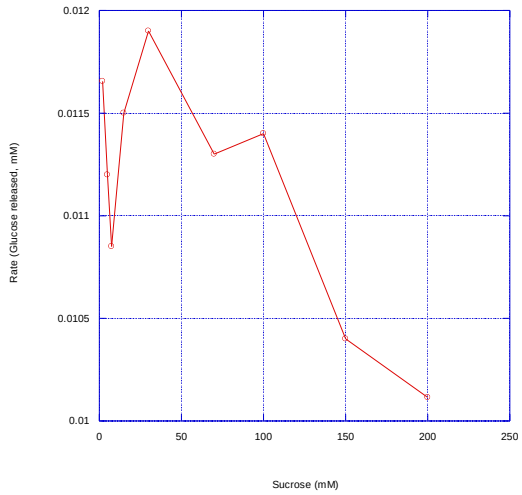


Figure S8 Michaelis-Menten plot for BT_0339 with sucrose as a substrate. Assay was completed with substrate in the range of 2-200mM.

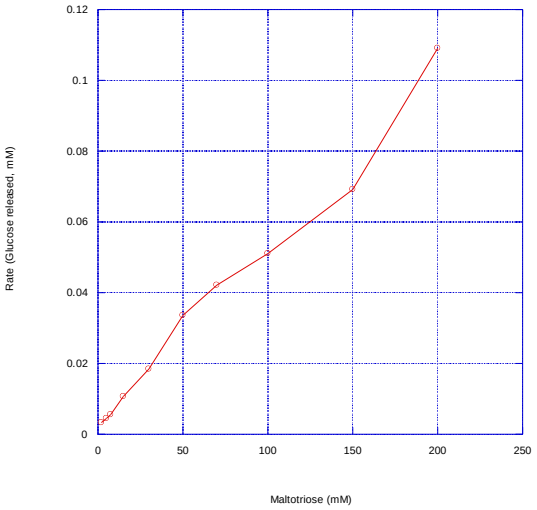


Figure S9 Michaelis-Menten plot for BT_3299 with maltotriose as a substrate. Assay was completed with substrate in the range of 2-200mM.

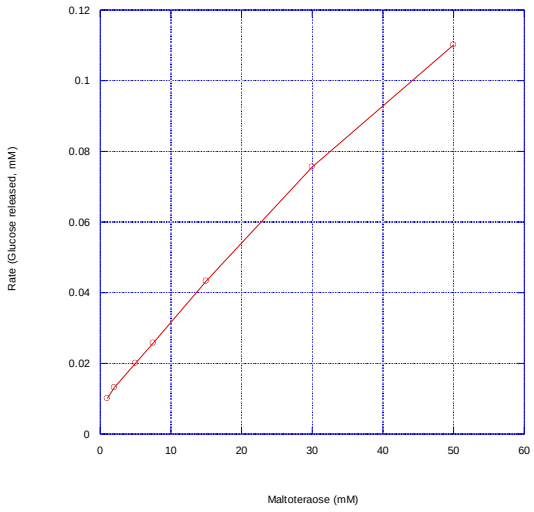


Figure S10 Michaelis-Menten plot for BT_3299 with maltotetraose as a substrate. Assay was completed with substrate in the range of 1-50mM.

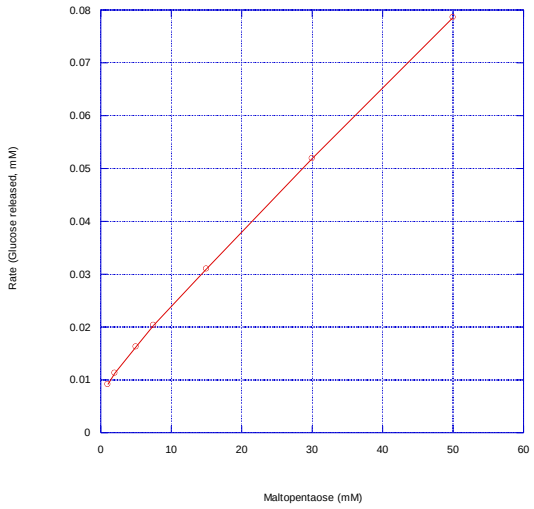


Figure S11 Michaelis-Menten plot for BT_3299 with maltopentaose. Assay was completed with substrate in the range of 1-50mM.

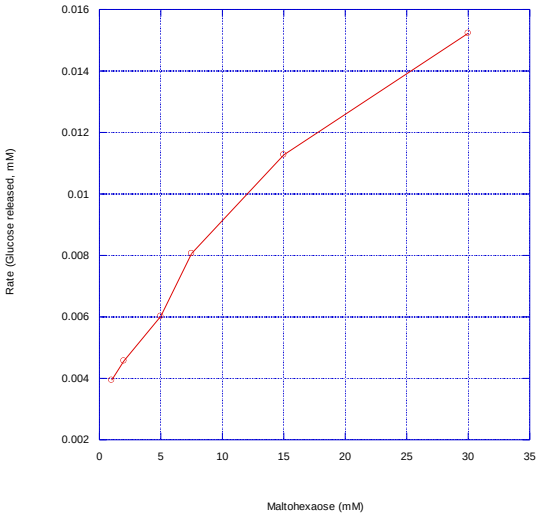


Figure S12 Michaelis-Menten plot BT_3299 with maltohexaose as a substrate. Assay was completed with substrate in the range of 1-30mM.

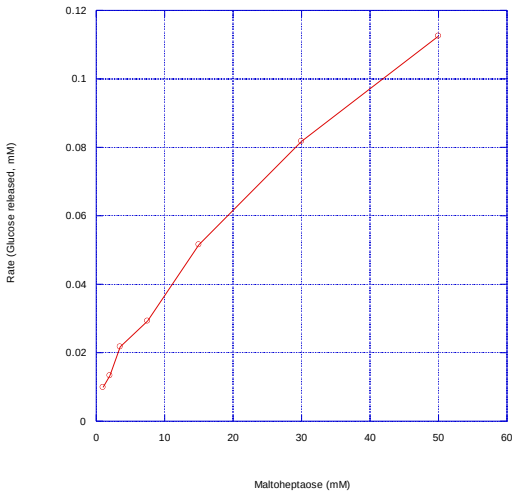


Figure S13 Michaelis-Menten plot BT_3299 with maltoheptaose as a substrate. Assay was completed with substrate in the range of 1-50mM.

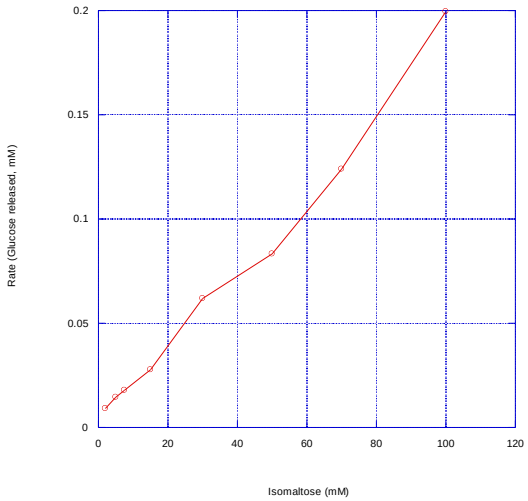


Figure S14 Michaelis-Menten plot BT_3299 with isomaltose as a substrate. Assay was completed with substrate in the range of 2-100mM.

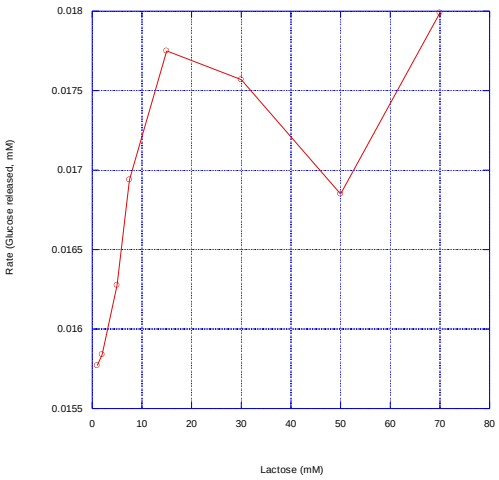


Figure S15 Michaelis-Menten plot for BT_3299 with lactose as a substrate. Assay was completed with substrate in the range of 1-70mM.

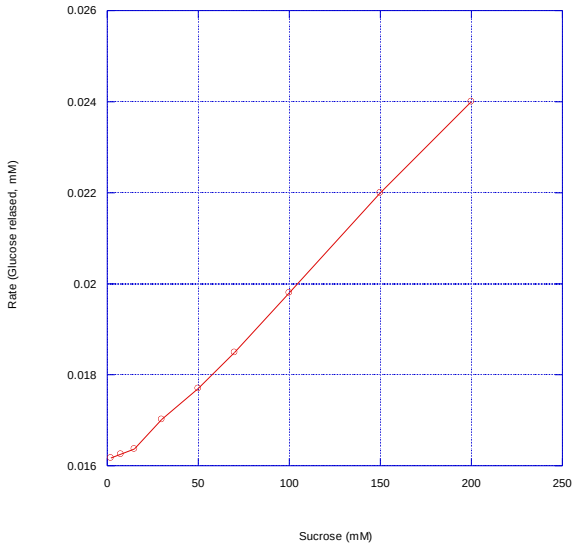


Figure S16 Michaelis-Menten plot for BT_3299 with sucrose as a substrate. Assay was completed with substrate in the range of 2-200mM

Table S1 Catalytic ability of BT_0339 and BT_3299 on various substrates.

	V/K	
Substrate	**BT_0339**	**BT_3299**
Maltose	0.00151 ±4.8E-05	0.00113±6.7E-05
Isomaltose	0.00217±2.8E-05	0.00206±1.6E-04
Malotriose	0.00046±1.0E-04	0.00061±2.3E-05
Maltotetraose	0.00114±5.0E-05	0.00252±1.7E-05
Maltopentaose	0.00187±1.2E-04	0.00173±6.4E-05
Maltohexaose	0.00041±1.7E-05	0.00051±6.6E-05
Maltoheptaose	0.00300±4.8E-04	0.00272±3.0E-04
Sucrose	0.00039±1.5E-05	0.00059±6.7E-06
Lactose	0.00012±1.2E-05	0.00058±4.1E-07

Table S2 Polysaccharide Utilization Loci # 55 for *Bacteroides thetaiotaomicron.*

Gene	Protein Name
bt_3292	β-glucosidase
bt_3293	β-galatosidase
bt_3294	β-glucosidase
bt_3295	Hypothetical protein
bt_3296	SusD-like
bt_3297	SusC-like
bt_3298	Hypothetical protein
bt_3299	α-glucosidase
bt_3300	β-glucosidase
bt_3301	α-1,6 mannase
bt_3302	two-component system regulator

Permissions

Selected figures were used in this thesis by means of permission that was granted to publish by the following permission statements.

1. Figure 1.1 Physiology of the gastrointestinal tract.

Dear Ms. Marcia Chaudet,

We hereby grant you permission to reprint the material detailed below at no charge in your thesis subject to the following conditions:

1. If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies.

2. Suitable acknowledgment to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

R20;This article was published in Publication title, Vol number, Author(s), Title of article, Page Nos, Copyright Elsevier (or appropriate Society name) (Year).”

3. Your thesis may be submitted to your institution in either print or electronic form.

4. Reproduction of this material is confined to the purpose for which permission is hereby given.

5. This permission is granted for non-exclusive world English rights only. For other languages please reapply separately for each one required. Permission excludes use in an electronic form. Should you have a specific electronic project in mind please reapply for permission.

6. This includes permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

Yours sincerely,
Kayleigh Harris
Rights Associate :: Global Rights :: ELSEVIER

2. Figure 1.4 *Bacteroides thetaiotaomicron's* role in the colon.

Dear Ms. Chaudet:

Thank you very much for your request and for your interest in our content.

Policies for reproducing AAAS material in your thesis or dissertation can be found on our Rights and Permissions Guidelines page at this URL:

http://www.sciencemag.org/site/about/permissions.xhtml#thesis

"Reproducing AAAS Material in your Thesis or Dissertation

NOTE**:** If you are the Original Author of the AAAS article being reproduced, please refer to your License to Publish for rules on reproducing your paper in a dissertation or thesis. AAAS permits the use of content published in its journals *Science, Science Technical Medicine*, and *Science Signaling*, but only provided the following criteria are met.

1. If you are using figure(s)/table(s), permission is granted for use in print and electronic versions of your dissertation or thesis.
2. A full text article may be used only in print versions of a dissertation or thesis. AAAS does not permit the reproduction of full text articles in electronic versions of theses or dissertations.
3. The following credit line must be printed along with the AAAS material: "From [Full Reference Citation]. Reprinted with permission from AAAS."
4. All required credit lines and notices must be visible any time a user accesses any part of the AAAS material and must appear on any printed copies that an authorized user might make.
5. The AAAS material may not be modified or altered except that figures and tables may be modified with permission from the author. Author permission for any such changes must be secured prior to your use.
6. AAAS must publish the full paper prior to your use of any of its text or figures.
7. If the AAAS material covered by this permission was published in *Science* during the years 1974–1994, you must also obtain permission from the author, who may grant or withhold permission, and who may or may not charge a fee if permission is granted. See original article for author's address. This condition does not apply to news articles.
8. If you are an Original Author on the AAAS article being reproduced, please refer to your License to Publish for rules on reproducing your paper in a dissertation or thesis.

Permission covers the distribution of your dissertation or thesis on demand by a third party distributor (e.g. ProQuest / UMI), provided the AAAS material covered by this permission remains in situ and is not distributed by that third party outside of the context of your Thesis/Dissertation.

Permission does not apply to figures/photos/artwork or any other content or materials included in your work that are credited to non-AAAS sources. If the requested material is sourced to or

references non-AAAS sources, you must obtain authorization from that source as well before using that material. You agree to hold harmless and indemnify AAAS against any claims arising from your use of any content in your work that is credited to non-AAAS sources.
By using the AAAS Material identified in your request, you agree to abide by all the terms and conditions herein.

AAAS makes no representations or warranties as to the accuracy of any information contained in the AAAS material covered by this permission, including any warranties of merchantability or fitness for a particular purpose.

Questions about these terms can be directed to the AAAS Permissions department at permissions@aaas.org."

In addition to sourcing the article, kindly credit the staff illustrator as well: Katharine Sutliff/SCIENCE

If you have any questions please just let me know.

Best regards,

Liz Sandler

(Ms) Elizabeth Sandler
Rights and Permissions
The American Association for the Advancement of Science

3. Figure 2.1 pET-29a vector and Figure 2.2 pET-15b vector.

Dear Marcia:

This e-mail confirms that you have permission to use the vector maps for pET29a and pET15b obtained from EMD Millipore Corporation in your Masters' degree thesis at the University of Waterloo, Canada.

Please let me know if you have any further questions.

Best regards,

Marie Azzaria, Ph.D.
Manager, Intellectual Property

EMD Millipore Corporation

References

1. Xu, J. *et al.* Evolution of symbiotic bacteria in the distal human intestine. *PLoS biology* **5**, e156 (2007).

2. Ouwehand, Arthur and Vaughan, Elaine, E. *Gastrointestinal microbiology*. 409 (CRC Press: London, UK, 2006).

3. Wilson, M. *Bacteriology of Humans. An Ecological Perspective*. (Blackwell Publishing: Oxford, UK, 2008).

4. Bäckhed, F., Ley, R.E., Sonnenburg, J.L., Peterson, D. a & Gordon, J.I. Host-bacterial mutualism in the human intestine. *Science (New York, N.Y.)* **307**, 1915-20 (2005).

5. DiBaise, J.K. *et al.* Gut microbiota and its possible relationship with obesity. *Mayo Clinic proceedings. Mayo Clinic* **83**, 460-9 (2008).

6. Comstock, L.E. & Coyne, M.J. Bacteroides thetaiotaomicron: a dynamic, niche-adapted human symbiont. *BioEssays : news and reviews in molecular, cellular and developmental biology* **25**, 926-9 (2003).

7. Flint, H.J., Duncan, S.H., Scott, K.P. & Louis, P. Interactions and competition within the microbial community of the human colon: links between diet and health. *Environmental microbiology* **9**, 1101-11 (2007).

8. Cho, Sung Soo, Susan and Finocchiaro, T.E. *Handbook of prebiotics and probiotics ingredients: health benefits.* 435 (CRC Press: Boca Raton, Florida, USA, 2010).

9. Anderson IH & , Levine AS, L.M. Incomplete absorption of the carbohydrate in all-purpose wheat flour. *New England Journal of Medicine* **304**, 891-892 (1981).

10. McNeil, N.I. The contribution of the large intestine to energy supplies in man. *The American journal of clinical nutrition* **39**, 338-42 (1984).

11. Pryde, S.E., Duncan, S.H., Hold, G.L., Stewart, C.S. & Flint, H.J. The microbiology of butyrate formation in the human colon. *FEMS microbiology letters* **217**, 133-9 (2002).

12. McIntyre, a, Gibson, P.R. & Young, G.P. Butyrate production from dietary fibre and protection against large bowel cancer in a rat model. *Gut* **34**, 386-91 (1993).

13. Louis, P., Scott, K.P., Duncan, S.H. & Flint, H.J. Understanding the effects of diet on bacterial metabolism in the large intestine. *Journal of applied microbiology* **102**, 1197-208 (2007).

14. Eckburg, P.B. *et al.* Diversity of the Human Intestinal Microbial Flora. *Science (New York, N.Y.)* **308**, 1635-1638 (2005).

15. Carolyn, M., Arbo, M.D.J., Grindlinger, J., Snydman, D.R. & Redondo, M.C. Associated with the Bacteroides Attributable fragilis Mortality of Bacteremia Group. **20**, 1492-1496 (2012).

16. Xu, J. & Gordon, J.I. Honor thy symbionts. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 10452-9 (2003).

17. Gilmore, M.S. & Ferretti, J.J. Microbiology. The thin line between gut commensal and pathogen. *Science (New York, N.Y.)* **299**, 1999-2002 (2003).

18. Salyers, A. a, Gupta, A. & Wang, Y. Human intestinal bacteria as reservoirs for antibiotic resistance genes. *Trends in microbiology* **12**, 412-6 (2004).

19. Wexler, H.M. Bacteroides: the good, the bad, and the nitty-gritty. *Clinical microbiology reviews* **20**, 593-621 (2007).

20. Mazmanian, S.K. & Kasper, D.L. The love-hate relationship between bacterial polysaccharides and the host immune system. *Nature reviews. Immunology* **6**, 849-58 (2006).

21. Mazmanian, S.K., Liu, C.H., Tzianabos, A.O. & Kasper, D.L. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122**, 107-18 (2005).

22. Cash, H.L., Whitham, C.V., Behrendt, C.L. & Hooper, L.V. Symbiotic bacteria direct expression of an intestinal bactericidal lectin. *Science (New York, N.Y.)* **313**, 1126-30 (2006).

23. Hooper, L.V., Stappenbeck, T.S., Hong, C.V. & Gordon, J.I. Angiogenins: a new class of microbicidal proteins involved in innate immunity. *Nature immunology* **4**, 269-73 (2003).

24. Sansonetti, P.J. War and peace at mucosal surfaces. *Nature reviews. Immunology* **4**, 953-64 (2004).

25. Upadhyay, N. & Moudgal, V. Probiotics: A Review. *Journal of Scientific Communication* **19**, 76-84 (2012).

26. Kneifel, W., Salminen, S. *Probiotics and Health Claims*. 360 (Blackwell Publishing: Oxford, UK, 2011).

27. Gareau, M.G., Sherman, P.M. & Walker, W.A. Probiotics and the gut microbiota in intestinal health and disease. *Nature reviews. Gastroenterology & hepatology* **7**, 503-14 (2010).

28. Varyukhina, S. *et al.* Glycan-modifying bacteria-derived soluble factors from Bacteroides thetaiotaomicron and Lactobacillus casei inhibit rotavirus infection in human intestinal cells. *Microbes and infection / Institut Pasteur* **14**, 273-8 (2012).

29. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59-65 (2010).

30. Noor, S.O. *et al.* Ulcerative colitis and irritable bowel patients exhibit distinct abnormalities of the gut microbiota. *BMC gastroenterology* **10**, 134 (2010).

31. Bäckhed, F. *et al.* The gut microbiota as an environmental factor that regulates fat storage. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 15718-23 (2004).

32. Mahowald, M. a *et al.* Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 5859-64 (2009).

33. Ley, R.E., Peterson, D. a & Gordon, J.I. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* **124**, 837-48 (2006).

34. Farooqi, S. & O'Rahilly, S. Genetics of obesity in humans. *Endocrine reviews* **27**, 710-18 (2006).

35. Martens, E.C. *et al.* Recognition and Degradation of Plant Cell Wall Polysaccharides by Two Human Gut Symbionts. *PLoS Biology* **9**, e1001221 (2011).

36. Anderson, K.L. & Salyers, a a Biochemical evidence that starch breakdown by Bacteroides thetaiotaomicron involves outer membrane starch-binding sites and periplasmic starch-degrading enzymes. *Journal of bacteriology* **171**, 3192-8 (1989).

37. D'Elia, J.N. & Salyers, a a Contribution of a neopullulanase, a pullulanase, and an alpha-glucosidase to growth of Bacteroides thetaiotaomicron on starch. *Journal of bacteriology* **178**, 7173-9 (1996).

38. Reeves, a R., Wang, G.R. & Salyers, a a Characterization of four outer membrane proteins that play a role in utilization of starch by Bacteroides thetaiotaomicron. *Journal of bacteriology* **179**, 643-9 (1997).

39. Wetzel, G., Heine, M., Rohwedder, A. & Naim, H.Y. Impact of glycosylation and detergent-resistant membranes on the function of intestinal sucrase-isomaltase. *Biological chemistry* **390**, 545-9 (2009).

40. Koropatkin, N.M., Martens, E.C., Gordon, J.I. & Smith, T.J. Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure (London, England : 1993)* **16**, 1105-15 (2008).

41. Koropatkin, N.M. & Smith, T.J. SusG: a unique cell-membrane-associated alpha-amylase from a prominent human gut symbiont targets complex starch molecules. *Structure (London, England : 1993)* **18**, 200-15 (2010).

42. Martens, E.C., Chiang, H.C. & Gordon, J.I. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell host & microbe* **4**, 447-57 (2008).

43. Okuyama, M. Function and structure studies of GH family 31 and 97 α-glycosidases. *Bioscience, biotechnology, and biochemistry* **75**, 2269-77 (2011).

44. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, H.B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic acids research* **37**, D233-238 (2009).

45. Sim, L. *et al.* Structural basis for substrate selectivity in human maltase-glucoamylase and sucrase-isomaltase N-terminal domains. *The Journal of biological chemistry* **285**, 17763-70 (2010).

46. Koshland, D.E. STEREOCHEMISTRY AND T H E MECHANISM OF ENZYMATIC REACTIONS. *Reactions* (1953).

47. Bourne, Y. & Henrissat, B. Glycoside hydrolases and glycosyltransferases: families and functional modules. *Current opinion in structural biology* **11**, 593-600 (2001).

48. Withers, S. Mechanisms of glycosyl transferases and hydrolases. *Carbohydrate Polymers* **44**, 325-337 (2001).

49. Jacobs, J. Characterization of BT3299: A Family GH31 Enzyme from a Prominent Gut Symbiont, Bacteroides thetaiotaomicron. 108 (2011).

50. Novagen pET System Manual. *BioSystems* **11**, 80 (2005).

51. Forbes, M. *Mass Spectrometry*. (Toronto, 2011).

52. Copeland, R. *Enzymes: A Practical Introduction to Structure, Mechanim and Data Analysis*. 412 (Wiley-VCH: 2000).

53. Dehal, P.S. *et al.* MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic acids research* **38**, D396-400 (2010).

54. Price, M.N., Huang, K.H., Alm, E.J. & Arkin, A.P. A novel method for accurate operon predictions in all sequenced prokaryotes. *Nucleic acids research* **33**, 880-92 (2005).

55. Hampton Research Index Screen. (2002).at <http://hamptonresearch.com/product_detail.aspx?cid=1&sid=24&pid=5>

56. Vaillancourt E., P. *E.coli gene expression protocols*. 347 (Humana Press: 2003).

57. Raaman, N. *Phytochemical Techniques*. 306 (Jai Bharat Printing Press: New Dehli, 2006).

58. Castaneda-Sceppa, C. & Castaneda, F. Sodium-dependent glucose transporter protein as a potential therapeutic target for improving glycemic control in diabetes. *Nutrition reviews* **69**, 720-9 (2011).

59. Rupp, B. *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*. 809 (Garland Science: New York, NY, 2010).