Starch Glycan Utilization by the Gut Microbiota:

Structural and Biochemical Insights Into Starch Digestion and the Role of the Colon Microbiota

by

Marcia Michele Chaudet

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Biology

Waterloo, Ontario, Canada, 2016

©Marcia Michele Chaudet 2016

Authors 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 maybe made electronically available to the public.

Abstract

The human digestive system is resident to a highly populated ecosystem of bacterial species known as the colon microbiota, which contributes significantly to our assimilation of dietary and mucosal carbohydrates. *Bacteroides thetaiotaomicron* is a member of this diverse colon environment and plays a large role in the break down of dietary complex carbohydrates. The process of sensing and responding to glycans present in the colon lumen is predicted to rely on the mechanisms of proteins that are part of complex gene loci deemed polysaccharide utilization loci. Members of these loci are responsible for acquiring substrates at the cell outer membrane, transporting oligosaccharides into the periplasm, degrading them into simple sugars and importing them into the cell. The objective of this thesis is to investigate specific components of the polysaccharide utilization locus 55 and investigate its role as an alternative starch utilization system in *B.thetaiotaomicron*. Our analysis of an outer membrane binding protein, a SusD homolog, highlights its contribution to this PUL by acquiring starch-derived glycans from the colon lumen. Through our structural characterization of two Family GH31 α-glucosidases we reveal the adaptability of this bacterium at hydrolyzing a range of starch-derived glycans with the emphasis on branched substrates. To understand the transcriptional control of this PUL, we investigated the interaction of this PUL's response regulator with *B.thetaiotaomicron*'s genomic DNA. With these results we demonstrate the predicted function of a locus of genes that is capable of contributing to starch hydrolysis in the human colon and the role this locus could play in upregulating related genes in other loci. Through evolutionary analysis of Family 31 α-glucosidases we were able we are able to observe a number of evolutionary changes that we believe resulted in duplicated genes that provided evidence of the colon microbiota being efficient at utilizing dietary starch. These results provide insight into the complex process of the colon microbiota sensing, acquiring and utilizing dietary glycans in the human colon.

.

Acknowledgements

Completing my PhD studies has been an experience, one that I will never forget. I will carry with me the ups and the downs that I experienced and really these experiences have allowed me to grow into the individual I am today. This journey is one that I could never of made on my own and the support of my family and friends is what has given me the strength to move forward and work hard in completing this degree.

I would like to thank my MacDonald family for the constant support and encouragement. There were many times when I wasn't sure I would be able to complete this, but they always encouraged me to keep going and to achieve my dreams. Thank you to my other family, the McGuire/Pearce family, for their constant support and even attempts to read my publications.

I would like to thank my past and present lab members who have always been supportive and who were always there to provide advice. This journey began with my good friend and my first Rose lab mate, Kyra Jones. Thank you for being a great friend and always being there for me, encouraging me throughout graduate school and always reminding me of the important things in life

I would like to thank my supervisor David Rose. Your encouragement throughout my studies is something I am truly grateful for. You have pushed me to become the researcher and academic that I am proud to be today. Thank you for giving me countless opportunities to grow as individual, but also as a scientist. Thank you also to my committee members, Todd Holyoak and Trevor Charles. Your guidance and advice over the last few years has been very helpful as this project developed over the years.

The members of the Holyoak lab have also been a great support during my graduate studies. Thank you to Will Lotoksy, Iain Wallace and Matt McLeod for your support and always willing to de-stress with a beer at the Grad house. A huge thanks to Kathy Lam from the Charles Lab. You were my *B.theta* confidant in the department and were always willing to chat about my experiments and provide guidance with my microbiology work.

Lastly, I would like to thank my husband, Matthew McGuire. Your support and encouragement I hold close to my heart. You were always willing to listen to my excitements of experimental results or my complaints of failed experiments. I've come to learn that crystallography is not just a difficult discipline for a scientist, but equally difficult for scientist's partner who bears the brunt of the emotional toll of 'trying to get crystals'. Thank you my love.

Table of Contents

AUTHOR'S DECLARATION ii
Abstract iii
Acknowledgments iv
Table of Contents vi
List of Figures ix
List of Tables xi

Chapter 1 Introduction and Thesis Outline 1
1.1 Human Microbiome 1
1.2 Human Colon Microbiota 2
1.3 *Bacteroides thetaiotaomicron* 3
1.4 Colon Microbiota and Glycan Metabolism 3
1.5 Starch Digestion 6
1.6 Starch Utilization System 8
1.7 Polysaccharide Utilization Loci 10
1.8 Glycoside Hydrolase Family 31 α-glucosidases 11
1.9 Experimental Approach 12

Chapter 2 Characterization of Starch Glycan Utilization by *Bacteroides thetaiotaomicron* 14
2.1 Chapter Overview 14
2.2 Contributions 14
2.3 Material and Methods 16
2.3.1 Expression and Purification of a SusD-homolog BT_3296 16
2.3.2 Binding Interaction Observed by Fluorescence Spectroscopy 17
2.3.3 Crystallization Screening of BT_3296-Se 18
2.3.4 Expression and Purification of BT_3299-Se and BT_0339-Se 18
2.3.5 Crystallization and Data Collection of BT_3299 and BT_0339 18
2.3.6 X-ray Structure Determination of BT_3299 and BT_0339 19
2.3.7 Molecular Modeling of BT_3299 and BT_0339 19
2.4 Results 20
2.4.1 Expression and Purification of BT_3296 and BT_3296-Se 20
2.4.2 Binding Activity of BT_3296 and Homology Modeling 20
2.4.3 Crystallization Screening of BT_3296-Se 24
2.4.4 Crystallization of Family GH31 α-glucosidases 25

2.4.5 Overall Structure of BT_3299 and BT_0339 26
2.4.6 Proposed Binding of Maltooligosaccharides 29
2.4.7 Comparison to Family GH31 α-glucosidases and SusB Structures 32
2.5 Discussion 34
2.6 Conclusion 36

Chapter 3 Evolution of Family 31 α-glucosidases 37
3.1 Chapter Overview 37
3.2 Contributions 38
3.3 Methods 39
3.3.1 Phylogenetic Analysis of Family 31 α-glucosidase Structures 39
3.3.2 Predicted Evolutionary Changes in Family 31 α-glucosidase Structures 39
3.3.3 Comparison of Predicted Evolutionary Changes and Acarbose Binding 39
3.3.4 Phylogenetic Analysis of Characterized Family 31 α-glucosidases 41
3.4 Results 43
3.41. Structural Evolution 43
3.4.2 Variability in Catalytic Region 47
3.4.3 Variability in Looped Substrate-Binding Region 50
3.4.4 Phylogenetic Relationships in Characterized Family 31 α-glucosidases 51
3.4.5 Eukaryote Signature Sequence 57
3.5 Discussion 59
3.6 Conclusion 63

Chapter 4 Transcriptional Control of PUL 55 64
4.1 Chapter Overview 64
4.2 Contributions 66
4.3 Material and Methods 67
4.3.1 Expression and Purification of the Response Regulator BT_3302-5 67
4.3.2 Electromobility Shift Assay 69
4.3.3 *Bacteroides thetaiotaomicron* Growth Experiments 70
4.4 Results 71
4.4.1 Growth Experiments of *Bacteroides thetaiotaomicron* on Starch Substrates 71
4.4.2 Expression and Purification of Response Regulator BT_3302-5 74
4.4.3 Electromobility Shift Analysis of BT_3302-5 74

4.5 Discussion 80
4.6 Conclusion 87

Chapter 5 Conclusion and Future Work 89
5.1 Concluding Remarks 89
5.2 Future Studies 92
5.2.1 BT_3296 SusD-Homolog 92
5.2.2 Structural Studies of Family GH31 α-glucosidases from *Bacteroides thetaiotaomicron* 92
5.2.3 Evolution of Family 31 α-glucosidases 93
5.2.4 Transcriptional control of PULs 94
5.3 Colon Microbiota and Digestive Disease States 95
5.4 Manipulation of the Colon Microbiota 96

Bibliography 98

Appendix 108

List of Figures

Figure 1.1 Simple and complex carbohydrates available in the colon for microbial fermentation as proposed by Koropatkin *et al,* 2012 5

Figure 1.2 α-glucans of starch 6

Figure 1.3 Proposed model of the starch utilization system by Cameron *et al,* 2014 9

Figure 1.4 Starch utilization components form the polysaccharide utilization locus 55.....12

Figure 2.4.1 Denaturing SDS-PAGE gel for BT_3296 and BT_3296-Se.....20

Figure 2.4.2 Emission scan of BT_3296 incubated with β-cyclodextrin 21

Figure 2.4.3 Emission scan of BT_3296 incubated with starch-derived sugars. 22

Figure 2.4.4 Predicted model of BT_3296 24

Figure 2.4.5 Crystallization of BT_3296-Se 25

Figure 2.4.6 Crystallization of family GH31 α-glucosidases 25

Figure 2.4.7 Crystal structures of BT_3299 and BT_0339. 30

Figure 2.4.8 Comparison of BT_3299 and BT_0339 active sites with modeled acarbose. 31

Figure 2.4.9 Comparisons of structurally solved family 31 GH31 α-glucosidases and SusB 33

Figure 3.4.1 Structural multiple sequence alignment of structurally solved family 31 α-glucosidases. 44

Figure 3.4.2 Phylogenetic tree of family GH31 α-glucosidases for experimentally solved structures 46

Figure 3.4.3 Predicted evolutionary changes in catalytic region 49

Figure 3.4.4 Predicted evolutionary changes in looped binding region.....51

Figure 3.4.5 Phylogenetic tree of characterized family 31 α-glucosidases. ..53

Figure 3.4.6 Multiple sequence alignment of characterized family 31 α-glucosidases..... 54-56

Figure 3.4.7 Predicted eukaryote signature sequence in characterized family 31 α-glucosidases. 55

Figure 4.1.1 Predicted domain organization of BT_3302. 65
Figure 4.4.1 Growth curve of wild-type *Bacteroides thetaiotaomicron* strain grown on starch derived substrates 71
Figure 4.4.2 Growth curve of Δtdk *Bacteroides thetaiotaomicron* strain grown on starch derived substrates 72
Figure 4.4.3 Growth curve of Δsus *Bacteroides thetaiotaomicron* strain grown on starch derived substrates 73
Figure 4.4.4 Denaturing SDS-PAGE gel for BT_3302-5 74
Figure 4.4.5 Electrophoresis mobility shift assay for BT_3302-5 75
Figure 4.5.1 Growth curves of *sus* mutants grown on various maltooligosaccharides 81
Figure 4.5.1 Proposed primary sensing of glycans by PUL 55. 86

List of Tables

Table 1.1.1 Polysaccharide utilization locus components related to starch utilization 13

Table 2.3.1 Primers for BT_3296 16

Table 2.4.2 Data collection and refinement statistics for BT_3299 and BT_0339 structures 28

Table 3.3.1 Family GH31 α-glucosidases used in the structural analysis of evolutionary change 40

Table 3.3.2 Characterized family 31 α-glucosidases used in evolutionary analysis 42

Table 3.4.1 Summary of predicted amino acid changes from DIVERGE 48

Table 4.3.1Construct design for the response regulator BT_3302 67

Table 4.3.2 Primers for BT_3302-5 68

Table 4.4.1 Fragments recovered from electromobility shift assay with BT_3302-5. 76

Table 4.4.2 Associated PULs of isolated fragments from electromobility shift assay 77

Table 4.4.3 Associated PULs of isolated fragments from electromobility shift assay 78

Table 4.4.4 Associated PULs of isolated fragments from electromobility shift assay 79

Table 4.5.1 PUL 55 Genes 84

Chapter 1
Introduction and Thesis Outline

1.1 Human Microbiome

There are ranges of factors that exist in nature that influence the physiology of the human body. Understanding these connections in relation to our health is of great importance in order to maintain our general health, but also for the future of preventative medicine. One contributing factor that has generated interest in the scientific community is the relationship between the human body and its microbiome. The human microbiome is a population of microbial cells that exists within or on the human body[1]. These cells have been shown to populate various areas of the body and in general form commensal relationships in symbiosis with the human host[2]. It has been estimated that microbial cells outnumber human somatic cells by a factor of 10^3. This relationship begins at infancy with the exposure to our mother's microbiota during birth followed by environmental exposure. This population is quite resilient and in general can remain quite stable over time. Changes can be observed in the community structure as result of alternations in environmental factors such as diet, disease or age[4].

The specifics of the human microbiome are not fully understood at this point in time though observations have been made in regards to its influence and direct interaction through its colonization of diverse sites of our body[5]. Populations occur in the respiratory system, the integumentary system, the urinary tract and the digestive system. The largest portion of this community can be found in the large intestine reaching numbers upwards of 10^{12} number of cells/g[6]. As a result of this predominant presence in the human digestive system this population is identified as a functioning entity or a metabolically active organ[7]. Understanding the functionality of this large population and its contribution to our gastrointestinal health has been a strong motivating factor for research in this area of microbiology.

1.2 Human Colon Microbiota

The rumen in ruminant animals and the human colon are sites that exhibit the highest populated and complex bacterial ecosystems on earth[8]. The human colon microbiota is believed to have co-evolved with humans and to possess physiological attributes that compensate for ones that we are lacking[6,9]. It has been observed to play an important role in the development of our immune system in the gastrointestinal tract, intestinal physiology and metabolism of dietary elements[8,10–12].

The composition of the colon microbiota primarily consists of the phyla Firmicutes, Bacteroidetes and Actinobacteria[13]. The level of diversity in this community is correlated with the general health of this population and subsequently the human host[14,15]. Individuals with a healthy colon microbiota have been observed to have a high level of diversity, where individuals suffering from disease state tend to have a lower level of diversity. Though this observation has been proven to be quite consistent, the specifics of this diversity (species level) can vary significantly amongst individuals[16].

This colon microbiota is also characteristically dynamic in nature with its composition continually fluctuating as a result of exposure to selective pressures that include diet, drug exposure and the overall health of the human host. Theses selective pressures combined with microbiota competition contribute to the overall shaping this population. Diet is specifically a factor that has been strongly observed to influence this population's stability as demonstrated with its initial development during infancy[13]. The colon microbiota composition has been observed to evolve during the transition from breast milk to solid food and provides evidence of the impact diet has on this community. Ingestion of breast milk creates an infant colon microbiota mainly being composed of Enterobacteria and Bacteroidetes with very little diversity. Transition to solid food causes a dramatic shift in the population resulting in a higher diverse community dominated by Bacteroides and Firmicutes[17].

1.3 *Bacteroides thetaiotaomicron*

One of the dominant members of the colon microbiota is *Bacteroides thetaiotaomicron*. *B.thetaiotaomicron* is a gram-negative anaerobe and historically has been the bacterium of choice for investigating the metabolic capabilities of the colon microbiota[18]. Genome sequencing of this bacterium demonstrated its dedicated metabolism to glycan utilization and provided a perspective of the specific role the colon microbiome has on the whole[19]. A large portion of its genome is specifically dedicated to the uptake and hydrolysis of a range of polysaccharides and is commonly identified as a highly saccharolytic bacterium[20]. *Bacteroides thetaiotaomicron* has classically been studied for the purpose of understanding the colon microbiota role in dietary sugar utilization[21]. Research has also provided evidence of its role in antibiotic resistance transfer in the colon microbiota and it has been identified as a causative agent of post-surgical infections[22,23].

1.4 Colon microbiota and Glycan Metabolism

It has become evident that in our pursuit to understand human nutrition and gastrointestinal health that we need to investigate the role of the colon microbiota in relation to our diet. The colon microbiota's influence on our physiology and metabolism was first observed with the investigation of the relationship between the colon microbiota and obesity. Analysis of lean and obese mice colon microbiota revealed significant differences in composition as well as diversity[24,25]. This was further supported with the analysis of the colon microbiota in lean and obese twins[26]. These research studies demonstrated that the metabolism of the colon microbiota can strongly impact our physiology but also is a variable that is interconnected with our intestinal health.

Humans ingest a broad range of carbohydrates, which comprise a large portion of our diet[27]. Our genome encodes an estimated 97 glycoside hydrolases, but only 8 are associated with hydrolyzing dietary carbohydrates. Our ability to digest carbohydrates is limited to the metabolism of lactose, starch and sucrose[28]. A consequence of this limitation is that large portions of ingested sugars enter the colon and are available for microbial fermentation. Our poorly evolved capability to fully utilize nutrient rich substrates is illustrated by the estimation that 20% of ingested dietary carbohydrates enter

the colon[29,30]. Available carbohydrates in the colon include resistant starches, complex polysaccharides and simple sugars that humans cannot absorb in our small intestine (Figure 1.1)[13]. The specific glycan composition of our diet as well as the subsequent availability of these substrates in the colon will influence microbial glycan metabolism. This contributes to the observed symbiosis that exists between humans and the colon microbiota. Humans provide a protected niche for colonization with a high nutrient composition and microbes provide metabolic capabilities that humans are otherwise lacking[31]. This relationship has ultimately allowed us to adapt to changes in our diet and effectively be able to use these substrates despite our limited enzymatic abilities.

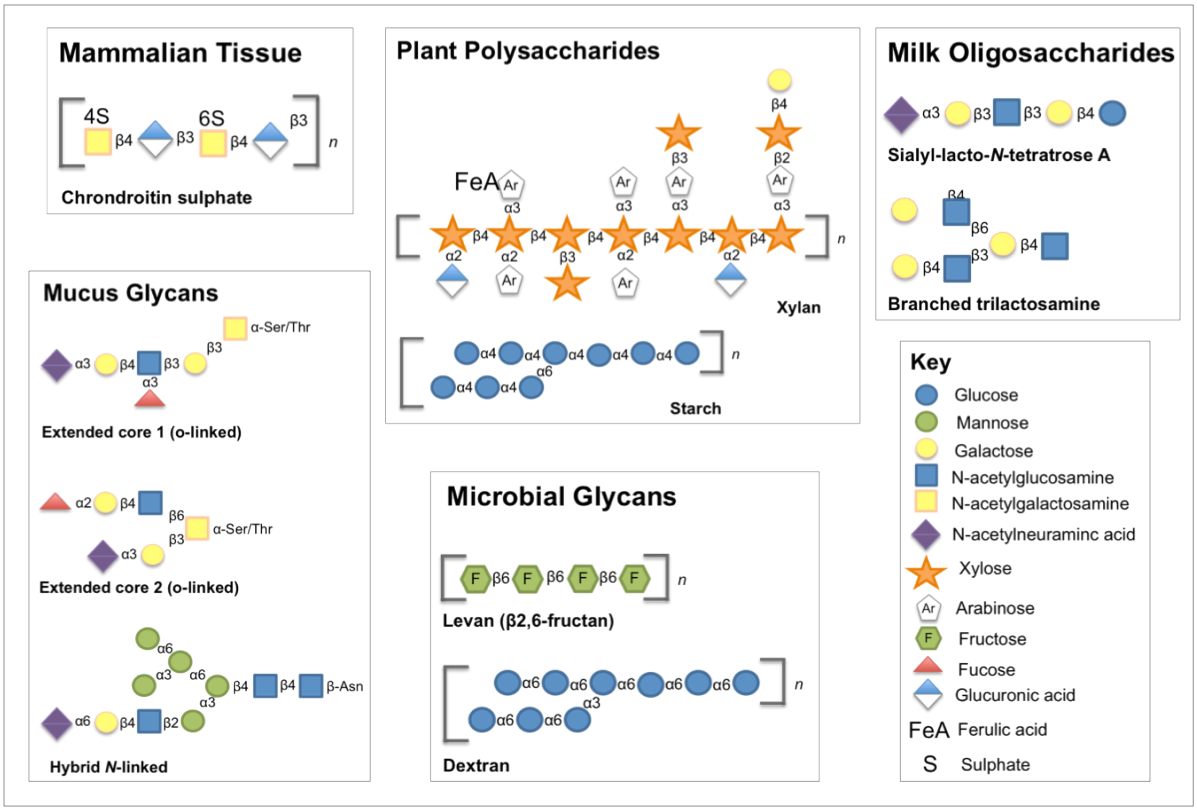


Figure 1.1 Simple and complex carbohydrates available in the colon for microbial fermentation as proposed by Koropatkin *et al*, 2012[32].

1.5 Starch Digestion

Starch is one of the main sources of glucose in the human diet, contributing a significant portion of our dietary food intake. Starch is a complex polymer consisting of repeating units of glucose linked by α-(1->4) glycosidic linkages and has branching points with α-(1->6) glycosidic linkages. The primary arrangement of starch is divided into two forms amylose and amylopectin (Figure 1.2). Amylose is a linear form of starch composed of only α-(1->4) glycosidic linkages. Amylopectin on the other hand is generally short in length and is highly branched in nature[33].

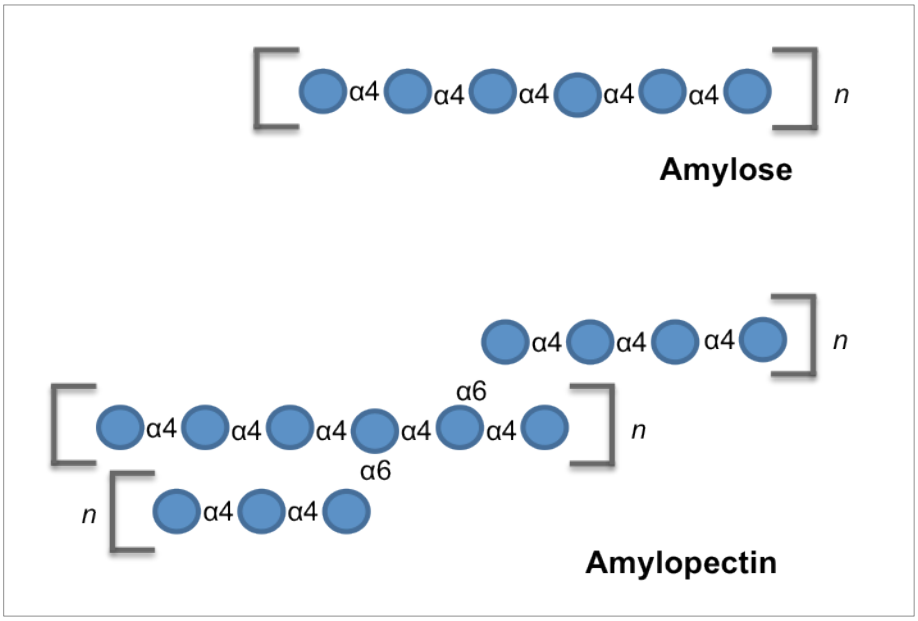


Figure 1.2 α-glucans of starch. The main forms of starch include amylose and amylopectin.

The release of glucose from starch involves a number of human enzymes. Initially starch is solubilized by salivary and pancreatic amylases resulting in the linearization of starch producing both amylose and amylopectin[34]. These substrates enter the small intestine and through the mechanisms of mucosal glycosidic enzymes glucose is liberated[35]. Mucosal glycosidic enzymes are identified as maltase glucoamylase (MGAM) and sucrase-isomaltase (SI)[36–39]. These enzymes each contain two independently functional domains (N-terminal and C-terminal) and are believed to have evolved from a single ancestral

gene[36,40]. It has been reported that the source of the starch polymer (i.e. maize verse potatoes) can influence the rate in which this substrate is hydrolyzed in the small intestine. The number of branches, the specific branch point and the number of glucose residues between branching are all factors that can influence rate of digestion by MGAM and SI in the small intestine[41,42]. Subsequently, this will influence the specific starch based substrates that enter the colon and become available for the colon microbiome[,31].

The majority of starch digestion is thought to occur in the small intestine; however, recent research has demonstrated that a significant portion of starch-derived glycans enter the colon and are available for microbial fermentation. These starch substrates are classified as resistant starches and are deemed as starch polysaccharides that are resistant to small-intestinal degradation[45]. Resistant starch can be classified into four groups; RS1, RS2, RS3, RS4[45]. RS1 includes starch substrates that are not physically accessible. RS2 includes substrates that α-amylases have difficulty accessing, leading to resistant granules. R3 includes certain cooked and cooled starches. R4 includes chemically modified starches[46]. As a result of starch being such a large portion of our diet, understanding its interaction with the colon microbiota will provide strong evidence regarding the relationship between our diet and this population.

1.6 Starch Utilization System

B.thetaiotaomicron's genes are organized in clusters, each co-regulated by a common substrate. This type of gene organization in *B.thetaiotaomicron* was first observed in the Starch Utilization System (Sus). This system was the first of its kind investigated in relation to glycan utilization. This well characterized metabolic system in *Bacteroides thetaiotaomicron* consists of eight genes (*susRABCDEFG*) (Figure 1.3). Outer membrane components are involved in starch glycan hydrolysis (*susG*), starch glycan binding (*susDEF)* and starch transportation into the cell (*susC*). There are two enzymatic components that are capable of hydrolyzing small starch oligosaccharides (*susAB*), and a gene regulator, which activates transcription of the system in response to starch molecules (*susR*)[47–49]. In vitro, in vivo and structural studies have provided evidence that the product of these genes allows this organism to grow efficiently on starch-derived glycans such as amylose, amylopectin and pullulan[21,49–52]. These studies provided evidence that *Bacteroides thetaiotaomicron* contributes to glycan utilization in the colon, specifically starch derived substrates and provided insight into the influence our diet has on such a metabolism.

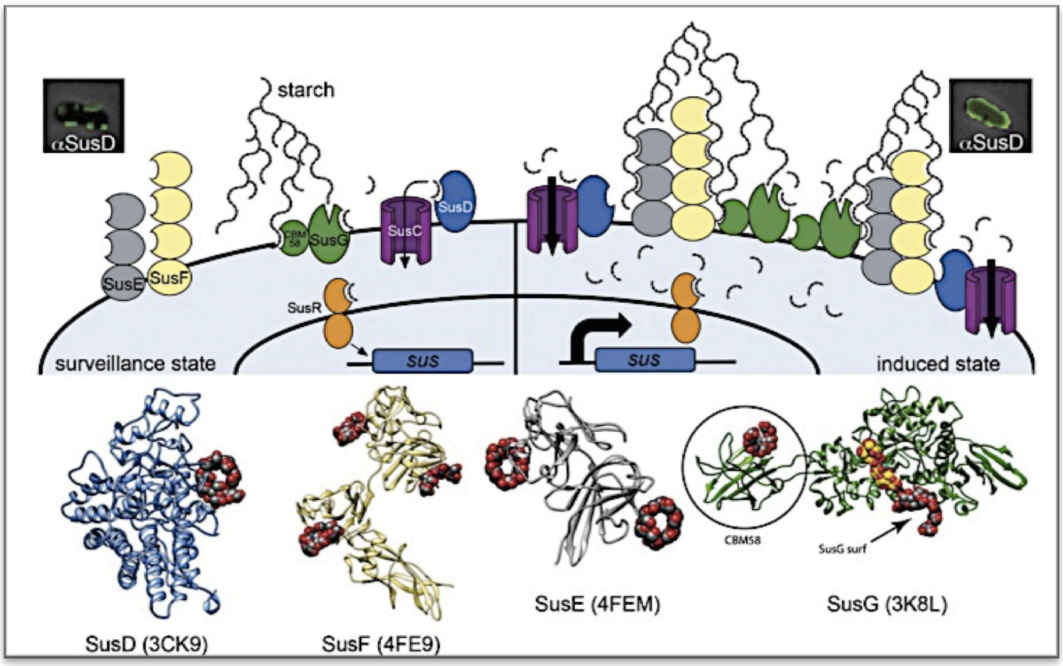


Figure 1.3 Proposed model of the starch utilization system by Cameron *et al,* 2014[50]. Four outer membrane proteins are involved in this system; SusD, SusE, SusF and SusG. SusE and SusF are multi-domain carbohydrate binding proteins. SusG contains a carbohydrate binding domain and catalytic amylase domain. SusD is a carbohydrate binding protein and in cooperation with SusC transport starch glycans into the cell. SusR senses starch glycans in the periplasm and increases expression of the *sus* locus. This increased expression of *sus* genes results in *B.thetaiotaomicron* being in an induced state specific for starch glycan utilization.

1.7 Polysaccharide Utilization Loci

In addition to the starch utilization system, *Bacteroides thetaiotaomicron* has additional structures that contribute to similar mechanisms with the overall goal of acquiring and utilizing dietary and host-derived glycans[53]. Sequencing of the *B.thetaiotaomicron* genome revealed an overall pattern of genetic clusters similar to that of the starch utilization system[19,54]. These 'sus-like' systems are classified as Polysaccharide Utilization Loci and constitute a significant portion of this bacterium's genetic material[55,56]. It has been reported that one-fifth of its genome is dedicated to these systems and it has been estimated that it as over 80 individual loci, reflecting a significant metabolic dedication to this overall function[32]. Similar to the starch utilization system, these gene clusters contain multiple glycoside hydrolases; a TonB-dependent, β-barrel-type outer membrane protein (SusC-homolog), an outer membrane glycan binding protein (SusD-homolog) and a transcriptional regulator[57,58]. Some PULs are regulated by specific transcription factors that belong to two-component systems or extracytoplasmic function-σ factor/anti- σ factor pairs[58]. The regulation of these gene systems is dependent on the specific substrates they have evolved to utilize. SusC-SusD-homologous pairs are observed to be functionally similar and as a result demonstrate sequence and structural homology. Differences in these pairs are believed to have evolved from acquired substrate preferences. The observed similarities are believed to have occurred from gene duplication and as a result it is hypothesized that some level of cross-talk may occur[59]. In general, it is believed though that the components of a PUL are involved in the utilization of a target polysaccharide that is specific to a PUL with all elements working cooperatively[60,61]. This metabolic organization and regulation is observed cross species in the colon microbiota and reflects a functional necessity in order to competitively inhabit this ecosystem[62].

1.8 Glycoside Hydrolase Family 31 α-glucosidases

Family GH31 is a group of glycoside hydrolases found in all three domains of life and can be observed in a variety of environments from volcano vents to the human digestive system. This family of enzymes encompasses a range of diverse functions that include α-glucosidase (EC 3.2.1.20); α-galactosidase (EC 3.2.1.22); α-mannosidase (EC 3.2.1.24); α-1,3-glucosidase (EC 3.2.1.84); sucrase-isomaltase (EC 3.2.1.48) (EC 3.2.1.10); α-xylosidase (EC 3.2.1.177); α-glucan lyase (EC 4.2.2.13); isomaltosyltransferase (EC 2.4.1.-) and oligosaccharide α-1,4-glucosyltransferase (EC 2.4.1.161). Family GH31 is of interest in relation to the human diet, specifically α-glucosidases, which are important molecules in primary metabolism. α-glucosidases are responsible for the hydrolysis of dietary starch molecules resulting in the release of free-glucose. In humans, two family GH31 α-glucosidases are known to play a key role in the final stages of starch digestion and are identified as maltase glucoamylase (MGAM) and sucrase-isomaltase (SI)[36,38,39,63,64]. These enzymes each contain two independently functional domains (N-terminal and C-terminal) and are believed to have evolved from a single ancestral gene[36,40]. The source of this environmental pressure is believed to be related to the structural complexity of its target substrate, starch[39,42,65]. Family GH31 α-glucosidases have also been identified in species found in the colon microbiota and are believed to contribute to starch utilization in the colon[63]. Family GH31 α-glucosidases have been shown to demonstrate differences in substrate preferences and reveal the ability of this colon bacterium to utilize a range of diverse starch glycans[66].

1.9 Experimental Approach

This study is focused on two perspectives of starch digestion by the human colon bacterium *Bacteroides thetaiotaomicron.* The first approach is to investigate the role a specific PUL, experimentally identified as PUL 55 and its role in starch utilization (Figure 1.4)[53,57].

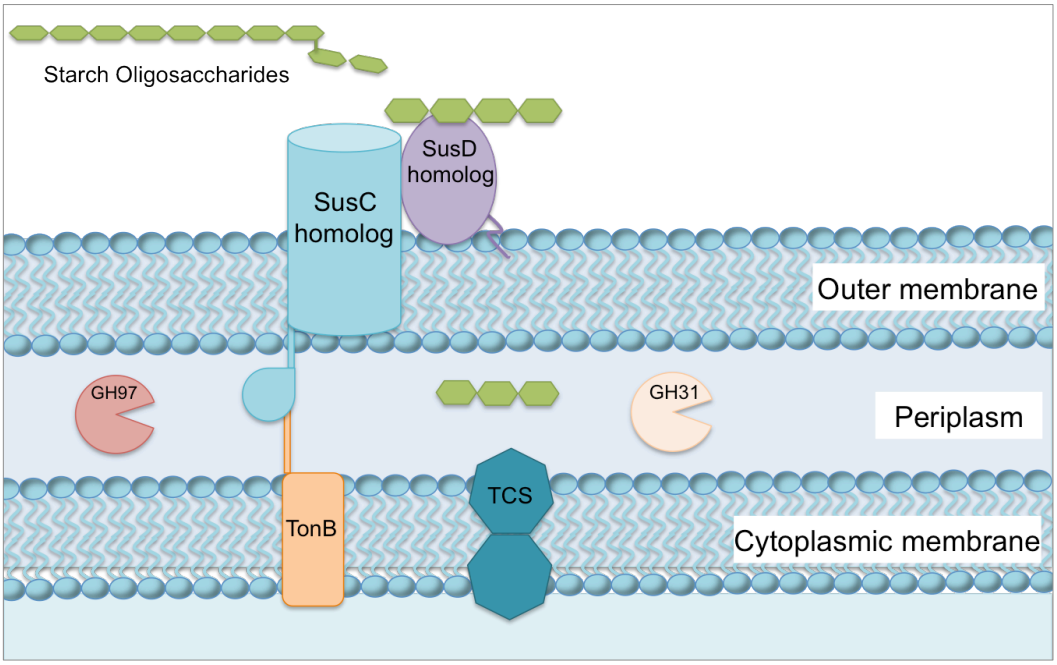


Figure 1.4 Starch utilization components from the polysaccharide utilization locus 55. PUL components that are predicted to be related in starch glycan utilization include an outer-membrane SusC/SusD-homolog pair (blue and purple respectively), periplasmic glycoside hydrolases from Families GH97 (red) and GH31 (beige), and lastly a Two-Component System (TCS, aqua). Starch glycans are predicted to bind to the SusD-homolog and through the cooperation with the SusC-homolog, a TonB-dependent transporter, are transported into the cell. Periplasmic glycoside hydrolases act on the starch glycans releasing glucose that can be transported into the cytoplasm. Periplasmic starch glycans are predicted to bind to the periplasmic domain of the TCS leading to an increased expression of the PUL 55 locus.

Table 1.1 Polysaccharide utilization locus 55 components related to starch utilization. Genes from PUL 55 locus predicted to be involved in starch utilization are identified in the following table.

Gene	NCBI Accession	Predicted Fucntion
BT_3294	AAO78400	Family GH 97 α-glucosidase
BT_3296	AAO78402	SusD Homolog
BT_3297	AAO78403	SusC Homolog
BT_3299	AAO78405	Family GH 31 α-glucosidase
BT_3302	AAO78408	Two-Component System

It is hypothesized for this work that some PULs have evolved specifically to process residual starch-based substrates undigested by the human intestinal enzymes, and that this is reflected in the regulation of those PULs and the characteristics of their enzymes

The second approach is to investigate the structural comparison between two family GH31 α-glucosidases and the differential roles in starch hydrolysis in *Bacteroides thetaiotaomicron.* This bacterium is predicted to have three family GH31 α-glucosidases as part of its genome and provides evidence that this organism has the ability to contribute to the utilization starch by means of more than one system. Our study will focus on investigating the structural features of two of the three predicted α-glucosidases. This analysis will provide insight into the role these enzymes play in starch hydrolysis in *Bacteroides thetaiotaomicron.* Comparative analysis will also demonstrate whether these enzymes provide discrete functions within this organism in relation to their prediction function. Evolutionary divergence analysis will determine if changes have occurred within glycoside hydrolase family 31 and will also provide insight into a rationale for the presence of multiple α-glucosidases as part of this bacterium's genome.

Chapter 2
Characterization of Starch Glycan Utilization by *Bacteroides thetaiotaomicron*

2.1 Chapter Overview

Here, we assess specific components of an identified polysaccharide utilization locus that is predicted to have a role in the utilization of starch-based glycans[53]. Overlapping predicted substrate preferences and the sheer number of PULs identified as part of *Bacteroides thetaiotaomicron*'s genome leads to ambiguity as to whether these PULs have a necessary discrete function or whether they represent a form of genetic redundancy in a bacterial species[53]. Investigating how these systems function independently and/or in cooperation with one another is important for understanding how *B.thetaiotaomicron* functions in the human colon but also how the colon microbiome utilizes dietary substrates. This chapter is focused on the investigation of the polysaccharide utilization locus 55 from the bacterium *Bacteroides thetaiotaomicron*, introduced briefly in Chapter 1, and its contribution specifically to starch utilization in the human colon. Additionally, this chapter discusses the structural investigation of two family GH31 α-glucosidases, BT_3299 (member of PUL 55) and BT_0339. The goal of this work is to assess the varied hydrolytic ability of this bacterium in relation to multiple α-glucosidases. Here we hypothesize that this bacterium has multiple systems, in addition to the Starch Utilization System, which are capable of extracting glucose from starch-based glycans.

2.2 Contributions

All work presented in this chapter was performed by the candidate. Sections of this chapter are modified or taken verbatim from the text written by the candidate in the following publication:

Chaudet, Marcia M. and Rose, David R. 2016. Alternative starch utilization system from the human gut bacterium *Bacteroides thetaiotaomicron.* Canadian Journal of Biochemistry and Cell Biology. DOI: 10.1139/bcb-2016-0002.

2.3 Material and Methods

2.3.1 Expression and Purification of a SusD-homolog BT_3296

SusD-homologs have been shown to contain an N-terminal region associated with the cell outer membrane. For ease of purification, this region was not included in the design of the construct and therefore, the first 25 amino acids were omitted. The gene encoding BT_3296 was amplified by the polymerase chain reaction using *Bacteroides thetaiotaomicron* genomic DNA (GenBank accession No. NC_004663.1). Primers were designed with a 5' *BamHI* restriction site and a 3' *XhoI* restriction site for directional cloning purposes (Table 2.3.1). The BT_3296 gene was cloned into pET29a vector by double digest and ligation using T4 ligase as described in the manufacturer's instructions.

Table 2.3.1 Primers for BT_3296. Sequence of primers used in the amplification of the BT_3296 gene for the plasmid pET29a. Restriction sites *BamHI* and *XhoI* were inserted for the forward primer BT_3296-F1.1 and the reverse primer BT_3296-R1.1 respectively as indicated by underlining.

Primer Name	Primer Sequence
BT_3296-F1.1	5' CGAGAGGATCCCTAAGTGTAGAACGTTTCTTCAACG 3'
BT_3296-R1.1	5' CGAGACTCGAGGAAAAAACAAACTTAAAATATATTTT 3'

Constructs were transformed into Nova Blue competent cells and screened on selective media. BT_3296 was transformed into BL21 (DE3) pLysS expression strain. Cells were grown at 37°C until an optical density of 0.6 at 595 nm was reached, followed by induction by 0.1 mM of isopropyl β-1-thiogalactopyranoside (IPTG). The culture was then incubated at 14°C overnight with shaking.

BT_3296 cells were lysed by sonication and re-suspended in lysis buffer (20 mM NaH_2PO_4, 300 mM NaCl and 10 mM imidazole). Cell lysate was then batch-bound overnight with Ni^{+}-NTA resin at 4°C with rotation. Affinity purification was completed

at 4°C by gravity flow. The column was washed twice with wash buffer (20 mM NaH_2PO_4, 300 mM NaCl and 25 mM imidazole) and the protein of interest was eluted from the column by an imidazole gradient (20 mM NaH_2PO_4, 300 mM NaCl and 50 mM to 500 mM imidazole). The buffer was subsequently exchanged into 20 mM MES pH 6.5. Protein sample was injected onto an Uno Q Ion Exchange Column (Bio-Rad) that was equilibrated in 20 mM MES pH 6.5. Elution of protein of interest was completed by a NaCl linear gradient. Fractions containing the protein of interest were concentrated and subsequently buffer exchanged into 20 mM MES pH 6.5.

Mass spectrometry analysis of the purified protein BT_3296 was completed in order to ensure the target protein was accurate. BT_3296 was concentrated to a final volume of 100µL at a concentration of 10µM using Amicon centrifugation filter units. Buffer exchange was completed for the protein to be stored in diH20. The sample was sent to Aims Mass Spectrometry Laboratory at the University of Toronto for analysis.

Selenomethionine prepared sample of BT_3296 was expressed and purified under the same conditions as described above with the exception of following the protocol of methionine inhibition pathway[68]. Mass spectrometry analysis was completed for selenomethionine substituted BT_3296 to ensure an acceptable level of incorporation. Samples were prepared in the same manner as described above for the native form of BT_3296.

2.3.2 Binding Interaction Observed by Fluorescence Spectroscopy

All measurements for fluorescence binding were completed using a Horiba Fluorolog FL-311. Emission scans were completed using purified BT_3296 (concentration was 12 µM) that was incubated with different glycans in 20 mM MES, pH 6.5, at 37°C prior to measurement. Excitation wavelength was set to 280 nm with emission intensities set to be collected at a wavelength range of 295-450 nm. The slit widths of the emission and excitation were set to 0.75 nm, integration was set for 0.2 seconds and experiments were run at 37°C. Three scans were collected and averaged to obtain an overall value. Reactions were completed in triplicate to ensure reproducibility. Substrates assessed

included maltose (5 mM), maltotriose (5 mM), maltoheptaose (5 mM), pullulan (10 mg/ml), β-cyclodextrin (10 mM) and amylopectin was ~0.2%. A negative control included the same conditions with no sugar present in the reaction.

2.3.3 Crystallization Screening of BT_3296-Se

Selenomethionine prepared sample of BT_3296 was sent for high-throughput crystallization screening at the Hauptman-Woodward Medical Research Institute. BT_3296-Se was at a concentration of 10 mg/ml and buffer exchanged into diH_20 prior to crystallization. 1536 microbatch-under-oil crystallization screening experiments were initiated using the soluble screen cocktails[69]. Experiments were stored at 23°C with observations completed at day 1, week 1, 2, 3, 4 and 6. All conditions were recorded with colour-photographed images at each time point. Images were are also assessed with a Formulatrix Rock Imager 1000 SONICC UV-TPEF, which will assist in the interpretation of crystals.

Optimizations of initial crystallization hits were completed in-house using hanging-drop vapor diffusion. Conditions were set-up in equal drop ratio between protein and crystallization solution.

2.3.4 Expression and purification of BT_3299-Se and BT_0339-Se

Expression and purification of native BT_3299 and BT_0339 was previously optimized and reported[66]. These conditions were then used for the expression of the selenomethionine protein following the protocol of methionine inhibition pathway[68]. Selenomethionine incorporation into the protein was confirmed by mass spectrometry, with the expected observed increase in molecular mass.

2.3.5 Crystallization and Data Collection of BT_3299 and BT_0339

Crystallization conditions were completed using the sitting drop method of vapor diffusion. BT_3299-SeMet and BT_0339-SeMet crystals were grown at room temperature using 7 mg/ml of protein. BT_3299 and BT_0339 were crystallized in 0.05 M Ammonium Sulfate 0.05 M Bis-Tris:HCl pH 6.5 30% (w/v) pentaerythritol ethoxylate

(15/4 EO/OH) and 0.1 M sodium formate pH 7.0, 12% w/v polyethylene glycol 3,350. Quality of crystals was improved by seeding crushed small crystals into mother liquor and the addition of the additives 0.1 M phenol and 0.1 M propanol for BT_3299 and BT-0339 respectively. Crystals were cryoprotected using Mitogen LV-CryoOil and flash frozen prior to data collection. Data was collected at the Canadian Light Source using the CLSI Beamline 08ID-1and a Rayonix MX-300 detector at 100K. X-ray data were processed and scaled with HKL2000[70,71]. It was observed that both data sets had R-merge values that are higher than normal (BT_3299 at 40.1 and BT_0339 at 31.9). After data collection, it was noted by the CLS operator that the shutter was malfunctioning and sporadically not opening completely. This resulted in poor statistics for the datasets.

2.3.6 X-ray Structure Determination of BT_3299 and BT_0339

The structure of BT_3299 and BT_0339 was determined by SAD-MR phasing from x-ray data collected from selenomethionine protein crystals. The program MRage from Phenix software was used to obtain preliminary phases for molecular replacement using five homologous models with an average 30% sequence identity. This preliminary model was then used as a partial model in SAD-MR using AutoSol. The location of the Se atoms was then used as 'heavy atom' sites in Phaser, resulting in improved phases. AutoBuild was then used to generate the model and followed by several rounds of refinement in phenix.refine[72].

2.3.7 Molecular Modeling of BT_3299 and BT_0339

Molecular modeling of substrates with BT_3299 structure was completed using Coot[73]. The homologous structure of N-terminal maltase-glucoamylase (PDB: 3L4W) was used in comparison with BT_3299 to model in predicted bound substrates. Acarbose was modeled in for comparative purposes by superimposing this substrate onto the bound acarbose in the structure 3L4W. Acarbose was also modeled into the model for BT_0339 by superimposing this structure onto BT_3299 with the modeled substrate as mentioned. Because of the difficulties in data collection (see above), care was taken to avoid over interpretation of atomic details.

2.4 Results

2.4.1 Expression and Purification of BT_3296 and BT_3296-Se

BT_3296 and BT_3296-Se were successfully expressed and purified, resulting in a considerable amount of soluble protein. Two-step chromatography purification was completed using nickel affinity and anion exchange yielding a clean, concentrated sample (Figure 2.4.1).

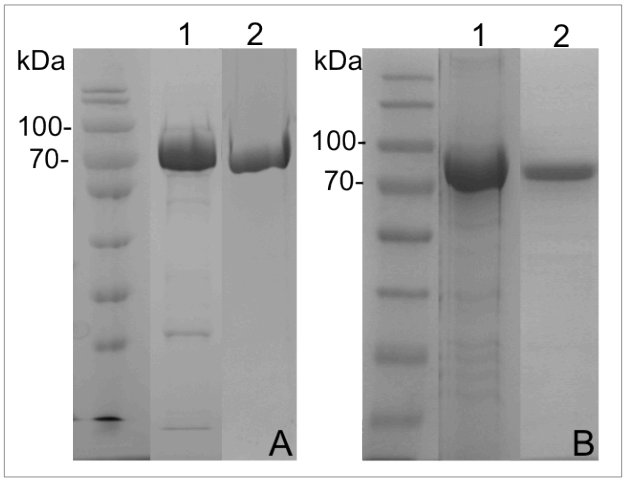


Figure 2.4.1 Denaturing SDS-PAGE gel for BT_3296 and BT_3296-Se. BT_3296 (A) was purified by nickel affinity (1) and anion exchange (2) chromatography. BT_3296-Se (B) was purified by nickel affinity (1) and anion exchange (2) chromatography**.**

2.4.2 Binding Activity of BT_3296 and Homology Modeling

The ability of BT_3296 to bind starch-derived glycans was examined in this study in order to determine if it has a similar level of activity as observed with the previously characterized SusD and SusD-homologs. Fluorescence spectrometry was used to observe the binding of BT_3296 and the preferred substrate was identified as β-cyclodextrin (Figure 2.4.2). In comparison to the additional substrates assessed (Figure 2.4.3), β-cyclodextrin demonstrated the greatest change in observed emission spectrum. The observed change in relative fluorescence by the addition of β-cyclodextrin was three times that of the BT_3296 alone. The change in the emission spectrum was observed to occur at ~300 nm, which corresponds to the emission of tyrosine. β-cyclodextrin was assessed alone in the absence of BT_3296 and was observed to not contribute to the

fluorescence intensity. This indicates that the binding of β-cyclodextrin influences the microenvironment of a specific tyrosine residue[74]. Attempt was made to identify the dissociation constant (Kd) for BT_3296 with β-cyclodextrin, but it could not be confirmed. This could be due to the energy association between BT_3296 and β-cyclodextrin being too low to specifically identify a Kd value. Alternatively, it could also be that BT_3296 prefers an alternative maltooligosacchride secondary structure similar to that of β-cyclodextrin, such as a helical form. Investigation into the glycan recognition of SusD revealed that it had weak binding for related analogues of its preferred substrate[75,76]. Based on this observation, we believe the preferred glycan BT_3296 recognizes has a similar structure and composition to that of β-cyclodextrin.

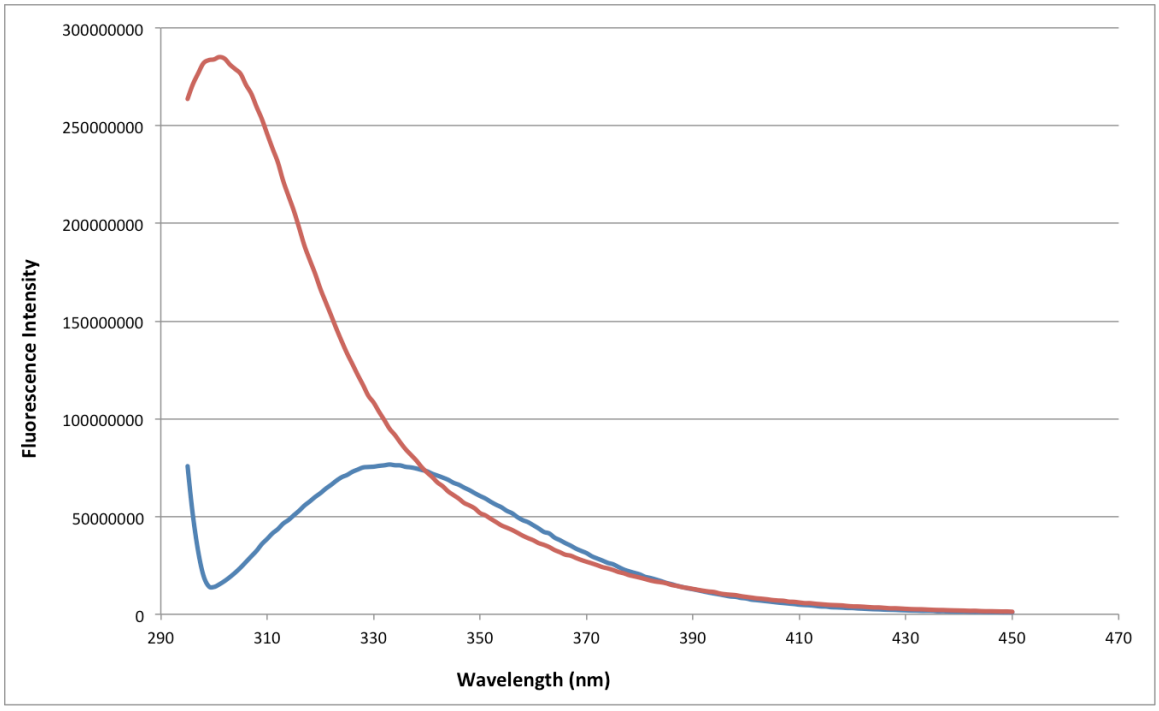


Figure 2.4.2 Emission scan of BT_3296 incubated with β-cyclodextrin. The concentration of BT_3296 (blue) was 12 μM and the concentration of β-cyclodextrin (red) was 10 mM.

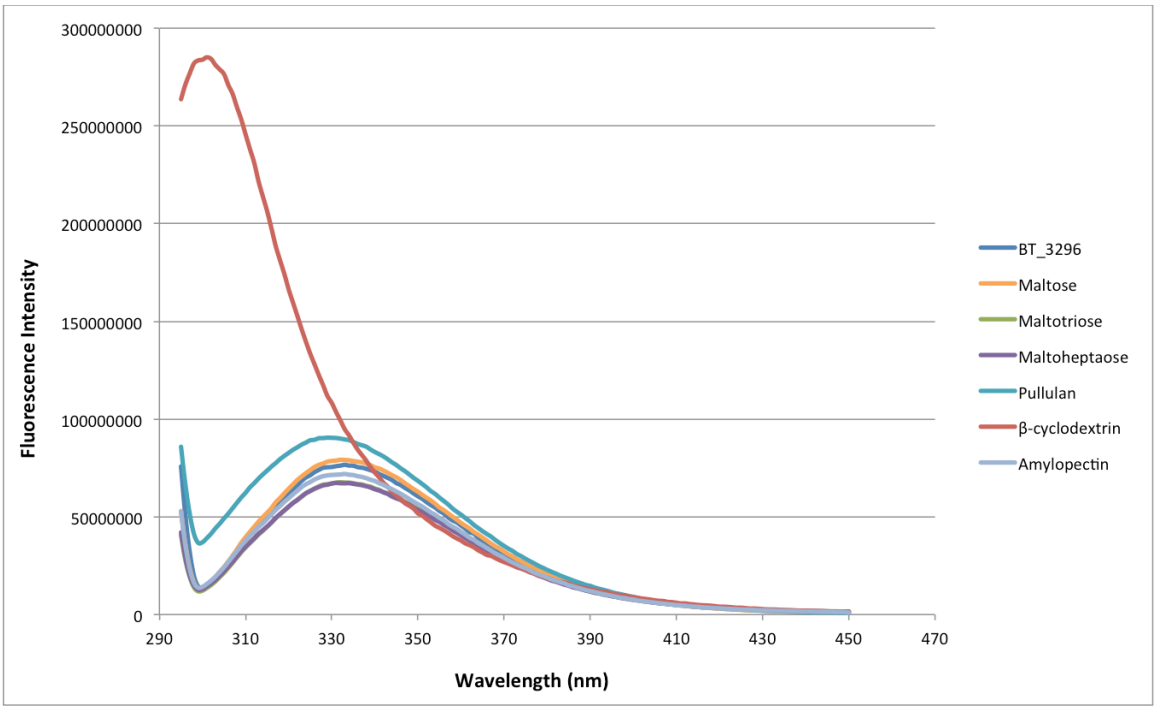


Figure 2.4.3 Emission scan of BT_3296 incubated with starch-derived sugars. The concentration of BT_3296 (navy blue) was 12 μM, where as the concentration of maltose (orange), maltotriose (green) and maltoheptaose (purple) was 5 mM. The concentration of pullulan (cyan) was ~10 mg/ml, β-cyclodextrin (red) was 10 mM and amylopectin (light blue) was ~0.2% (w/v).

In order to gain insight into the conserved structural and functional elements of SusD homologs, a bioinformatics approach was used to investigate the binding of β-cyclodextrin. A three-dimensional model of BT_3296 was built using Phyre2[77]. Using this predicted model, we were able to hypothesize the conserved residues involved in substrate binding. This predicted model was then input into the bioinformatics tool ConSurf, which allows for a conservation analysis of the predicted structure with homologous sequences[78,79]. Analysis was completed on 84 identified homologous sequences and demonstrated a range of sequence conservation as illustrated in Figure 2.4.4B. The predicted model of BT_3296 was aligned with the structure for SusD (PDB: 3CK8) and allowed for the guidance of a predicted binding site and modeling of the substrate β-cyclodextrin into the model for BT_3296 (Figure 2.4.4A). Sequence similarity of BT_3296 with SusD is ~18%. Taking into account aromatic residues, structural comparison of SusD and SusD-homolog structures with the predicted structure of BT_3296 we believe the conserved residue Tyr335 is one of the key residues involved in binding β-cyclodextrin and is responsible for the observed change in fluoresce.

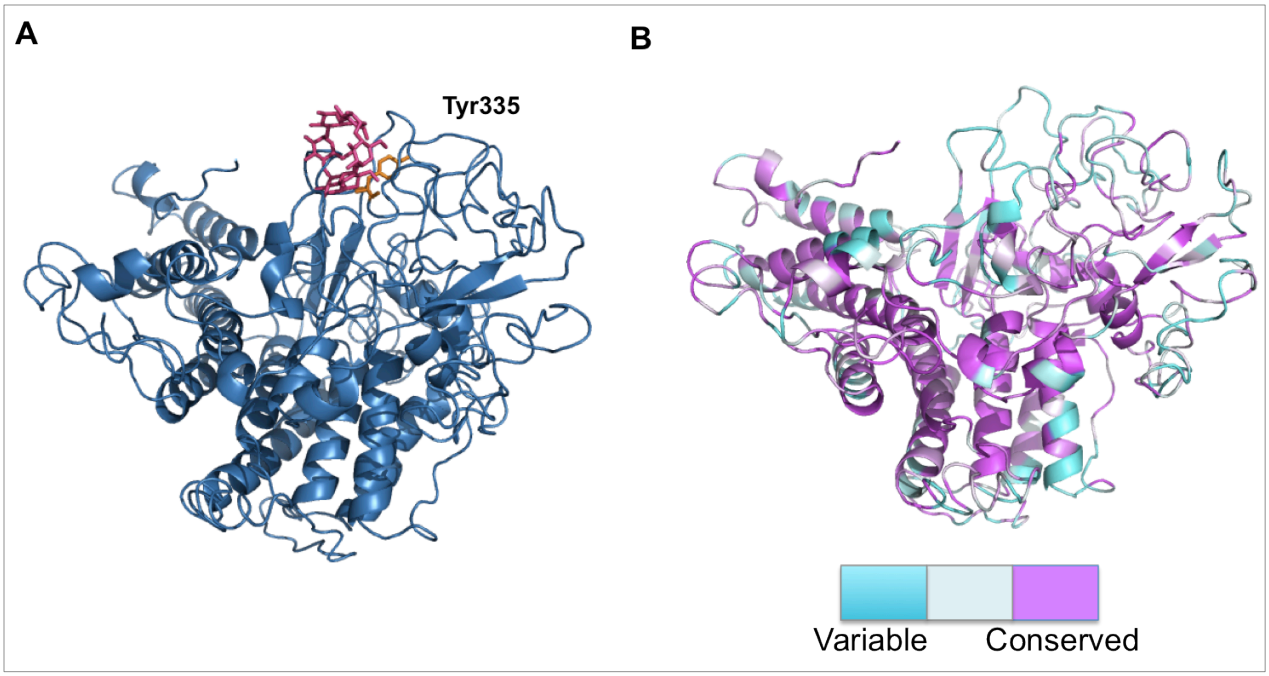


Figure 2.4.4 Predicted model of BT_3296. Predicted model of BT_3296 with the modeled substrate of β-cyclodextrin (A). Tyr335 is believed to be one of the main residues involved in binding the substrate β-cyclodextrin. Overall sequence based homology for BT_3296 in comparison to identified SusD-homologs sequences in Consurf (B).

2.4.3 Crystallization Screening of BT_3296-Se

A number of conditions were observed to have preliminary crystals during high throughput screening (Figure 2.4.5 A-C). In house screening was completed using the preliminary conditions as a starting point for optimization using the hanging-drop vapor diffusion method. BT_3296-Se crystals were grown at room temperature using 10 mg/ml of protein. BT_3296-Se crystallized in 2% 1,4 Dioxane, 0.1M Tris pH 8.0 and 15% (w/v) polyethylene glycol 3,350 (Figure 2.4.5D).

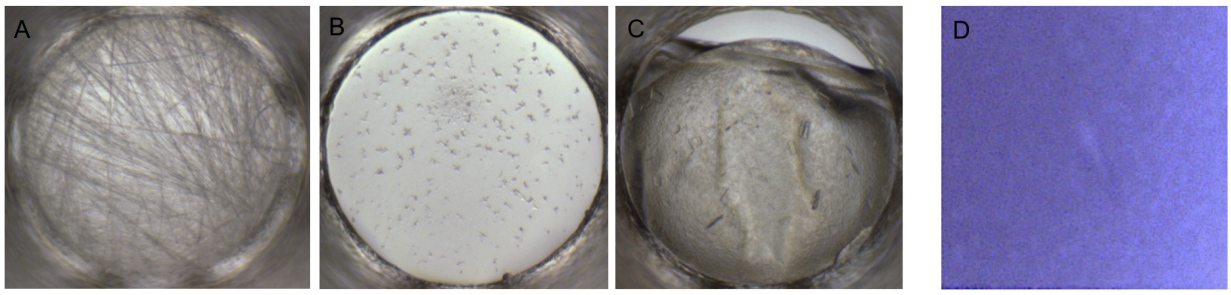


Figure 2.4.5 Crystallization of BT_3296-Se. Initial crystallization hits for BT_3296-Se from high throughput screening (A-C). Optimized in-house screening by handing-drop vapor diffusion (D).

Optimization and replication of the crystallization condition for BT_3296-Se were continued by another student.

2.4.4 Crystallization of Family GH31 α-glucosidases

BT_3299 and BT_0339 were successfully crystallized (Figure 2.4.6).

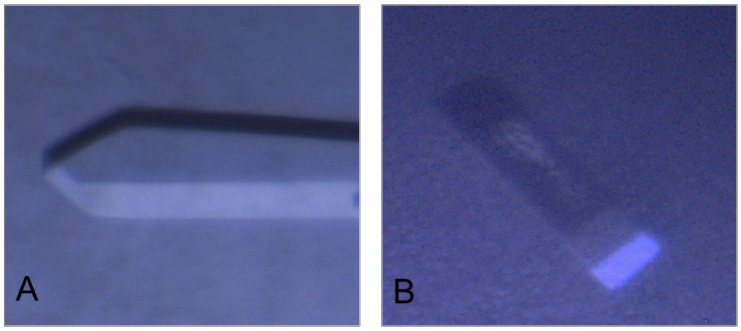


Figure 2.4.6 Crystallization of family GH31 α-glucosidases. Crystallization of BT_3299-Se (A) and BT_0339-Se (B).

2.4.5 Overall Structure of BT_3299 and BT_0339

The crystal structure of the α-glucosidase BT_3299 was successfully solved to 2.7Å using x-ray crystallography (PDB ID: 5DJW)[80]. BT_3299 was expressed as a selenomethionine substituted protein and through combined methods of molecular replacement and single wavelength anomalous dispersion we were successfully able to obtain phases. Assessment of the crystal packing revealed two molecules were present in the asymmetric unit (chain A and B). Comparison of the two monomeric chains reveals consistent regions of unresolved electron density in both the N-terminal and C-terminal ends, suggesting that these regions are of dynamic disorder. Data collection and refinement statistics can be found in Table 2.4.1.

While not part of the same PUL, an additional family GH31 α-glucosidase structure, BT_0339, was solved with the goal of providing additional comparative insight into this family of enzymes from *Bacteroides thetaiotaomicron*. The structure of BT_0339 was successfully determined from selenomethionine-substituted protein from x-ray data collected to 2.6 Å (PDB ID: 5F7C)[80]. Similarly to BT_3299, a combined method approach for phasing included molecular replacement and single-anomalous dispersion. Analysis of the crystal packing indicated four molecules in the asymmetric unit. Again, a common region is unresolved in all four chains. Data collection and refinement statistics can be found in Table 2.4.1.

The commonality of an unresolved region is believed to reflect a dynamic loop that is previously observed in related prokaryote models in family GH31 α-glucosidases. Crystal structures of α-glucosidases from *Blausti obeum* (formally *Ruminnococcus obeum*) and

Sulfolobus solfataricus were both observed to have helical bundles outside of the catalytic site and is believed to be the unresolved region absent in both BT_3299 and BT_0339 models[67,81].

Table 2.4.1 Data collection and refinement statistics for BT_3299 and BT_0339 structures.

Parameter	BT_3299	BT_0339
Data Collection		
Space Group	P1211	P1211
Unit Cell Dimensions		
a (Å)	89.26	82.34
b (Å)	74.86	74.99
c (Å)	94.42	232.86
α (°)	90	90
β (°)	95.69	95.14
γ (°)	90	90
mol/ASU	2	4
Wavelength (Å)	0.979	0.979
Resolution range (Å)	47.8-2.7 (2.73-2.66)	48.81-2.6 (2.83-2.76)
Unique reflections	34303	87516
I/σ	4.79	6.67
Completeness (%)	100 (99.9)	99.98 (99.9)
Redundancy	7.6	7.6
Rmerge	0.4	0.31
Refinement		
Rwork (%)	32.99	30.25
Rfree (%)	38.97	34.91
No. of atoms	9118	19996
Solvent atoms	23	67
Deviations from ideal (RMSD)		
Bond lengths (Å)	0.02	0.007
Bond angles (°)	2.18	1.05
Ramachandran plot		
Most favoured (%)	80	86
Outliers (%)	8.4	4.1
PDB ID	5DJW	5F7C

2.4.6 Proposed Binding of Maltooligosaccharides

Overall, BT_0339 and BT_3299 are very similar in structure (Figure 2.4.7) despite the sequence similarity being quite low at about ~28%. Analysis of the active site reveals key differences in amino acids that are believed to interact with bound substrates (Figure 2.4.8). To gain insight into the binding mechanism and specific residues involved in hydrolysis, the substrate acarbose was modeled into both structures. The catalytic residues are expected to be conserved in this family and are confirmed by the structural data[82]. The catalytic nucleophile is identified as Asp321 and the acid/base catalyst is identified as Asp401in BT_3299; in BT_0339 they are identified as Asp460 and Asp523 respectively. However, the specific residues believed to be involved in binding substrates differ between these models. The structure of BT_3299 has a number of aromatic side chains (Phe183 and Trp285) in the -1 sugar subsite, which we suggest allows for accommodation of short, branched maltooligosaccharides. However, BT_0339 is observed to have fewer aromatic resides in this -1 location, suggesting that this enzyme's active site preferentially binds small, linear maltooligosaccharides.

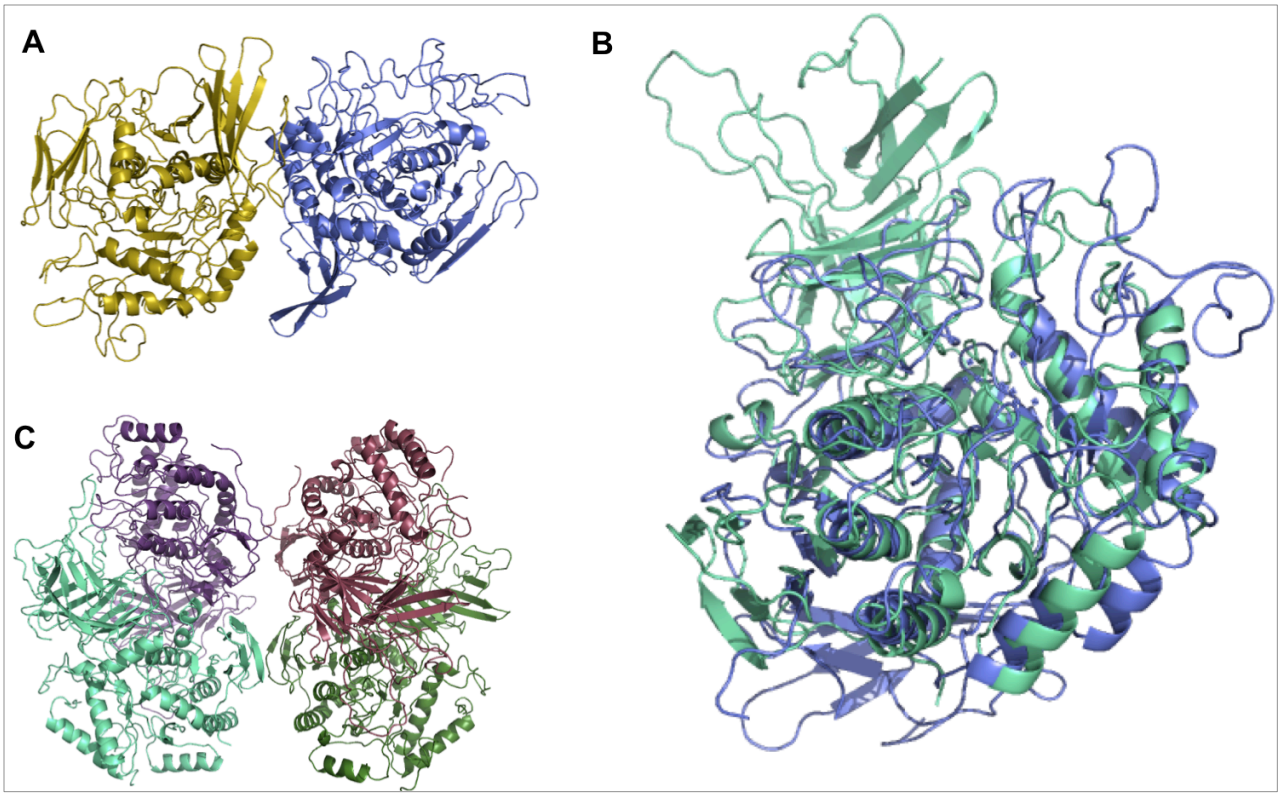


Figure 2.4.7 Crystal structures of BT_3299 and BT_0339. Arrangement of molecules in the asymmetric unit of BT_3299 (A). Alignment of structures BT_3299 (slate blue) and BT_0339 (cyan green) (B). Arrangement of molecules in the asymmetric unit of BT_0339 (C).

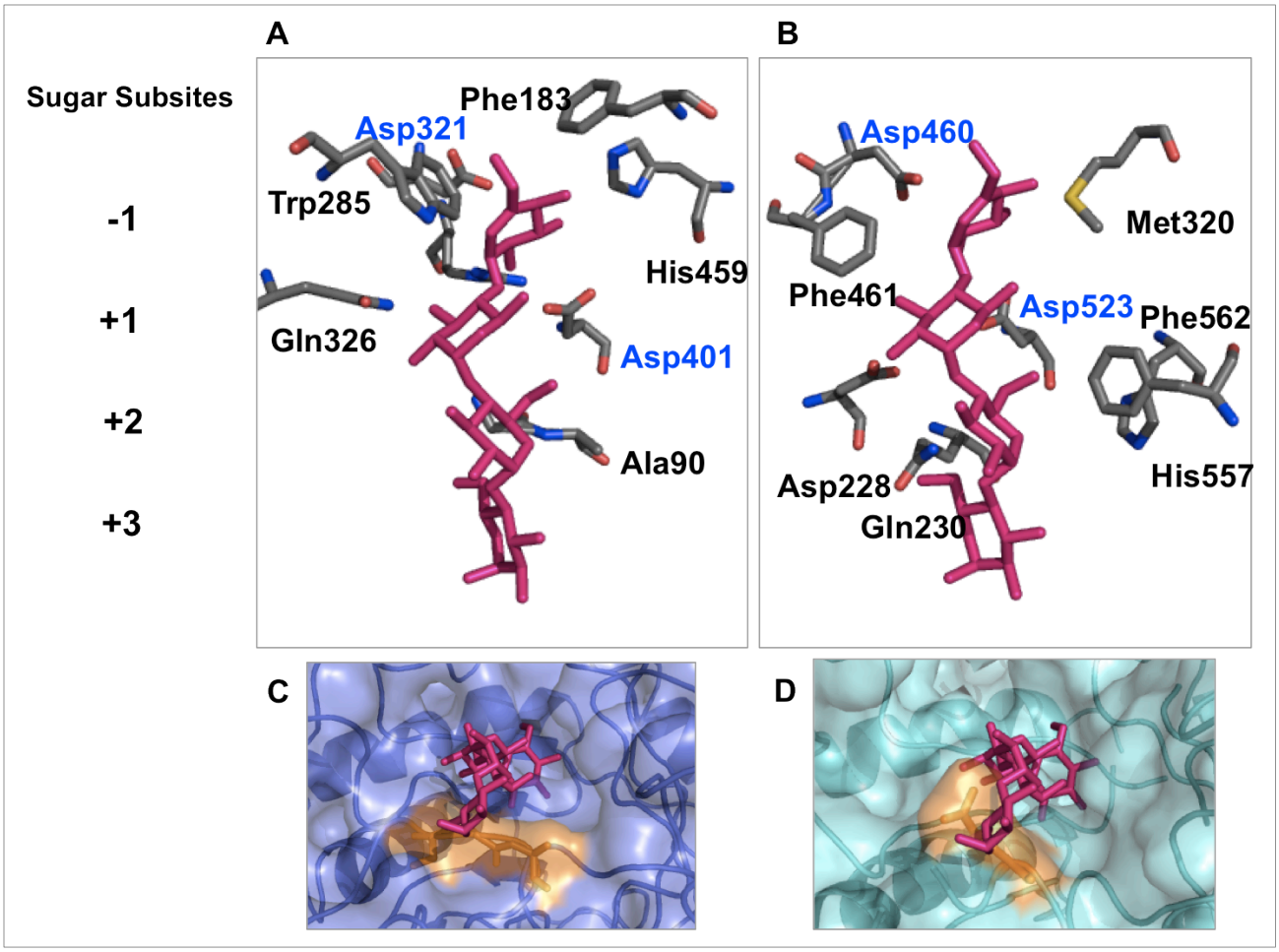


Figure 2.4.8 Comparison of BT_3299 and BT_0339 active sites with modeled acarbose. The modeled acarbose molecule (magenta) was modeled into BT_3299 and BT_0339 models by superimposing the solved structure for NT_MGAM with a bound acarbose (PDB: 3L4W). Highlighted residues believed play a role in the binding substrates in the active sites of BT_3299 (A) and BT_0339 (B). Catalytic residues are indicated by blue lettering (A,B). Surface representation of the catalytic sites of BT_3299 (C) and BT_0339 (D) with catalytic residues coloured in orange.

2.4.7 Comparison to Family GH31 α-glucosidases and SusB Structures

The overall structural architecture of BT_3299 is closely related to previously solved family GH31 α-glucosidases structures. This family of glycosidic enzymes has been reported to demonstrate an overall conservation of structural features despite there being a low sequence similarity. The active site in particular is observed to have a high level of conservation. The identified catalytic residues of the human small intestinal α-glucosidase structures (N-terminal subunit of maltase-glucoamylase, PDB: 2QMJ; C-terminal subunit of maltase-glucoamylase, PDB: 3TON; N-terminal subunit of sucrose-isomaltase, PDB: 3LPO) in comparison with BT_3299 and BT_0339 demonstrate this conservation (Figure 2.4.9A). Differences in the active site are observed for residues involved in binding substrates, which reflects reported differences in the preferred substrates for these α-glucosidases[35,37,83,84].

Structural and biochemical characterization of SusB demonstrated its preference for small α-1,4 linked substrates, having the highest activity towards maltotriose and revealing low activity on isomaltose[85]. Alternatively, both BT_3299 and BT_0339 have been shown to prefer shorter maltooligosaccharides, but have the ability to utilize both isomaltose and maltose[66]. Comparison by superimposing the structures of BT_3299 and SusB, allowed for the observation of structural difference in their active sites, which can account for the differing substrate preferences (Figure 2.4.9B,C). The active site of SusB is narrow and enclosed, accommodating linear molecules. Where as BT_3299 is wide and open, allowing it to bind short, branched substrates (composed of both α-1,4 and α-1,6 linked glucose monomers).

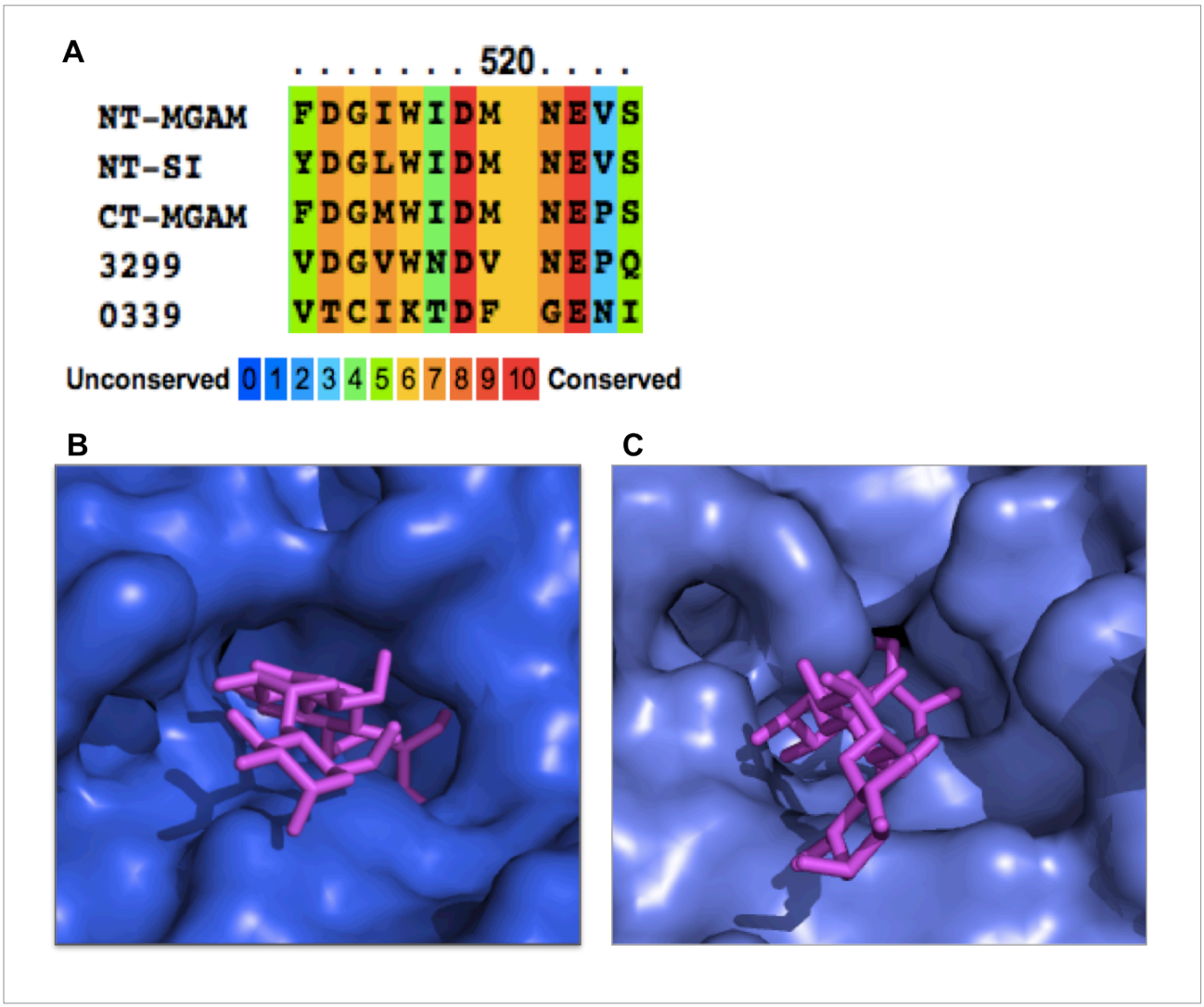


Figure 2.4.9 Comparisons of structurally solved family GH31 α-glucosidases and SusB. Multiple sequence alignment of previously solved Family GH 31 small intestinal α-glucosidases aligned with BT_3299 and BT_0339 (A). Active site comparison between SusB (B) and BT_3299 (C).

2.5 Discussion

In this study we investigated the activity of a SusD-homolog in the context of its potential contribution to starch digestion in the human colon. Our qualitative results predict that BT_3296 has the function of a carbohydrate binding protein and, like previously characterized SusD/SusD-homologs, preferentially binds cyclical oligosaccharides. Specifically, we believe the preferred substrate of the SusD-homolog BT_3296 is a resistant starch that is capable of forming secondary structures such as a cyclical form. Through predicted protein modeling the specific residue we believe is involved in binding β-cyclodextrin is a structurally conserved tyrosine, Tyr335. This residue is not identified as conserved in the multiple sequence alignment, but comparison with solved SusD-homolog structures it can be observed to be structurally conserved. It was previously identified in the structure of SusD that the residue Tyr296 is directly involved in the hydrophobic stacking with the bound β-cyclodextrin[75]. This structural homology can be observed in the predicted model for BT_3296 and all *Bacteroides thetaiotaomicron* SusD-homolog structures and supports our prediction that Tyr335 in BT_3296 represents a necessary residue in the role in binding preferred substrates. These results together with previously characterized SusD and SusD homologs support a preferential role for these molecules to bind cyclical substrates at the cell surface through the interaction of conserved aromatic side chains.

The crystal structures of BT_3299 and BT_0339 with modeled acarbose provide insight into the specific residues involved in substrate binding and hydrolysis. Structural differences observed in these enzymes highlight the flexibility this organism has for

utilizing the array of maltooligosaccharides it could encounter in the colon. It has been reported that the specific source of starch can result in differences in branching patterns[41]. As a result, this bacterium could encounter a significant range of maltooligosaccharides that can differ based on length, degree of branching and/or secondary structure[,38]. Through the mechanisms of these two enzymes, they are able to contribute to the overall process in *Bacteroides thetaiotaomicron* at effectively utilizing diverse starch-derived substrates that are available in the colon.

Comparison of BT_3299 with the crystal structure of SusB suggests that BT_3299 has the ability to hydrolyze an alternative form of maltooligosaccharides than observed for SusB. By having the ability utilize short-branched maltooligosaccharides, BT_3299 is fulfilling a role that is possibly missing from the Starch Utilization System. Based on this observation it is believed the acquiring and utilizing starch glycans does not involve Sus alone, but could involve multiple systems, such as the PUL discussed here.

2.6 Conclusion

Overall, these results provide evidence that an alternative system to the Starch Utilization System in *Bacteroides thetaiotaomicron* contributes to the acquisition and hydrolysis of starch-derived substrates. Further investigation of this PUL and understanding the additional components not addressed in this study would be beneficial and help provide a better understanding of the specific target glycan of this PUL. Continued exploration of this and other bacteria will provide insight into the colon microbiome and its contribution to human nutrition. Ultimately manipulation of this population is a promising approach to improving general gastrointestinal health and minimizing colonic related disease states.

Chapter 3
Evolution of Family 31 α-glucosidases

3.1 Chapter Overview

Here we assess structurally solved Family GH 31 α-glucosidases and infer amino acid changes that reflect the substrate specificity of these enzymes. Using bioinformatics tools we were able to observe discrete changes that are believed to have evolved at different rates within this family of enzymes, which ultimately resulted in changes that influenced functional differences between enzymes. A number of family 31 α-glucosidase structures have been solved, providing important structural information that can be related to function. These models include one from the Archaea domain, an α-glucosidase from *Sulfolobus solfataricus*[81]. Three bacterial α-glucosidases that come from organisms inhabiting the human colon; two from the gram-negative bacterium *Bacteroides thetaiotaomicron* and one from the gram-positive bacterium *Blausti obeum*[67,86]. Four structures from the eukaryote domain that include α-glucosidases from *Homo sapiens* and an α-glucosidase from the plant species *Beta vulgaris*[83,84,87,88]. Overall, these structures share a conserved three-dimensional architectural homology characterized by four domains and 2 subdomains; N-terminal β-sandwich domain, $(\beta/\alpha)_8$ barrel domain, insertion subdomain 1 and 2, and proximal C-terminal domain[88,89].

Despite this family of enzymes having a commonly observed mechanism of hydrolyzing starch glycans, it has been reported that these enzymes demonstrate distinct functions. These distinct functions can range from having the ability to hydrolyze α, 1-4 and/or α, 1-6 linkages between glucose residues or hydrolyzing short and/or long linear maltooligosaccharides. The ability of this family to accommodate a range of starch glycans is believed to result from the variations in starch structure that these enzymes are exposed to in their selective environments.

In this current study we are interested in investigating specific evolutionary changes that may have occurred over time that influence their ability to utilize starch substrates and

ultimately result in observable differences in secondary structure. Here we investigate the phylogenetic evolution of structurally solved Family 31 α-glucosidases and assess the phylogenetic relationships within this family. We assessed structural differences that have evolved within Family GH31 α-glucosidases and provided evidence of single amino acid changes that contribute to functional divergences.

3.2 Contributions

All work presented in this chapter was completed by the candidate. Dr. Laura Hug in the Department of Biology at the University of Waterloo provided guidance in selecting appropriate phylogenetic analysis tools and assistance with the analysis. Sections of this chapter are modified or taken verbatim from the text written by the candidate in the following publication:

Chaudet, M.M. and Rose, D.R. 2016. Divergent evolution for diverse substrate recognition by family 31glycoside hydrolases. Canadian Journal of Biochemistry and Cell Biology. DOI: 10.1139/bcb-2016-0022.

3.3 Methods

3.3.1 Phylogenetic Analysis of Family 31 α-glucosidase Structures

Table 3.3.1 indicates the structures that were investigated in this study. We focused our analysis on experimentally solved structures of α-glucosidases that are part of the family 31 glycoside hydrolases. The computer program MUSTANG was used to generate a structure based multiple sequence alignment by aligning the protein structures[90]. FastTree was used to generate a maximum-likelihood phylogenetic tree using the structure sequence alignment from MUSTANG[91,92]. The Geneious software program was used to analyze alignments and phylogenetic trees[93]. Constructed trees were un-rooted, as no assumption was made regarding which sequence is closest to the common ancestor.

3.3.2 Predicted Evolutionary Changes in Family 31 α-glucosidase Structures

Phylogenetic analysis was completed in the program DIVERGE by the neighbor-joining method using the structure sequence alignment and phylogenetic tree[94,95]. Analysis of predicted changes were then mapped to the Protein Data Bank files for the solved structures[80]. These evolutionary differences were then assessed manually with a focus on regions related to catalysis and substrate binding (Table 3.4.1). Analysis of these predicted residues was done with care due to the limitations of the sample population of 8 structures as reported in the PDB.

3.3.3 Comparison of Predicted Evolutionary Changes and Acarbose Binding

To aid in the analysis of predicted evolutionary changes, structures were analyzed in relation to a bound acarbose molecule. Structures that were experimentally solved with a bound acarbose included NT_MGAM (PDB code: 2QMJ), CT_MGAM (PDB code: 3TOP) and SBG (PDB code: 3W37)[83,87,88]. The structures of BT_3299 (PDB code: 5DJW) and BT_0339 (PDB code: 5F7C) were previously reported with a modeled acarbose molecule[86]. Molecular modeling of acarbose was completed using Coot for structures that were not experimentally solved with a bound acarbose[73]. The structure of Ro-αG1 (PDB code: 3MKK) was previously solved with a bound isomaltose molecule,

which was used as comparison to guide the placement of an acarbose molecule[67]. The homologous structure of NT_MGAM (PDB: 2QMJ) was also used in comparison with Ro-αG1. The structure of NT_SI (PDB code: 3LPP) was previously solved with a bound inhibitor kotalanol, which was used as comparison to guide the placement of an acarbose molecule[84]. The structure 2QMJ was used to guide the placement of any remaining acarbose rings that are absent in kotalanol. Acarbose was modeled into the structure of MalA (PDB code: 2G3M) by using the structure NT_MGAM as guidance. In all cases the acarbose molecule was modeled in by superimposing the substrate onto the bound molecules in the comparative structures and used as a guide to place the inhibitor molecule.

Table 3.3.1 Family GH31 α-glucosidases used in the structural analysis of evolutionary changes.

Organism	PDB ID	Domain	Common Name	Preferred Substrate
Sulfolobus solfataricus	2G3M	Archaea	MalA	Maltose and small maltooligosaccharides (Rolfsmeier and Blum 1995)
Ruminococcus obeum	3MKK	Bacteria	Ro-αG1	Isomaltose (Tan et al. 2010)
Bacteroides thetaiotaomicron	5DJW	Bacteria	BT_3299	Maltose and isomaltose (Chaudet et al. 2012)
Bacteroides thetaiotaomicron	5F7C	Bacteria	BT_0339	Maltose and isomaltose (Chaudet et al. 2012)
Beta vulgaris	3W37	Eukaryota	SBG	Long linear maltooligosacchrides (>G6) (Tagami et al. 2013)
Homo sapiens	2QMJ	Eukaryota	NT_MGAM	Maltose (Sim et al. 2010, Lee et al. 2012)
Homo sapiens	3TOP	Eukaryota	CT_MGAM	Maltose and maltooligosaccharides (Lee et al. 2012)
Homo sapiens	3LPP	Eukaryota	NT_SI	Maltose (Sim et al. 2010, Lee et al. 2012)

3.3.4 Phylogenetic Analysis of Characterized Family 31 α-glucosidases

Sequences used in the evolutionary analysis of characterized family 31 α-glucosidases are shown in Supplementary Table 1. Selected sequences include those that have been characterized as α-glucosidases and have the designated enzyme commission number of 3.2.1.20. The computer program MUSCLE was used to generate the multiple sequence alignment[96]. Gaps were removed prior to phylogenetic tree construction to ensure accurate evolutionary analysis. FastTree was used to generate a maximum-likelihood phylogenetic tree using the modified multiple sequence alignment[91,92]. As referred to above, the software program Geneious was used to analyze alignments and phylogenetic trees[93] with the constructed tree being un-rooted, due to no assumption regarding which sequence is closest to the common ancestor.

Table 3.3.2 Characterized family 31 α-glucosidases used in evolutionary analysis. Table includes organisms the protein belongs to, the common name associated with the protein and the accompanied accession number.

Common Name	Organism	Accession Number
α-glucosidase 1 (Aglu1;Aglu-1;At5g11720)	*Arabidopsis thaliana*	AAB82656.1
α-glucosidase (α-G I;α-GI)	*Aspergillus awamori*	BAD06006.1
α-glucosidase B (AgdB)	*Aspergillus nidulans*	BAB39856.1
α-glucosidase (AgdA;AN2017.2)	*Aspergillus nidulans*	EAA64849.1
α-glucosidase (Agl;ANG)	*Aspergillus niger*	BAM72725.1
α-glucosidase (AgdA;Agl; AO090003001209)	*Aspergillus oryzae*	BAA08125.1
α-glucosidase (AG31A;BspAG31A)	*Bacillus sp.* SW20	BAQ19546.1
α-glucosidase II / oligo-α-1,6-glucosidase (BthG II)	*Bacillus thermoamyloliquefaciens*	BAA76396.1
α-glucosidase (BT0339;BT_0339)	*Bacteroides thetaiotaomicron*	AAO75446.1
α-glucosidase (BT3299;BT_3299)	*Bacteroides thetaiotaomicron*	AAO78405.1
α-glucosidase (Agl;SBG)	*Beta vulgaris*	BAM74081.1
α-glucosidase / maltase-glucoamylase (RUMOBE_03919)	*Blautia obeum (Ruminococcus obeum)*	ZP_01966167.1
α-glucosidase II α-subunit (BmGII?)	*Bombyx mori*	BAM78681.1
acidic α-glucosidase	*Bos taurus*	AAF81637.1
α-glucosidase (Aagr-2:Msp-31;R05F9.12;ZK546.10)	*Caenorhabditis elegans*	AAA83174.3
α-glucosidase (glucoamylase) (Gca1)	*Candida albicans*	AAC31968.1
α-glucosidase II (Rot2;CaO19.8589)	*Candida albicans* SC5314	EAK97887.1
α-glucosidase (Gaa1)	*Coturnix japonica*	BAA25884.1
α-glucosidase (HdAgl1;HdAgl)	*Haliotis discus hannai*	BAN67474.1
maltase-glucoamylase (MgaM;Mga;MGAML)	*Homo sapiens*	AAC39568.2
maltase-glucoamylase (MgaM;Mga;MGAML)	*Homo sapiens*	AAC39568.2
sucrase-isomaltase (SI)	*Homo sapiens*	AAT18166.1
sucrase-isomaltase (SI)	*Homo sapiens*	AAT18166.1
α-glucosidase (Agl1;Agl97)	*Hordeum vulgare subsp. vulgare*	AAF76254.1
α-glucosidase	*Litopenaeus vannamei*	CAB85963.1
α-glucosidase (AgdA)	*Mortierella alliacea*	BAC57563.1
α-glucosidase (lysosomal)	*Mus musculus*	AAH10210.1
sucrase-isomaltase (SI)	*Oryctolagus cuniculus*	AAA31459.1
α-glucosidase (Ong1;OngX-H1;Os06g0676700)	*Oryza sativa Japonica Group*	BAD45913.1
α-glucosidase (Ong2;Ong-2;Ong3;OngX-H3;Os06g067570(	*Oryza sativa Japonica Group*	BAD45910.1
α-glucosidase (AglA;PTO0092)	*Picrophilus torridus*	AAT42677.1
α-glucosidase	*Pseudozyma tsukubaensis*	CAA39501.1
sucrase-isomaltase (SI)	*Rattus norvegicus*	AAA65097.1
α-glucosidase II (Gls2;Rot2;YBR229c;YBR1526) (ER)	*Saccharomyces cerevisiae* S288c	CAA85192.1
α-glucosidase (Glz)	*Saccharomycopsis fibuligera* KZ	CAF31354.1
α-glucosidase (Agl1;Agl;SPAPB24D3.10c)	*Schizosaccharomyces pombe* 972h-	CAC36906.1
α-glucosidase	*Schwanniomyces occidentalis* ATCC 2607	BAE20170.1
α-glucosidase (glucoamylase) (Gam1;GAM1)	*Schwanniomyces occidentalis* ATCC 2607	AAA33923.1
α-glucosidase (MalA;Saci_1160)	*Sulfolobus acidocaldarius* DSM 639	AAY80507.1
α-glucosidase (MalA;SSO3051)	*Sulfolobus solfataricus*	AAK43151.1
α-glucosidase (ST2525)	*Sulfolobus tokodaii*	BAB67639.1
α-glucosidase	*Thermoanaerobacter ethanolicus* JW 200	ABR26230
α-glucosidase (GluA;TTE0006)	*Thermoanaerobacter tengcongensis* MB4	AAM23323.1
α-glucosidase (AglA;Ta0298)	*Thermoplasma acidophilum*	CAC11443.1

3.4 Results

3.4.1 Structural Evolution

The first step in the evolutionary analysis was the generation of a structural-based multiple sequence alignment of the structures as experimentally determined for family 31 α-glucosidases (Figure 3.4.1). Overall, there is little sequence similarity with some regions demonstrating high conservation. One of the common reported features of family 31 α-glucosidases is their signature catalytic sequence WiDMNE, where 'i' is variable[81]. Observing this sequence in the sequence alignment we see there is a large variation in this region in *B.thetaiotaomicron* BT_0339 (highlighted in Figure 3.4.1). BT_0339's signature catalytic sequence is KTDFGEN. These differences were also observed structurally and were reported to reflect differences in substrate preferences[86]. This initially gives us some insight into possible evolutionary differences that may exist within this family and interestingly within *B.thetaiotaomicron.*

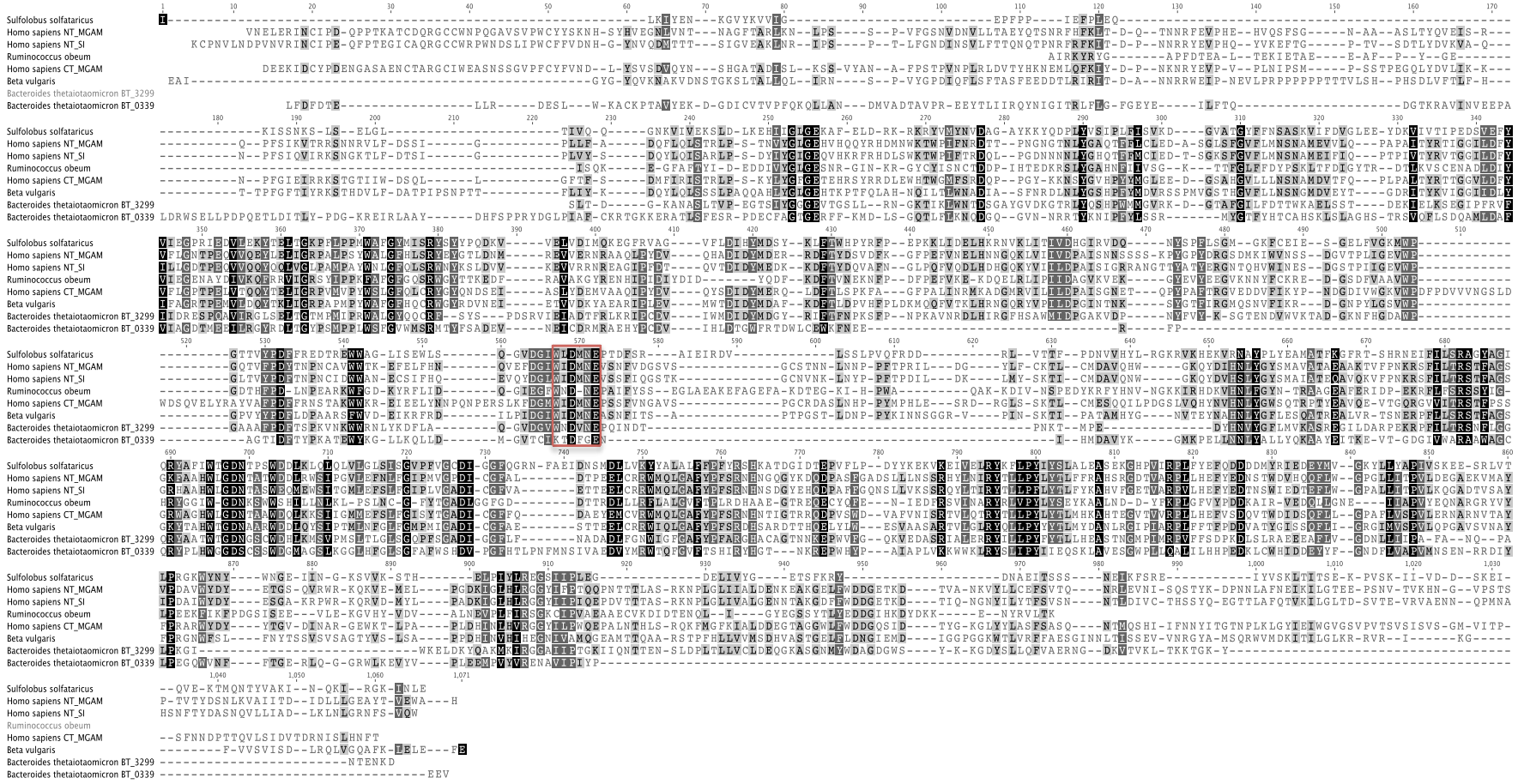


Figure 3.4.1 Structural multiple sequence alignment of structurally solved Family 31 α-glucosidases. Signature catalytic sequence WiDMNE is highlighted in red.

A phylogenetic tree was constructed by the maximum-likelihood method using the above alignment (Figure 3.4.2). The sequence alignment was manually corrected to remove regions of low similarity. Overall, the branch pattern in this tree was not surprising with eukaryotic and prokaryotic clusters. Interestingly, *B.thetaiotaomicron* BT_3299 and BT_0339 are not located close to one another in this phylogenetic tree with BT_0339 showing close relatedness to the archeal organism *S.solfataricus*. Based on this observation, there appears to be some evolutionary change that occurred within this organism in relation to the mechanism of an α-glucosidic enzyme. It also worth noting that *R.obeum* appears to clustering independently and is can be observed that this difference has arisen from the discrete substrate preference of this α-glucosidase by preferring isomaltose.

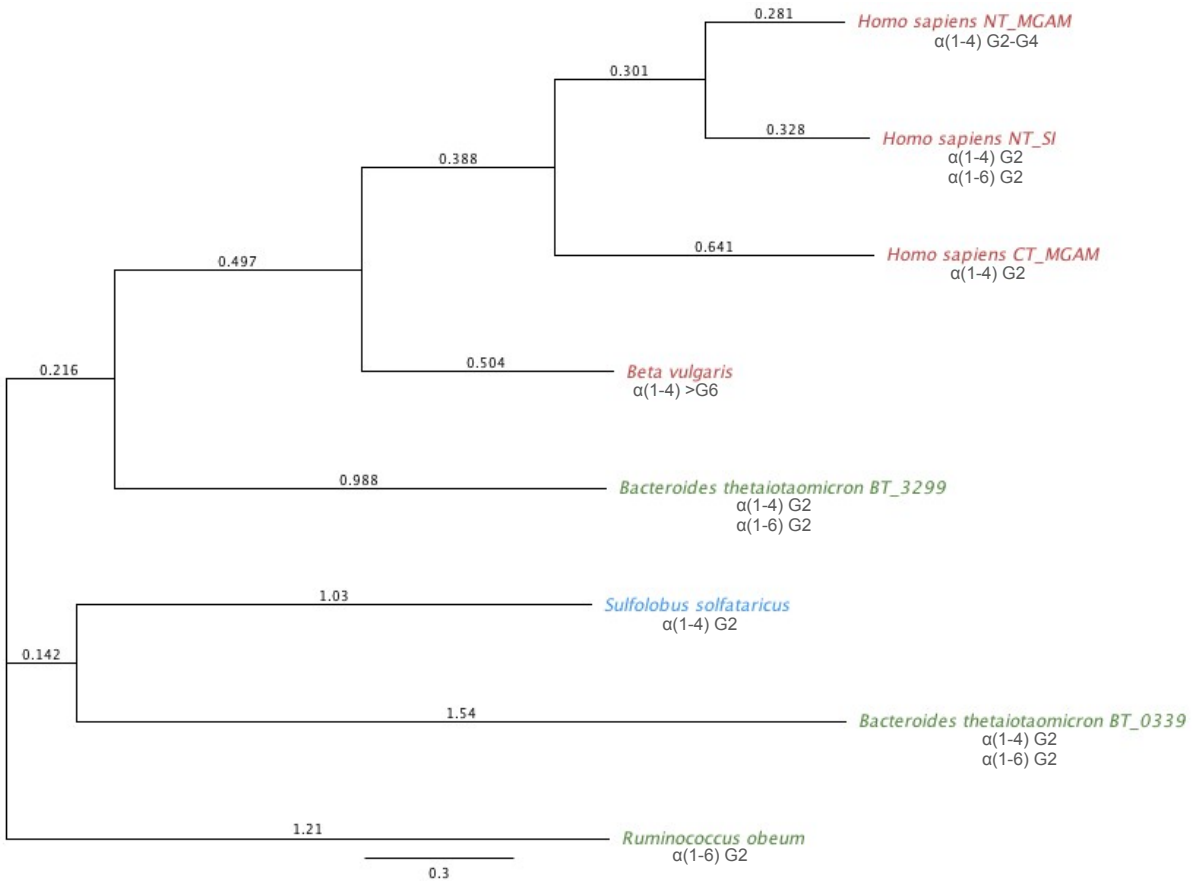


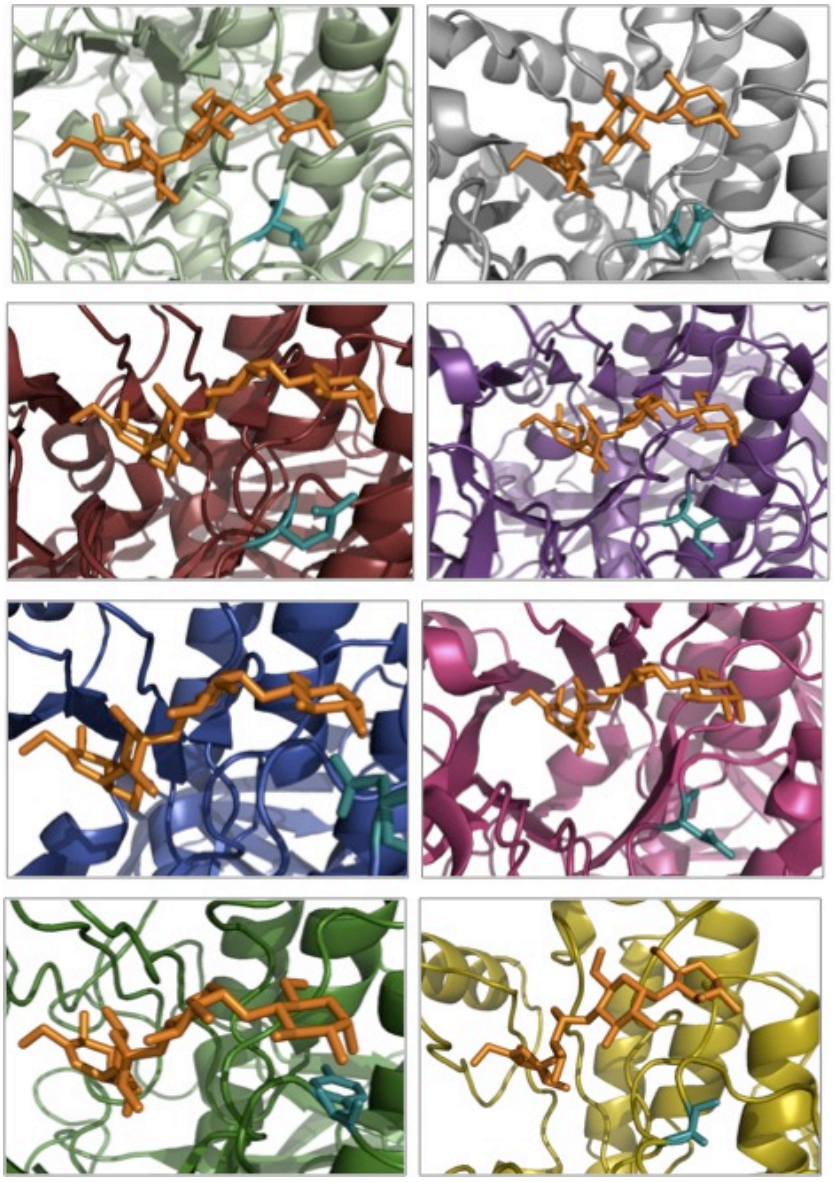
Figure 3.4.2 Phylogenetic tree of family GH31 α-glucosidases for experimentally solved structures. Tree is colour coded specifically for eukaryote (red), bacterial (green) and archaea (blue) organisms. Bootstrap values are shown on tree. Scale bar for branch length is shown at the base of the tree. Substrate preference has been identified for all structures as shown on tree. Preferred linkage(s) derived from starch (α(1-4) and/or α(1-6)) are provided with the length of oligosaccharide as identified by the number of glucose residues (G2:maltose, G4:maltotetraose, G6:maltohexaose).

3.4.2 Variability in Catalytic Region

The program DIVERGE reported a number of amino acids within the analyzed structures that are predicted to have evolved and analysis of these predictions resulted in these two regions occurring in the catalytic site and a looped substrate-binding region (Table 3.4.1). Observing the first predicted change that occurs in the catalytic site allowed for the observation that this change could be involved binding substrates in the +1 and +2 subsites. When we assess this amino acid change individually in each structure we find support of this residue having a role in substrate binding (Figure 2). The three human structures all demonstrate residues that interact with bound acarbose as observed with Thr204, Gln1158 and Gln232 in NT_MGAM, CT_MGAM and NT_SI respectively. The α-glucosidase structure from *Beta vulgaris* was observed to have a predicted change to the amino acid Ile233 and structurally is in close proximity to the bound acarbose inhibitor in the +2 subsite. The structure from MalA has the predicted amino acid Asp346 observed to closely interact with a modeled acarbose in the +1 sub-site. Ro-αG1 structure was observed to have a predicted to change to the residue Asp74 and directly interacts with the bound acarbose substrate in the +1 subsite. Both *B.thetaiotaomicron* structures demonstrate a predicted amino acid change that is believed to influence substrate binding as observed through residues Tyr91 and Gly229 for BT_3299 and BT_0339 respectively. Tyr91 in the BT_3299 model is specifically believed to be involved in stabilizing bound substrates in the active site in the +1 subsite.

Table 3.4.1 Summary of predicted amino acid changes from DIVERGE. PDB ID represents that model used in prediction assessment. Two regions were identified as having site-specific changes in amino acids arising from differences in evolutionary rates. First region is the catalytic region and the -looped substrate-binding region. Indicated residue number is derived from the original structural amino acid sequence as found on the PDB.

Organism	PDB ID	Catalytic Region	Looped Substrate-Binding Region
Sulfolobus solfataricus	2G3M	Asp346	Ala88
Ruminococcus obeum	3MKK	Asp74	Asp333
Bacteroides thetaiotaomicron BT_3299	5DJW	Tyr91	Ala349
Bacteroides thetaiotaomicron BT_0339	5F7C	Gly229	Met474
Beta vulgaris	3W37	Ile233	Asn495
Homo sapiens NT_MGAM	2QMJ	Thr204	Arg471
Homo sapiens CT_MGAM	3TOP	Gln1158	His1449
Homo sapiens NT_SI	3LPP	Gln232	Asp500



PDB ID	Amino Acid
NT-MGAM	Thr204
CT-MGAM	Gln1158
NT-SI	Gln232
SBG	Ile233
MalA	Asp346
Ro-αG1	Asp74
BT_3299	Tyr91
BT_0339	Gly229

Figure 3.4.3 Predicted evolutionary changes in catalytic region. NT_MGAM (light green), CT_MGAM (grey), NT_SI (red), SBG (violet), MalA (blue), Ro-αG1 (pink) BT_3299 (green) and BT_0339 (golden) are shown in ribbon diagrams. Observed amino acid changes are shown in cyan and the tetrasaccharide inhibitor acarbose is shown in orange.

3.4.3 Variability in Looped Substrate-Binding Region

The second predicted evolutionary change as observed in DIVERGE occurs in a looped substrate-binding region in the assessed structures. The observed single amino acid change resides in the second subdomain just outside of the active site and is proposed to be involved in stabilizing longer bound substrates. Predicted evolutionary changes to MalA and Ro-αG1 (*B.thetaiotaomicron* structures were observed to have unresolved regions that were not visible) are observed to occur in a helix bundle that is explicitly observed in these prokaryote structures. These predicted changes include Arg334 and Asp333 in Mala and Ro-αG1 respectively (Figure 3.4.4A).

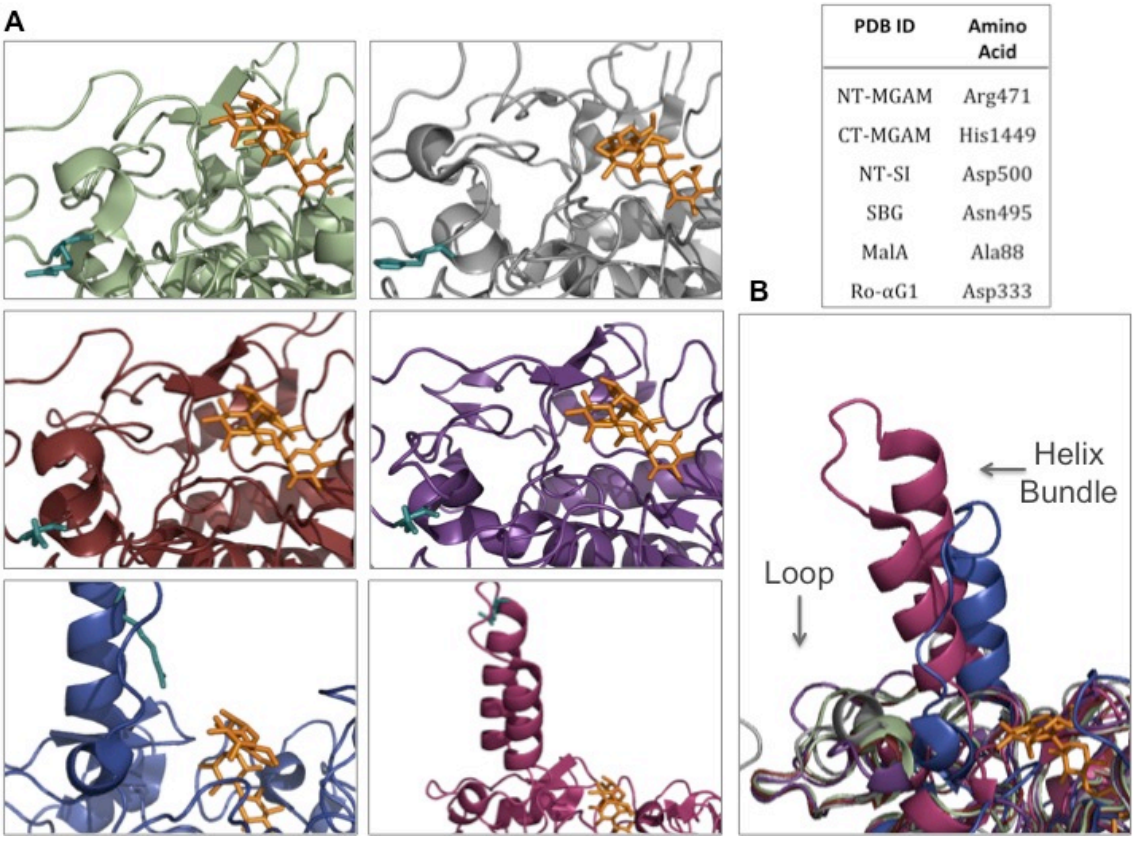


Figure 3.4.4 Predicted evolutionary changes in looped binding region.
NT_MGAM (light green), CT_MGAM (grey), NT_SI (red), SBG (violet), MalA (blue) and Ro-αG1 (pink) are shown in ribbon diagrams. (A). Secondary structural changes predicted to have evolved within family 31 (B). Observed amino acid changes are shown in cyan and the tetrasaccharide inhibitor acarbose is shown in orange.

Predicted changes in the looped binding region were also observed in the eukaryote structures and are positioned to also play a key role in substrate binding. Predicted changes include Arg471, His1449, Asp500 and Asn493 for NT_MGAM, CT_MGAM, NT_SI and SBG respectively (Figure 3.4.4A). All of these changes reside in a loop that is observed to be absent in the four prokaryote structures (Figure 3.4.4B). This observation gives insight into the possibility that this region evolved to accommodate specific starch glycans that are unique to eukaryote α–glucosidases. As a result of the action of salivary and pancreatic α–amylase, small intestinal α–glucosidases are exposed to amylose (linear starch glycan) and amylopectin (branched starch glycan)[97,98].

3.4.4 Phylogenetic Relationships in Characterized Family 31 α-glucosidases

A phylogenetic tree was constructed using characterized α-glucosidase protein sequences using the maximum-likelihood method (Figure 3.4.5). This analysis resulted in a number of branch clusters. The first cluster is composed of eukaryotic organisms consisting of mammals and plant species. The second eukaryotic cluster mainly includes fungal organisms. Archaea and bacteria form a cluster independently of the eukaryotic organisms. There are a number of interesting features of this tree. BT_3299 α-glucosidase from *B.thetaiotaomicron* is distant phylogenetically from BT_0339 and clusters more closely with the archaea group. Another interesting quality of this tree is the location of the *Homo sapiens* α-glucosidases. NT_MGAM and NT_SI cluster together, where as CT_MGAM and CT_SI are more distant. It is observed that CT_SI demonstrates an evolutionary relatedness to *Oryctolagus cuniculus* (rabbit) and NT_MGAM shows relatedness to *Rattus norvehicus* (rat).

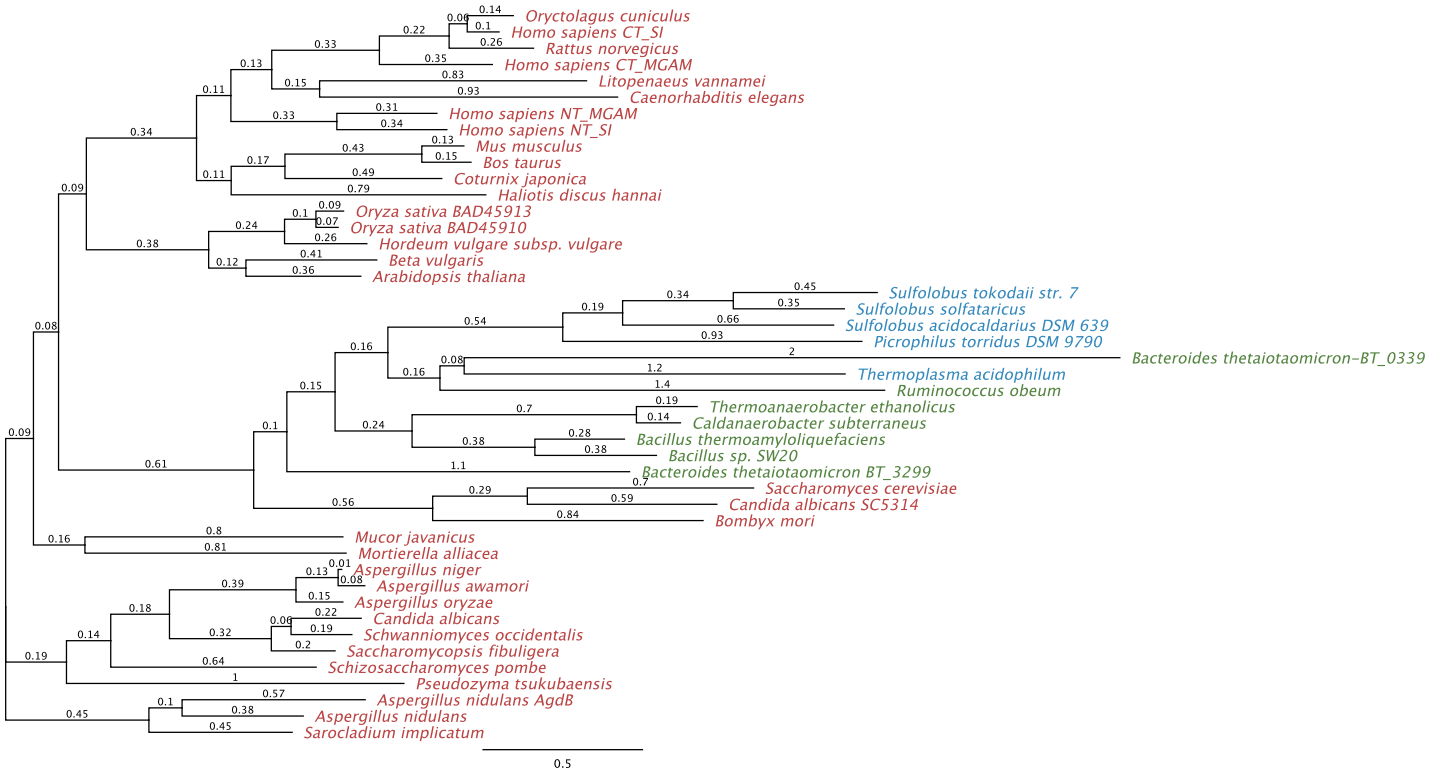


Figure 3.4.5 Phylogenetic tree of characterized Family 31 α-glucosidases. Tree is colour coded specifically for eukaryote (red), bacterial (green) and archaea (blue) organisms. Bootstrap values are shown on tree.

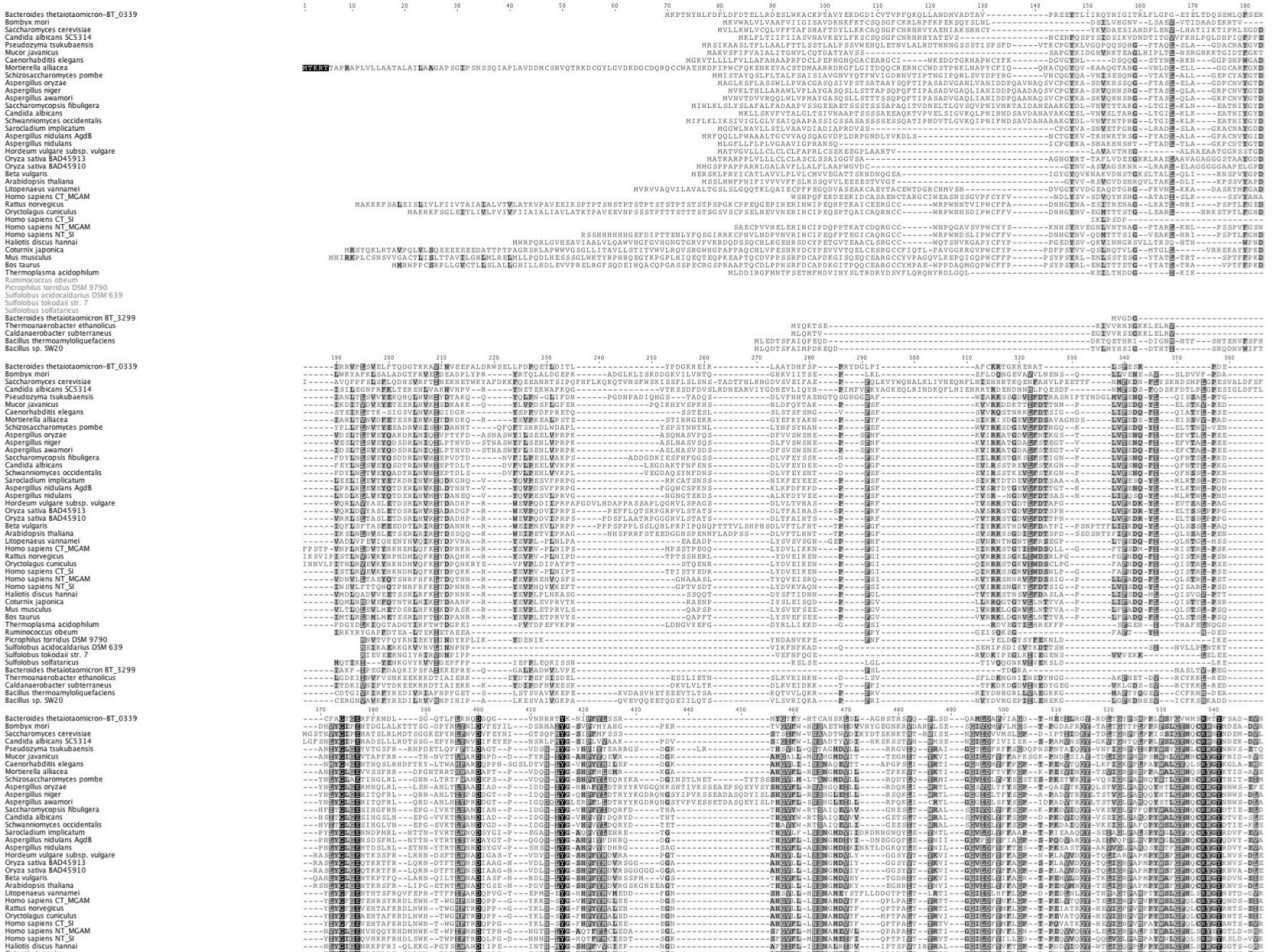


Figure 3.4.6 Multiple sequence alignment of characterized Family 31 α-glucosidases.

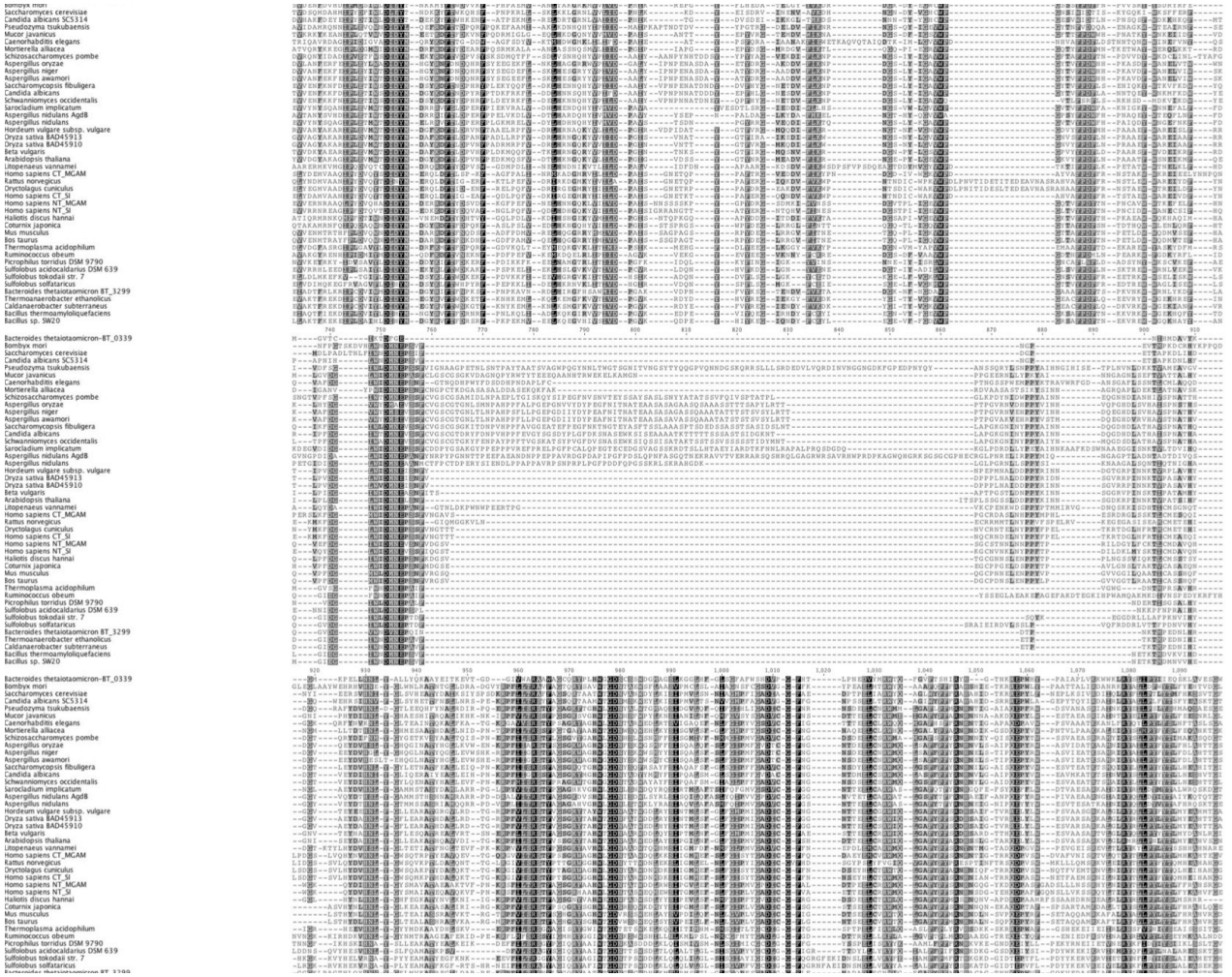


Figure 3.4.6 Multiple sequence alignment of characterized Family 31 α-glucosidases. Figure Continued.

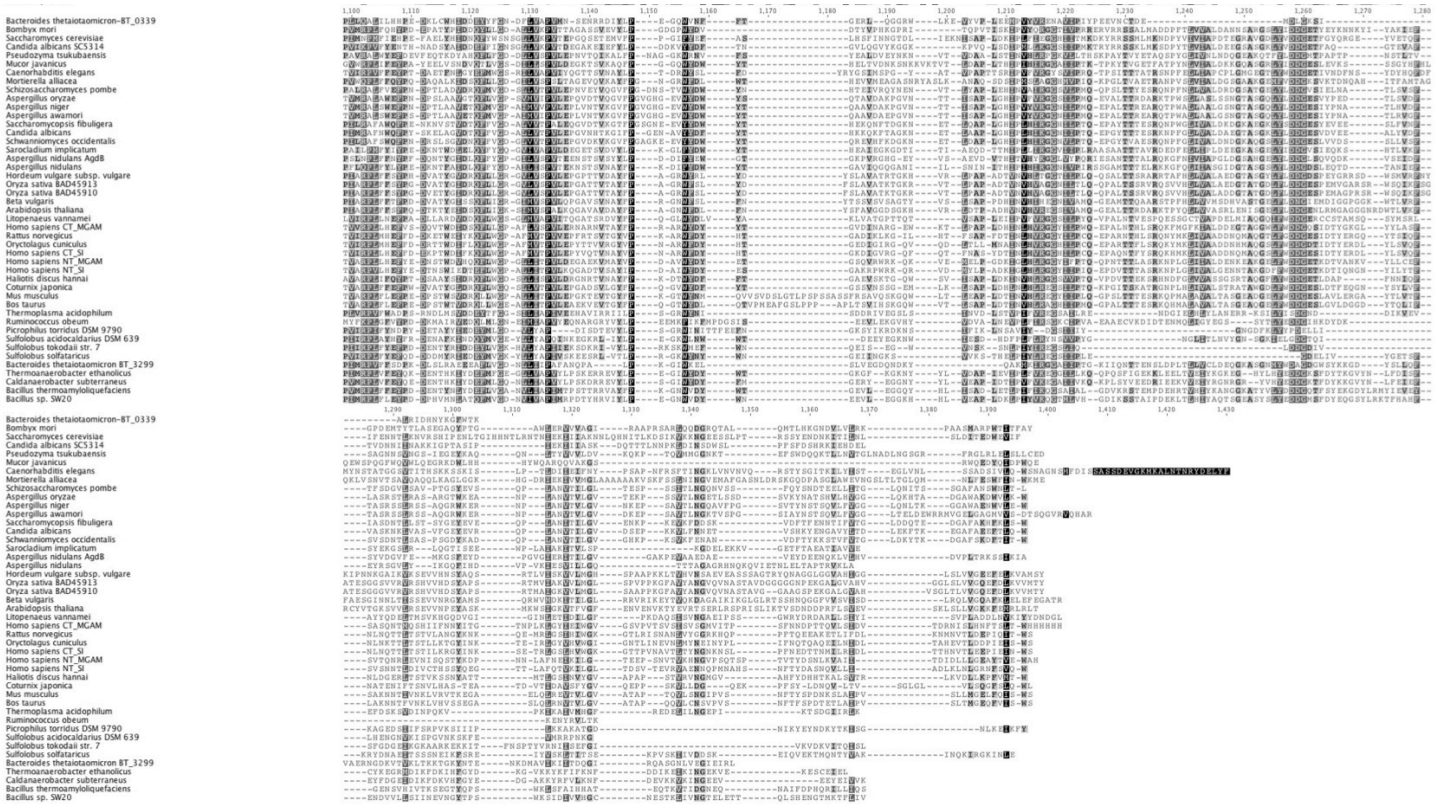


Figure 3.4.6 Multiple sequence alignment of characterized Family 31 α-glucosidases. Figure Continued.

3.4.5 Eukaryote Signature Sequence

Comparison of the phylogenetic tree of characterized α-glucosidases with the solved structures allowed for the observation of an additional signature sequence within this family (Figure 3.4.6A). The sequence occurs at positions 877-879 in the multiple sequence alignment and is identified as 'ppY, where p is variable but predominantly proline. Interestingly, it is observed that this sequence is conserved in only characterized eukaryotic α-glucosidases and is not present in characterized prokaryote α-glucosidases. This region of eukaryotic conservation happens to fall in the evolved loop in subdomain2 and is related to the second predicted evolutionary change that occurs in the looped binding region (Figure 3.4.6B).



Figure 3.4.7 Predicted eukaryote signature sequence in characterized Family 31 α-glucosidases. (A) Multiple sequence alignment for the region 876-881 with highlighted conserved residues of characterized family 31 α-glucosidases. (B) NT_MGAM structure (PDB code: 2QMJ) with bound acarbose (orange) and highlighted 'ppY' eukaryote signature sequence (pink).

3.5 Discussion

Phylogenetic analysis derived from structural sequence alignment allowed for observations of family 31 α-glucosidases in relation to their ability to utilize maltooligosaccharides. The α-glucosidase from *R.obeum* is previously reported to preferentially hydrolyze isomaltose in comparison to starch-derived glycans[67]. This specificity is unique amongst the structurally analyzed family 31 α-glucosidases and we believe is reflected in its branching in the phylogenetic tree. This organism resides in the human colon, which is a diverse, competitive environment. In order to out compete surrounding organisms perhaps *R.obeum* developed a distinct α-glucosidic ability to use starch-derived glycans in the human colon. The observation of BT_0339 having a higher structural relatedness to MalA reveals that this enzyme may have some ancient origin and that BT_3299 is more closely related to the eukaryote organisms. As previously reported, NT_MGAM is more closely related to NT_SI and as evident in the tree, these enzymes cluster closer together than to CT_MGAM. Though these N-terminal domains have slightly different substrate preferences, it can be observed that they both in general are able to bind and hydrolyze bulkier maltooligosaccharides (linear starch glycans and branched starch glycans) and this could reflect their evolutionary relatedness. This observation contributes to hypothesis that these domains originate from a common ancestral gene[36].

A number of changes that we propose influence substrate binding in family 31 α-glucosidases can be observed from this analysis. The first change involves residues in the +1 and +2 sugar binding site in the catalytic centre. This was observed in NT_MGAM with evidence from previous structural observations. It was previously observed that Asp203 was involved in binding acarbose in the +1 sub-site of NT_MGAM and provides evidence that a change at the amino acid Thr204 could influence substrate binding[83]. The structure for NT_SI demonstrated a water molecule interacting with Gln232 similarly to what we are predicting here. A change to this residue would have specifically influenced sugars binding in the +1 subsite of

NT_SI. Similarly, it was observed in CT_MGAM that D1157 interacted directly with acarbose in the +1 subsite and therefore provides support to our prediction that a change to the neighboring amino acid Gln1158 could influence the binding of starch molecules. Structural studies of Ro-αG1 mutants resulted in observed structural differences in the residue Asp73 resulted in specific disorder causing a lack of bound substrate in the active site[67]. Enzyme assays involving this mutant (D73A) resulting in significant decrease in enzyme activity in the presence of maltooligosaccharides[67]. These observations provide support for our proposal that the neighboring residue Asp74 could have a key role in substrate binding and an evolutionary change to this residue would have influenced the catalytic abilities of this enzyme. It is also thought that a similar interaction exists within the *Bacteroides* structures based on modeling of acarbose within the models. Based on these results we suggest that a change in the +1 and +2 sugar binding site could strongly impact how substrates bind overall in the catalytic site of family 31 α-glucosidases.

The second predicted evolutionary change involves a secondary site that is proposed to be involved in substrate binding. Predicted changes to the helix bundle observed in MalA and Ro-αG1 are supported by structural observations. In the structure of Ro-αG1 researchers reported that the helix bundle was an important element in the dimerization of the crystal structure and could reflect a dimerization necessary for function based on substrate preference observations[67]. This structural feature was also observed with MalA and was reported that the helix bundle was an integral component to stabilizing the crystal packing[81]. These crystallographic observations provide evidence that the helix bundle mediates protein-protein interactions and are believed to be necessary for substrate access in the active site[67]. These experimental observations support our prediction that the evolutionary change of a specific residue in the helix bundle is believed to influence the secondary structure of these enzymes and ultimately could, play a role in stabilizing substrates in the catalytic site. A similar predicted change was observed within a loop specific to the eukaryotic structures and is also supported by previous structural observations. This structural feature was predicted to be involved in

binding long starch glycans in the +4 subsite in SBG structure[88]. NT_MGAM, CT_MGAM and NT_SI are all observed to have the ability to utilize linear maltooligosaccharides and therefore, are structurally thought to involve additional binding sites outside the catalytic site of -1/+1 subsites (Figure 3B). CT_SI has been observed to utilize both sucrose and maltose-derived substrates; further supporting the evolved ability of this group of enzymes[64]. This predicted evolutionary divergence is further supported in the recognition of a eukaryotic specific signature sequence of 'ppY', as our results suggest that this characteristic region allows for the binding larger starch-based sugars. This gives support to the previous observation of the SBG structure and the prediction of additional sugar binding subsites existing in this region, which allows this plant α-glucosidase to bind and hydrolyze long-chain maltooligosaccharides, that include maltoheptaose[88]. We believe this evolved ability to accommodate longer starch substrates is reflected in the evolution of our dietary starch. Through evolution the availability of dietary starch as increased and as a result our digestive system is exposed to an increase in the diverse starch based polymers through dietary starch sources (ie. corn verse potato). This environmental change has allowed for our α-glucosidases (MGAM and SI) to adapt in order to extract the maximum amount of glucose from our diet.

In general, our phylogenetic analysis demonstrated typical clustering of branches amongst domains and phyla, but also provided evidence of interesting patterns of evolution. BT_0339 and BT_3299 from *B.thetaiotaomicron* were observed to not cluster together as one would imagine with BT_0339 being closer related to certain archeal organisms. Based on this result we believe these enzymes did not arise from a gene duplication event and that perhaps one of these paralogs was obtained by means of horizontal gene transfer. Alternatively, this could reflect some form of specie divergence that is specific for *B.thetaiotaomicron*. It is also believed that BT_0339 being closely related to genes derived from the archaea domain reflects the origin of BT_0339 being evolutionarily ancient as stated above. These paralogs in *B.thetaiotaomicron* could be evolution artifacts, gene redundancies that have yet to optimally removed from the genome, or alternatively they could perform discrete

functions that are necessary for utilizing diverse forms of starch that exist in our diet. Focusing on the specific functions observed for the α-glucosidases involved in human digestion, we propose that these enzymes adapted alternative mechanisms in order to be able to utilize available substrates derived from dietary starch. The colon microbiome is a highly populated and competitive environment and as a result inhabiting organisms must acquire or evolve enzymatic mechanisms in order to thrive in such a population. The phylogenetic relationship between the human enzymes MGAM and SI is one that is interesting, and our results support the notion that the divergence of these enzymes likely occurred prior to mammalian evolution as a result of the relatedness to distant species[40]. It is likely from our analysis that the common ancestor of these enzymes is one that occurs quite distantly in the phylogenetic tree of life in relation to *H.sapiens* and the sequence differences that exist between MGAM and SI likely occurred prior to the emergence of *H.sapiens*.

3.6 Conclusion

The observed evolutionary changes that have occurred within family 31 glycoside hydrolases are focused on changes that influence substrate binding. Overall, we propose that these changes influence the accessibility of these enzymes to utilize longer starch glycans in addition to smaller oligosaccharides[88]. Particularly, the ability of eukaryote organisms to degrade long linear maltooligosaccharides and prokaryotes purposefully hydrolyzing smaller branched polymers. Changes involved in catalysis would likely result in alterations in conserved motifs that are signatures of family 31. As a result of the evolutionary adaptions to the members of this family, there is an overall ability to utilize the polymer starch, but also there are observed discrete functions that reflect phylogenetic specific substrate preferences. Furthermore, this family of enzymes in the context of human digestion reveals an evolutionary symbiosis that developed over time, resulting in the mutual relationship at utilizing dietary starch. Understanding the evolutionary changes that occur within an enzyme family can help define structural studies and provide evidence for biochemical studies of substrate utilization within carbohydrate-active families. The results shown here provide brief insight into the phylogenetic relationship within this family and highlight the need for further analysis into the relationship between ancestor and descendants of family GH31 α-glucosidases.

Chapter 4
Transcriptional Control of PUL 55

4.1 Chapter Overview

In this chapter we investigated the transcriptional control of the polysaccharide utilization locus 55. Understanding the catabolic regulation of this specific glycan locus is important in order to understand its functionality and relationships it may have with other loci. We decided to focus on two perspectives to assess the transcriptional control of PUL 55. The first approach is using in-vivo experiments that consist of observing the level of growth *B.thetaiotaomicron* is able to achieve in the presence of specific substrates. The metabolism of starch glycans will be assessed in *B.thetaiotaomicron* by the use of a strain lacking the starch utilization system. If there is an observed level of growth in this knockout strain, we can make an assumption that there are additional genes involved in this metabolic activity, such as the genes from PUL 55.

The second approach is using in-vitro techniques that allow for the assessment of a response regulator as part of the two-component system from PUL 55 and determine its level of regulation. Two-component systems are a well-documented mechanisms that bacteria employ in order to recognize and adapt to changes in their surrounding environment[99]. This system consists of two elements that include a histidine kinase and a response regulator (Figure 4.1).

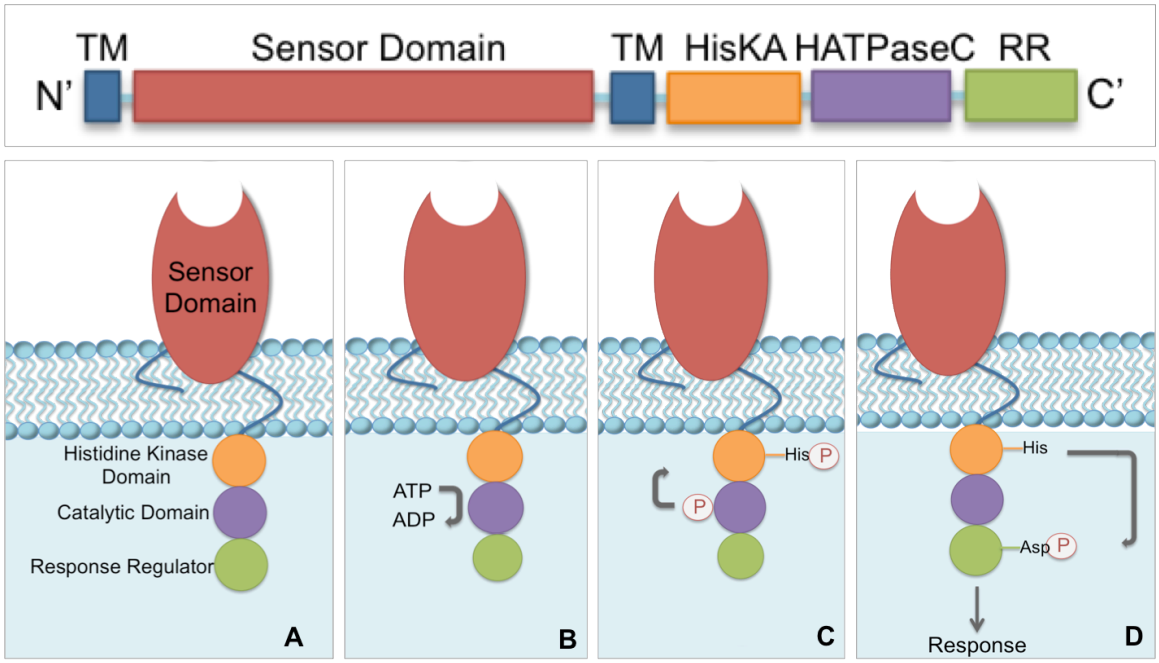


Figure 4.1.1 Predicted domain organization of BT_3302. (A) BT_3302 consists of transmembrane domain (TM), periplasmic sensing domain, another transmembrane domain, histidine kinase domain (HisKA), catalytic domain (HATPase_c) and a response regulator (RR). (B) An ATP molecule binds to the catalytic domain. (C) The catalytic domain transfers the phosphate to a conserved histidine residue in the histidine kinase domain. (D) This phosphate is then transferred to a conserved aspartic acid residue in the response regulator, which can then produce a response.

A two-component system functions by sensing a stimulus, which triggers a response by the response regulator. The two-component system as part of PUL 55 is presumed to detect small oligosaccharides in the periplasm by the sensor domain. The histdine kinase domain is phosphorylated by the catalytic domain, which has a bound ATP molecule. A phosphotransfer event occurs between the histidine kinase domain and the response regulator. This phosphorylated response regulator then binds to regions of DNA in the genome upregulating the expression of specific genes[100]. Up-regulated regions are generally involved in the utilization or processing of the targeted glycan. Understanding the role of these two-component systems will provide insight into the ability of colon microorganisms to respond to changes in the colon environment.

4.2 Contributions

All work presented in this chapter was completed by the candidate.

Dr. Jin Duan provided guidance with construct design and cloning as part of the work in this chapter. Dr. Kathy Lam provided guidance with experimental procedure for the growth experiments and Patrick Diep assisted with the cloning of fragments recovered from electromobility shift assay. Dr. Nicole Koropatkin provided the *Bacteroides thetaiotaomicron* strains used in the growth experiments.

4.3 Material and Methods

4.3.1 Expression and Purification of the Response Regulator BT_3302-5

Two-component systems in *B.thetaiotaomicron* in general consist of a number of domains that cross the cytoplasmic membrane multiple times[99]. BT_3302 specifically has two transmembrane domains that straddle the sensor domain and allow it to be anchored into the cytoplasmic membrane. It is important to consider these characteristics in relation to the expression of the targeted protein. Based on this prediction, multiple constructs were designed based on the predicted structural situation in the cytoplasmic membrane (Table 4.3.1). The constructs for BT_3302 were designed using prediction tools available from NCBI-Blast. The Conserved Domain Database was used to predict functional units within the annotated protein of BT_3302.

Table 4.3.1 Construct design for the response regulator BT_3302. 'Region' in table indicates the targeted domain in BT_3302. Construct boundaries relate to the genomic context in *Bacteroides thetaiotaomicron* VPI-5482.

Region	Construct ID	Construct Boundaries
Entire Transcript	3302-1	4235444 to 4239544
Sensor Domain	3302-2	4235444 to 4238746
Histidine Kinase Domain	3302-3	4238294 to 4239541
Catalytic Domain	3302-4	4238294 to 4238737
Response Regulator	3302-5	4238747 to 4239541

Due to time limitations, only the response regulator, BT_3302-5, will be investigated during this project. The gene encoding BT_3302-5 was amplified by the polymerase chain reaction using *Bacteroides thetaiomicron* genomic DNA (GenBank accession No. NC_004663.1). Primers were designed with a 5' *BamHI* restriction site and a 3' *XhoI* restriction site for directional cloning purposes (Table 4.3.2). The BT_3302-5 gene was cloned into pET29a vector by double digest and ligation using T4 ligase as described in the manufacturer's instructions. Constructs were transformed into Nova Blue competent

cells and screened on selective media. BT_3302-5 was transformed into BL21 (DE3) pLysS expression strain. Cells were grown at 37°C until an optical density of 0.6 at 595nm was reached, followed by induction by 0.1 mM of isopropyl-β-1-thiogalactopyranoside (IPTG). The culture was then incubated at 14°C overnight with shaking.

Table 4.3.2 Primers for BT_3302-5. The sequence of primers used in the amplification of the BT_3302-5 gene for the plasmid pET29a. Restriction sites *SacI* and *NotI* were inserted for the forward primer BT_3302-F5.1 and the reverse primer BT_3302-R1.1 respectively as indicated by underlining.

Primer Name	Primer Sequence
BT_3302-F5.1	5' CGAGAGAGCTCATGAAAAAACACAATGAATGG 3'
BT_3302-R1.1	5' CGAGAGCGGCCGCTACGTTCATAGTTCCTCCTCC 3'

BT_3302-5 cells were lysed by sonication and re-suspended in lysis buffer (20 mM NaH_2PO_4, 300 mM NaCl and 10 mM imidazole). The cell lysate was then batch-bound overnight with Ni^+-NTA resin at 4°C with rotation. Affinity purification was completed at 4°C and by gravity flow. The column was washed twice with wash buffer (20 mM NaH_2PO_4, 300 mM NaCl and 25 mM imidazole) and the protein of interest was eluted from the column by imidazole gradient (20 mM NaH_2PO_4, 300 mM NaCl and 50 mM to 500 mM imidazole). The buffer was subsequently exchanged into 20 mM MES pH 6.0. Protein sample was injected onto an Uno Q Ion Exchange Column (Bio-Rad) that was equilibrated in 20 mM MES pH 6.0. Elution of protein of interest was completed by a NaCl linear gradient. Fractions containing the protein of interest were concentrated and subsequently buffer exchanged into 20 mM MES pH 6.0.

These conditions were then used for the expression of the selenomethionine protein following the protocol of methionine pathway inhibition[101].

4.3.2 Electromobility Shift Assay

Analysis of the binding of BT_3302-5 was completed using the electromobility shift assay. BT_3302-5 is predicted to bind to an unknown promoter region in *B.thetaiotaomicron* genomic DNA. *B.thetaiotaomicron* genomic DNA was previously purified by the Charles Lab at the University of Waterloo and obtained for experimental purposes. Purified *B.thetaiotaomicron* genomic DNA was (5 μg) was digested overnight by the restriction enzyme Sau3AI as indicated by the manufacturer (New England Biolabs). Purified BT_3302-5 (0.5 mg) was incubated with digested genomic DNA, 9% glycerol, 50 mM lithium potassium acetyl phosphate, 1 mM Tris-HCl pH 8.0, 0.25 mM magnesium acetate 1.6 mM potassium acetate and 2BSA (2.5 μg/ml). Reaction was incubated at 37°C for 1 hr. Phosphorylated BT_3302 is the expected active form of this regulator; however, phosphorylation was not confirmed prior to completing this assay. It can be assumed BT_3302-5 is phosphorylated in the presence of the phosphate donor lithium potassium acetyl phosphate.

Reactions were analyzed by 5% native polyacrylamide gel electrophoresis. Protein samples were directly loaded onto gel following incubation with equal loading for both the negative control and BT_3302-5. Gel was run in 1x sodium boric acid buffer at 40 mAmps for 7 hours at room temperature. Gel was stained for 30 min in RedSafe nucleic acid stain (FroggaBio) and visualized on a UV transilluminator.

The band of interest was excised from the gel and dissolved overnight in 0.5 M ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA pH 8.0 and 0.1% SDS at 50°C. Excised genomic DNA was purified using the Qiagen Gel Purification Kit. Purified fragments were ligated into pBlueScript vector and sent for sequencing. Sequences were analyzed in NCBI-Blast to confirm the location within the genome. Experiment was repeated in triplicate in order to assess binding reproducibility.

4.3.3 *Bacteroides thetaiotaomicron* Growth Experiments

B.thetaiotaomicron ATCC 29148 (VPI-5482) strains were used as the wild-type strain in growth experiments. *B.thetaiotaomicron* strain Δ-tdk was used in the gene deletion and growth experiments. This strain is a not a true wild-type strain as it has a deletion of its tdk gene. This strain will be used for counter selection purposes and demonstrates no auxotrophic characteristics. *B.thetaiotaomicron* strain Δ-sus was also used in the gene deletion and growth experiments. Δ-sus has a full deletion of the entire locus *susRABCDEFG* in the Δ-tdk strain. This deletion includes all enzymes, carbohydrate binding proteins, transporters, regulators and intergenic regions of the *sus* locus. This strain allows for an observation of starch utilization in the absence of the starch utilization system. All strains were obtained from Nicole Koropatkin's research group and were grown in the following conditions based on their instruction.

Strains were stored in 50% glycerol in -80°C until cultured. Strains were routinely grown in Brain-Heart Infusion broth (Becton Dickson) or minimal medium. Minimal media included salt source (1 mM KH_2PO_4, 1.5 mM NaCl, 0.85 mM $(NH_4)_2SO_4$; pH 7.2), 40 μM L-cysteine, Histidine-hematin solution (0.2 M histidine, 1.9 mM hematin) 10 μM MgCl2, 0.02 μM $FeSO_4$, 5 μM menadione solution, 0.8% $CaCl_2$ and 4 nM vitamin B12.

Overnight cultures of each strain were grown in BHI broth at 37°C with minimal shaking. Cultures were pelleted and washed three times in minimal media. The cells were re-suspended in minimal media and diluted to an equal OD reading at 600nm to ensure equal cell density for experiment initiation as previously descibed[50]. Experimental conditions consisted of growing strains in minimal media supplemented with 0.5% carbon source. Carbon sources included glucose, maltose, amylopectin pullulan and β-cyclodextrin. Cultures were grown anaerobically in test tubes. Test tubes were plugged with a cotton ball and set on fire. The cotton ball was allowed to burn for a few seconds. 35% pyrogallol and 20% saturated sodium carbonate was added to the top of the burnt cotton ball. The test tube was then tightly sealed with a rubber stopper. Test tubes were

incubated at 37°C with minimal shaking. Absorbance of each culture was measured every 12 hours over 6 days at 600nm.

4.4 Results

4.4.1 *Growth Experiments of Bacteroides thetaiotaomicron on Starch Substrates*

Growth experiments for *B.thetaiotaomicron* were successfully completed and provided evidence of multiple systems involved in the utilization of starch-derived substrates. The wild-type strain was used as a control strain for comparison with Δtdk and Δsus strains. The wild-type strain demonstrated the ability to utilize all the provided substrates and was observed to have a higher level of growth on maltose, followed by β-cyclodextrin, pullulan, glucose and lastly amylopectin (Figure 4.4.1). A longer lag was observed for β-cyclodextrin in comparison to the other substrates.

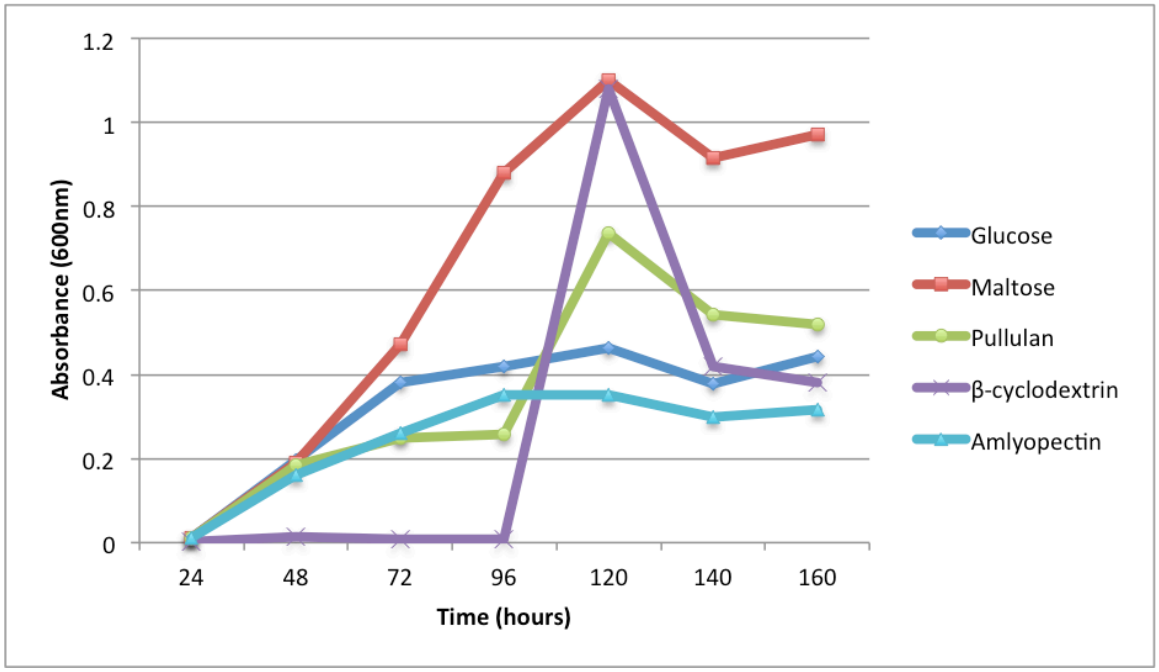


Figure 4.4.1 Growth curve of wild-type *B.thetaiotaomicron* strain grown on starch derived substrates.

In comparison, Δtdk demonstrated a similar level of growth for the same substrates, but it was observed to prefer pullulan to β-cyclodextrin (Figure 4.4.2). There was also an observed longer lag time for β-cyclodextrin in comparison to the other substrates.

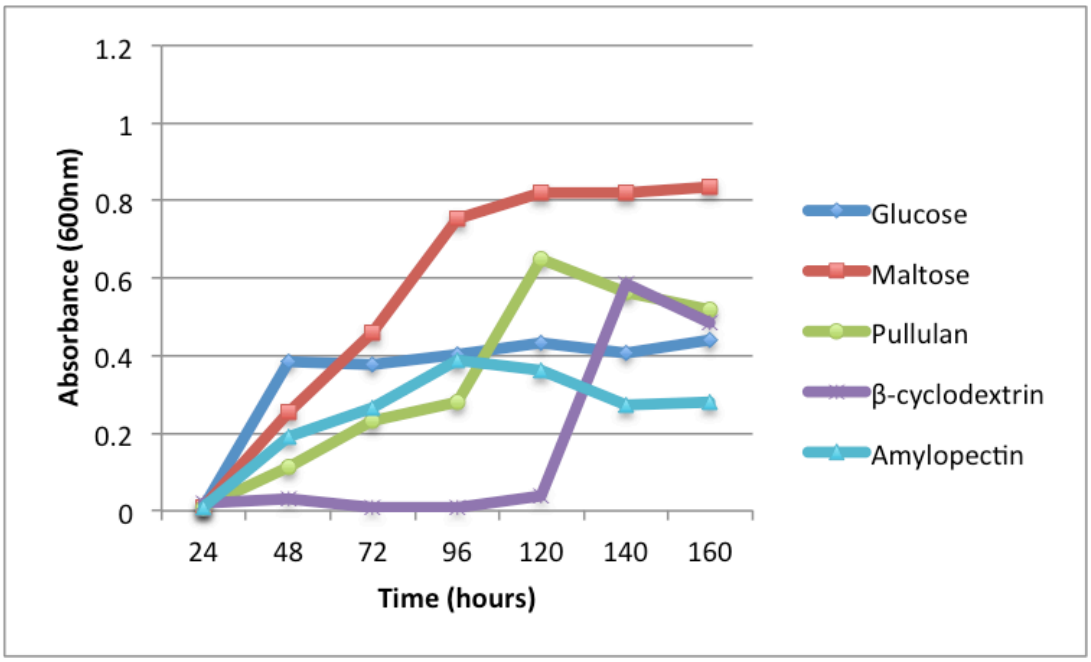


Figure 4.4.2 Growth curve of Δtdk *B.thetaiotaomicron* strain grown on starch derived substrates.

The strain Δsus was observed to utilize maltose at the highest rate, followed by glucose, amylopectin and pullulan. This strain was not able to utilize β-cyclodextrin as demonstrated by observed little/no growth (Figure 4.4.3).

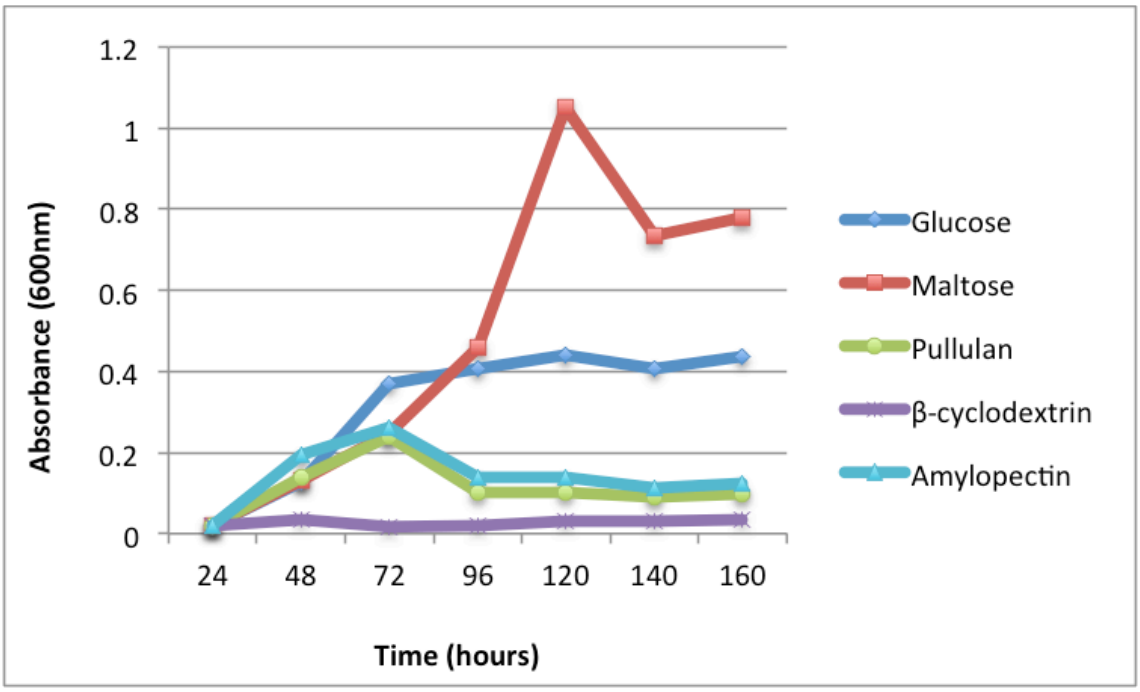


Figure 4.4.3 Growth curve of Δsus *B.thetaiotaomicron* strain grown on starch derived substrates.

4.4.2 Expression and Purification of Response Regulator BT_3302-5

BT_3302-5 was successfully expressed and demonstrated a highly level of soluble protein in the cell lysate. Two-step chromatography purification was completed using nickel affinity and anion exchange yielding a clean concentrated sample (Figure 4.4.1).

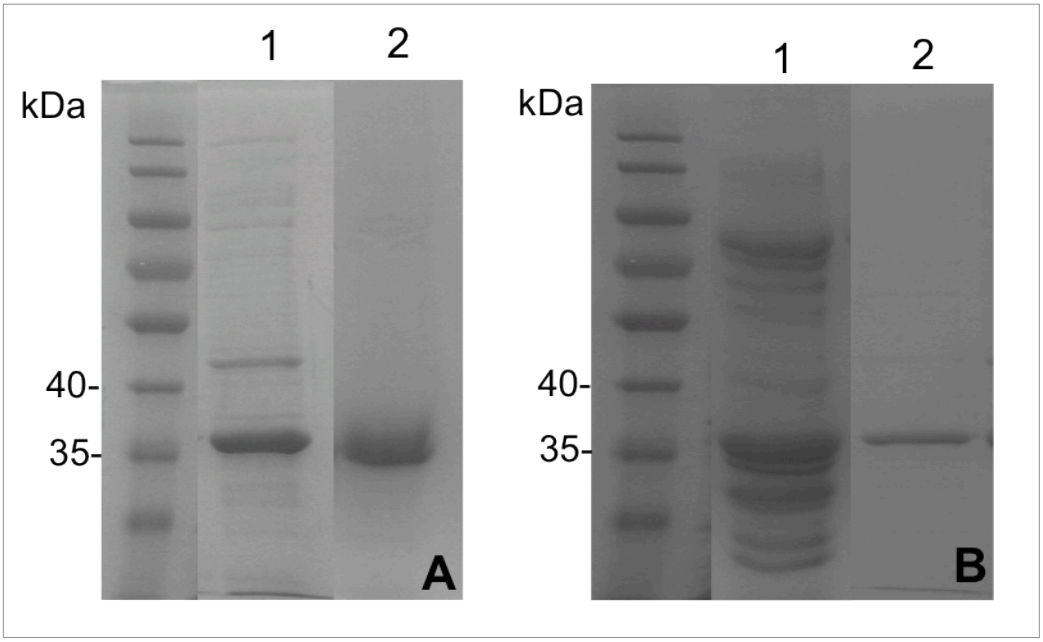


Figure 4.4.4 Denaturing SDS-PAGE gel for BT_3302-5. BT_3302-5 (A) was purified by nickel affinity (1) and anion exchange (2) chromatography. BT_3302-5-Se (B) was also purified by nickel affinity (1) and anion exchange (2) chromatography.

4.4.3 Electromobility Shift Analysis of BT_3302-5

An electromobility shift assay was used to observe the interaction of phosphorylated BT_3302-5 with digested *B.thetaiotaomicron* genomic DNA. A portion of the digested DNA was observed to bind to BT_3302-5 as demonstrated by the trapped DNA at the top of the non-denaturing gel (Figure 4.4.5). A faint band is observed in the negative control lane, but it is believed that this is genomic DNA that has not fully moved through the gel. Electrophoresis conditions were improved by running the non-denaturing gel for an additional 2 hours, from 5 to 7 (Refer to Appendix). This minimized the amount of DNA

in the top of the gel in the negative gel. The region at the top of the gel was successfully excised, purified and a number of clones were sent for sequencing (Refer to Appendix-Supplementary Figure S1).

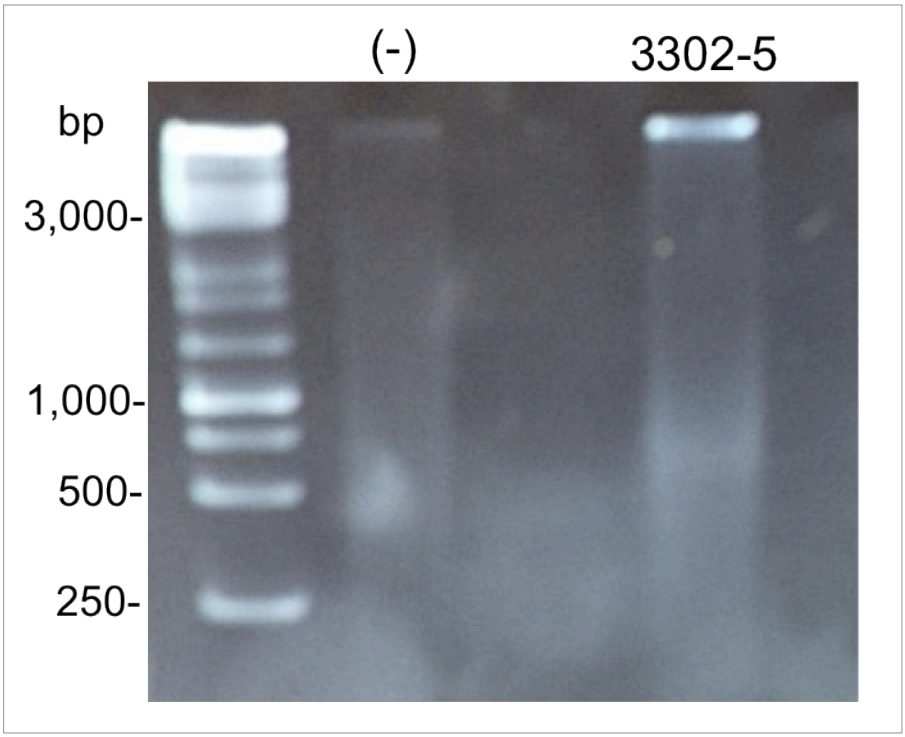


Figure 4.4.5 Electrophoresis mobility shift assay for BT_3302-5. Lane one is the 1 kB DNA ladder, lane two is the negative control and the third lane is the experimental condition for 3302-5.

Analysis of sequenced clones allowed for the observation of genetic regions the response regulator BT_3302-5 could regulate in-vivo (Table 4.4.1). A number of genomic regions of *B.thetaiotaomicron* were observed to interact with phosphorylated BT_3302-5 (Supplementary Table S1). A majority of these genetic regions are contained within genes that are members of PULs as identified by Martens *et al,* 2011[53] (Table 4.4.2-4). The genetic regions that BT_3302-5 was observed to bind to were all intragenic regions.

Table 4.4.1 Fragments recovered from electromobility shift assay with BT_3302-5. Table indicates the fragments that were isolated from the electromobility shift assay. Position identifies the region in which sequencing identified the fragments is located in *B.thetaiotaomicron's* genome. Predicted gene function, associated PUL and predicted inducing glycans are as described by Martens *et al,* 2011[53]. Selected regions in Table 4.4.1 are fragments that were observed in more than one clone as summarized in Supplementary Table S1.

Position		Size	Gene	Predicted Gene Function	Associated PUL	Predicted Inducing Glycans
5'	3'	(bp)				
235653	236051	398	BT_0212	protease	4	host/residual dietary glycans (unknown type)
4320944	4321066	122	BT_3350	chondroitinase	57	chondroitin sulfate, hyaluronan (dermatan sulfate)
4903147	4903304	157	BT_3778	hypothetical protein	68	alpha-mannan, host N-glycans
1587812	1587911	99	BT_1277	fucose permease	L-fucose locus	Fucosylated glycans
2396348	2396442	94	BT_1907	RNA polymerase sigma factor RpoS	N/A	N/A
4954057	4954246	189	BT_3806	hypothetical protein	N/A	N/A
3421511	3421657	146	BT_2761	hypothetical protein	N/A	N/A

Table 4.4.2 Associated PULs of isolated fragments from electromobility shift assay. Description of PUL components and predicted inducing glycans are as described by Martens *et al,* 2011[53].

PUL Identification	Gene Identification	Predicted Function	Predicted Inducing Glycans
L-fucose locus	fucR	DNA-binding transcriptional regulator	fucosylated glycans
	fucI	L-fucose isomerase	
	fucA	L-fucose phosphate aldoase	
	fucK	L-fucose kinase	
	fucP	Fucose permase	
16	BT1278	ECF-type sigma factor	mucin *O*-glycans
	BT1279	anti-sigma factor	
	BT1280	susC-like	
	BT1281	susD-like	
	BT1282	hypothetical protein	
	BT1283	hypothetical protein	
	BT1284	putative endo-beta-N-acetylglucosaminidase	
	BT1285	Glycoside Hydrolase Family 18: endo-β-N'acetylglucosaminidase	

Table 4.4.3 Associated PULs of isolated fragments from electromobility shift assay. Description of PUL components and predicted inducing glycans are as described by Martens *et al,* 2011[53].

PUL Identification	Gene Identification	Predicted Function	Predicted Inducing Glycans
57	BT3324	Polysaccharide Lyase Family 8: chondroitinase	chondroitin sulfate, hyaluronan (dermatan sulfate)
	BT3328	hypothetical protein	
	BT3329	hypothetical protein	
	BT3330	hypothetical protein	
	BT3331	susD-like	
	BT3332	susC-like	
	BT3333	arylsulfatase A precursor (ASA)	
	BT3334	Hybrid two-component system regulator	
	BT3348	Glycoside Hydrolase Family 88: unsaturated glucuronyl hydrolase	
	BT3349	putative sulfatase yidJ	
	BT3350	Polysaccharide Lyase Family 8: chondroitinase	

Table 4.4.4 Associated PULs of isolated fragments from electromobility shift assay. Description of PUL components and predicted inducing glycans are as described by Martens *et al,* 2011[53].

PUL Identification	Gene Identification	Predicted Function	Predicted Inducing Glycans
68	BT3773	Glycoside Hydrolase Family 92: α-1,2-mannosidase	alpha-mannan, host *N*-glycans
	BT3774	Glycoside Hydrolase Family 38: α-mannosidase	
	BT3775	hypothetical protein	
	BT3776	hypothetical protein	
	BT3777	hypothetical protein	
	BT3778	hypothetical protein	
	BT3779	hypothetical protein	
	BT3780	hypothetical protein	
	BT3781	hypothetical protein	
	BT3782	Glycoside Hydrolase Family 76: putative membrane attached hydrolase	
	BT3784	Glycoside Hydrolase Family 92: α-1,2-mannosidase	
	BT3786	Hybrid two-component system regulator	
	BT3787	hypothetical protein	
	BT3788	susC-like	
	BT3789	susD-like	
	BT3790	hypothetical protein	
	BT3791	hypothetical protein	
	BT3792	Glycoside Hydrolase Family 76: α-1,6-mannosidase	

4.5 Discussion

In this study we investigated the transcriptional control of the polysaccharide utilization locus 55. Our results provide insight into other loci PUL 55 potentially interacts with in the context of the human colon. Assessing different growth conditions for *B.thetaiotaomicron* in a wild-type strain, a Δtdk and a *sus* knockout strain, we were able to observe growth differences in relation to the provided substrates. The wild-type and Δtdk strain demonstrated similar growth curves for glucose, maltose, pullulan and amylopectin. There were observed differences in the ability of these strains to use β-cyclodextrin, with Δtdk demonstrating a longer lag time. However, both strains were observed to have an overall longer lag time for β-cyclodextrin in comparison to other assessed substrates. This observation was also previously observed in growth experiments for the wild-type and *sus* mutants, where cultures were observed to have an increase in lag time in relation to an increase in substrate size (Figure 4.5.1)[50]. This difference in response time with regards to substrate length could reflect the time required for the cell to transition into an active induced state; thus allowing it to effectively utilize the sensed substrate in its environment.

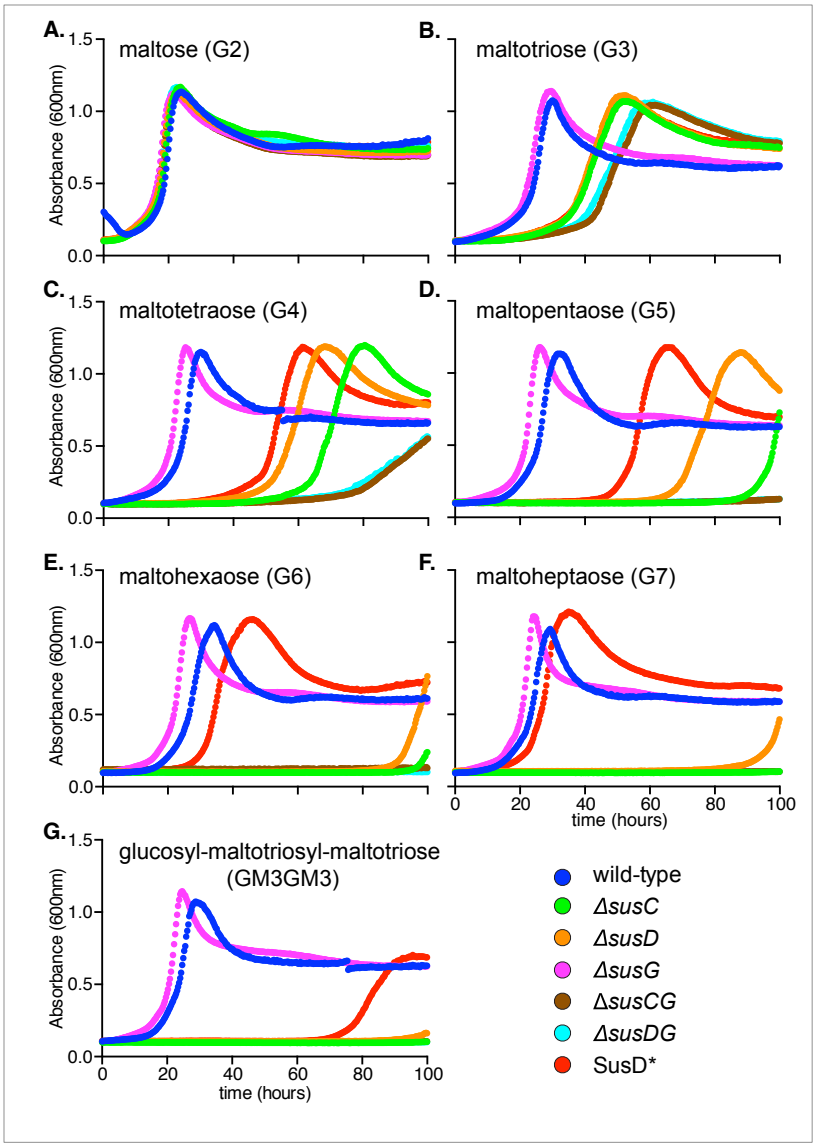


Figure 4.5.1 Growth curves of *sus* mutants grown on various maltooligosaccharides. These growth curves are reported in the supplementary material from Cameron *et al,* 2014[50]. Comparison of wild-type and mutant strains it can be observed that an increase in substrate length results in an increase in lag time.

The ability of the Δsus strain to grow moderately on amylopectin and pullulan reveals that additional genes are involved in utilizing these substrates and this function is derived from outside of the starch utilization system. At this point, we cannot definitively state that this ability arises from functions of genes as part of PUL 55. However, it does

indicate that creating knockouts of PUL 55 in the Δsus strain could provide further information into this uncertainty.

Using the electromobility shift assay we were able to observe regions of *B.thetaiotaomicron*'s genome that the response regulator BT_3302-5 is believed to bind to. Based on the interaction of BT_3302-5 with these observed genomic regions it is hypothesized that this regulator is involved in up-regulating proteins outside of PUL 55. It was expected that this regulator would also bind to its own PUL and therefore, it was unusual that we did not see this result in our experiment. Based on this preliminary observation, it is suggested that the function of PUL 55 does not necessarily lie in the discrete function of utilizing the predicted substrates in relation to its genetic elements (Table 4.5.1). Rather, we believe the product of this locus is to generate sugars that activate BT_3302-5, which then has a cascade effect of up-regulating other PULs.

BT_3302-5 was observed to bind to the gene BT_1277, a fucose permease that is a member of the characterized L-fucose locus[102]. This locus is predicted to be involved in utilizing fucosylated glycans and has been shown to play a role in microbial colonization of the colon[103]. The ability of BT_3302-5 to bind to this operon lends to the idea that PUL 55 contributes to the ability of this bacterium to retain its place in this ecosystem. Upregulating the L-fucose operon could also contribute to the effective utilization of fucosylated glycans that are derived from mucin glycans as an energy source[104].

BT_3302-5 was observed to bind to the gene BT_3778, a hypothetical protein that is a member of the PUL 68. Transcriptional profiling of *B.thetaiotaomicron* revealed that PUL 68 is predicted to utilize mannan oligosaccharides[53]. The gene BT_3301 (member of PUL 55) is also predicted to utilize mannan oligosaccharides and is a predicted GH76 α-1,6 mannanase (Table 4.5.1). Previous characterization of the gene BT_3773 (member of PUL 68) revealed its ability to hydrolyze mannobiose[105]. Based on this relationship it can be suggested that the product of the glycoside hydrolase BT_3301 could activate a locus that is primarily involved in utilizing this specific substrate. This ultimately would result in multiple pathways enabling this bacterium to effectively utilize mannose, which is commonly found in the colon due to its presence in plant cell walls and glycoproteins that are derived from host intestinal epithelium[106]. Alternatively, the relationship between PUL 55 and PUL 68 could exist because of crosstalk that is predicted to exist in these systems in *B.thetaiotaomicron*[107].

BT_3302-5 was observed to bind to the gene BT_3350, which is a predicted chondroitinase. This gene is a member of PUL 57 and is predicted to utilize the substrates chondroitin sulfate and hyaluronan (dermatan sulfate). This substrate is predicted to be used as well by a glycoside hydrolase in PUL 55. The gene BT_3292 is a predicted a β-glucuronidase and it believed to hydrolyze the substrate heparan sulfate (Table 4.5.1). It is believed that the product of BT_3292 could activate BT_3302 and ultimately upregulate genes, such as those from PUL 57 that would be able to hydrolyze proteoglycans.

BT_3302-5 was observed to bind to the gene BT_0212, a hypothetical protein and a member of PUL 4. This locus does not have a defined overall function and the preferred substrates are not predicted at this point in time. But it can be suggested that similar to the above loci, the preferred substrate of PUL 4 could be similar to the predicted substrates of PUL 55. This evidence could be used to investigate PUL 4 and define the functions of the hypothetical genes that make up this locus.

Lastly, BT_3302-5 was observed to bind to the gene BT_1907, a RNA polymerase sigma factor RpoS. RpoS is a transcription regulator found in gram-negative bacteria and has been shown to play a role in their ability to adapt stress and contributes to biofilm formation[108–110]. These functions contribute overall to the ability of *B.thetaiotaomicron* to successfully colonize the human colon.

Table 4.5.1 PUL 55 Genes. Predicted function and proposed glycans as described by by Martens *et al,* 2011[53].

Gene Identification	Predicted Function	GH Family	Predicted Glycan
BT_3292	β-glucuronidase	2	heparan sulfate
BT_3293	β-galatosidase	2	lactose
BT_3294	α-glucosidase	97	maltose
BT_3296	SusD-like	N/A	N/A
BT_3297	SusC-like	N/A	N/A
BT_3299	α-glucosidase	31	maltose
BT_3300	β-glucosidase	3	cellobiose
BT_3301	α-1,6 mannanase	76	mannan oligosaccharides
BT_3302	two-component system regulator	N/A	N/A

The specifics of the regulation of BT_3302-5 are not fully understood at this point. The observation of BT_3302-5 binding intragenetic regions of *B.thetaiotaomicron* genomic DNA, specifically regions not part of PUL55, reveal that lack of understanding of the specific transcriptional role it plays in the cell. The number of genes and operons that BT_3302-5 is potentially capable of regulating is not necessarily surprising considering reported levels of regulation by other response regulators. For example, analysis of the two-component system PhoR-PhoP from *Bacillus subtilus* it was observed that the response regulator controls the level of expression of 31 genes and 10 operons[111]. Based on the evidence here, we propose that PUL 55 is involved in a primary sensing of substrates (Figure 4.5.1). If substrates activate the two-component system in PUL 55, genes related to diverse substrate preference of PUL are unregulated. We believe the results here contribute to the understanding of the evolved capabilities that *B.thetaiotaomicron* has at sensing stimuli, but also more specifically responding to substrates in its environment.

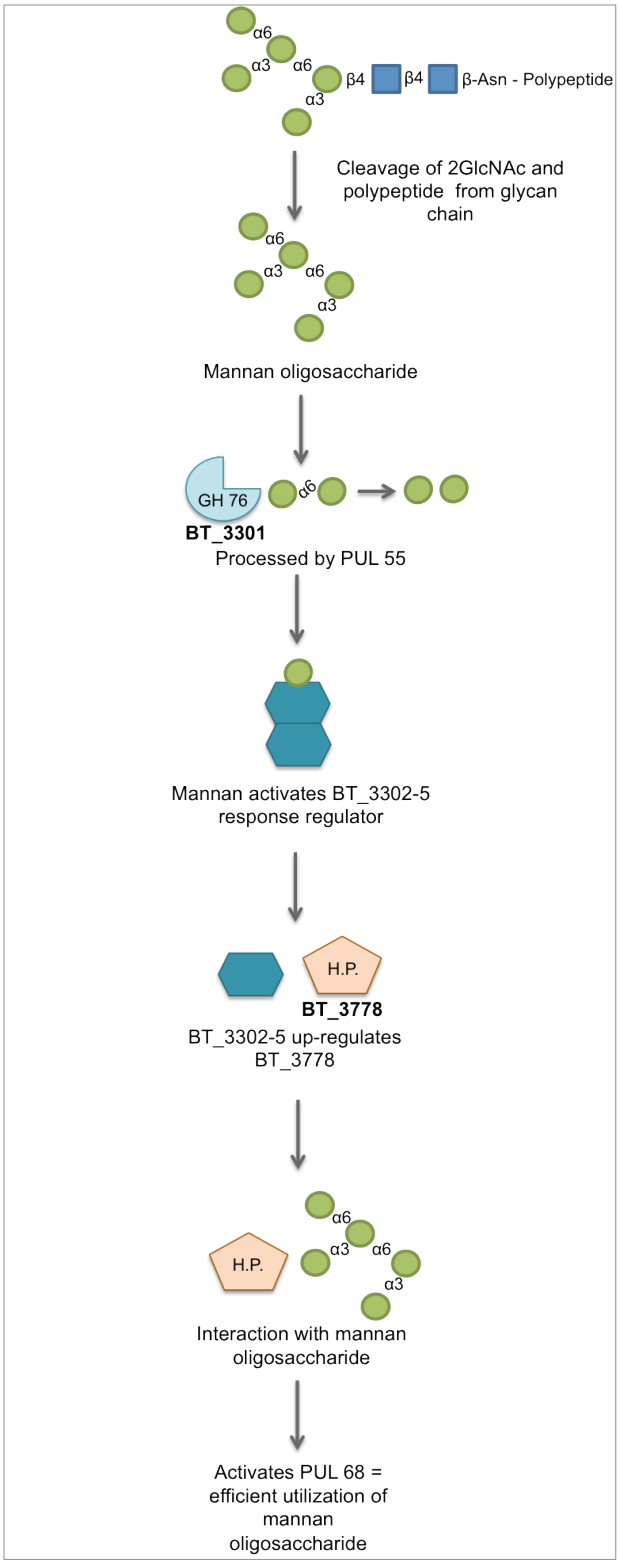


Figure 4.5.1 Proposed primary sensing of glycans by PUL 55. Example demonstrated through the regulation of PUL 68 predicted to be involved in mannan oligosaccharide sensing.

4.6 Conclusion

The human colon is dynamic environment with constant changing selective pressure influencing the colon microbiota. Changes in colon host physiology, derived substrates and population composition are factors that can define the metabolism of this population. It can be observed that as a result of this dynamic environment in order for an organism to thrive in this population it is necessary to have the abilities to adapt to these changes quickly and discretely. The investigation of the response regulator of the polysaccharide utilization locus 55 contributes evidence of the transcriptional abilities necessary in order to adapt to changes in dietary glycans. Our results suggest that the role of this specific locus in *B.thetaiotaomicron* is not necessarily solely dedicated to metabolizing glycans specific to that locus. Rather, we hypothesize that this locus is capable of working in cooperation with additional loci and is believed to contribute to the regulation of those loci. Growth curves of *B.thetaiotaomicron* and the assessment of a response regulator provide evidence that the utilization of dietary and host glycans is not necessarily dependent on a single locus. Observations of growth curves for the *B.thetaiotaomicron* Δsus strain on pullulan and amylopectin indicate that multiple systems may be required to utilize specific forms of starch. In addition to, observations of the interaction of the response regulator form PUL 55 demonstrates that utilization of substrates in this bacterium is believed to occur through multiple levels of transcriptional control involving multiple loci. This appears to be a genetic checkpoint for this bacterium. Allowing it the ability to sense a substrate that is available in its environment prior to dedicating energy to regulating genes for its utilization. Ultimately, this allows this bacterium to effectively utilize substrates in the surrounding colon environment. Alternatively, these results can

raise the possibility of crosstalk between multiple loci, resulting in multiple loci sensing and responding to a common substrate. Further investigation into the transcriptional function of PULs and their relationships to one another can provide valuable insight into how microbes respond to dietary elements in our colon.

Chapter 5
Conclusion and Future Work

5.1 Concluding Remarks

The work presented here has contributed to the understanding of the role the colon microbiota may have in relation to our diet. By investigating a specific PUL from the colon bacterium *B.thetaiotaomicron* we are able to infer the general function of bacterial organisms in the human colon and provide insight into how glycans are acquired and processed during this stage of digestion. The biochemical and modeling data here provides evidence for the hypothesis that multiple loci are believed to be involved in the utilization of starch in the colon. These loci are believed to work in cooperation with one another and function in complement with one another. Starch is a diverse polymer dependent on its source (rice, potato) and our results suggest that in order to be able to efficiently be able to use this polymer, *B.thetaiotaomicron* has developed discrete mechanisms in varying PULs. Our work on PUL 55 from *B.thetaiotaomicron* has helped to answer questions about how this bacterium acquires and process starch-derived glycans in the human colon.

Our analysis of the SusD-homolog BT_3296 provides evidence that this outer membrane protein has a similar ability as SusD by binding starch-based glycans. We suggest that its preferred substrate is summarily a cyclical starch glycan, but is an alternative to the preferred β-cyclodextrin of SusD. Structural differences were observed in two GH31 α-glucosidases and it is suggested that these differences contribute to this organism having the ability to efficiently utilize the complexity of starch oligosaccharides. BT_3299 has a number of residues in the active site that interacts with the modeled acarbose allowing for the observation that this family GH31 α-glucosidase has the ability to accommodate short-branched maltooligosaccharides. Where as the active site of BT_0339 with a modeled acarbose allows for the observation that this enzyme prefers short maltooligosaccharides. Collectively, these results suggest that this PUL is capable of acquiring starch-glycans from the colon lumen and contribute to the overall utilization of

this common colonic substrate. Its been a common belief that the ability to utilize starch in *B.thetaiotaomicron* is solely dependent on the starch utilization system[58]. However, our results here suggested an alternative scenario that when this bacterium encounters starch-derived glycans multiple loci are activated allowing or the optimal acquisition and processing.

In Chapter 3 we investigated the evolution of family 31 α-glucosidase with the goal of having a better understanding of the role these enzymes in starch digestion. Our evolutionary analysis explored evolutionary changes that occurred within this family of glycoside hydrolases and to related these divergences to observed structural differences in relation to predicted substrate preferences. Different rates of evolution at single amino acids are believed to be directly related to the ability of these enzymes to bind different starch-based glycans. Through phylogenetic analysis we are able to observe a number of evolutionary adaptions that we believe resulted in unique functions amongst this family of enzymes (MGAM and SI; BT_3299 and BT_0339) that allow for the efficient utilization of dietary starch in the small intestine and in the colon microbiota.

Lastly, in Chapter 4 we investigated the transcriptional control of PUL 55 through biochemical studies of a response regulator as part of its two-component system. This transcriptional element was observed to bind to a number of genetic regions in *B.thetaiotaomicron's* genomic DNA. These interactions reflect the ability of this bacterium to sense and respond to glycans in its environment. BT_3302-5 was observed to bind to a number of regions that have overlapping function with PUL 55 and thus it is suggested that the role of PUL 55 is to act as a primary sensing locus.

Kinetic studies of BT_3299 revealed this glycoside hydrolase has a general low level of activity and we were unable to determine individual kinetic values[66]. This activity maybe due to the control of this locus and reflect its true function in the cell. We hypothesize that PUL 55 has the ability to sense a range of glycans and posses a number of glycoside hydrolases and through this glycosidic activity release simple sugars that interact with

this PUL's regulatory system. Perhaps a by-product of the function of these hydrolases in PUL 55 is the initiation of PULs that are also able to utilize the sensed substrate. We hypothesize that this PUL is an example of the transcriptional control necessary for this bacterium to effectively utilize the array of glycans that are available in the colon. Further experimental evidence is necessary in order to support this proposed role of PUL 55.

The results presented here provide insight into the diverse colon environment and the mechanisms that are necessary for microorganisms to colonize and thrive in this populated system. Through the proposed mechanisms of PUL 55 we are able gain insight into the possible genetic control this bacterium must employ in order to sense, respond and acquire dietary and host glycans. This knowledge can provide guidance with future experiments and indicates areas that require further research to better understand *B.thetaiotaomicron's* role in the colon microbiota and the study of the human colon microbiome.

5.2 Future Studies

5.2.1 BT_3296 SusD-homolog

Biochemical and computational investigations of the SusD-homolog BT_3296 revealed its similarity to the characterized SusD protein. Both proteins demonstrate a common ability of binding starch-derived glycans with the predicted model of BT_3296 relating structural features to SusD. Experimental structural data would be a beneficial addition to this project, as it would demonstrate whether the reported modeled of BT_3296 is correct and confirm the structural relatedness of these proteins. Structural data would also provide further insight into the SusD-homolog group of enzymes and their conserved structural features. Investigating the transcriptional control of SusD-homologs in relation to their preferred substrates is also recommended as future studies. It is not well understood whether these SusD homologs function independently when binding colonic derived substrates or if they are capable of working in cooperation with one another. Investigating SusC-SusD homolog pairs is also recommended for future studies. It is believed that SusC-SusD pairs arise from gene duplication within *B.thetaiotaomicron's* genome; however, the crosstalk between these pairs and how they evolved is not well understood[58]. The interaction between SusC and SusD homologs is also not well understood at this time and is acknowledged as a gap in the understanding of the starch utilization system and polysaccharide utilization loci[50]. An understanding of the function of SusC and how it is capable of working in cooperation with SusD (as well as other SusC/SusD homolog pairs) will allow us to understand how substrates bind and interact with the outside of the bacterial cell. Ultimately, these proposed studies would allow us to understand how this bacterium functions in relation to acquiring dietary substrates from the colon lumen and how they are transported into the cell where they are used for energy.

5.2.2 Structural Studies of Family GH31 α-glucosidases from *Bacteroides thetaiotaomicron*

Structural studies of the α-glucosidases BT_3299 and BT_0339 provided insight into the capabilities of *B.thetaiotaomicron* at utilizing starch glycans. These structures

demonstrated the ability of this bacterium's enzymes to utilize a range of starch glycan forms and provided structural evidence that these enzymes are able of contributing to starch utilization. Further studies into these enzymes could provide necessary evidence of the specifics of the functionality of these enzymes, which will further support the reported biochemical data[66]. It is recommended that future structural studies include investigating the structure of these enzymes co-crystallized with various starch-derived glycans. This evidence could provide insight into how these enzymes bind starch-derived substrates and further support how these enzymes structurally differ in relation to their reported activity.

It is also suggested that the transcriptional control of these enzymes would be worth investigating. Understanding the specific substrate that induces the expression of these enzymes could be beneficial in understanding their role independently and in relation to one another. BT_0339 was previously observed to be up-regulated when grown on rhamnoglacturonan I and II[53]. These substrates have not been reported to be utilized by α-glucosidases and this demonstrates the complexity of the utilization of dietary glycans[53]. This reported activity could be a place to start in the assessment of substrates that regulate this gene. BT_3299 has not been reported to be up-regulated by a specific dietary or mucosal glycan, but assessing glycans derived from substrates predicted to be used by PUL 55 provides some context for the possible glycans that could regulate its activity. Investigating the specific substrates that induce the expression of these family GH31 enzymes could provide insight into additional glycans these enzymes are capable of utilizing.

5.2.3 Evolution of Family 31 α-glucosidases

Assessing the structural evolution of family GH31 α-glucosidases provided predictions for structurally solved enzymes within this family. We proposed a number of structural divergences in the comparison of eukaryote and prokaryote GH31 α-glucosidases. These observations could be further supported with additional structures for family GH31 α-glucosidases that come from organisms of interest based on our generated phylogenetic tree of characterized α-glucosidases. Another layer to this story would have been relating

the activity of this family of enzymes in relation to how they cluster in the phylogenetic tree. Unfortunately, there is not a good level of understanding of substrate preference and specificity within this family, despite a fair number of characterized GH31 α-glucosidases. Investigating the specific function and substrate preferences of GH31 α-glucosidases from key areas of the phylogenetic tree could provide further evidence of how these enzymes have evolved in relation to the ability to hydrolyze specific sugars. Understanding the relationship between the selective pressures that influences evolutionary change is important in understanding of the evolution functional abilities.

Mutational studies are also recommended for future studies. The introduction of single amino acid changes into the structurally solved α-glucosidases in relation to the predicted evolutionary changes could provide evidence for these predictions. For example, introducing a region to similar to the eukaryotic signature sequence of 'ppY' into the prokaryotic enzymes could provide interesting biochemical data to support this hypothesis.

5.2.4 Transcriptional control of PULs

Growth experiments provided evidence for the hypothesis that multiple operons are involved in the utilization of starch-derived glycans. Future studies are recommended to include knockout experiments of targeted PUL 55 genes and investigating this influence in growth experiments. These experiments should complement the completed growth experiments using starch-derived glycans; however, it is also recommended that additional substrates should be explored. Based on the results for the up-regulated genes from BT_3302-5, it is suggested that additional substrates should be explored in growth experiments in relation to PUL 55. For example, BT_3302-5 is believed to interact with a hypothetical gene from PUL68, which is predicted to utilize mannan oligosaccharides. It is then suggested that this would be an example of an additional sugar to investigate in relation to created PUL 55 knockout strains.

Our results of the response regulator from PUL 55 suggest the function of PULs may lie in cooperation or interaction with additional PULs. PUL 55 was observed to up-regulate

operons that are not included in its own loci boundaries. Currently, there is no evidence from this study that suggests BT_3302-5 up-regulates its own PUL. Further studies are necessary to fully understand the interaction of the response regulator from PUL 55 with the observed unregulated genes and determine its specific role in regulating PUL 55. It is suggested that transcriptional profiling experiments be explored in relation to BT_3302-5 and to confirm the level of regulation this transcription factor plays in relation to the suggested genes. Furthermore, it is unclear if the regulations of these genes are dependent on BT_3302-5 or if it contributes to their regulation in presence of specific sugars. Understanding this connection would be beneficial to understanding the complexity of the regulation of *B.thetaiotaomicron's* genes in relation to the exposure of dietary and mucosal glycans. Furthermore, continued investigations of the other components of PUL 55 will provide further insight into this particular PUL but also how they relate to other co-regulated loci.

5.3 Colon microbiota and Digestive Disease States

The focus of this thesis has been on the relationship that exists between the colon microbiota and our diet and has highlighted the complexity of this interaction. The profound presence of the colon microbiota in the digestive system results in a strong influence on a number of metabolic syndromes. This interaction is believed to lie in the relationships between our diet, the colon microbiota and digestive diseases (Figure 6.3.1). This population is quite complex and diverse and its composition is greatly influenced by numerous variables. The health of this population can rely on exposure to dietary foods, competition in the colon and the overall health of the colon environment. A number of digestive disease states have been shown to be influenced by the general health of the colon microbiota. This includes obesity, irritable bowel syndrome and inflammatory bowel disease[10,24,112–115].

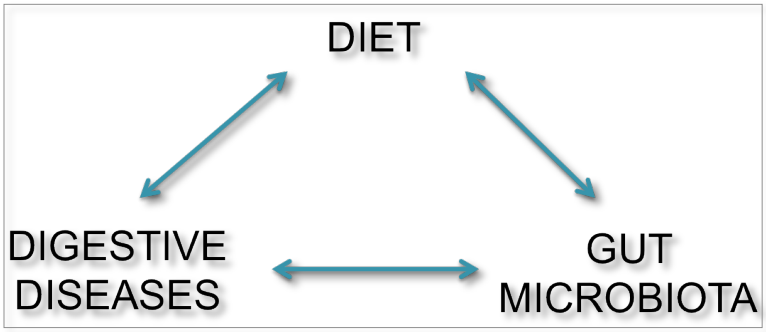


Figure 6.3.1 Relationship between diet, digestive diseases and the colon microbiota.

Understanding the connection between the colon microbiota and digestive disease is highly dependent upon having a good understanding of the relationship between our diet and this population. The research presented here contributes to the body of evidence that this population is capable of utilizing a number of dietary glycans, which has been reported to influence this population's health and composition. It also reveals that further research is necessary into understanding this connection between the colon microbiota, our diet and ultimately digestive diseases. It is believed that understanding these connections will aid in mitigating and minimizing these ailments and providing an understanding that will contribute to a well rounded health care approach to gastrointestinal illnesses.

5.4 Manipulation of the Colon Microbiota

The knowledge of the interaction of the colon microbiota with our diet allows for a number of opportunities in regards to manipulating this population to improve general gastrointestinal health, but also to limit the occurrence or treatment of the previous mentioned digestive diseases. An emerging area of treatment for the colon microbiota and digestive diseases is fecal microbiota transplant. A fecal microbiota transplant (FMT) is the administration of a microbial solution or fecal solution from a donor that is transplanted into a recipient's lower gastrointestinal tract[116] The purpose of this therapy is to re-establish or change the recipient's colon microbiota to improve the health of this population as a treatment for a gastrointestinal illness[117]. FMT has been shown to an

effective at treating a variety of gastrointestinal diseases that include inflammatory bowel disease, irritable bowel syndrome, metabolic disorders and *Clostridium difficile* infection[117–119]. An important component of this treatment is the recovery of the colon microbiota following this transplantation. Ensuring the transplanted microbiota's metabolism is functional through diet is one controllable variable that allows for a successful recovery. Understanding the complex metabolism of the strains used in this treatment and their relationship with our diet is an important element of using this treatment[120]. Continued research needs to be completed on this treatment in order for there to be concrete evidence that this is an effective treatment for these illnesses and preventing reoccurrences.

Another area of research that has a good possibility for treatment of gastrointestinal illness related to the colon microbiota is the use of 'designer glycans'. This treatment involves individuals ingest glycan specific foods or food molecules with the purpose of stimulating the growth of specific strains in the colon microbiota[121]. This approach of using prebiotics to stimulate the colon microbial population and to shape its composition has been historically controversial with it effectiveness at this role[122]. In order, for this approach to beneficial for individuals suffering from gastrointestinal illnesses there needs to continued research supporting that the specific glycans can encourage specific metabolic traits a species and encouraging its presence in the colon microbiota.

A focus of colon microbiota research that has prompted interest in the general public is personalized colon microbiota characterization. Companies such as uBiome provide a service to the general public where they will sequence an individual's colon microbiota[123]. This is an interesting opportunity to have a catalogue of the species present in a range of the human population's colon microbiota. This business venture demonstrates the importance of continued research in understanding the metabolic capabilities of bacterial species present in the human colon microbiota and the relationship with our diet. This knowledge will allow us to understand the interaction our diet has in relation to the specific composition of a profiled colon microbiota and how we can create customizable therapeutic options through glycan metabolism

Bibliography

1. Turnbaugh, P. J. *et al.* The human microbiome project: exploring the microbial part of ourselves in a changing world. *Nature* **449,** 804–810 (2007).

2. Cho, I. & Blaser, M. J. The human microbiome: at the interface of health and disease. *Nat. Rev. Genet.* **13,** 260–70 (2012).

3. Gill, S. R. *et al.* Metagenomic analysis of the human distal gut microbiome. *Science (80-.).* **312,** 1355–1359 (2006).

4. Lozupone, C. a, Stombaugh, J. I., Gordon, J. I., Jansson, J. K. & Knight, R. Diversity, stability and resilience of the human gut microbiota. *Nature* **489,** 220–30 (2012).

5. 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).

6. Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. Host-Bacterial Mutualism in the Human Intestine. *Science (80-.).* **307,** 1915–1920 (2005).

7. O'Hara, A. M. & Shanahan, F. The gut flora as a forgotten organ. *EMBO Rep.* **7,** 688–693 (2006).

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

9. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. **464,** (2010).

10. Round, J. L. & Mazmanian, S. K. The gut microbiota shapes intestinal immune responses during health and disease. *Nat. Rev. Immunol.* **9,** 313–23 (2009).

11. 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).

12. Wardwell, L. H., Huttenhower, C. & Garrett, W. S. Current concepts of the intestinal microbiota and the pathogenesis of infection. *Curr. Infect. Dis. Rep.* **13,** 28–34 (2011).

13. Scott, K. P., Gratz, S. W., Sheridan, P. O., Flint, H. J. & Duncan, S. H. The influence of diet on the gut microbiota. *Pharmacol. Res.* (2012).

14. Turnbaugh, P. J., Bäckhed, F., Fulton, L. & Gordon, J. I. Diet-Induced Obesity Is Linked to Marked but Reversible Alterations in the Mouse Distal Gut Microbiome. *Cell Host Microbe* **3,** 213–223 (2008).

15. Ley, R., Turnbaugh, P., Klein, S. & Gordon, J. Microbial ecology: human gut microbes associated with obesity. *Nature* **444,** 1022–3 (2006).

16. Human Microbiome Project Consortium, T. Structure, function and diversity of the healthy human microbiome. *Nature* **486,** 207–214 (2012).

17. Clemente, J. C., Ursell, L. K., Parfrey, L. W. & Knight, R. The impact of the gut microbiota on human health: an integrative view. *Cell* **148,** 1258–70 (2012).

18. Comstock, L. E. & Coyne, M. J. Bacteroides thetaiotaomicron: a dynamic, niche-adapted human symbiont. *Bioessays* **25,** 926–9 (2003).

19. Xu, J. *et al.* A Genomic View of the Human – Bacteroides thetaiotaomicron Symbiosis. *Science (80-.).* **299,** 2074–2076 (2003).

20. Zocco, M. A., Ainora, M. E., Gasbarrini, G. & Gasbarrini, A. Bacteroides thetaiotaomicron in the gut: Molecular aspects of their interaction. *Dig. Liver Dis.* **39,** 707–712 (2007).

21. D'Elia, J. N. & Salyers, A. A. Contribution of a neopullulanase, a pullulanase, and an alpa-glucosidase to growth of Bacteroides thetaiotaomicron on starch. *J. Bacteriol.* **178,** 7173–7179 (1996).

22. Wexler, H. M. Bacteroides: the Good, the Bad, and the Nitty-Gritty. *Clin. Microbiol. Rev.* **20,** 593–621 (2007).

23. Salyers, A. A., Gupta, A., Wang, Y., D'Elia, J. N. & Salyers, A. A. Human intestinal bacteria as reservoirs for antibiotic resistance genes. *Trends Microbiol.* **12,** 412–416 (2004).

24. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444,** 1027–31 (2006).

25. Turnbaugh, P. *et al.* Obesity alters gut microbial ecology. *PNAS* (2005).

26. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457,** 480–484 (2009).

27. Wilson G. Pond, Buford L. Nichols, D. L. B. *Adequate Food for All: Culture, Science and Technology in the 21st Century*. (CRC Press Taylor and Francis Group, 2009).

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

29. Englyst, H. N., Cummings, J. H., Trowell, M. & Southgate, D. Dietary fiber and resistant starch. *Am. J. Clin. Nutr.* **46,** 873–874 (1987).

30. McNeil, N. I. The contribution of the large intestine to energy supplies in man. *Am. J. Clin. Nutr.* **39,** 338–42 (1984).

31. Hooper, L. V., Midtvedt, T. & Gordon, J. I. How host-microbial interactions shape the nutrient environment of the mammalian intestine. *Annu. Rev. Nutr.* **22,** 283–307 (2002).

32. Koropatkin, N. M., Cameron, E. A. & Martens, E. C. How glycan metabolism shapes the human gut microbiota. *Nat. Rev. Microbiol.* **10,** 323–35 (2012).

33. Englyst HN, Kingman SM, C. J. Classification and measurement of nutritionally important starch fractions. *Eur. J. Clin. Nutr.* S33–50 (1992).

34. Butterworth, P. J., Warren, F. J. & Ellis, P. R. Human alpha-amylase and starch digestion: An interesting marriage. *Starch/Staerke* **63,** 395–405 (2011).

35. Ao, Z. *et al.* Evidence of native starch degradation with human small intestinal maltase-glucoamylase (recombinant). *FEBS Lett.* **581,** 2381–8 (2007).

36. Nichols, B. L. *et al.* The maltase-glucoamylase gene: common ancestry to sucrase-isomaltase with complementary starch digestion activities. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 1432–7 (2003).

37. Quezada-Calvillo, R. *et al.* Luminal substrate 'brake' on mucosal maltase-glucoamylase activity regulates total rate of starch digestion to glucose. *J. Pediatr. Gastroenterol. Nutr.* **45,** 32–43 (2007).

38. Van Beers, E. H. *et al.* Lactase and sucrase-isomaltase gene expression during Caco-2 cell differentiation. *Biochem. J.* **308 (Pt 3),** 769–775 (1995).

39. Hunziker, W., Spiess, M., Semenza, G. & Lodish, H. F. The sucrase-isomaltase complex: primary structure, membrane-orientation, and evolution of a stalked, intrinsic brush border protein. *Cell* **46,** 227–34 (1986).

40. Naumoff, D. G. Structure and evolution of the mammalian maltase-glucoamylase and sucrase-isomaltase genes. *Mol. Biol.* **41,** 962–973 (2007).

41. Lin, A. H.-M. H.-M. *et al.* Starch Source Influences Dietary Glucose Generation at the Mucosal -Glucosidase Level. *J. Biol. Chem.* **287,** 36917–21 (2012).

42. Lin, A. H. *et al.* Branch pattern of starch internal structure influences the glucogenesis by mucosal Nt-maltase-glucoamylase. *Carbohydr. Polym.* **111,** 33–40 (2014).

43. Zhang, C. *et al.* Dietary modulation of gut microbiota contributes to alleviation of both genetic and simple obesity in children. *EBioMedicine* **2,** 966–982 (2015).

44. Ze, X., Duncan, S. H., Louis, P. & Flint, H. J. Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. *ISME J.* **6,** 1535–1543 (2012).

45. Topping DL, C. P. Short-Chain Fatty Acids and Human Colonic Function: Roles of resistant starch and nonstarch polysaccharides. *Physiol. Rev.* 1031–1064 (2001).

46. Englyst, H., Kingman, S. & Cummings, J. Classification and measurement of nutritionally important starch fractions. *Eur. J. Clin. Nutr.* **46,** Suppl 2:S33–50 (1992).

47. Reeves, A. R. *et al.* A Bacteroides thetaiotaomicron outer membrane protein that is essential for utilization of maltooligosaccharides and starch . These include : A Bacteroides thetaiotaomicron Outer Membrane Protein That Is Essential for Utilization of Maltooligosaccharides. *J. Bacteriol.* **178,** 823–830 (1996).

48. Salyers, A. A. & Cho, K. H. Biochemical Analysis of Interactions between Outer Membrane Proteins That Contribute to Starch Utilization by Bacteroides thetaiotaomicron Biochemical Analysis of Interactions between Outer Membrane Proteins That Contribute to Starch Utilization by Bacter. *J. Appl. Microbiol.* **183,** 7224–7230 (2001).

49. Shipman, J. A., Berleman, J. E., Abigail, A. & Salyers, A. A. Characterization of Four Outer Membrane Proteins Involved in Binding Starch to the Cell Surface of Bacteroides thetaiotaomicron Characterization of Four Outer Membrane Proteins Involved in Binding Starch to the Cell Surface of Bacteroides thetaiotaomicron. *J. Bacteriol.* **182,** 5365–5372 (2000).

50. Cameron, E. A. *et al.* Multifunctional Nutrient-Binding Proteins Adapt Human Symbiotic Bacteria for Glycan Competition in the Gut by Separately Promoting. *MBio* **5,** 1–14 (2014).

51. 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* **18,** 200–15 (2010).

52. Bolam, D. N. & Koropatkin, N. M. Glycan recognition by the Bacteroidetes Sus-like systems. *Curr. Opin. Struct. Biol.* **22,** 563–9 (2012).

53. Martens, E. C. *et al.* Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol.* **9,** 1–16 (2011).

54. Cameron, E. A. The Structure and Function of the Starch Utilization System in Bacteroides thetaiotaomicron. **29148,** (2009).

55. Martens, E. C., Roth, R., Heuser, J. E. & Gordon, J. I. Coordinate regulation of glycan degradation and polysaccharide capsule biosynthesis by a prominent human gut symbiont. *J. Biol. Chem.* **284,** 18445–57 (2009).

56. Hemsworth, G. R., Dejean, G., Davies, G. J. & Brumer, H. Learning from microbial strategies for polysaccharide degradation. *Biochem. Soc. Trans.* **44,** 94–108 (2016).

57. 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).

58. Martens, E. C., Koropatkin, N. M., Smith, T. J. & Gordon, J. I. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J. Biol. Chem.* **284,** 24673–7 (2009).

59. Bjursell, M. K., Martens, E. C. & Gordon, J. I. Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, Bacteroides thetaiotaomicron, to the suckling period. *J. Biol. Chem.* **281,** 36269–36279 (2006).

60. Larsbrink, J. *et al.* A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature* **506,** 498–502 (2014).

61. Cuskin, F. *et al.* Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature* **517,** 165–169 (2015).

62. Martens, E. C., Kelly, A. G., Tauzin, A. S. & Brumer, H. The Devil Lies in the Details: How Variations in Polysaccharide Fine-Structure Impact the Physiology and Evolution of Gut Microbes. *J. Mol. Biol.* **426,** 3851–3865 (2014).

63. Nichols, B. L. *et al.* Mucosal Maltase-Glucoamylase Plays a Crucial Role in Starch Digestion and Prandial Glucose Homeostasis of Mice 1 – 3. (2009).

64. Quezada-Calvillo, R. *et al.* Luminal starch substrate 'brake' on maltase-glucoamylase activity is located within the glucoamylase subunit. *J. Nutr.* **138,** 685–92 (2008).

65. Pontremolil, C. *et al.* Natural Selection at the Brush-Border: Adaptations to Carbohydrate Diets in Humans and Other Mammals. *Genome Biol. Evol.* **7,** 2569–2584 (2015).

66. Chaudet, M. M., Allen, J.-L. & Rose, D. R. Expression and purification of two Family GH31 α-glucosidases from Bacteroides thetaiotaomicron. *Protein Expr. Purif.* **86,** 135–41 (2012).

67. Tan, K. *et al.* Novel α-glucosidase from human gut microbiome: substrate specificities and their switch. *FASEB J.* **24,** 3939–49 (2010).

68. Van Duyne, G. D., Standaert, R. F., Karplus, A. P., Schreiber, S. L. & Clardy, J. Atomic Structures of the Human Immunphilin FKBP-12 Complexes wtih FK506 and Rapamycin. *J. Mol. Biol.* **229,** 105–125 (1993).

69. Luft, J. R., Collins, R. J., Fehrman, N. A., Lauricella, A. M., Veatch, C. K. & DeTitta, G. T. A deliberate approach to screening for initial crystallization conditions of biological macromolecules. *J. Struct. Biol.* **142,** 170–179 (2003).

70. Grochulski, P., Fodje, M., Gorin, J., Labiuk, S. & Berg, R. Beamline 08ID-1, the prime beamline of the Canadian Macromolecular Crystallography Facility. *J. Synchrotron Radiat.* **18,** 681–4 (2011).

71. Otwinowski, Z. & Minor, W. Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol.* **276,** 307–326 (1997).

72. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 213–221 (2010).

73. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.* **60,** 2126–32 (2004).

74. Lee, Y. C. Fluorescence spectrometry in studies of carbohydrate-protein interactions. *J. Biochem.* **121,** 818–25 (1997).

75. 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* **16,** 1105–15 (2008).

76. Koropatkin, N., Martens, E. C., Gordon, J. I. & Smith, T. J. Structure of a SusD Homologue , BT1043 , Involved in Mucin O-Glycan Utilization in a Prominent Human Gut Symbiont. *Biochemistry* **48,** 1532–1542 (2009).

77. Kelley, L. a, Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10,** 845–858 (2015).

78. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38,** W529–W533 (2010).

79. Celniker, G. *et al.* ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. *Isr. J. Chem.* **53,** 199–206 (2013).

80. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28,** 235–242 (2000).

81. Ernst, H. a. *et al.* Structure of the Sulfolobus solfataricus α-Glucosidase: Implications for Domain Conservation and Substrate Recognition in GH31. *J. Mol. Biol.* **358,** 1106–1124 (2006).

82. Lovering, A. L., Lee, S. S., Kim, Y.-W., Withers, S. G. & Strynadka, N. C. J. Mechanistic and structural analysis of a family 31 alpha-glycosidase and its glycosyl-enzyme intermediate. *J. Biol. Chem.* **280,** 2105–2115 (2005).

83. Sim, L., Quezada-Calvillo, R., Sterchi, E. E., Nichols, B. L. & Rose, D. R. Human intestinal maltase-glucoamylase: crystal structure of the N-terminal catalytic subunit and basis of inhibition and substrate specificity. *J. Mol. Biol.* **375,** 782–92 (2008).

84. Sim, L. *et al.* Structural basis for substrate selectivity in human maltase-glucoamylase and sucrase-isomaltase N-terminal domains. *J. Biol. Chem.* **285,** 17763–70 (2010).

85. Kitamura, M. *et al.* Structural and functional analysis of a glycoside hydrolase family 97 enzyme from Bacteroides thetaiotaomicron. *J. Biol. Chem.* **283,** 36328–37 (2008).

86. Chaudet, M. M. & Rose, D. R. Alternative starch utilization system from the human gut bacterium Bacteroides thetaiotaomicron. *Can. J. Biochem. Cell Biol.* **94,** 1–6 (2016).

87. Ren, L. *et al.* Structural insight into substrate specificity of human intestinal maltase-glucoamylase. *Protein Cell* **2,** 827–36 (2011).

88. Tagami, T. *et al.* Molecular basis for the recognition of long-chain substrates by plant alpha-glucosidases. *J. Biol. Chem.* **288,** 19296–19303 (2013).

89. Janeček, Š., Svensson, B. & MacGregor, E. A. A remote but significant sequence homology between glycoside hydrolase clan GH-H and family GH31. *FEBS Lett.* **581,** 1261–1268 (2007).

90. Konagurthu, A. S. *et al.* MUSTANG-MR structural sieving server: Applications in protein structural analysis and crystallography. *PLoS One* **5,** (2010).

91. Price, M. N., Dehal, P. S. & Arkin, a. P. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Mol. Biol. Evol.* **26,** 1641–1650 (2009).

92. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS One* **5,** e9490 (2010).

93. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28,** 1647–9 (2012).

94. Gu, X. *et al.* An Update of DIVERGE Software for Functional Divergence Analysis of Protein Family. *Mol. Biol. Evol.* **30,** 1713–1719 (2013).

95. Gu, X. & Vander Velden, K. DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family. *Bioinformatics* **18,** 500–501 (2002).

96. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acid Res.* **32,** 1792–1797 (2004).

97. Dhital, S., Lin, A. H. M., Hamaker, B. R., Gidley, M. J. & Muniandy, A. Mammalian Mucosal α-Glucosidases Coordinate with α-Amylase in the Initial Starch Hydrolysis Stage to Have a Role in Starch Digestion beyond Glucogenesis. *PLoS One* **8,** 1–13 (2013).

98. Merritt, D. A. & Karn, R. C. in *Advances in Human Genetics* 135–234 (1977).

99. Krell, T. *et al.* Bacterial sensor kinases: diversity in the recognition of environmental signals. *Annu. Rev. Microbiol.* **64,** 539–59 (2010).

100. Jung, K., Fried, L., Behr, S. & Heermann, R. Histidine kinases and response regulators in networks. *Curr. Opin. Microbiol.* **15,** 118–24 (2012).

101. Van den Ent, F. & Lowe, J. Expression of selenomethionine substituted proteins in non-methionine auxotrophic E.coli. 8–9 (2003).

102. Hooper, L. V, Xu, J., Falk, P. G., Midtvedt, T. & Gordon, J. I. A molecular sensor that allows a gut commensal to control its nutrient foundation in a competitive ecosystem. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 9833–9838 (1999).

103. Beccerra, J. E., Yebra, M. J. & Monedero, V. An l-Fucose Operon in the Probiotic Lactobacillus rhamnosus GG Is Involved in Adaptation to Gastrointestinal Conditions. *Appl. Environ. Microbiol.* **81,** 3880–8 (2015).

104. Coyne, M. J., Reinap, B., Lee, M. M. & Comstock, L. E. Human symbionts use a host-like pathway for surface fucosylation. *Science* **307,** 1778–1781 (2005).

105. Zhu, Y. *et al.* Mechanistic insights into a Ca2+-dependent family of alpha-mannosidases in a human gut symbiont. *Nat. Chem. Biol.* **6,** 125–32 (2010).

106. Ladeveze, S. *et al.* Role of glycoside phosphorylases in mannose foraging by human gut bacteria. *J. Biol. Chem.* **288,** 32370–32383 (2013).

107. Perry, J., Koteva, K. & Wright, G. Receptor domains of two-component signal transduction systems. *Mol. Biosyst.* **7,** 1388–98 (2011).

108. Battesti, A., Majdalani, N. & Gottesman, S. The RpoS-Mediated General Stress Response in *Escherichia coli* *. *Annu. Rev. Microbiol.* **65,** 189–213 (2011).

109. Schellhorn, H. . Elucidating the function of the RpoS regulon. *Future Microbiol.* **9,** 497–507 (2014).

110. Lévi-Meyrueis, C. *et al.* Repressor activity of the RpoS/σS-dependent RNA polymerase requires DNA binding. *Nucleic Acids Res.* 1–13 (2015).

111. Birck, C., Chen, Y., Hulett, F. M. & Samama, J. P. The crystal structure of the phosphorylation domain in PhoP reveals a functional tandem association mediated by an asymmetric interface. *J. Bacteriol.* **185,** 254–261 (2003).

112. Shanahan, F. The colonic microbiota and colonic disease. *Curr. Gastroenterol. Rep.* **14,** 446–52 (2012).

113. Ding, H. *et al.* The gut microbiota as an environmental factor that regulates fat storage. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 15718–23 (2004).

114. Manco, M., Putignani, L. & Bottazzo, G. F. Gut microbiota, lipopolysaccharides, and innate immunity in the pathogenesis of obesity and cardiovascular risk. *Endocr. Rev.* **31,** 817–844 (2010).

115. Seksik, P. *et al.* Review article: the role of bacteria in onset and perpetuation of inflammatory bowel disease. *Aliment. Pharmacol. Ther.* **24 Suppl 3,** 11–8 (2006).

116. Gupta, S., Allen-Vercoe, E. & Petrof, E. O. Fecal microbiota transplantation: in perspective. *Therap. Adv. Gastroenterol.* **9,** 1–11 (2015).

117. Matsuoka, K., Mizuno, S., Hayashi, A. & Hisamatsu, T. Fecal Microbiota Transplantation for Gastrointestinal Diseases. *Keio J Med* 69–74 (2014).

118. Drekonja, D. *et al.* Fecal microbiota transplantation for clostridium difficile infection a systematic review. *Ann. Intern. Med.* **162,** 630–638 (2015).

119. Anderson, J. L., Edney, R. J. & Whelan, K. Systematic review: Faecal microbiota transplantation in the management of inflammatory bowel disease. *Aliment. Pharmacol. Ther.* **36,** 503–516 (2012).

120. Seekatz, A. M., Aas, J. & Gessert, C. E. Recovery of the Gut Microbiome following Fecal Microbiota. **5,** 1–9 (2014).

121. Gibson, G. R., Probert, H. M., Loo, J. Van, Rastall, R. a & Roberfroid, M. B. Dietary modulation of the human colonic microbiota: updating the concept of prebiotics. *Nutr. Res. Rev.* **17,** 259–75 (2004).

122. Licht, T. R., Ebersbach, T. & Frokiaer, H. Prebiotics for prevention of gut infections. *Trends Food Sci. Technol.* **23,** 70–82 (2012).

123. Costandi, M. Citizen microbiome. *Nat. Biotechnol.* **31,** 90–90 (2013).

Appendix

A1.1 Electromobility shift assay fragment isolated clones.

Table S1 Clone identification from BT_3302-5 electromobility shift assay. Clone ID refers to isolated fragments cloned and sent for sequencing analysis. Multiple clone hits of the same genetic region are highlighted and grouped by colour.

	Position			
Clone ID	5'	3'	Size	Gene
3302-5P.1	1587813	1587911	98	BT_1277
3302-5P.2	1587813	1587911	98	BT_1277
3302-5P.10	1587812	1587911	99	BT_1277
3302-5P.11	1587812	1587911	99	BT_1277
3302-5P.13	1587812	1587911	99	BT_1277
3302-5P.4	2138915	2138975	60	BT_1734
3302-5P.5	3814427	3814631	204	BT_3016
3302-5P.6	4903147	4903304	157	BT_3778
3302-5P.9	4903147	4903305	158	BT_3778
3302-5P.7	4320944	4321066	122	BT_3350
3302-5P3.7	4320943	4321066	123	BT_3350
3302-5P3.8	4320943	4321066	123	BT_3350
3302-5P.12	2470143	2470289	146	BT_1967
3302-5P.16	936546	936664	118	BT_0754
3302-5P2.8	14047	14207	160	BT_0015
3302-5P3.1	235653	236051	398	BT_0212
3302-5P3.10	235653	236051	398	BT_0212
3302-5P3.16	235653	236051	398	BT_0212
3302-5P3.4	2396348	2396442	94	BT_1907
3302-5P3.15	2396348	2396442	94	BT_1907
3302-5P3.12	4954057	4954246	189	BT_3806
3302-5P3.13	4954057	4954246	189	BT_3806
3302-5P3.14	4954057	4954246	189	BT_3806
3302-5P4.11	1349127	1349213	86	BT_1062
3302-5P4.1	2513060	2513165	105	BT_1998
3302-5P2.6	2723016	2723128	112	BT_2165
3302-5P3.9	3421511	3421657	146	BT_2761
3302-5P4.12	3421511	3421657	146	BT_2761
3302-5P4.14	3421511	3421657	146	BT_2761
3302-5P4.15	3421511	3421657	146	BT_2761
3302-5P4.16	3421511	3421657	146	BT_2761
3302-5P4.3	3421511	3421657	146	BT_2761
3302-5P4.4	3421511	3421657	146	BT_2761
3302-5P4.7	3421511	3421657	146	BT_2761
3302-5P4.8	3421511	3421657	146	BT_2761
3302-5P3.2	3421511	3421657	146	BT_2761
3302-5P3.3	3421511	3421657	146	BT_2761

A1.2 Raw Data from the *Bacteroides thetaiotaomicron* Growth Experiments

Table S2 Optical density of wild-type strain of *Bacteroides thetaiotaomicron.*

Time (hours)	Carbon Source				
	Glucose	Maltose	Pullulan	β-cyclodextrin	Amylopectin
24	0.01	0.01	0.01	0.001	0.01
48	0.195	0.189	0.183	0.013	0.162
72	0.382	0.472	0.249	0.007	0.262
96	0.42	0.88	0.259	0.008	0.353
120	0.462	1.101	0.736	1.08	0.351
140	0.378	0.915	0.542	0.42	0.3
160	0.442	0.97	0.518	0.38	0.315

Table S3 Optical density of Δtdk strain of *Bacteroides thetaiotaomicron.*

Time (hours)	Carbon Source				
	Glucose	Maltose	Pullulan	β-cyclodextrin	Amylopectin
24	0.01	0.01	0.01	0.02	0.01
48	0.384	0.253	0.113	0.029	0.191
72	0.376	0.46	0.232	0.01	0.264
96	0.402	0.754	0.282	0.007	0.39
120	0.432	0.82	0.648	0.04	0.364
140	0.408	0.82	0.562	0.586	0.274
160	0.442	0.835	0.518	0.486	0.28

Table S4 Optical density of Δsus strain of *Bacteroides thetaiotaomicron.*

	Carbon Source				
Time (hours)	Glucose	Maltose	Pullulan	β-cyclodextrin	Amylopectin
24	0.02	0.02	0.02	0.02	0.02
48	0.129	0.131	0.14	0.036	0.195
72	0.37	0.248	0.24	0.017	0.262
96	0.408	0.46	0.101	0.019	0.14
120	0.442	1.05	0.102	0.03	0.14
140	0.408	0.736	0.091	0.0311	0.114
160	0.438	0.778	0.098	0.036	0.124

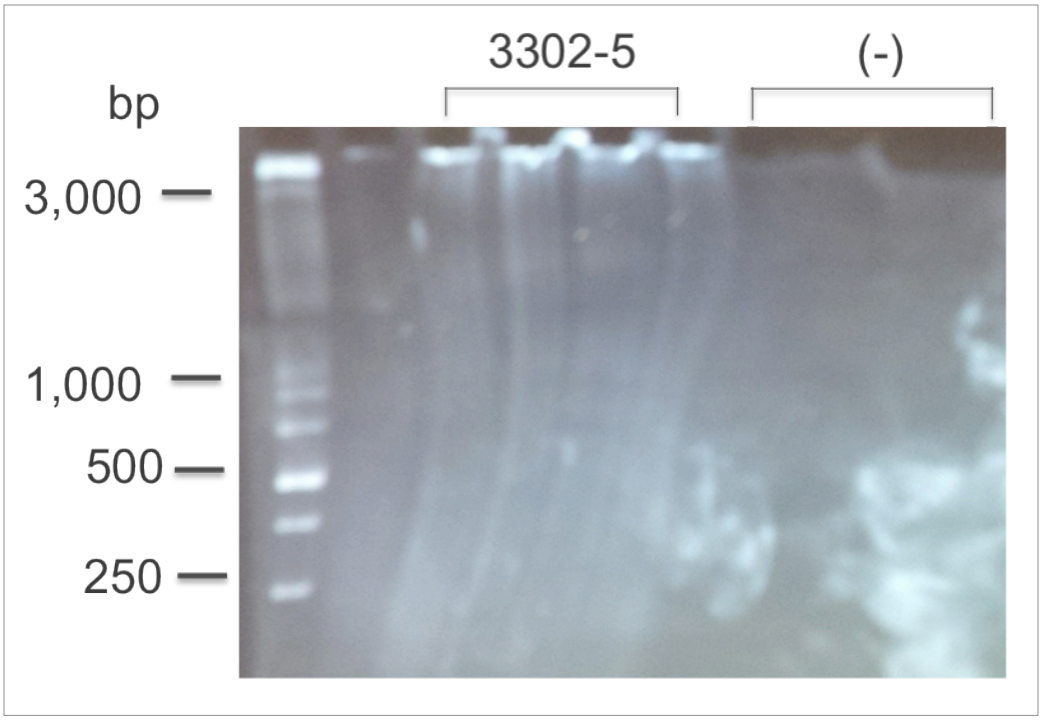


Figure S1 Electrophoresis mobility shift assay for BT_3302-5 run at 7hours. Lane one is the 1 kB DNA ladder, lane three to six is the experimental condition for 3302-5 and lanes 7 to 10 are the negative control.