GLYCOSIDE HYDROLASES AND MODIFICATION OF GLYCANS

MICROBIAL GLYCOSIDE HYDROLASE MEDIATED MODIFICATION OF HOST CELL SURFACE GLYCANS

By AARTHI PASUPATHI, B.Sc., MS

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree

Master of Science

Master of Science (2023)

McMaster University

(Biochemistry & Biomedical Sciences)

Hamilton, Ontario

TITLE: Microbial glycoside hydrolase mediated modification of host cell surface glycans

AUTHOR: Aarthi Pasupathi B.Sc., Biochemistry, University of Madras, Chennai, India

MS, Medical Laboratory Technology, Birla Institute of Technology and Science, Pilani, India

SUPERVISOR: Dr. Michael G. Surette

NUMBER OF PAGES: XIV, 108

LAY ABSTRACT

All the cell surfaces of animals, plants, and microbes are coated with sugars, also known as glycans. These sugars on the cell surface act as a barrier and protect them from the external environment. Glycans on the cells of both microbes and humans are essential for basic interactions between them. Many bacteria produce enzymes such as glycoside hydrolases to obtain nutrients from dietary sugars and alter the sugars on host proteins. There are various families of these enzymes, and they act on specific sugars and cleavage sites**.** The substrate specificities and characterization of these enzymes from most bacteria found in the human microbiome have not been studied in detail. My work focuses on developing standard enzyme assays for determining specific substrate specificities. This tool can be used to reshape glycans and understand their role in cell processes.

ABSTRACT

All cells and extracellular matrices of prokaryotes and eukaryotes are made up of glycans, the carbohydrate macromolecules that play a predominant role in cell-to-cell interaction, protection, stabilization, and barrier functions. Glycans are also central to human microbiome-host interactions where bacterial glycans are recognized by innate immune signaling pathways, and host mucins are a major nutrient source for various gut bacteria. Many microorganisms encode glycoside hydrolases (GHs) to utilize the available host cell surface glycans as a nutrient source and to modulate host protein function. The GHs are divided into families having conserved linkage specificity within each family and individual family members can be specific for dramatically divergent macromolecular substrates. In general, within a given GH family very few members have been biochemically characterized and the substrate specificity is poorly understood. GH genes are abundant in the human gut microbiome and culture-enriched metagenomics identified more than 10,000 distinct bacterial GH genes in an individual. The focus of this thesis is **endo-β-N-acetylglucosaminidases (ENGases)** encoded by GH18 and GH85 families. Bioinformatic analysis shows that the predicted proteins within each of these GH families fell into separate clusters in the Sequence Similarity Networks of each family. The hypothesis of this project is that human microbiome-encoded ENGases from the same GH family differ in their substrate specificities and within the SSN network of the same GH family, enzymes with similar substrate specificity may fall in the same cluster. In this work, I established conditions for overexpression of GH18 and GH85 proteins and investigated the activity of these enzymes on various substrates.

ACKNOWLEDGEMENTS

This thesis would not have been achievable without the help of wonderful people I met before and after my studies at McMaster University. First, I want to thank my parents and my brother Pasupathi, Uma Pasupathi, and Shiva Narayanan who, with love and affection, inspired me every day to be a better person. During the past two years, we had a lot of heartening talks that encouraged me to keep working for my dreams regardless of being away from them during the COVID-19 pandemic. Thank you for always believing in me.

To Dr. Michael G. Surette, who gave me the invaluable opportunity to be part of his laboratory at McMaster University. Mike, you are one of the kindhearted and most admirable mentors I have ever had. I am very grateful for all the support that you have extended to me during my time here at McMaster University. I deeply appreciate how you have been continuously encouraging and guiding me in the last two years, and how you have always been so supportive of all of my efforts and struggles. I grew as a scientist under your supervision, and undoubtedly it was one of the most challenging and joyful moments of my life that allowed me to grow as a person.

To my committee meeting members, Dr. Lesley MacNeil, Dr. Jennifer Stearns, and Dr. Sara Andres - Thank you for sharing your knowledge and for all the feedback and patience during the past two years of my project. I have learned and grown more than I could imagine; I will be forever grateful.

To my husband, Santhosh Narayanan, who always supported me and encouraged me with kindness and love, giving me wise advice through all the random situations during my studies at McMaster university.

To Dr. Evan Mann, who mentored me all through the project. Thank you, Evan, for all your support and guidance all the time. You are one of the smartest people I've had the privilege to work with and also a gem of a person who is always ready to provide whatever support a student needs.

To Laura Rossi, Blerina Kadiu, and Michelle Shah - Thank you for all the support and wise advice in and outside the laboratory. I am so glad I had the opportunity to meet wonderful people like you who worked with integrity and kindness.

To my lab mates: Nancy ElChaar, Dominique Tertigas, Grace Kim , Sakshi Karwasra, Paige Hopkinson, Dr. Salah Uddin Khan, Dr. Sharok Shekarriz, and Dr. Nick Dimonaco from the Surette Laboratory for their friendship and constant feedback on my project. You made the time I spent in the laboratory fun and joyful.

Finally, I want to Thank you, God, for everything in my life. Some of them were blessings, and some were lessons, and I'm grateful for everything you sent my way.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABBREVIATION	DEFINITION
Gal	Galactose
Glc	Glucose
Man	Mannose
Fuc	Fucose
Xyl	Xylose
GalNAc	N-Acetylgalactosamine
GlcNAc	N-Acetylglucosamine
GlcA	Glucuronic Acid
IdoA	Iduronic Acid
SA	Sialic Acid
Neu5Ac	N-Acetylneuraminic Acid
Asn	Asparagine
Ser	Serine
Thr	Threonine
OGTs	O-Linked GlcNAc Transferases
OGAs	O-Glcnasases
PG	Peptidoglycan
CPS	Capsular Polysaccharide
WTA	Wall Teichoic Acid
LTA	Lipoteichoic Acid
LPS	Lipopolysaccharides
PAMPs	Pathogen-Associated Molecular Patterns
PULs	Polysaccharide Utilization Loci
SUS	Starch Utilization System
SGBP	Surface Glycan-Binding Proteins
GH	Glycoside Hydrolase
HMNG	High Mannose N-Glycan

CAZymes	Carbohydrate-Active Enzymes
GT	Glycosyltransferases
PL	Polysaccharide Lyases
CE	Carbohydrate Esterases
CBM	Carbohydrate-Binding Modules
HMO	Human Milk Oligosaccharide
ENGases	Endo-B-N-Acetyl Glucosaminidases
CNGs	Complex N-Glycans
SSNs	Sequence Similarity Networks
IPTG	Isopropyl B- D-1-Thiogalactopyranoside
ER	Endoplasmic Reticulum
4MU- β-GlcNAc	2’-(4-Methyl- Umbelliferyl)- B-D-N-Acetyl Glucosamine
DMSO	Dimethyl Sulfoxide
MALDI-TOF MS	Matrix-Assisted Laser Desorption/Ionization Coupled Time Of Flight Mass Spectrometry
HCCA matrix	A- Cyano- 4- Hydroxycinnamic Acid
RNAse B	Ribonuclease B
Guan HCl	Guanidine Hydrochloride

DECLARATION OF ACADEMIC ACHIEVEMENT

The bioinformatic analysis of the human gut microbiome and identification of uncharacterized endo-β-N acetylglucosaminidase (GH18 and 85) were performed by Dr. Evan Mann from the Surette Laboratory.

The author completed the rest of the work in this study.

CHAPTER ONE: INTRODUCTION

Many trillion microbes live on and within human beings and play an important role in human health and diseases. Gut microbiota predominate the overall human microbiome composition and are known to carry over a million genes compared to our entire human genome which has ~ 24,000 (Ursell et al., 2014; Wang et al., 2017). The human epithelial cell surfaces interact with symbiotic microbes that play an important role in human biological processes, including modulating the metabolic phenotype, regulating epithelial development, and influencing innate immunity (Ley et al., 2006). In addition, microbial metabolic processes that are beneficial to the host are involved in the nutritional acquisition and also provide a physical barrier by protecting the host from pathogens (Tailford et al., 2015 & Varki & Gagneux, 2015). In eukaryotes, all cells and extracellular matrices are covered with complex carbohydrates called glycans that mediate a variety of events (Varki & Kornfeld, 2015). Many bacteria in the gut, including those that offer health benefits, rely on dietary fibers, largely hemicelluloses (eg. xylans, galactans, mannans) from plants, to sustain growth. Alternatively, some community members utilize host-derived glycans, including those found on extracellular and membrane-embedded proteins primarily as O-glycans (largely from mucins) and N-glycans (Bell et al., 2021 &Varki A., 2017). Thus, glycans are central to human microbiome-host interactions and in this thesis, I investigate the role of these carbohydrate-metabolizing enzymes produced by gut microbiota and explore their diverse set of substrate specificities.

1.1 Role of Glycans in health and disease

The intestinal epithelia glycocalyx and mucosal layers (a loose outer layer and a firm inner layer) form a glycan-rich barrier that acts as a first line of host defense against commensal bacteria (pathobionts) and invading pathogens (Turner et al., 2009 & Martens et al., 2018). The composition of gut microbiota is unique among individuals and is shaped by various factors, including food intake from early childhood onward, maternal microorganisms, and external factors such as antibiotics (Palmer et al., 2007). These microbes include bacteria, fungi and viruses with Firmicutes and Bacteroidetes being the dominant gut bacterial phyla (Turner et al., 2009 & Valdes et al.,2018). One major factor shaping the composition and physiology of the microbiota is the influx of glycans into the intestine from diet and host mucosal secretions. Microbial glycan fermentation into short chain fatty acids (SCFAs), such as acetate, propionate, and butyrate, are critical for host nutrient metabolism, maintenance of structural integrity of the gut mucosal barrier, immunomodulation, and protection against pathogens (Valdes et al., 2018 and Khosravi A &Mazmanian SK, 2013). In recent years, various studies suggest that abnormal bacterial composition (dysbiosis) is associated with various inflammatory diseases, such as inflammatory bowel disease, colorectal cancer, and obesity (Valdes et al., 2018 and Rinninella et al., 2019).

The intestinal epithelium is home to various cell types - enterocytes, enteroendocrine cells, Paneth cells, tuft cells, goblet cells and microfold (M) cells - which have roles in absorption, hormone secretion, antimicrobial peptide (AMP) secretion, taste-chemosensory responses, mucus production and antigen sampling (Gerbe et al., 2012). Intestinal mucus secreted by goblet cells is a major store for mucin glycoproteins, secretory IgA (Immunoglobulin A), and AMPs, and fosters components of the host immune response. Many *N*-glycosylated proteins are accessible to gut microbiota, including mucin, antibodies, and AMPs. Alterations in *N*-glycan structures are involved in disease development and aggravation (Cummings RD., 2019). *N*-glycans on T cells, B cells, and antibodies play an important role in development of immune cells and adaptive immune responses (Rudd et al., 2001 & Gornik et al., 2012).

1.2 Glycans – the structural backbone of glycoconjugates

Glycans are saccharides (or sugar units) that can be found attached to a protein or a lipid to form glycoconjugates which contribute important functions, prominently as ligands for glycan-binding proteins, including receptors, antibodies, and lectins. Although more than one hundred different monosaccharides in nature, just ten are combined to form myriad unique glycan structures in humans: galactose (Gal), glucose (Glc), mannose (Man), fucose (Fuc), xylose (Xyl), *N*-acetylgalactosamine (GalNAc), *N*-acetylglucosamine (GlcNAc), glucuronic acid (GlcA), iduronic acid (IdoA) and sialic acid (*N*-acetylneuraminic acid (Neu5Ac)) (Varki A et al., 2017). The initial sugar (monosaccharide) unit linked to a protein or lipid aglycone is further elongated with other monosaccharides (glycones) through glycosidic linkages. The resulting oligosaccharide or polysaccharide structures, which in mammals include *N*-linked glycans, *O*-linked glycans, proteoglycans, glycosaminoglycans and glycolipids (Kudelka et al., 2020 and Varki et al., 2015) (Figure 1), exhibit enormous structural diversity and biological properties. Diversity arises not only from the types of sugars but also the way they are linked. Unlike linear polymers like nucleic acids and proteins that contain only one basic type of linkage, sugars are linked with an α- or a β-linkage to any one of several positions on another monosaccharide in a chain or to another type of molecule forming varying branching structures. Branching is a prime characteristic of many glycans found on mammalian cell surfaces. Functionally, this diversity facilitates precise interaction with specific protein receptors involved in a variety of essential biological processes, such as cell adhesion, cell growth, cell death, cell migration, embryonic development, homeostasis, and immunity (Varki A. 2017).

1.2.1 *N*-glycans

In humans, more than 7000 distinct proteins are *N*-glycosylated, a process that is often critical for protein structure and function. All ***N*-glycans** share the same $Man_3GlcNac_2$ core structure, but *N*-glycans that terminate in mannose (high-mannose *N*-glycans), galactose or sialic acid (complex *N*-glycans) or some combination (hybrid *N*-glycans) can be found on the cell surface (Figure 2) (Kudelka et al., 2020). *N*-glycosylation occurs in the endoplasmic reticulum (ER) and Golgi apparatus, with most of the terminal processing occurring in the cis-, medial- and trans-Golgi compartments. Unlike other cell processes such as transcription or translation, glycosylation is non-templated, and the cells rely on host enzymes such as **glycosyltransferases** and **glycosidases** to form carbohydrate structures in a series of steps that are controlled by substrate availability, enzyme specificity, levels of gene transcription and enzyme location within the organelles (Martens et al.,2018).

N-glycan biosynthesis starts with the formation of a lipid precursor in which GlcNAc and Man form a branched carbohydrate structure that is attached to dolichol phosphate (Dol-P) on the cytoplasmic side of the ER. This lipid precursor is flipped to face the ER lumen, where Man and Glc units are added to form a 14-sugar precursor glycan, $Glc_3Man_9GlcNAc_2$. This is transferred *en bloc* from the carrier lipid to the nitrogen atom of an Asn (Asparagine) side chain within an Asn–X–Ser/Thr (Serine/Threonine) consensus sequence (where X is any amino acid except Proline) by the oligosaccharyltransferase complex in the endoplasmic reticulum (Schjoldager et al., 2020). The nascent carbohydrate– protein conjugate undergoes further processing in the ER, which usually involves removal of the glucose residues as part of a quality-control process. The structure then moves to the Golgi for final processing by a series of specific mannosidases and glycosyltransferases which dictate the ultimate glycan structure (Varki at al.,2017).

1.2.2 O-glycans

Serine and threonine glycosylation of proteins occurs in the secretory pathway of all eukaryotic cells. O**- glycans** are sugars attached to the oxygen atom of serine or threonine in glycoproteins (Figure 3) (Watson et al., 2015). In humans, the most common sugars linked to Ser or Thr are GlcNAc and GalNAc (Varki et al.,2017 and Watson et al., 2015). GalNAc-linked glycans, called mucin-type O-glycans, which are attached to mucins are the most abundant membrane associated glycans (Drickamer K., 2006 and Vasudevan D & Haltiwanger RS., 2014). Mucin-type O-glycans consist of thousands of unique structures built out of nine core structures that are attached to more than eighty percent of the cell surface proteins and secreted proteins. All mucin-type O-glycans are built from a single GalNAc attached via α-linkage to serine or threonine in a glycoprotein by GalNAc-transferases to form GalNAc-α-Ser/Thr (also known as Tn antigen) (Bennett et al., 2012).

The Tn-antigen can be elongated and branched to create eight distinct O-glycan core structures, of which cores 1–4 are the most common structures. The core 1 and 2 are found in all cell types, and core 3 and 4 predominant in the intestinal epithelium (Higel et al., 2016). These sugars are added as the protein moves through the cis-, medial- and trans-Golgi compartments and, unlike N-glycans, processing that trims existing sugar structures does not occur. Instead, the glycopeptide O-glycan chains are modified by distinct glycosyltransferases that can expand the existing structure with galactose, GlcNAc, sialic acid and, in some instances, fucose (Drickamer K., 2006). Non-templated sequential addition of glycans to the initial GalNAc gives rise to a diverse set of carbohydrate structures that are often highly clustered on certain glycoproteins, including mucins and human immunoglobulin A1 (IgA1). GlcNAc linked to Ser or Thr is typically found on intracellular glycoproteins present in nuclear, mitochondrial and cytoplasmic compartments. Unlike the mucin-type O-glycans, which are GalNAc-linked, addition of GlcNAc does not typically occur in the Golgi apparatus and is regulated through O-linked GlcNAc (O-GlcNAc) transferases (OGTs) and O-GlcNAsases (OGAs) (Drickamer K., 2006 and Vasudevan D & Haltiwanger RS., 2014). This dynamic process seems to be unique to this glycosylation motif and is thought to regulate many cellular functions, including cellular metabolism (Higel et al., 2016).

1.3 Bacterial glycans – role in adhesion and invasion

Bacterial glycans play a critical role in bacterial invasion, colonization, and host glycan utilization. Glycan interactions of the host and bacteria are critical for bacterial pathogenesis and host immune response to pathogens. Surface glycans expressed on the bacteria aids in successful colonization, adhesion and invasion into the host. The cell surface and membranes of Gram positive and Gram negative bacteria are made up of varying glycoconjugates such as peptidoglycan (PG), capsular polysaccharide (CPS), wall teichoic acid (WTA), lipoteichoic acid (LTA), and lipopolysaccharides (LPS). These molecules can act as pathogen-associated molecular patterns (PAMPs) or act as molecular shields preventing host cell recognition of PAMPs modulating immune responses for successful colonization and invasion (Maldonado et al., 2016 and Turner et al., 2018). LPS, peptidoglycans and teichoic acids are attractive targets for the invention of narrow-spectrum therapeutics and diagnostics as they are connected to bacterial survival and pathogenesis (Prado Acosta M & Lepenies B., 2019). As a part of host pathogen interaction, host cells produce lectins such as C-type, Siglecs, and galectins that bind to specific oligosaccharides (or glycans) that are present on the surfaces of pathogens to induce inflammatory responses and pathogenesis (Prado Acosta M & Lepenies B., 2019 and Li et al., 2020).

1.4 Host glycans – essential source for commensal bacteria and their role in bacterial pathogenesis

Glycans are found on the surfaces of both host and bacterial cells, providing them with a wide range of biological functions and are vital for bacterial pathogenesis and host responses to invading pathogens. Glycans on the surfaces of cell membrane serve as a major nutrient source for various commensal bacteria and influx of glycans are in turn essential for maintaining the microbiome composition (Berkhout et al., 2022 and Koropatkin et al., 2012). Glycans which are essential molecules for glycoconjugates have various structural and functional roles such as glycoprotein folding, membrane organization, and receptor signaling (Martens et al., 2018 and Palmer et al., 2007). Physical barriers in respiratory tract, gastrointestinal tract, and urogenital tract of the host are all heavily glycosylated. The mucus layer, often the site of pathogen attachment and entry, are made up of mucin glycans which protect these tissues from pathogens (Tailford et al., 2015). These mucins serve as a site for colonization of many pathogens such as *Salmonella typhimurium, Helicobacter pylori* and *Vibrio* cholerae (De et al., 2018 and Suwandi et al.,2019). Also, host glycans play crucial roles in innate and adaptive immunity including lectin binding to cytokines regulation, and autophagy (Radovani B & Gudelj I., 2022).

1.5 Polysaccharide Utilization Loci (PULs) – host glycan utilization machinery

Microorganisms belonging to *Bacteroidetes* phylum are the primary degraders of polysaccharides and predominant commensals in the gut. In *Bacteroidetes*, genes encoding carbohydrate degrading enzymes are often arranged in gene clusters called *Polysaccharide Utilization Loci* (PULs), with the archetypal PUL being that of the Starch Utilization System, encoding *susA-G*. A specific PUL encodes the requisite machinery for saccharification and uptake of a specific glycan structure (Grondin et al., 2017 and Martens et al., 2009). PULs are characterized by the presence of genes encoding *SusC* and *SusD* homologs that form an outer membrane oligosaccharide transport system, as well as a complete complement of cell surface glycan-binding proteins (SGBPs), intracellular and extra cellular carbohydrate degrading enzymes and transcriptional regulators (Lapebie et al., 2019). More frequently, single PULs encode all the framework necessary to access a specific glycan structure, while there are multiple PULs necessary to access some highly complex glycans. The current working model of Bacteroidetes saccharification involves capture of glycans by SGBPs, which are cleaved into oligosaccharides by endo-acting GHs. These oligosaccharides are imported to the periplasm by SusCD for further catabolism by exo-acting GHs. Studies with the *Bacteroides thetaiotaomicron (Bt)* uses a complex multi locus degradation apparatus to access complex *N*-glycans (CNGs) (Briliūtė et al., 2019) of which two are classed as PULs that encodes predicted Carbohydrate Active enZymes (CAZymes), SusC/D homologues and sensor regulators whereas, a single PUL is involved in the case of high mannose *N*-glycans (HMNGs) (Cuskin et al., 2015).

Although distinct in structure from that of Gram negative PULs, the Gram-positive human gut symbionts such as *Eubacterium rectale* and *Roseburia species* deploy substrate specific gene clusters targeting a variety of plant based and host derived glycans. These Gram-positive PULs (gpPULs) contain a minimum of one CAZyme, a carbohydrate transport system (most commonly ATP-binding cassette transporters), and a LacI- or AraC-like transcriptional regulator (Grondin et al., 2017 and O Sheridan et al., 2016).

1.6 Glycoside hydrolase – classification and mechanism

The enzymes that break down glycans are known as Carbohydrate-Active Enzymes (CAZymes) and often display exquisite specificity that distinguishes the carbohydrate moieties and type of glycosidic bond. Thus, the diversity of complex glycans has resulted in the evolution of numerous and highly diverse families of CAZymes (Briggs et al., 2021). CAZy (www.cazy.org) is a database that catalogues genomic, structural, and biochemical information for different families of CAZymes that catabolize, modify, or form new glycosidic bonds (Drula et al., 2022 and Cantarel et al., 2009). This database contains different modules of families such as glycoside hydrolases (GH), glycosyltransferases (GT), polysaccharide lyases (PL), carbohydrate esterases (CE), auxiliary activities, and non-catalytic carbohydrate-binding modules (CBM) (CAZypedia Consortium 2018). GHs, the enzyme focus of this thesis, are classified by CAZy into 184 families based on sequence and folding similarity (as of Aug 20, 2023). The catalytic machinery and reaction mechanism are conserved within the GH families while the precise substrate specificity is not strictly defined (Henrissat et al., 1995 and 1996).

Glycoside hydrolases (EC 3.2.1) are enzymes that catalyze the hydrolysis of the glycosidic bond between two sugars or between a sugar and an aglycone, leading to the formation of a sugar hemiacetal or hemiketal and the corresponding free aglycone. The IUBMB enzyme nomenclature of GHs is based on their substrate specificities. They are classified in different ways such as based on subsite-cleavage (exo- or endo-acting), catalytic mechanism (retaining or inverting) and amino acid sequences. Exo-acting GHs act on the reducing or non-reducing end of a glycan chain whereas endo-acting GHs cleave at some point along the oligosaccharide chain (Gloster et al., 2008). These subsite differences help to differentiate substrate binding in glycosidase active sites. Inverting and retaining glycosidases are distinguished by their configurational, anomeric specificity, and stereochemical outcome of the hydrolysis reaction which is conserved within each GH family (Stütz AE., & Wrodnigg TM., 2011).

Most GH families employ classical Koshland inverting or retaining mechanisms. The hydrolysis of the glycosidic bond is catalyzed by two amino acid residues (generally glutamic or aspartic acids) of the enzyme: a general acid (proton donor) and a nucleophile/base. Depending on the spatial position of these catalytic residues, hydrolysis occurs via overall retention or overall inversion of the anomeric configuration (Henrissat et al., 1995 and Vuong & Wilson, 2010). Inverting GHs catalyze the hydrolysis of a glycoside with a net inversion of anomeric configuration through a single step reaction forming a transition state with the assistance of a general acid and a base located 6-11 Å apart. Retaining GHs catalyze the reaction by retaining the actual anomeric configuration of the glycoside through a two-step reaction via a glycosyl-enzyme intermediate, with the general acid/base and nucleophile located ~5.5 Å apart (Vuong & Wilson, 2010).

The organization of GHs into families is based on amino acid sequence similarities and tertiary structure. There are nearly 184 GH families classified based on sequence and folding similarity by CAZy database and new families of enzymes are continuously being discovered (Kaoutari et al., 2013). The catalytic machinery and reaction mechanism are conserved within the GH families while the precise substrate specificity is not strictly defined (Henrissat et al., 1995 and Vuong and Wilson, 2010). Enzymes with different substrate specificities are sometimes found in the same family; on the other hand, enzymes that hydrolyze the same substrate are sometimes found in different families (Henrissat et al., 1995). Each GH family contains characteristic proteins and domains with similar fold and enzymatic reactions. However, they may vary in their actual macromolecular substrate within the family which is the primary focus of this thesis. A few GH families of enzymes are characterized by known functions and substrate specificities. (McCarter & Stephen Withers, 1994).

1.7 Role of microbial glycoside hydrolases in glycan utilization

Carbohydrates have an important role in shaping the microbiota and the ability to utilize complex dietary and host glycans is central to the survival of prominent members of the gut microbiota. The gut is a competitive environment for microorganisms, and the ability to degrade carbohydrates (whether found in food or derived from the host or other members of the human gut microbiota) provides a competitive advantage to certain bacteria (Martens et al., 2014). Microbial carbohydrate-degrading activities provide substantial benefits to the host as well. Humans lack most of the enzymes necessary to degrade dietary carbohydrates. Instead, we outsource these enzymatic needs to the resident gut microbiota (Martens et al.,2014 and Kaoutari et al.,2013). Glycoside hydrolases are one such class of carbohydrate metabolizing enzymes that are widely distributed among commensal and pathogenic microorganisms. These microbes make use of these enzymes for the degradation of dietary and mucus glycans for nutrition (Kaoutari et al.,2013). Microbes can also modify host surface glycans to facilitate attachment and internalization (Koropatkin et al., 2012). Many pathogens such as *Streptococcus pneumoniae* encode more than 50 GHs for both nutrition and modulating glycan binding sites for colonization of the airways (Hobbs et al., 2018). Human milk is known to contain indigestible carbohydrates called human milk oligosaccharides (HMOs) which are essential during the early stages of life and are being degraded by GHs such as β-hexosaminidases/β-1,6-N-acetylglucosaminidases (GH20), α-L- fucosidase (GH29) and Fucosidases/Sialidases (GH33) produced by many bacteria including *Bifidobacterium longum subsp. infantis*, *Bifidobacterium bifidum*, and *Bifidobacterium breve* (Ioannou et al., 2021). The commensal microorganisms such as

Bacteroides thetaiotaomicron and many other Bacteroidetes and Firmicutes use a wide range of GHs such as exo-β-glucosaminidase (GH2), β-mannosidase(GH5), endo-β-1,4-galactosidase(GH16), endo-β-N-acetylglucosaminidase (GH18, GH85), sialidases (GH33, GH156), β-galactosidase (GH42), and α-galactosidase (GH110) to access mucin glycans in the gut (Raba et al., 2023 & Wardman et al., 2022).

In the preliminary bioinformatic analysis performed by Dr. Evan Mann in the Surette laboratory, ~10,000 GH genes were identified in individual human gut samples using culture-enriched metagenomics. This includes a wide range of GH families encoded within the microbiome which can aid in modulating various sites of microbiota-host interaction. The overarching aim of this research project was to characterize a specific family of GHs that could modify N-glycans found on the human cell surface and extracellular matrix glycome. The focus of the project is the endo-β-N-acetyl glucosaminidases, which can act on N-glycans (GH18 and GH85 families).

1.8 Glycoside hydrolase 18 (GH18) and Glycoside hydrolase 85 (GH85)

Glycoside hydrolase 18 (GH18) include both catalytically active chitinases (EC 3.2.1.14) and endo-β-N-acetylglucosaminidases (ENGases) (EC 3.2.1.96) and also sub-families of non-hydrolytic proteins that function as carbohydrate-binding modules / lectins or as xylanase inhibitors. Chitinases hydrolyze the beta-1,4-linkages in chitin, an abundant N-acetyl-beta-D-glucosamine polysaccharide that is a structural component of protective biological matrices such as insect exoskeletons and fungal cell walls (Shrivastava, 2020). ENGases follow a substrate-assisted hydrolysis mechanism with retention of the anomeric configuration. The catalytic domain of GH18 may be connected to one or several substrate binding modules (CBMs), and contain peptide signals for localization such as an N-terminus secretion peptide, a C-terminus glycosyl-phosphatidylinositol (GPI) anchor signal for attachment to the plasma-membrane, or N-linked or O-linked glycosylation sites for oligosaccharide modifications (Stals et al., 2012). The active site motif DxxDxDxE is essential for the activity of the GH18 catalytic domain where the Glu (E) in this motif acts as the catalytic proton donor, and the last Asp (D) contribute to the stabilization of the essential distortion of the substrate (Tsuji et al., 2010).

Glycoside hydrolase 85 (GH85) are glycoside hydrolases that are catalytically active endo-β-N-acetylglucosaminidases (ENGases) (EC 3.2.1.96). Both the GH18 and GH85 ENGases act on β 1,4 glycosidic linkage between two GlcNAc residues present in the conserved core of all human N-linked glycans (Fujita et al., 2001 & Wang LX., 2008). These enzymes are secretory enzymes and work on a broad spectrum of substrates. They exhibit a retaining type of catalytic mechanism with carbonyl oxygen of C-2 acetamido group of substrates acting as catalytic nucleophile/base and glutamic acid residues as proton donor. The transglycosylation property of ENGases makes them a potential biocatalyst for the convergent synthesis of glycopeptides, and the re-modeling of the heterogeneous mixtures of glycoprotein glycoforms, including IgGs (Immunoglobulin G) (Fairbanks AJ., 2017).

1.9 ENGases (endo-β-*N*-acetylglucosaminidases) – catalytic site and mechanism

Endo-β-*N*-acetylglucosaminidases are endo-acting GHs that act on the chitobiose core (GlcNAc-β-1,4-GlcNAc) of *N*-linked glycans (Wang LX., 2008 and Fairbanks AJ., 2017) produced by organisms ranging from bacteria to high order species including humans (Figure 4). All ENGases have a similar $(\beta/\alpha)_8$-TIM barrel structure consisting of eight α-helices and eight parallel β-strands. All ENGases are retaining glycosidases that follow a substrate-assisted hydrolysis mechanism (in contrast to the Koshland mechanism described above) with retention of the anomeric configuration. In this mechanism, a general acid/ base residue (Asp or Glu) first acts as a general acid to protonate the glycosidic oxygen. Upon activation, the 2-acetamide group of the (-1) GlcNAc in the substrate acts as a nucleophile to attack the anomeric center, resulting in the breakdown of the glycosidic bond with simultaneous formation of an oxazolinium ion intermediate, instead of a glycosyl-enzyme intermediate. The oxazolinium intermediate then undergoes hydrolysis or transglycosylation by reacting with a water molecule or an alcoholic acceptor activated by the general acid/base residue. This mechanism also identifies another important residue, which is located at 1 or 2 amino acid residues upstream from the general acid/base catalytic residue. This key residue was shown to be an Asp residue for the GH18 ENGases or an Asn residue for the GH85 ENGases which was required for the proper orientation of the acetamide group to promote the oxazolinium ion formation (Tomabechi et al., 2014 & Umekawa et al., 2010).

ENGases are classified into two GH families in the CAZy database: GH18 and GH85. These widely (but not comprehensively) studied and characterized enzymes have different subsite specificities. Endo-S GH18 ENGases (Collin M and Olsén A., 2001) produced by *Streptococcus pyogenes* has the specificity for CNGs (complex N-glycans) of only human IgG (Figure 4) and EndoS2that belongs to M49 serotype of *S. pyogenes* acts on hybrid and

HMNGs (Sjögren et al., 2013). Endo F1-3 GH18 ENGases are produced by *Flavobacterium meningosepticum* (Tarentino et al., 1992) with varying specificity for *N*-glycans: Endo F1 cleaves only high mannose and hybrid oligosaccharides, Endo F2 acts on biantennary complex glycans, and Endo F3 is specific for bi- and triantennary complex oligosaccharides. Endo F2 and F3 can hydrolyze substrates with core fucosylation whereas fucose reduces the activity of Endo F1(Tarentino et al., 1992 and 1993).

GH85 family includes ENGases from both prokaryotes and eukaryotes. Endo M GH85 ENGases produced by *Mucor hiemalis* (Yamamoto et al., 1994) acts sialylated CNGs in addition to high mannose and hybrid oligosaccharides (Figure 4). Endo D (Robb et al., 2017) GH85 ENGases isolated from *Streptococcus pneumoniae* have specificity for CNGs and HMNGs (Figure 4). In addition to the hydrolytic activity, some of the ENGases possess transglycosylation activity, capable of using them in chemoenzymatic synthesis of oligosaccharides, glycopeptides, and glycoproteins (Wang et al., 2008).

1.10 RESEARCH HYPOTHESIS AND OBJECTIVES

The overarching goal of this research program was to investigate the activity of GHs from the human gut microbiome on host cell activity using both purified enzymes and bacteria. My project focuses on characterizing enzyme activity of cloned GHs using multiple protein and small molecule substrates. The objective was to identify endo-β N acetyl glucosaminidases, specifically GH18 and GH85 family enzymes, that could act on the chitobiose core of N-linked glycans. The foundation for my thesis was a bioinformatic analysis carried out by Dr. Evan Mann. A comprehensive analysis of the sequence space of GH18 and GH85 proteins was carried out as described previously for the GH156 family by Dr. Mann (Mann et al, 2022). GH18 and GH85 enzymes identified in the Surette lab human microbiome strain collection (~1400 human gut isolates with whole genome sequences) were included in the analysis. Visualizing the sequence space using sequence similarity networks (SSNs) shows that many of these enzymes fall into distinct clusters, some of which are not closely related to characterized family members and may have a distinct substrate specificity (Figure 5 and Figure 6). Sixty-six previously uncharacterized and non-redundant putative ENGases from GH18 and GH85 were selected for functional characterization.

I hypothesize that:

1) **Different ENGases from the same GH family differ in their *substrate specificity* and have specific substrate structural preferences at positions different from the point of cleavage.**
2) **Within the SSN networks for each GH family, enzymes with similar substrate specificity will fall in the same cluster.**

I further hypothesize that some ENGases from gut bacteria modify host cell surface glycans to alter cellular signaling and function.

The main goal of this project is to establish standard expression conditions and functional assays to characterize GH18 and GH85 enzymes from the human microbiome. With that, my thesis has two aims:

Aim 1 Develop and standardize protein expression and purification conditions for various GHs for high throughput screening.

Aim 2 Develop assays to characterize putative ENGases for their substrate specificity.

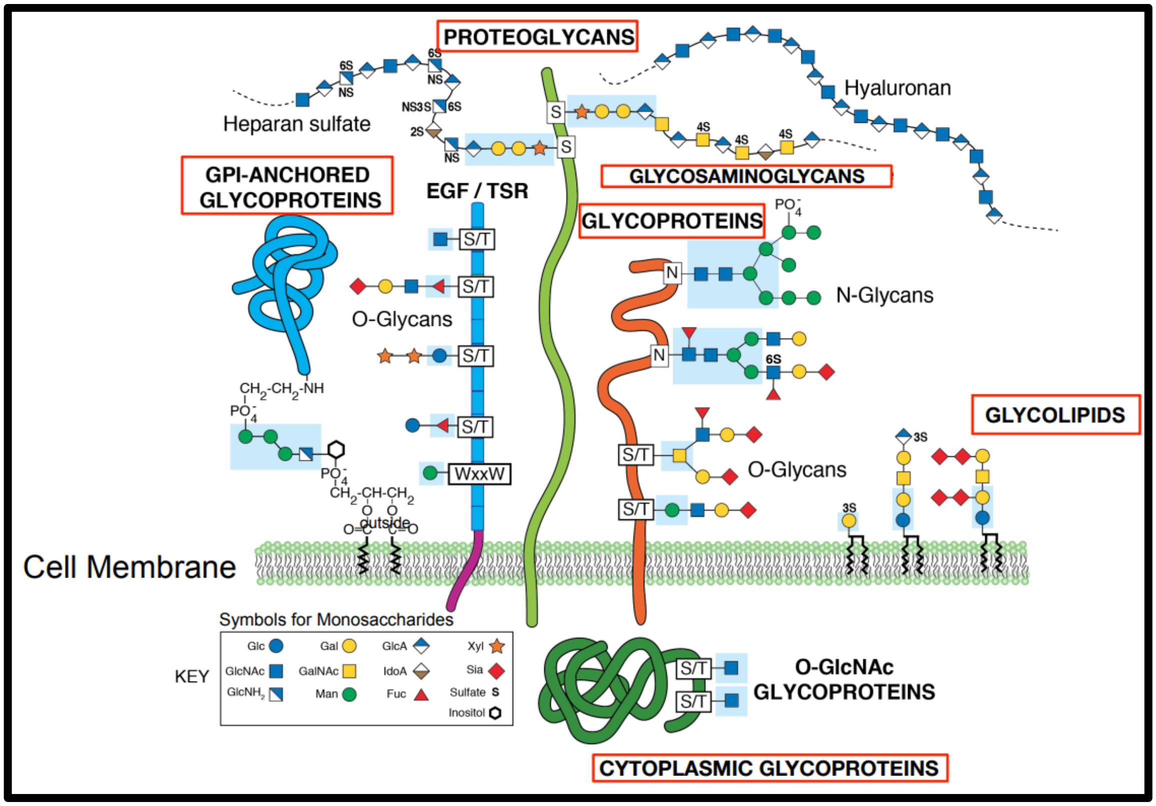


Figure 1 Schematic representation of different glycoconjugates that are found on cell surfaces

This diagram is a schematic representation of different types of glycans found on the surface of the cells of all mammals. Each monosaccharide unit is represented using a specific color and symbol. (Figure originally reported in Essentials of Glycobiology, 4th edition by Varki et al., 2015 published by Cold Spring Harbor Laboratory Press licensed under CC BY-NC-ND 4.0).

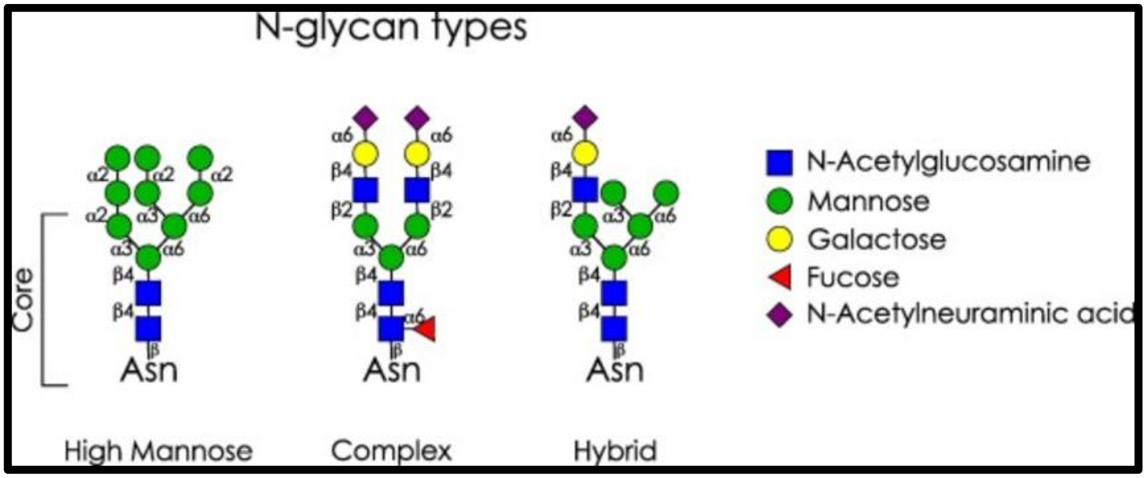


Figure 2. Core N glycan structure, types of N-glycans, and symbolic representation of sugars

The figure shows examples of different types of N-glycans. The three different types (High Mannose, Complex and Hybrid) share a common core structure including the first two N-acetylglucosamine residues and the first three mannose residues. (Figure originally reported in European Journal of Pharmaceutics and Biopharmaceutics by Higel et al., 2016 published under the terms of Creative commons License CC BY NC ND).

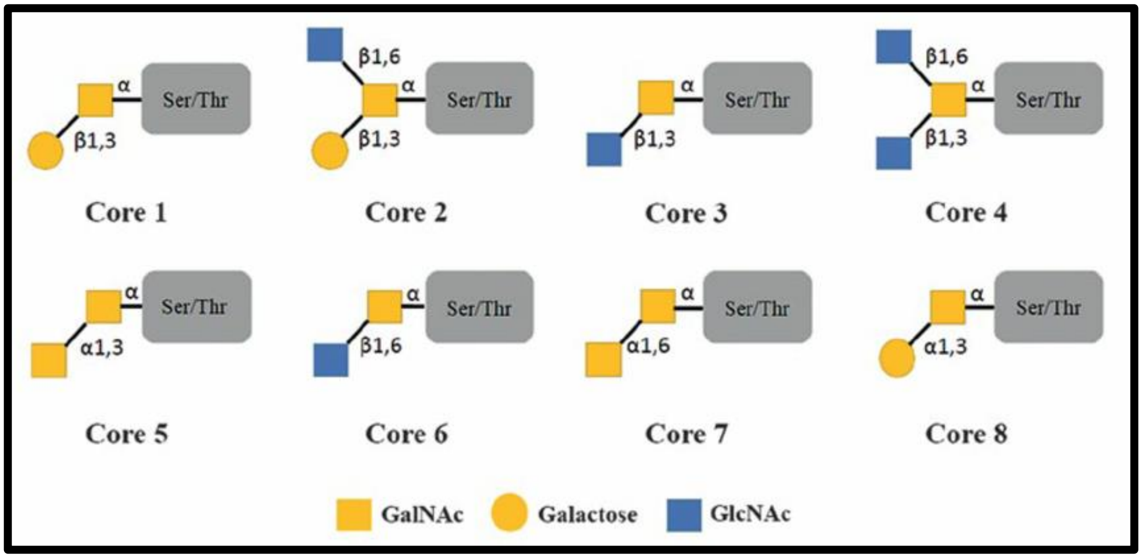


Figure 3 Different core structures of O-glycans

In humans, the most common sugars linked to Ser or Thr are GlcNAc and N-acetylgalactosamine (GalNAc). This diagram shows the schematic representation of different GalNAc-linked glycans, called mucin-type O-glycans, attached to mucins as the most abundant membrane-associated glycans. (Figure originally reported in Journal of gastrointestinal and liver diseases : JGLD by Watson et al., 2015 published under the terms of Creative commons License CC BY-NC-ND 4.0).

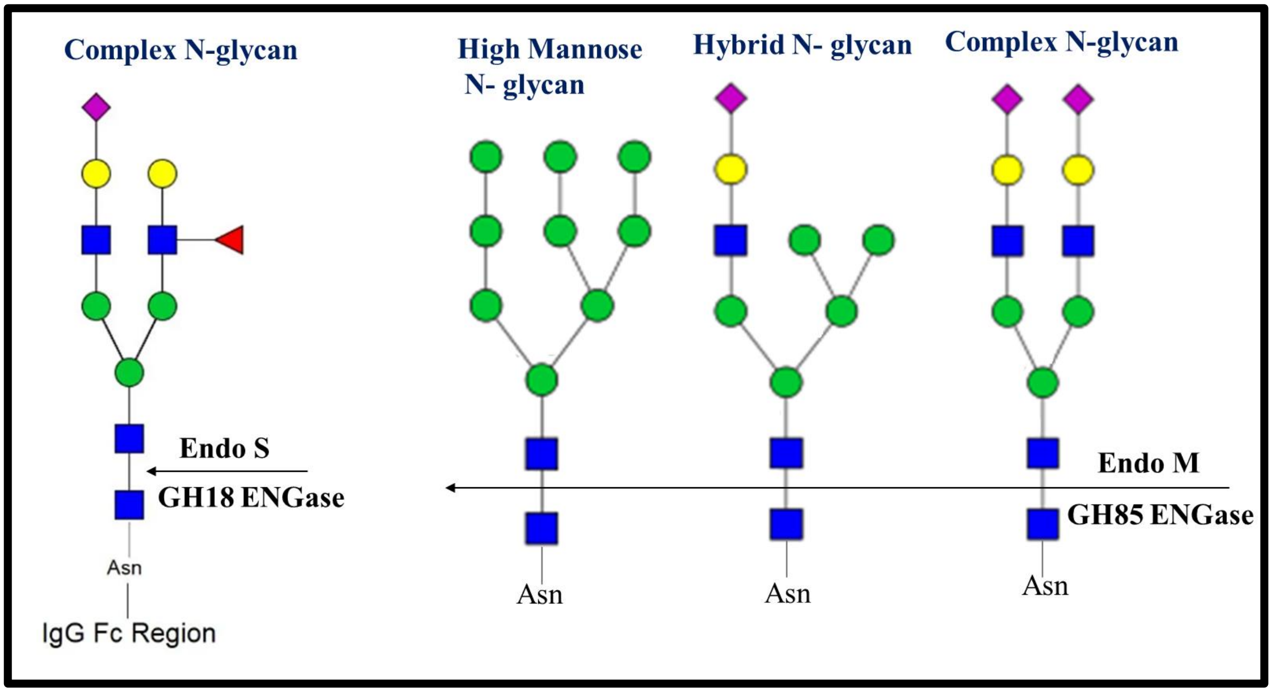


Figure 4 Action of GH18 and GH85 ENGases on N - glycans

This figure represents the activity of GH18 and GH85 ENGases that act on the β 1,4 glycosidic linkage between the two GlcNAc residues present in the chitobiose core of N-glycans having specificity for unique N-glycan structures. As an example, Endo S - GH18 ENGase has specificity for IgG-carrying CNGs and Endo M – GH85 ENGase has specificity for all three types of N-glycans. (Figure has been adapted from Frontiers in Cell and Developmental Biology by Krautter and Iqbal, 2021 published under the terms of Creative commons Attribution License CC BY).

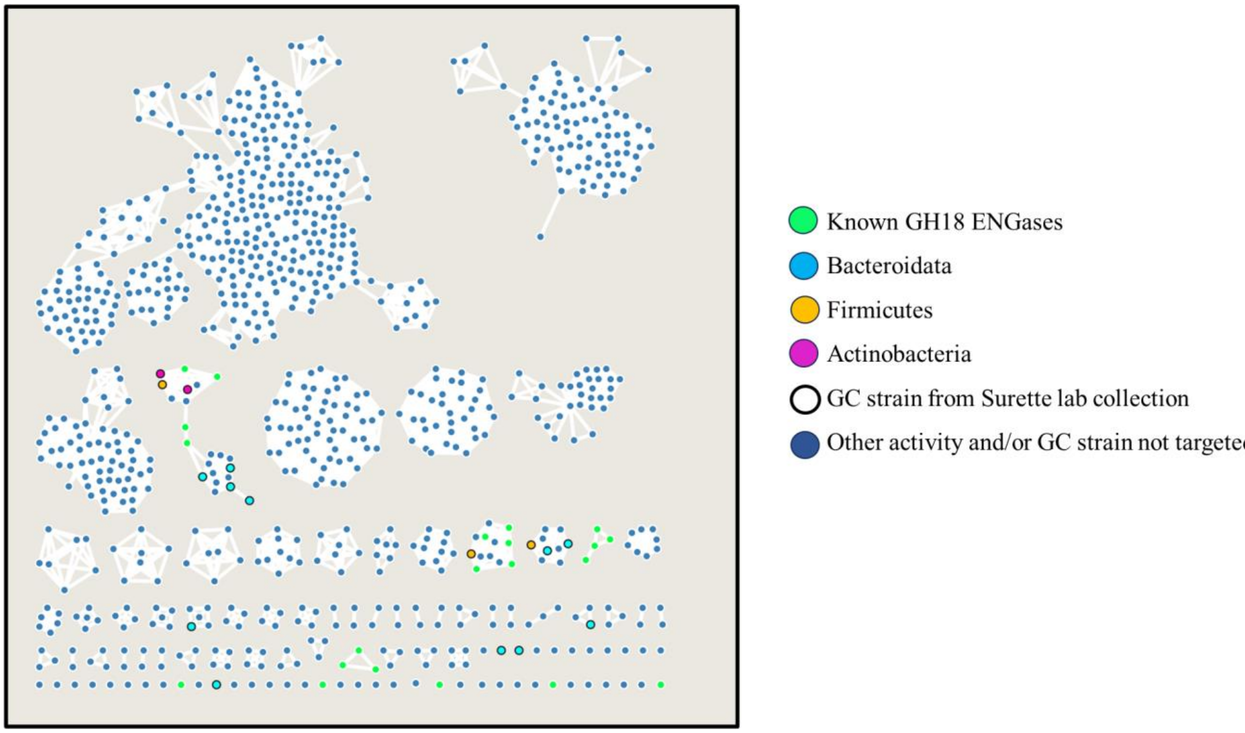


Figure 5 Sequence similarity network using GH18 ENGases

The available GH18 protein sequences (from a public database and Surette lab genome sequences) were clustered with cdhit to 95% identity and used to generate a sequence similarity network clustering at <45% edge identity threshold. Each circle represents a node (indicating protein sequence) and white lines connecting the nodes were the edges (indicating similarity in amino acid sequence). Proteins from different phyla were indicated by a specific color and putative GH18 ENGases used from the Surette lab collection were represented by a black circle around the node. Known ENGase active proteins were colored green. Those proteins that are not characterized and/or GC strain not targeted were colored blue. Data and figure was generated by Dr. Evan Mann.

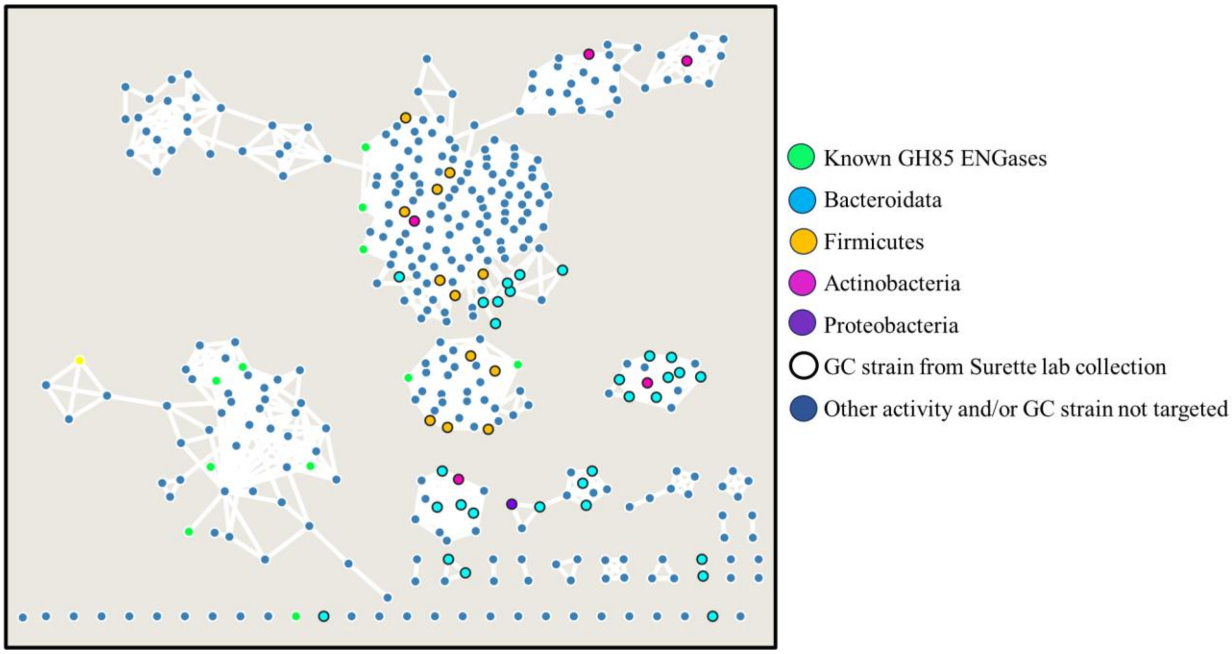


Figure 6 Sequence similarity network using GH85 ENGases

The available GH85 protein sequences (from a public database and Surette lab genome sequences) were clustered with cdhit to 95% identity and used to generate a sequence similarity network clustering at <45% edge identity threshold. Each circle represents a node (represents protein sequence) and white lines connecting the nodes were the edges (indicating similarity in amino acid sequence). Proteins from different phyla were indicated by a specific color and putative GH85 ENGases used from the Surette lab collection were represented by a black circle around the node. Known ENGase active proteins were labeled with their respective protein name and are colored green. Those proteins that are not characterized and/or GC strain not targeted were colored blue. Data and figure was generated by Dr. Evan Mann.

CHAPTER TWO: MATERIALS AND METHODS

2.1 Bacterial strains and growth conditions

Bacterial cultures were grown at 37°C with aeration in lysogeny broth (LB) supplemented with 50 µg/ mL kanamycin. ***E. coli*** **TOP10** {*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F9 proAB lacIq ZDM15 Tn10* (Tet^r)]} (Stratagene) was used for general cloning, and ***E. coli*** **BL21-CodonPlus(DE3) RIL strain** (*E. coli* B $F^–$ *ompT hsdS*(r_B^- m_B^-) *dcm*$^+$ Tet^r *gal* λ(DE3) *endA* I *[argU ileY leuW* Cam^r]) (Stratagene) was used for recombinant protein expression. An overview of the cloning and expression strategy is outlined in Figure 7.

2.2 DNA methods - Phusion PCR for amplification of ENGase domains

Custom oligonucleotide primers for intact genes specific for GH18 and 85 ENGase domains were obtained from Integrated DNA Technologies (Table 1). PCR amplification was performed using Phusion polymerase (New England Biolabs). Hot-start PCR was performed using a 50 µL reaction containing 5X Phusion reaction buffer, 10mM dNTPs, 10µM primers specific for the plasmid and 0.5µL of Phusion polymerase. The following thermocycler profile was used for the amplification: initial activation at 98°C for 30 seconds followed by 12 cycles of touchdown PCR with denaturation at 98°C for 10 seconds, annealing at 62°C (-1°C for each cycle) at 20 seconds and extension at 72°C for 30 seconds, then 25 cycles of standard PCR with denaturation at 98°C for 10 seconds, annealing at 50°C at 20 seconds and extension at 72°C for 30 seconds and a final extension at 72°C for 5 minutes. The amplification of products were visualized by agarose gel electrophoresis. The Monarch DNA gel extraction and Monarch PCR and DNA cleanup kits (New England Biolabs) were used to clean PCR products according to the manufacturer's instructions.

2.3 Cloning of GH18 and GH85 ENGase domains

PCR amplicons were cloned into IPTG inducible expression vectors to facilitate recombinant polyhistidine-tagged protein production. Initially, the TOPO cloning kit was used and later due to the unavailability of the TOPO cloning kit, the cloning was performed using GeneArt™ Gibson Assembly® HiFi Master Mix (Cat. No. A46628). Selective clones were validated by Sanger sequencing and kept in a non-T7 RNA polymerase strain (*E. coli* Top10) as frozen glycerol stocks and purified plasmids stocks.

2.3.1 Gibson assembly cloning

Gibson assembly method allows multiple overlapping DNA fragments to be seamlessly linked in a one-step single tube isothermal reaction. GeneArt™ Gibson Assembly®HiFi Master Mix contains a proprietary mixture of enzymes and reagents optimized to facilitate one-step assembly of overlapping DNA fragments. PCR amplicons can be easily assembled onto the desired vector using Gibson assembly where homology regions between 20 and 40 bp are required at the terminal ends to join fragments to the linearized vector. Independent assembly reactions were set up using 0.08pmol of each insert and of the vector (pET28b(+)) in 1X GeneArt Gibson assembly master mix. After 1 hour of incubation at 50°C, the entire reaction was transformed onto TOP10 *E. coli* competent cells and 100µl of transformed cells were plated onto LB agar plates containing kanamycin and incubated overnight at 37°C. Ten to twelve colonies from each plate were screened next day by colony PCR and those that were carrying the inserts were used for protein expression.

2.3.2 Colony PCR for selection of plasmids with inserts

After the transformations, colonies were screened for plasmids carrying the cloned inserts. The colonies were resuspended in 10μL of sterile water using toothpick and the same samples was plated onto new LB Agar with Kanamycin and the plate incubated at 37°C overnight. Positive clones were recovered from the plates. PCR tubes were boiled to 100°C for ten minutes and DNA was used for the PCR. Hot-start PCR was performed using a 50 μL reaction containing 10X reaction buffer, 25mM MgCl2,10mM dNTPs, 10μM primers specific for the plasmid (Table 1) and 0.25μL of Taq polymerase. The following thermocycler profile was used for the amplification: initial denaturation at 95°C for 5 minutes followed by 12 cycles of touchdown PCR with denaturation at 95°C for 30 seconds, annealing at 62°C (-1°C for each cycle) at 30 seconds and extension at 72°C for 1 minute, then 25 cycles of standard PCR with denaturation at 95°C for 30 seconds, annealing at 50°C at 30 seconds and extension at 72°C for 1 minute and a final extension at 72°C for 10 minutes. Meanwhile, glycerol stocks were made for the plasmids carrying the clones and were stored at -80°C for future use. The Monarch Plasmid Miniprep Kit (New England Biolabs) was used according to manufacturer's instructions for high-quality plasmid purification.

2.4 Protein expression and Solubilization of inclusion bodies

2.4.1 IPTG induction and growth conditions

The recombinant plasmids were transformed into *E. coli* BL21 cells – **BL21-CodonPlus(DE3)-RIL strain** (*E. coli* B F^{-} *ompT hsdS*(r_B^{-} m_B^{-}) *dcm*$^{+}$ Tet^{r} *gal* λ(DE3) *endA* I *[argU ileY leuW* Cam^{r}]). The transformed cells were then plated onto LB agar with kanamycin (50µg/mL) containing 0.4% glucose, and incubated overnight at 37°C. The recombinant strains were grown on LB broth with kanamycin (50µg/mL) containing 0.4% glucose and incubated overnight at 37°C. 1/100 of overnight cultures were diluted into 40ml of LB broth with kanamycin (50µg/mL) and allowed to grow to a minimum absorbance A_{600} of 0.6. Different IPTG induction conditions were carried out to standardize protein expression, and I decided to express all the clones at 37°C for 2 hours with 0.1 mM IPTG. To harvest the cells, cultures were centrifuged on a cooling centrifuge at 3000 x g for 15 minutes. The supernatant was discarded, and the cell pellets were lysed using 500µL Bugbuster (Millipore) reagent containing protease inhibitor cocktail, 20µM EDTA and 30 KU/µL of rLysozyme solution. Cleared cell lysates were prepared by centrifuging the whole cell lysate on a cooling centrifuge at 3000 x g for 15 minutes. The whole cell lysates and cleared lysates(supernatants) were visualized by SDS-PAGE to confirm protein expression in the clones. Proteins expressed as soluble forms are present in both whole-cell lysates and cleared lysates. In those cases, the cleared lysates were used for protein purification using Ni-IMAC magnetic beads. Those in which proteins aggregate in the form of inclusion bodies, the pellets were used for the solubilization protocol.

2.4.2 Solubilization of inclusion bodies

The protocol to solubilize the inclusion bodies was similar to Jevesar et al[53] with a few modifications using different concentrations of reagents that are optimal for all the expressed proteins. The concentrations of N-lauroyl sarcosine ranging from 0.1% - 1% were used for standardization of optimal protocol that can be used for all the proteins. The lysed pellet from the previous step that contains the inclusion bodies were used for this solubilization protocol. This method involves three washes using 1ml of 10mM Tris-HCl buffer, pH 8.0, and overnight mild solubilization using 0.2% N-lauroyl-sarcosine in 40mM Tris-HCl buffer, pH 8.0. The pellets at each step were centrifuged at 21000 x g for 20 minutes. The final supernatant after the overnight solubilization contains recovered proteins. All the steps for protein purification and solubilization were performed on ice. The solubilized inclusion body obtained using this protocol was further used for Ni-IMAC protein purification and measurement of enzyme activity. All the fractions from the solubilization protocol were visualized using SDS-PAGE to identify the presence of solubilized proteins.

2.4.3 Protein Purification using Ni-IMAC magnetic beads

Recombinant protein purification was performed using Pierce™ High-Capacity Ni-IMAC Magnetic Beads. 500 µL of the cleared lysate (for soluble proteins) and solubilized inclusion body fractions were allowed to bind to 20 µL of Pierce™ High-Capacity Ni-IMAC Magnetic Beads for 1 hour on a shaker. Before adding protein fractions, beads were equilibrated using 200 µL of equilibration buffer (50 mM sodium phosphate, 300 mM sodium chloride, 10 mM imidazole in water; pH 7.4). The tubes were placed on a magnetic stand until the beads were separated and the supernatant was discarded. Remove the tubes from the magnet and added 200 µL of wash buffer (50 mM sodium phosphate, 300 mM sodium chloride, and 20 mM imidazole in water; pH 7.4). The tubes were placed once again on the magnetic stand to allow the beads to separate and remove the supernatant. The wash step was repeated three times and final elution was done using 50 µL of elution buffer (50 mM sodium phosphate, 300 mM sodium chloride, 500 mM imidazole in water; pH 7.4). The purified proteins were visualized on SDS-PAGE to assess purity and further used for enzyme activity assays.

2.5 Enzyme activity assays

2.5.1 Activity assay using small molecule substrate – β-GlcNAc assay

Glucosidase activity was fluorometrically monitored using the substrate 2'-(4-methyl-umbelliferyl)- β-D-N-acetyl glucosamine (4MU- β-GlcNAc) (in DMSO). A working solution of 1mM 4MU-N-acetyl-β-D-GlcNAc was prepared from 30mM stock of 4MU-N-acetyl-β-D-GlcNAc (in DMSO) using nuclease-free water. To 5µL of solubilized inclusion body fractions (or purified proteins), 4MU- β-GlcNAc was added at a final concentration of 100µM in 50mM sodium acetate buffer (pH 5.2) and final volume made up to 50µL using water. The reaction mixture was incubated at 37°C for 2 hours. The endpoint assay was done by stopping the reaction after 2 hours using 250mM glycine buffer (pH10.2) as the released 4MU fluoresce at basic pH. Fluorescence intensity (360nm excitation wavelength (EX) / 460nm emittance wavelength (EM) was measured in a Synergy H1 plate reader (BioTek) at 37°C. The fluorescent units result for each assay were analyzed and expressed as fluorescence change per minute. Assays were performed on black-walled 96 well plates and in technical duplicates. pET28b(+) empty vector was used as a control. 4MU-α-L fucopyranoside was also added as a negative control to check for background hydrolysis.

2.5.2 Activity assay using glycoprotein substrates

Glycoproteins carrying different N-glycan structure were used as substrates for the cloned enzymes. These included bovine IgG (50mg/mL) (Sigma Aldrich, Catalogue # I5506), Ribonuclease B(RNase B) (5mg/mL) (New England Biolabs, Catalogue # P7817S), Fetuin (100mg/ml) (Sigma Aldrich, Catalogue # F3385), and α1 acid glycoprotein (100mg/mL) (Sigma Aldrich , Catalogue # G3643). The reaction mixture consists of 2µL assay buffer (final concentration of 10mM sodium phosphate buffer, pH 7.4), 10µg of individual glycoprotein substrates, 2µL enzymes (purified protein using IMAC magnetic beads and solubilized inclusion body), and 10µL nuclease-free water. The reaction mixture of RNase B and IgG was incubated at 37°C for 4 hours, and reaction mixtures of Fetuin and α1 acid glycoprotein were incubated at 37°C for 16 hours. 2µL from each reaction was loaded onto SDS-PAGE gel for visualizing changes in the glycosylation.

2.5.3 MALDI-TOF assay and sample preparation

Ground steel MALDI target plates were used for the assay. Initially, 1ul of the assay mix was spotted onto the target plate, dried completely and 0.6 ul of the HCCA matrix (α-Cyano- 4- hydroxycinnamic acid) in Bruker solvent as per manufacturer's instructions

(Bruker) was added and allowed to dry. Some of the assay reactions gave no visible spectrum due to the interference of detergents in the reaction. These assay reactions were diluted 1:2 dilutions using Bruker solvent, 0.5µL of diluted mix was spotted on the ground plate target, dried completely and 0.5µL of the matrix was added and allowed to dry. Using this matrix, RNase B reactions show better resolution of peaks compared to other glycoproteins.

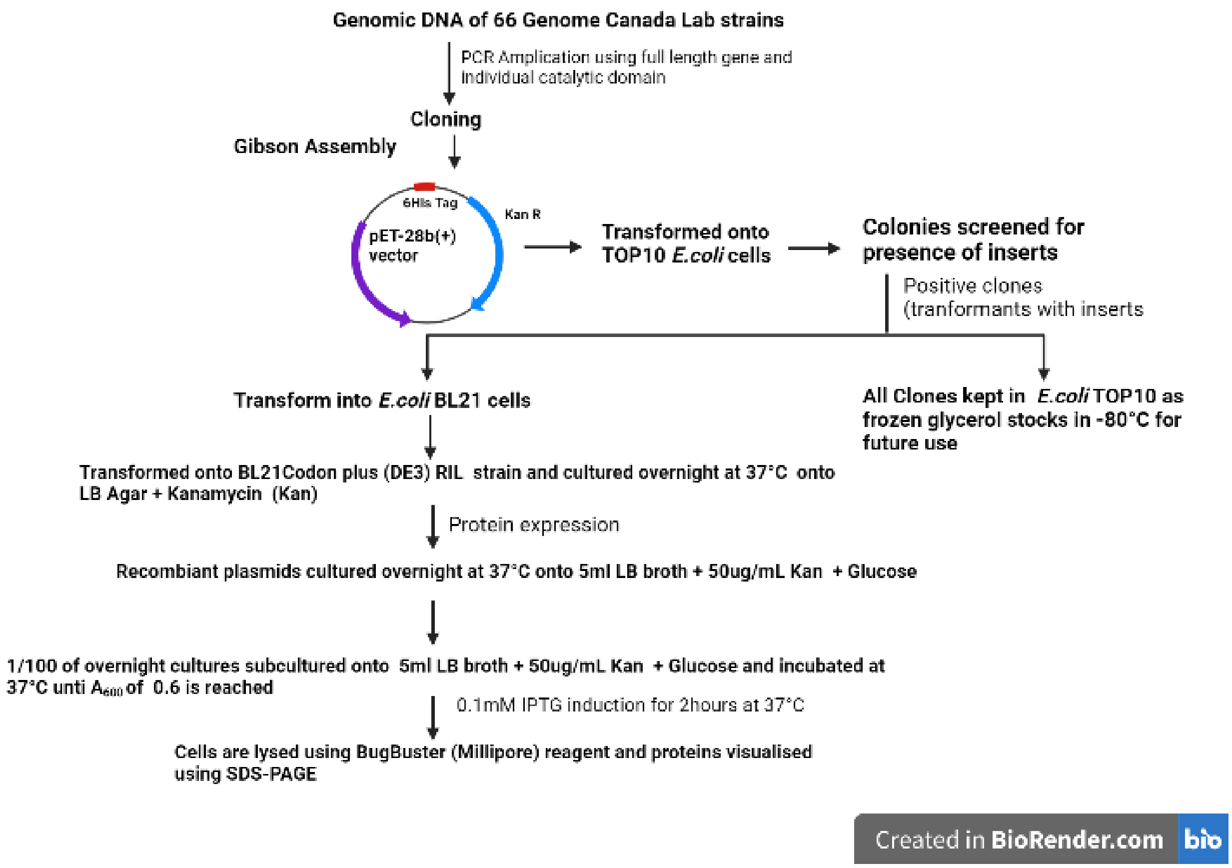


Figure 7 Description of experimental protocol for cloning and expression of recombinant proteins

Flow chart explaining the outline of the experimental protocol. Type of expression vectors, various strains of *E. coli* used for protein expression, and different protein expression conditions are outlined.

TABLE 1

Oligonucleotide primers used for generating plasmids

Primer ID	Sequence*
GH18-1$_{GC504}$	**gccgcgcggcagccatatggct**CAGACTTATGCTGTTGCC
GH18-1$_{GC561}$	**gccgcgcggcagccatatggct**AGTACATTCTGGGGATGC
GH18-1$_{GC722}$	**gccgcgcggcagccatatggct**GAAGTGATTACAGTATTTTATG
GH18-1$_{GC73}$	**gccgcgcggcagccatatggct**TCCAAACGCGACTCCATC
GH18-2$_{GC73}$	**gccgcgcggcagccatatggct**TACGAGAATCTGAGAGCATATAA
GH18-1$_{GC97}$	**gccgcgcggcagccatatggct**TTAACTTTCCCAGAAGAGC
GH18-1$_{GC123}$	**gccgcgcggcagccatatggct**GCTTTCAAAGTCGTCGGATATG
GH18-1$_{GC124}$	**gccgcgcggcagccatatggct**CAGGAACAGGTGTTCACG
GH18-1$_{GC408}$	**gccgcgcggcagccatatggct**GTATATTTTGAATTGACCAAG
GH18-1$_{GC453}$	**gccgcgcggcagccatatggct**ACGGCACAGACGCCCG
GH18-1$_{GC1062}$	**gccgcgcggcagccatatggct**ACAGACATGGAACCAAAATTC
GH18-1$_{GC241}$	**gccgcgcggcagccatatggct**AATGCTTTGCCCAATACGG
GH18-1$_{GC603}$	**gccgcgcggcagccatatggct**GCAGCTGTGCCTGATTC
GH18-2$_{GC97}$	**gccgcgcggcagccatatggct**TTTGAAGTACAAGCAGCATC
GH18-1$_{GC765}$	**gccgcgcggcagccatatggct**AATAATCAATCTGAAGTAGTC
GH85-1$_{GC962}$	**gccgcgcggcagccatatggct**AATCCTGAATGCAGTC
GH85-1$_{GC897}$	**gccgcgcggcagccatatggct**AACTCAAAGGCTAAGG
GH85-1$_{GC112}$	**gccgcgcggcagccatatggct**AACTCTAAAGCCAAGG

GH85-1$_{GC84}$	**gccgcgcggcagccatatggct**AACTCCAAAGCACTTG
GH85-1$_{GC108}$	**gccgcgcggcagccatatggct**ACCAATTCTAAAGCCAAAGAC
GH85-2$_{GC962}$	**gccgcgcggcagccatatggct**TGCGGTGCTTGTCCG
GH85-1$_{GC1038}$	**gccgcgcggcagccatatggct**TGCGGTGCTTGTCCGTC
GH85-2$_{GC1038}$	**gccgcgcggcagccatatggct**AATGCATGCAGCAC
GH85-1$_{GC768}$	**gccgcgcggcagccatatggct**AATCCGATGTGTAGCC
GH85-1$_{GC155}$	**gccgcgcggcagccatatggct**TGTAGTTTATGCCCTTCGC
GH85-2$_{GC155}$	**gccgcgcggcagccatatggct**TGCAGTTTGTGCCCTTCTC
GH85-1$_{GC960}$	**gccgcgcggcagccatatggct**GAGATGACCAAACAATGGG
GH85-1$_{GC1181}$	**gccgcgcggcagccatatggct**GACACTTACTGGCAATCG
GH85-1$_{GC1072}$	**gccgcgcggcagccatatggct**CTTCCCCGTTACAATTTTG
GH85-3$_{GC962}$	**gccgcgcggcagccatatggct**GAGTGGGGACCTATGCC
GH85-1$_{GC144}$	**gccgcgcggcagccatatggct**TATTGCTTTGAGGGCGATAAC
GH85-3$_{GC155}$	**gccgcgcggcagccatatggct**TATTCGCTCGAAGGAGAC
GH85-4$_{GC155}$	**gccgcgcggcagccatatggct**TATTGCTTCGAAGGCGAC
GH85-5$_{GC155}$	**gccgcgcggcagccatatggct**TATTGTTTTGAGGGCGATAA
GH85-2$_{GC1181}$	**gccgcgcggcagccatatggct**GGAAATGCCAGCAATGGTA
GH85-6$_{GC155}$	**gccgcgcggcagccatatggct**ACCAACGCACTTCCCGATG
GH85-2$_{GC768}$	**gccgcgcggcagccatatggct**CTTGAAGGAGATAATTTCAG
GH85-1$_{GC1065}$	**gccgcgcggcagccatatggct**GCTACCGGAGGGAATGTG
GH85-1$_{GC329}$	**gccgcgcggcagccatatggct**AAGAACGCTCTCCCTGATG

GH85-1$_{GC149}$	**gccgcgcggcagccatatggct**AAGAACGCTCTCCCTGATG
GH85-4$_{GC962}$	**gccgcgcggcagccatatggct**AATCCATCTTGTAGCC
GH85-2$_{GC960}$	**gccgcgcggcagccatatggct**AAAGCGTGGGGCCCTTTC
GH85-1$_{GC109}$	**gccgcgcggcagccatatggct**AATTCTAAAGCTGCTAC
GH85-1$_{GC751}$	**gccgcgcggcagccatatggct**GCTGGCGACTACGGCAAC
GH85-1$_{GC1045}$	**gccgcgcggcagccatatggct**CAGGGCGATTACGGAAAC
GH85-3$_{GC768}$	**gccgcgcggcagccatatggct**GCAATTGGTGGGTATCCAAG
GH85-1$_{GC665}$	**gccgcgcggcagccatatggct**GGACAAGGTGAAAGTGTC
GH85-5$_{GC962}$	**gccgcgcggcagccatatggct**GGAAAAGACGGTTCG
GH85-1$_{GC331}$	**gccgcgcggcagccatatggct**GGAAAGAACCTCGGTGG
GH85-2$_{GC144}$	**gccgcgcggcagccatatggct**GGAAAGAACCTCGGTGG
GH85-3$_{GC960}$	**gccgcgcggcagccatatggct**AGTACCGGTAAAGACGGAG
GH85-7$_{GC155}$	**gccgcgcggcagccatatggct**GGCGGTTATCCATCGAAAG
GH85-8$_{GC155}$	**gccgcgcggcagccatatggct**TCAACAATCGGTGGTTATC
GH85-3$_{GC1038}$	**gccgcgcggcagccatatggct**AGTACAGGTAAAGACGGA
GH85-1$_{GC263}$	**gccgcgcggcagccatatggct**GCCAATAAAACAACCAGCTC
GH85-1$_{GC982}$	**gccgcgcggcagccatatggct**AACTCACAAACTTCTGG
GH85-1$_{GC4}$	**gccgcgcggcagccatatggct**AATAGTTCAACTAGTGG
GH85-1$_{GC447b}$	**gccgcgcggcagccatatggct**AACAGCAGTACCAGTG
GH85-1$_{GC570}$	**gccgcgcggcagccatatggct**AATTCCAGCACCAGTG
GH85-1$_{GC96}$	**gccgcgcggcagccatatggct**AACTCCTCGACAAGC

GH85-1$_{GC27}$	**gccgcgcggcagccatatggct**AATAGTAGTACAAGCGG
GH85-1$_{GC1078}$	**gccgcgcggcagccatatggct**AATCCAACGACGAGCGG
GH85-1$_{GC466}$	**gccgcgcggcagccatatggct**GGTGACGGGAAAAAAGAATG

* upper case bases are homologous to domain regions being amplified and boldface bases are restriction enzyme recognition sites

CHAPTER THREE: RESULTS

3.1 Background

The human microbiome encodes many glycoside hydrolases that are involved in nutrition, production, and turnover of their own glycans, and modification of host glycans. The latter include enzymes used for immune evasion and modulation of host response but overall have not been thoroughly investigated. A bioinformatic analysis of GH18 and GH85 enzymes predicted in our lab's collection of human microbiota was used to identify specific genes of interest. The enzymes of interest fell into distinct clusters in the Sequence Similarity Network of each family (Figures 5 and 6). This included 15 members of the GH18 family and 48 members of the GH85 family. The list of isolates for cloning and protein expression along with their respective protein sizes (whole protein and domains) are detailed in Tables 2 and 3. An overview of the cloning and expression strategy is outlined in Figure 7.

3.2 Prioritizing ENGases

1149 GC + 6 WAC prokaryote genomes were annotated using dbCAN2 v. 2.0 (CAZy sequences obtained July 31, 2019) with default settings. Initial GH85 and GH18 candidates were included based on the consensus of at least 2 of the 3 dbCAN2 tools used during annotation. Since the GH18 family includes enzymes specific for substrates other than N-glycans, only sequences with higher similarity to known ENGases than to enzymes with divergent substrate specificities were retained. Since the catalytic residues for GH18 and GH85 members are known, the presence of catalytic residues was confirmed by manual inspection with the aid of a multiple sequence alignment generated by Clustal Omega that compared candidates and representative biochemically/structurally characterized enzymes. Representative sequences for functional characterization were chosen using cdhit with default settings and identity thresholds of 0.9 for GH18 and 0.8 for GH85. Domain boundaries were determined using HMMsearch with the family specific HMMs provided with dbCAN2. Domain boundaries were assigned as the 'env coord' start and end positions of the output tsv.

3.3 Results of Aim 1

3.3.1 PCR amplification of target genes and Gibson assembly cloning

Coding sequences of 63 individual catalytic domains of ENGases were cloned into IPTG-inducible vectors for recombinant protein expression which incorporate His_6-tags for Ni-affinity purification. These genes were amplified using Phusion PCR and the amplicons were purified using Monarch gel extraction and PCR clean-up kits. PCR primers were listed in Table 1. The amplicon size varies for each of the genes as the size of domain regions were different for each protein (Tables 2 and 3). The inserts were amplified by Phusion PCR and representative examples of the inserts amplified were shown in Figure 8 and the PCR cleanup products were used for cloning. The cloning was performed using GeneArt™ Gibson Assembly® HiFi Master Mix (Cat. No. A46628). Plasmids carrying the inserts were confirmed by colony PCR and an example of amplified inserts was shown in Figure 9. Clones were validated by Sanger sequencing and maintained in *E. coli* Top10 (a non-T7 RNA polymerase strain) as frozen glycerol stocks and purified plasmid prep stocks.

3.3.2 Standardization of protein expression

The recombinant plasmids were transformed into *E. coli* BL21 cells – **BL21-CodonPlus(DE3)-RIL strain** (*E. coli* B F^{-} *ompT hsdS*$(r_B^{-}\ m_B^{-})$ *dcm*$^{+}$ Tet^{r} *gal* λ(DE3) *endA* I *[argU ileY leuW* Cam^{r}]). Bacterial cultures were grown at 37°C overnight with aeration in lysogeny broth (LB) supplemented with 50 µg/ mL kanamycin. Protein expression was done by growing 1/100th of the overnight cultures on lysogeny broth supplemented with 50 µg/ mL kanamycin until absorbance of A_{600} of 0.6 (2 hours) followed by 2 hours of IPTG induction.

To standardize protein expression, different IPTG induction growth conditions and concentrations were carried out at different temperatures (37°C, 30°C, room temperature, and 15°C), for 2 hours, 4 hours, 8 hours, 12 hours, and 16 hours. The various IPTG concentrations that were used for standardization were 0.05mM, 0.1mM, 0.2mM, 0.5mM, and 1mM. From these preliminary results, I decided to express all the clones at 37°C for 2 hours with 0.1 mM IPTG. The cells were lysed using Bugbuster (Millipore) reagent. The whole cell lysates and cleared lysates (supernatant after centrifugation at 21,000 x g for 20 minutes) were visualized by SDS-PAGE to confirm protein expression in the clones as shown in Figure 10 (selected GH85 ENGases were used in the figure and data not shown for others).

To use proteins for various high throughput screening, the recombinant proteins should be stable and expressed in soluble active form. Of the 63 expressed recombinant proteins, only two were found to be soluble proteins – GH18-1_{GC123} and GH18-1_{GC124} (Figures 11 and 12) and others are insoluble proteins (high-level expression of recombinant proteins can lead to the formation of protein aggregation called inclusion bodies) (Freedman R. B., & Wetzel R., 1992 and Singh et al., 2015). Representative data is shown in Figure 11.

To obtain a biologically active and soluble protein of high yield, inclusion bodies must be solubilized and refolded *in vitro*. The use of strong expression promoters and high inducer concentrations results in high protein concentrations that may lead to protein aggregation before folding. Reducing the rates of transcription and translation facilitates folding by allowing the newly synthesized protein to fold before it aggregates. To improve the solubility of the protein, I started to manipulate different expression parameters such as lowering expression temperature and inducer concentration. Initially started with lowering the expression temperature (15-25°C) to improve the solubility of recombinantly expressed proteins. At lower temperatures, cell processes slow down, and thus lead to reduced rates of transcription, translation, cell division, and reduced protein aggregation. Lowering the expression temperature also results in a reduction in the degradation of proteolytically sensitive proteins. Along with this, I also used varying concentrations of IPTG (0.1mM – 0.5mM induction agent) which reduces the expression, thereby improving the solubility and activity of recombinant proteins.

In addition to various expression conditions, I used denaturation protocol with different denaturants such as Sodium chloride, Triton X-100, and Guanidine hydrochloride (Guan HCL) which helps in the solubilization of proteins and an additional refolding step on a Ni-NTA affinity column for protein recovery. This involves five wash steps using 50mM Sodium phosphate buffer, pH 8.0 with additional denaturants at every wash step and final solubilization step using 6M Guan HCL. This approach was not successful in recovering soluble proteins from the inclusion body (Figure 13) (Randomly selected proteins were shown on the figure).

As an alternative approach, I adapted a protocol to solubilize inclusion bodies using mild solubilization (Jevesar et al., 2005). The main advantage of this method was that it does not require a refolding step and I made a few modifications using different concentrations of reagents that are optimal for all the recombinant proteins. Traditional inclusion body solubilization involves the use of strong denaturants resulting in the denaturation of native protein and thus requires renaturation to recover properly folded protein. In this current protocol, mild denaturant such as N-lauroyl-sarcosine was used that maintains proper folding of native protein without denaturing the protein throughout the solubilization process.

To standardize optimal solubilization conditions that are suitable for all the proteins used in the study, I tested varying concentrations of N-lauroyl-sarcosine ranging from 0.1% to 1%, different buffers such as 10mM sodium phosphate buffer pH 7 and 10mM Tris-HCl pH 8, and varying solubilization time such as 2 hours , 4 hours, 8 hours, 12 hours and overnight. I was able to recover all the proteins from the inclusion body at a starting concentration of 0.2% N-lauroyl-sarcosine in 10mM tris-HCl buffer under overnight solubilization. I then decided to choose the minimum concentration of 0.2% to avoid further interference of detergents on the enzyme activity assays. All the incubation and reaction conditions were performed under cold conditions.

This method of inclusion body solubilization involves simple washes using 10mM Tris-HCl buffer, pH 8.0, and mild solubilization using 0.2% N-lauroyl-sarcosine in 40mM Tris-HCl buffer, pH 8.0. The solubilized inclusion body (Figure 14) obtained using this protocol was further used for Ni-IMAC protein purification and measurement of enzyme activity.

3.3.3 Protein Purification using Ni-IMAC Magnetic Beads

Recombinant protein purification was performed using Pierce™ High-Capacity Ni-IMAC Magnetic Beads. Initially to assess the binding capacity and to standardize the purification for all the proteins, I started with manual purification of one recombinant GH18 protein – GH18-$2_{GC\ 97}$ (*Enterococcus faecalis*) and used the same for the enzyme activity assays. The solubilized inclusion bodies from the previous step were allowed to bind to the magnetic beads after equilibration of the magnetic beads using 50mM sodium phosphate, 300 mM sodium chloride, 10 mM imidazole in water, pH 7.4. The tubes were placed on the magnetic stand and recombinant proteins were purified through repeated wash steps using 50mM monosodium phosphate, 300 mM sodium chloride, 20 mM imidazole in water; pH 7.4 followed by elution 50mM monosodium phosphate, 300 mM sodium chloride, 500 mM imidazole in water; pH 7.4. The purified proteins were visualized on SDS-PAGE (Figures 12 and 15) and used for further functional characterization. Not all the recombinant proteins were purified by this method as the stability of the proteins was a major drawback. Proteins from GH18-1_{GC123} and GH18-2_{GC73} that stored as glycerol stocks after purification and dialysis protocol were lost and degradation of proteins was also observed (Figure 16). However, the solubilized inclusion body fractions were stable.

Hence, apart from the two soluble proteins, for all the other recombinantly expressed GH18 and 85 ENGase, solubilized inclusion body fractions were used for characterization assays.

3.4 Results of Aim 2

The purpose of Aim 2 was to develop assays to characterize putative ENGases for their substrate specificity. The purified recombinant proteins and solubilized inclusion body fractions were tested *in vitro* for their enzyme activity against a small molecule substrate and glycosylated bovine proteins to test the ability of enzymes to act on *N*-glycans of the host proteins. The activity of enzymes on protein substrates was analyzed using SDS-PAGE and MALDI-TOF mass spectrometry.

3.4.1.Enzyme activity assay using small molecule substrate – β-GlcNAc assay

A modified protocol from Chen et al., 2014 was used to measure glucosidase activity using 4MU- β-GlcNAc as a substrate. The endpoint assay was performed *in vitro* at 2 hours measuring the release of fluorophore as fluorescence units. When these enzymes cleave β-GlcNAc from the substrate, the released 4MU shows fluorescence at basic pH (250mM Glycine HCl pH 10.2. Of the fifteen predicted GH18 ENGases, thirteen show activity on the substrate 4MU- β-GlcNAc. (Figure 17) Similarly, of the forty-eight predicted GH85 ENGases, thirty-three show activity on the small molecule substrate 4MU- β-GlcNAc (Figure 18 A and B). Since the 4MU substrates show some background fluorescence, I used 4MU fucose as a negative control to test for background hydrolytic activity in the resolubilized inclusion body fractions as these predicted ENGases should not act on this substrate (Figure 19). The empty vector was processed the same way as the other enzymes and was used as a blank for this assay. Two soluble GH18 ENGases (GH18-1_{GC123} and GH18-1_{GC124}) purified using Ni-IMAC magnetic beads were used for activity assays (Figures 12 and 17). No activity was detected against this substrate. The details of ENGases with their activity against various substrates are tabulated in Tables 5 and 6.

3.4.2 Enzyme activity assay using complex glycoproteins

The activity on small molecule substrates alone does not mean that they can be active on all types of higher molecular weight glycoproteins carrying N-glycans. In order to test the activity of these predicted ENGases on N-glycans, those that show activity on 4MU- β-GlcNAc were tested for their ability to act on complex high molecular weight glycoproteins carrying N-glycans on their structure. The purified recombinant proteins or solubilized inclusion bodies were tested *in vitro* for their enzyme activity against various glycosylated bovine proteins. The different glycoproteins used as substrates for the assays were bovine IgG, bovine Fetuin, bovine α_1 acid glycoprotein, and bovine RNAse B (Figure 20). These glycoproteins carry different types of N-glycans (complex, hybrid, and high mannose) in their structure. GH 18 and 85 ENGases act on the β 1,4 glycosidic linkage between the two N-acetyl glucosamine residues present on the chitobiose core of the N-glycan structure (Figures 2 and 4). The reaction mixture consists of assay buffer (100mM sodium phosphate buffer, pH 7.4), individual glycoprotein substrates, enzymes (purified protein using IMAC magnetic beads or solubilized inclusion body), and water. The reaction mixtures were incubated at 37°C for 4 hours. All the fifteen GH18 ENGases and forty-eight GH85 ENGases were tested for their activity against all four glycoprotein substrates.

RNAse B is a 15kDa glycoprotein carrying high mannose N-glycans on its structure. The enzyme activity on the RNAse B was observed as a change in the size of the glycoprotein as a result release of the N-glycan structure by active ENGases. Of the thirteen GH18 ENGases that were active on the small molecule substrate 4MU-β GlcNAc, five of the GH18 ENGases (GH18-1_{GC504} - *Bacteroides uniformis*, GH18-1_{GC408} - *Bacteroides uniformis*, GH18-1_{GC603} - *Bacteroides intestinalis*, GH18-2_{GC97} - *Enterococcus faecalis*, GH18-1_{GC73} - *Bacteroides fragilis*) showed activity against bovine RNAse B. There was complete deglycosylation of RNAse B with GH18 ENGase – GH18-2_{GC97} (*Enterococcus faecalis*) that was visualized on SDS-PAGE as a shift in the protein band (Figure 21), whereas an incomplete reaction was observed with the other four GH18 ENGases under these reaction conditions.

α1 acid glycoprotein is ~41kDa protein that is heavily glycosylated with biantennary complex N-glycans-carrying glycoprotein where the glycans occupy nearly 45% of the protein mass. The activity was visualized on SDS- polyacrylamide gel electrophoresis (SDS-PAGE) by visualizing the change in the size of the glycoprotein substrates. Of the 15 GH18 ENGases, only one putative protein - GH18-1_{GC73} (*Bacteroides fragilis*) , showed activity on α1 acid glycoprotein as visualized as a thin band around 20kDa on SDS-PAGE which is hypothesized to be the result of a change in the glycosylation of the glycoprotein (Figure 22). This enzyme was also very active on RNAse B (Figure 21). None of the GH18 ENGases showed activity against bovine IgG and Fetuin (Figure 23). There was no activity with the GH85 ENGases on any of the glycoprotein substrates (Figures 24 and 25).

3.4.3 MALDI-TOF assay for analysis of enzyme activity

Mass spectrometry is a sensitive tool for monitoring glycoside hydrolase activity. It can be used to detect changes in the protein (change in mass due to removal of glycans) or detection of released glycans. MALDI-TOF is the preferred method to look at changes in protein size and access to an instrument in our lab facilitated the development of methods for looking at GH activity of protein substrates. The same assay reactions that were used for visualization on SDS PAGE were analyzed using MALDI-TOF for change in the spectrum.

For initial standardization, I focused on RNAse B activity assay. MALDI-TOF requires the application of the substrate to an assay plate and mixing it with a matrix solution. I used equal volumes HCCA matrix (a-Cyano-4-hydroxycinnamic acid) and RNAseB substrate. RNAse B is a high mannose N-glycan-carrying glycoprotein that has one N-glycosylation site at Asn 34 and can exist as five different glycoforms ranging from 5 to 9 mannose residues on them. These glycoforms were clearly visible on the spectrum with m/z of ~14kDa (Figure 26). In addition, a second glycosylate peak was observed with m/z of ~7.4kDa. This fragment was not observed on SDS-PAGE indicating it was not a proteolytic fragment of RNAseB. This suggests that the 7.4kDa form arose from *in-source* fragmentation.

MADLI-TOF can be affected by the composition of the sample and the use of solubilization inclusion body fractions for activity assay with the presence of detergents, resulting in poor spectrums. To overcome this issue the reaction mixture was diluted (1:2 dilution) using

Bruker solvent. 0.5µL of the diluted mix was then loaded on the target plate, dried completely and 0.5µL of HCCA was spotted, dried, and analyzed using MALDI-TOF. The assay reaction with individual substrates without enzymes was used as a control. RNAse B reactions gave visible spectrums for all the GH18 and 85 ENGases. The five GH18 ENGases (GH18-1_{GC504}, GH18-1_{GC408}, GH18-1_{GC603}, GH18-$2_{GC97,}$ and GH18-1_{GC73}) that showed activity on the SDS-PAGE were analyzed on the MALDI TOF. The change in the spectrum was visualized as a shift in the peak of the actual glycoprotein. The results were shown in Figures 26A-C. With one of the GH18 ENGase - GH18-2_{GC97} (*Enterococcus faecalis*) where there was complete deglycosylation, was visualized as a complete shift in the mass spectrum on MALDI-TOF assay. Similarly, the remaining four active GH18 ENGases (GH18-1_{GC504} - *Bacteroides uniformis*, GH18-1_{GC408} - *Bacteroides uniformis*, GH18-1_{GC603} - *Bacteroides intestinalis*, and GH18-1_{GC73} - *Bacteroides fragilis*) on RNAse B that were visualized on SDS PAGE were analyzed on MALDI-TOF. Partial deglycosylation was observed as a shift in the mass spectrum along with the RNAse B peak (Figures 26A and 26B). Those where no enzyme activity was observed have the same spectrum as that of the control RNAse B (Figure 26C).

Using these protocols, it was not able to visualize a proper spectrum using other glycoprotein substrates. Unlike RNAse B, these are high molecular-weight glycoproteins that were difficult for ionization using the HCCA matrix. More suitable matrices for such glycoprotein substrates must be used for further analysis.

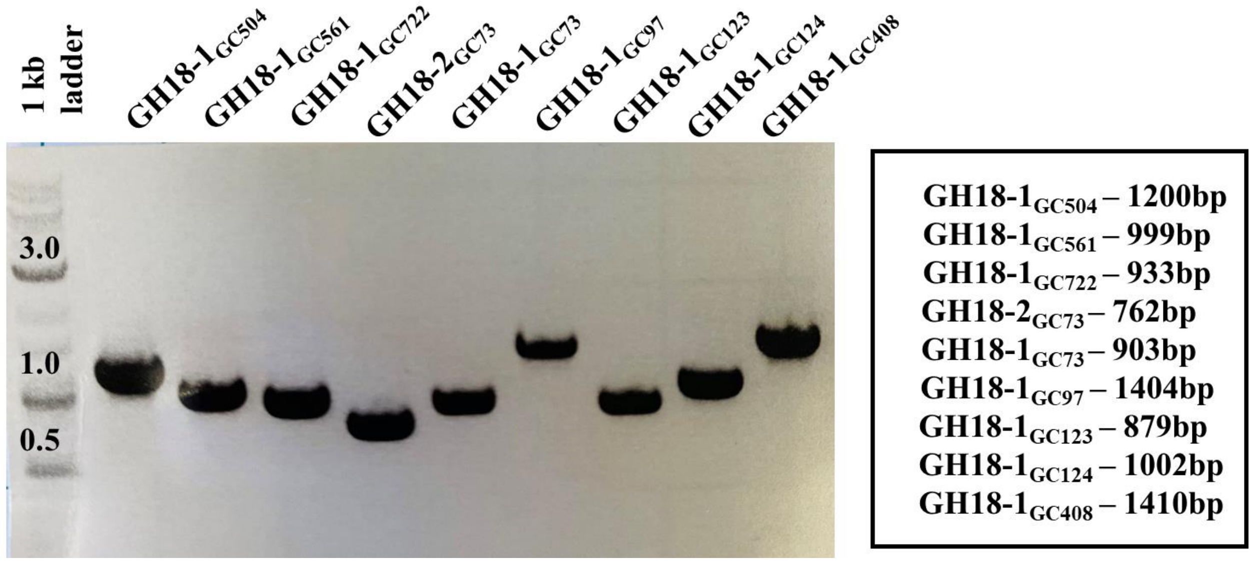


Figure 8 Phusion PCR of individual catalytic domains of GH18 ENGases

Lanes are labeled with strain IDs. The desired amplicon size of the domain region was labeledon the right along with the strain IDs. The protein IDs are: GH18-1$_{GC504}$ – *Bacteroides uniformis*, GH18-1$_{GC561}$ – *Bacteroides vulgatus*, GH18-1$_{GC722}$ – *Bacteroides ovatus*, GH18-1$_{GC73}$ & GH18-2$_{GC73}$ – *Bacteroides fragilis*, GH18-1$_{GC97}$ – *Enterococcus faecalis*, GH18-1$_{GC123}$ – *Alistipes shahii*, GH18-1$_{GC124}$ – *Alistepes finegoldii*, and GH18-1$_{GC408}$ – *Bacteroides uniformis*.

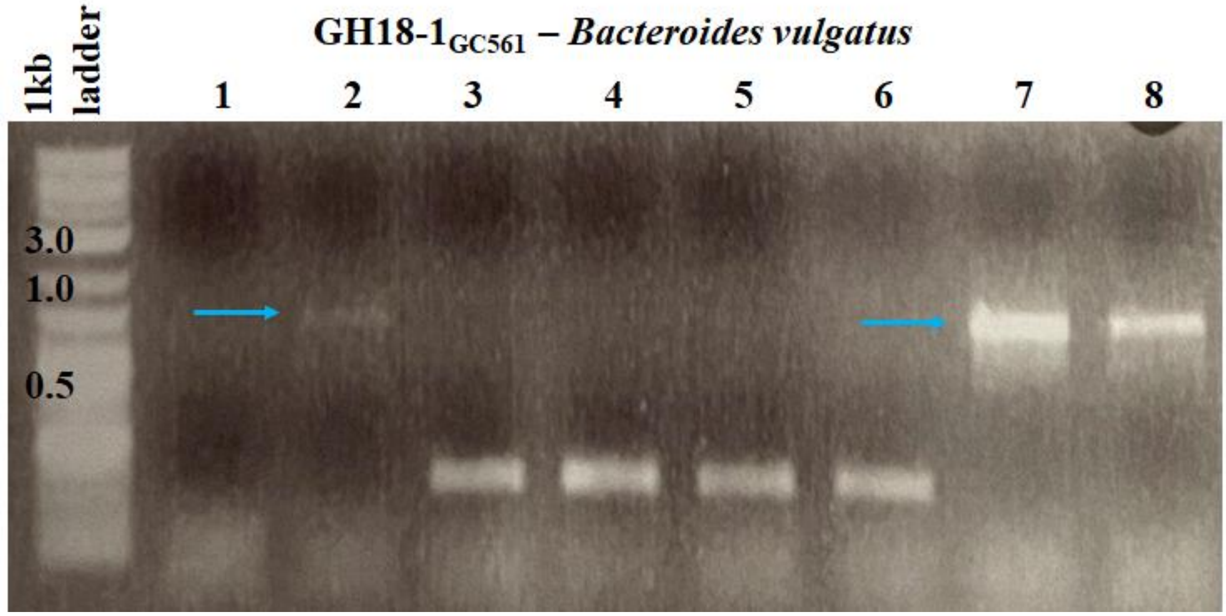


Figure 9 Agarose gel electrophoresis for Colony PCR of ENGAse domain region of GH18-1$_{GC561}$ cloned on to pET28b(+) plasmid

Eight individual colonies of the transformants of the ENGase catalytic domain region of GH18-1$_{GC561}$ (*Bacteroides vulgatus*) were screened for the positive clones carrying the inserts. Lanes were numbered with the colony number that was screened. Arrows represent the clones carrying the inserts. Of the eight screened colonies, three showed the presence of inserts.

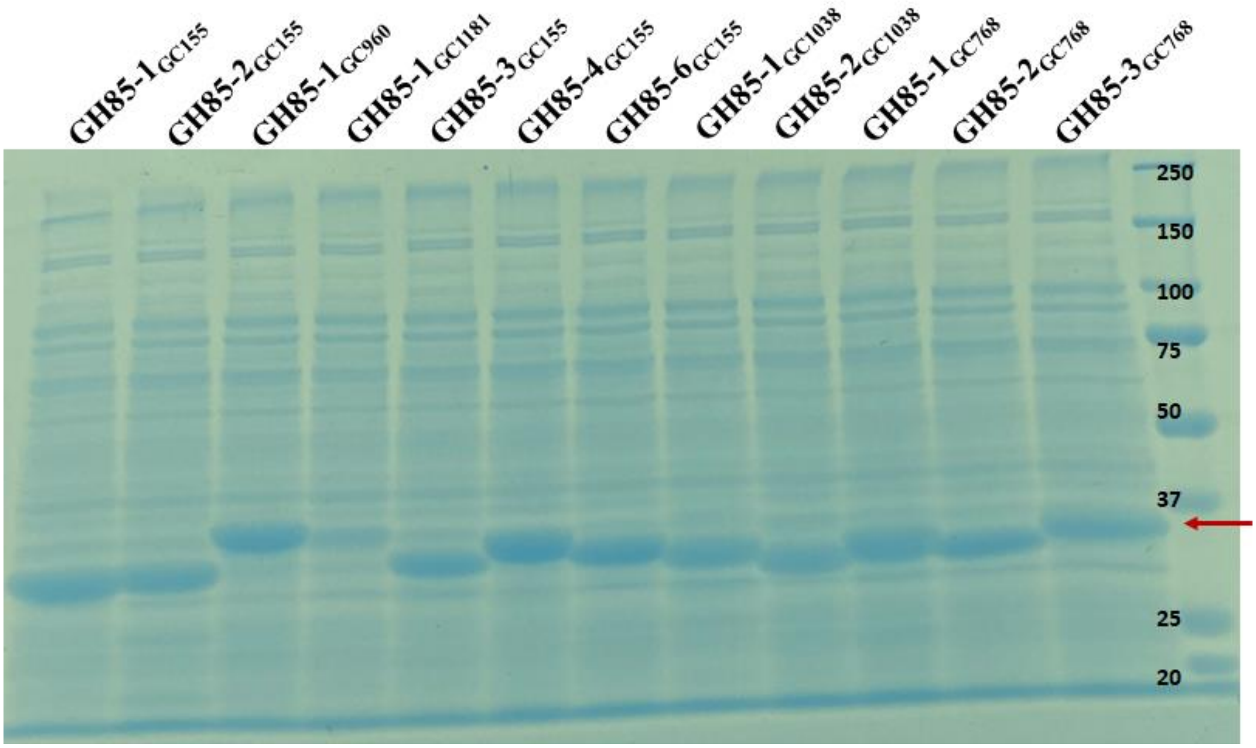
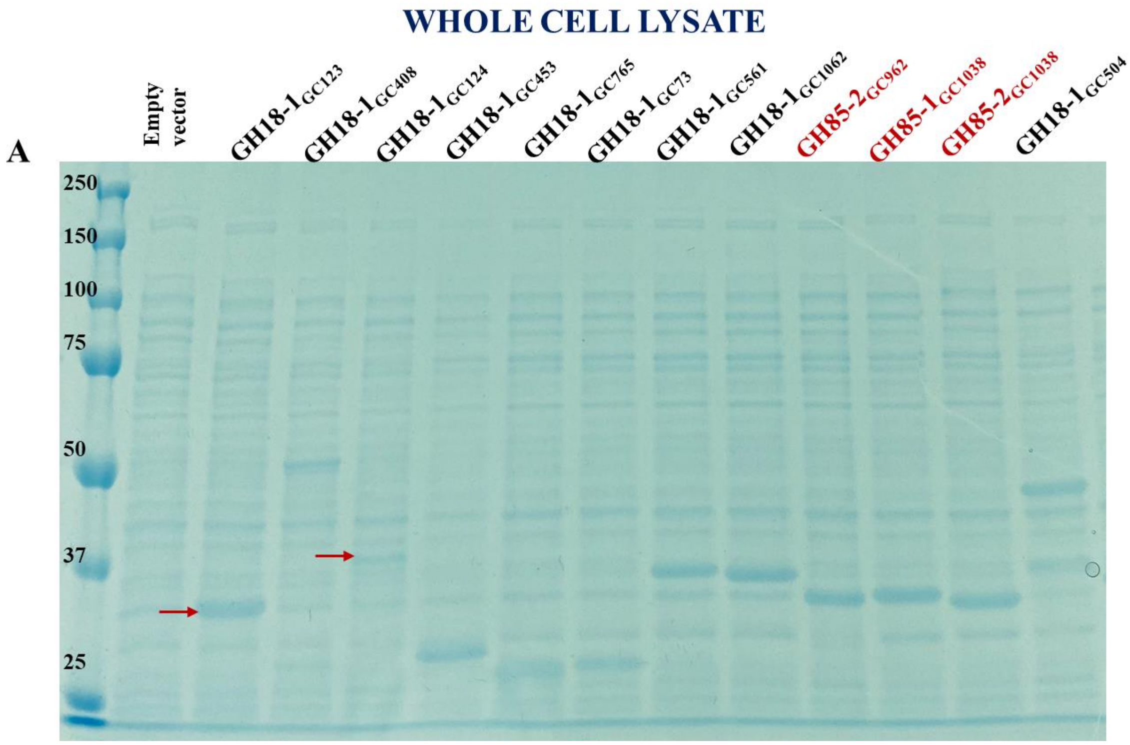


Figure 10 SDS PAGE of catalytic domains of recombinant GH85 ENGases that were expressed at 37°C with 0.1mM IPTG induction

The lanes were numbered with protein IDs (Table 3), and the arrow represents each clone's predicted protein size. Randomly selected GH85 ENGases were loaded on the gel. The predicted domain protein sizes and their respective taxonomic identities are listed in Table 3.


WHOLE CELL LYSATE
A
Empty vector
GH18-1GC123
GH18-1GC408
GH18-1GC124
GH18-1GC453
GH18-1GC765
GH18-1GC73
GH18-1GC561
GH18-1GC1062
GH85-2GC962
GH85-1GC1038
GH85-2GC1038
GH18-1GC504
250
150
100
75
50
37
25

B

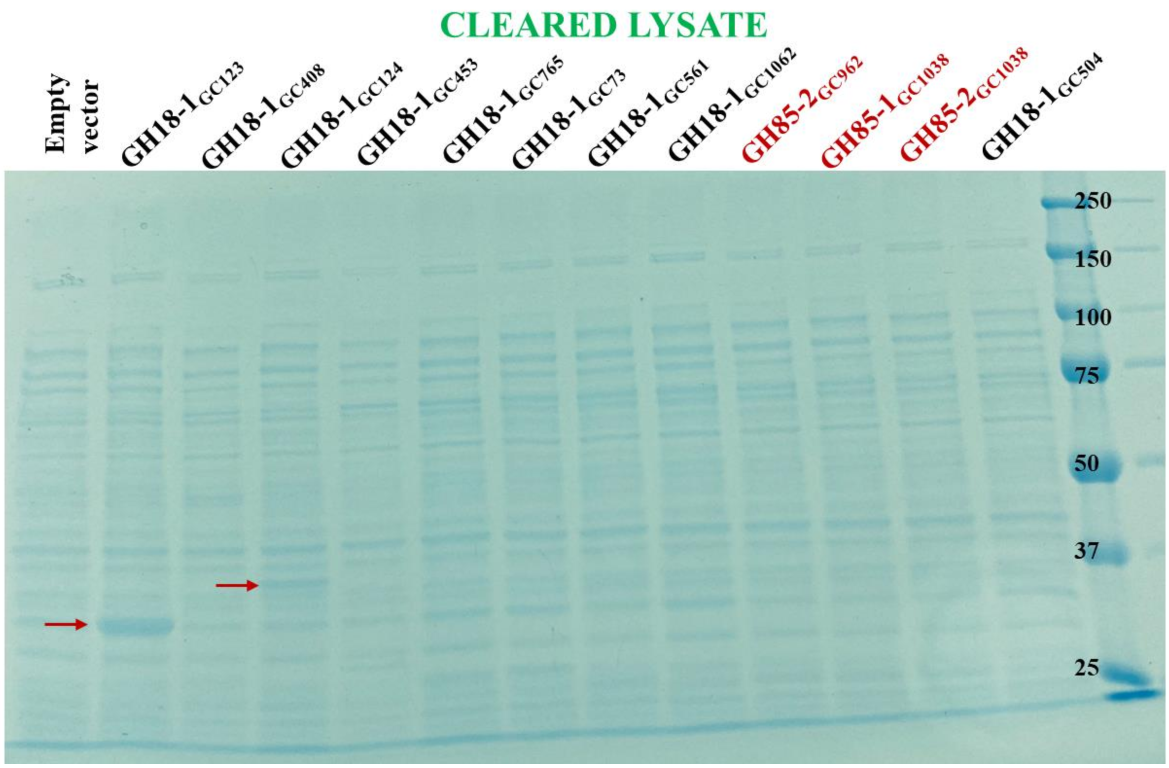


Figure 11 Protein expression of catalytic domains of recombinant GH18 and GH85 ENGases

The lanes are labeled with protein IDs. Whole-cell lysates(A) and cleared lysates(B) were loaded on the SDS-PAGE gel to visualize protein expression. The arrows represent the predicted protein size of each clone. pET28b(+) without the inserts (empty vector) was used as a control. Results of randomly selected GH18 and GH85 ENGases were shown in the figures. GH18 ENGases were labeled in black and GH85 ENGases were labeled in maroon.

The predicted domain sizes are: GH18-1$_{GC123}$(*Alistipes shahii)* :32.58 kDa, GH18-1$_{GC408}$(*Bacteroides uniformis*): 52.7 kDa, GH18-1$_{GC124}$*(Alistipes finegoldii*): 37.65 kDa, GH18-1$_{GC453}$(*Cutibacterium acnes*): 29.36 kDa, GH18-1$_{GC765}$(*Bacteroides* *thetaiotaomicron*): 29.12 kDa, GH18-1$_{GC73}$(*Bacteroides fragilis*): 33 kDa, GH18-1$_{GC561}$ (*Bacteroides vulgatus*): 37 kDa, GH18-1$_{GC1062}$ (*Bacteroides xylanisolvens*): 37kDa,GH85-2$_{GC962}$(*Coprobacter fastidiosus*): 35.98 kDa, GH85-1$_{GC1038}$ and GH85-2$_{GC1038}$ (*Coprobacter fastidiosus*): 35 kDa, GH18-1$_{GC504}$(*Bacteroides uniformis*): 45.31 kDa. Of the 12 proteins expressed, only two of the GH18 proteins were expressed as soluble proteins which can be visualized on both whole cell and cleared lysates whereas those that are in the form of insoluble aggregates (inclusion bodies) were visualized only on the whole cell lysates. The soluble proteins are indicated by a red arrow and include the following proteins: GH18-1$_{GC123}$(*Alistipes shahii)* : 32.58 kDa and GH18-1$_{GC124}$*(Alistipes finegoldii*): 37.65 kDa.

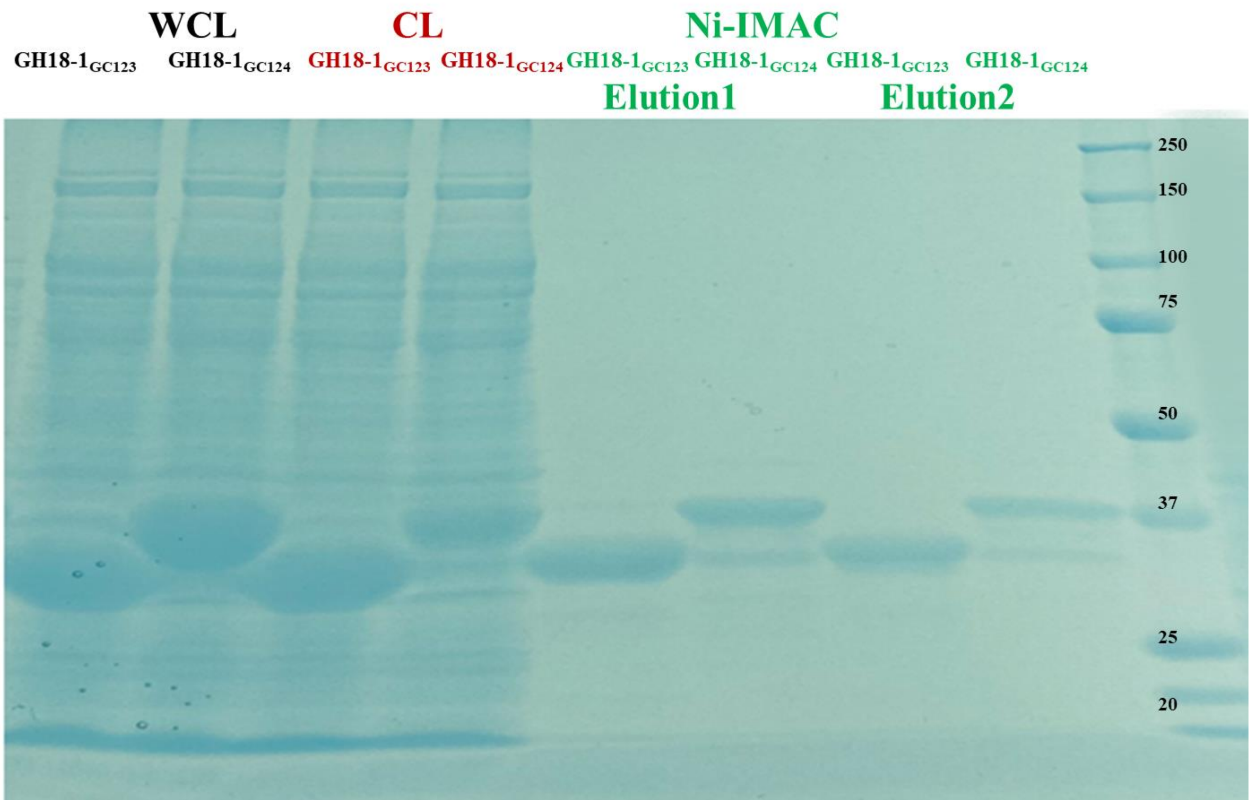


Figure 12 Protein expression and purification of catalytic domains of two soluble recombinant GH18 ENGases

Soluble putative GH18 ENGases - GH18-1$_{GC123}$ (*Alistipes shahii* :32.58 kDa) and GH18-1$_{GC124}$ GC124*(Alistipes finegoldii*: 37.65 kDa) were loaded on SDS-PAGE for visualization. WCL represents the whole cell lysate, CL represents the cleared lysate and Ni-IMAC represents the purification of proteins using Pierce™ High-Capacity Ni-IMAC Magnetic Beads.

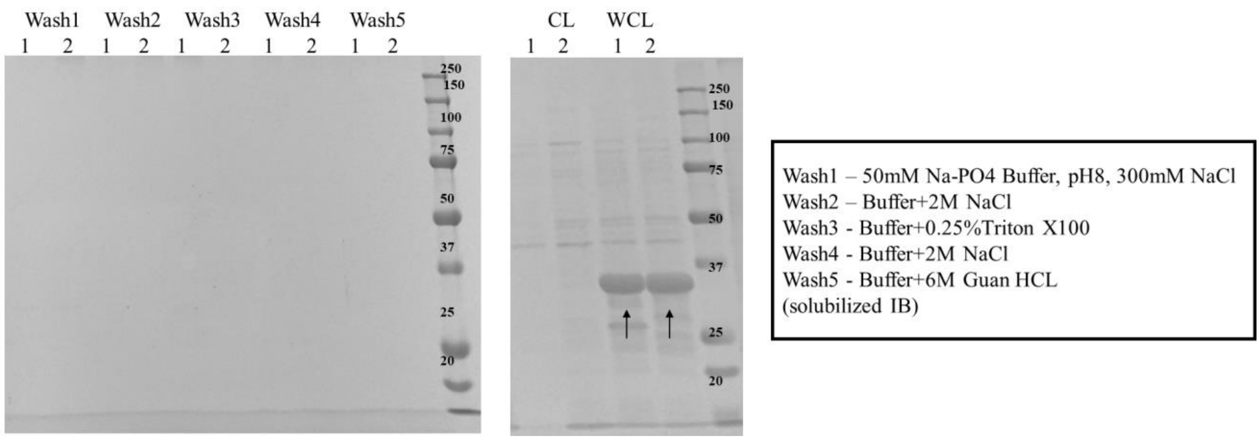


Figure 13 Denaturing protocol for inclusion body solubilization

Lanes are labelled with wash steps and WCL represents whole cell lysate and CL represents cleared lysate. Numbers 1 and 2 refers to randomly selected GH85 ENGases – GH85-1_{GC768} – *Bacteroides nordii* and GH85-2_{GC155} – *Prevotella nanceiensis.* 50mM Sodium phosphate buffer, pH 8 with 300mM sodium chloride (NaCl) was used in each wash step along with additional denaturants that are boxed on the right. Protein expressions are visualized in the whole cell lysates indicated by a black arrow but were not recovered by this inclusion body solubilization protocol.

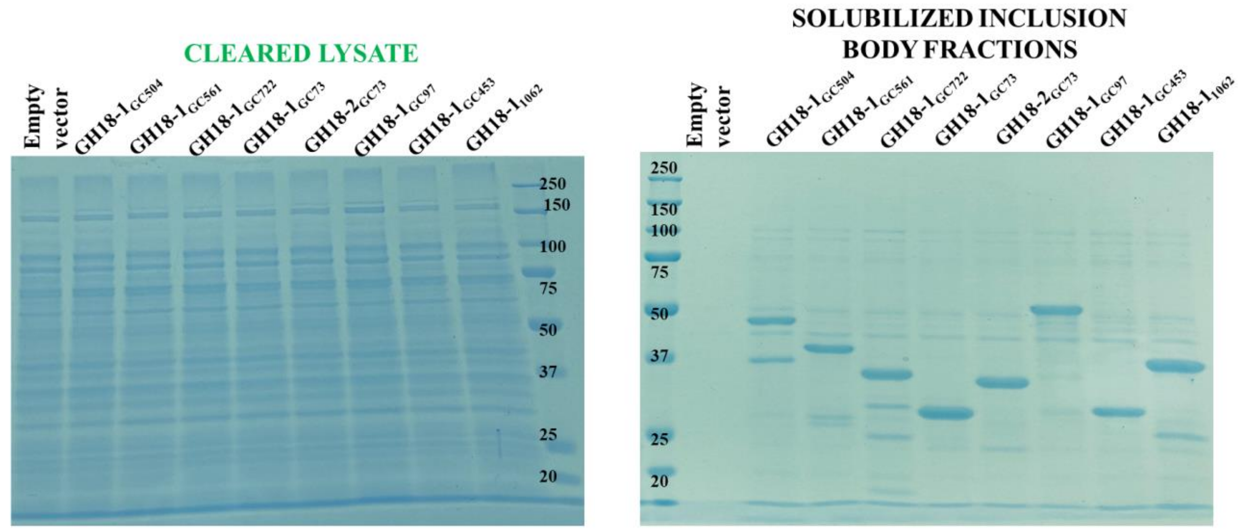


Figure 14 Solubilization of Inclusion bodies using 0.2% N-Lauroyl Sarcosine in 40mM Tris-HCl, pH 8.0

The lanes were labeled with protein IDs. The protein ladder was marked with the protein standard size in kDa. Cleared lysates and the solubilized inclusion bodies carrying the soluble proteins were run on SDS-PAGE. Randomly selected GH18 ENGases were shown on the figure. The predicted catalytic domain sizes were: GH18-1$_{GC504}$(*Bacteroides uniformis*): 45.31 kDa, GH18-1$_{GC561}$ (*Bacteroides vulgatus*): 37kDa, GH18-1$_{GC722}$ (*Bacteroides ovatus*): 35 kDa, GH18-1$_{GC73}$ and GH18-2$_{GC73}$(*Bacteroides fragilis*): 29 and 33 kDa, GH18-1$_{GC\ 97}$ (*Enterococcus faecalis*): 52 kDa, GH18-1$_{GC\ 453}$ (*Cutibacterium acnes*): 29kDa and GH18-1$_{GC1062}$ (*Bacteroides xylanisolvens*): 37 kDa.

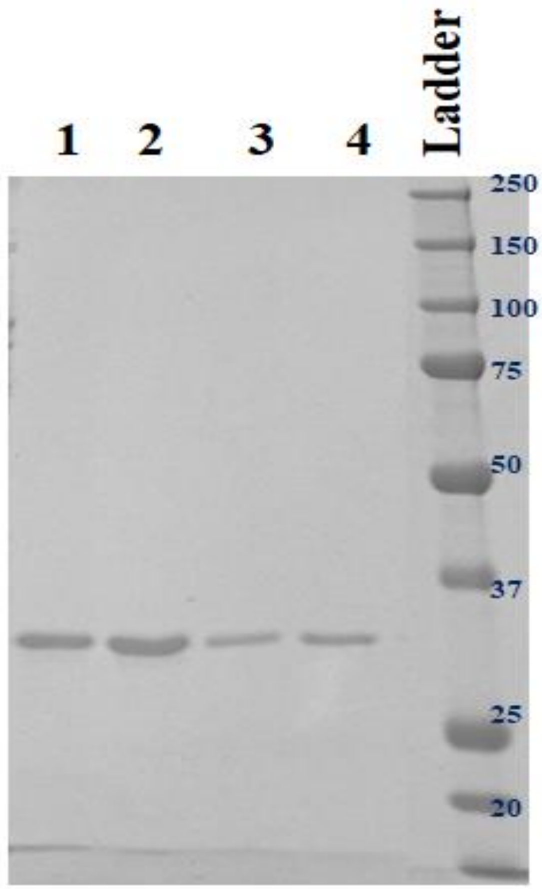


Figure 15 Protein purification using Ni-IMAC magnetic beads

Recombinant ENGases that were expressed in soluble form and solubilized inclusion body fractions were purified using Ni-IMAC magnetic beads and visualized on SDS-PAGE. Lane 1: Purified protein from GH18-1_{GC123}(*Alistipes shahii)* :32.58 kDa (GH18 ENGase expressed as soluble protein); Lane 2 : Purified protein from GH18-2_{GC73}(*Bacteroides fragilis*): 33 kDa; Lane 3 : GH85-2_{GC1038} (*Coprobacter fastidiosus*); Lane 4 : Purified protein from GH85-1_{GC155} (*Prevotella nanceiensis*). Lane 2, 3 and 4 were from solubilized inclusion body fractions.

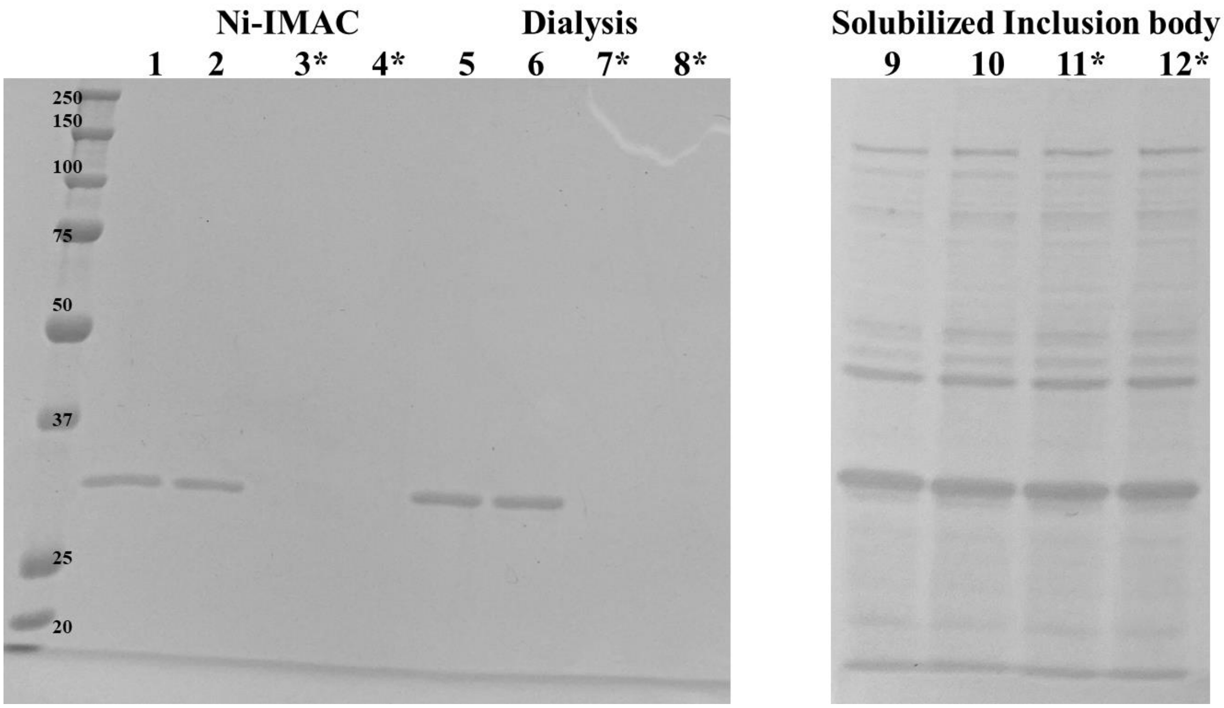


Figure 16 Gel picture to show the loss of stored protein

This figure shows the protein loss after storing them for a week at -20°C for one week. Here we see that the proteins are completely lost in the purified and Dialysis aliquots after storage whereas the solubilized inclusion body fractions were still stable. The gel on the left indicates purified protein using NI-IMAC magnetic beads and dialysis using dialysis membrane whereas the gel on the right indicates the solubilized inclusion body fractions using N-lauroyl sarcosine. Lanes 1,3,5,7,9 and 11 are putative ENGase of GH18-1_{GC123}whereas the lanes 2,4,6,8,10 and 12 are putative ENGase of GH18-2_{GC73}. The * indicates stored protein aliquots.

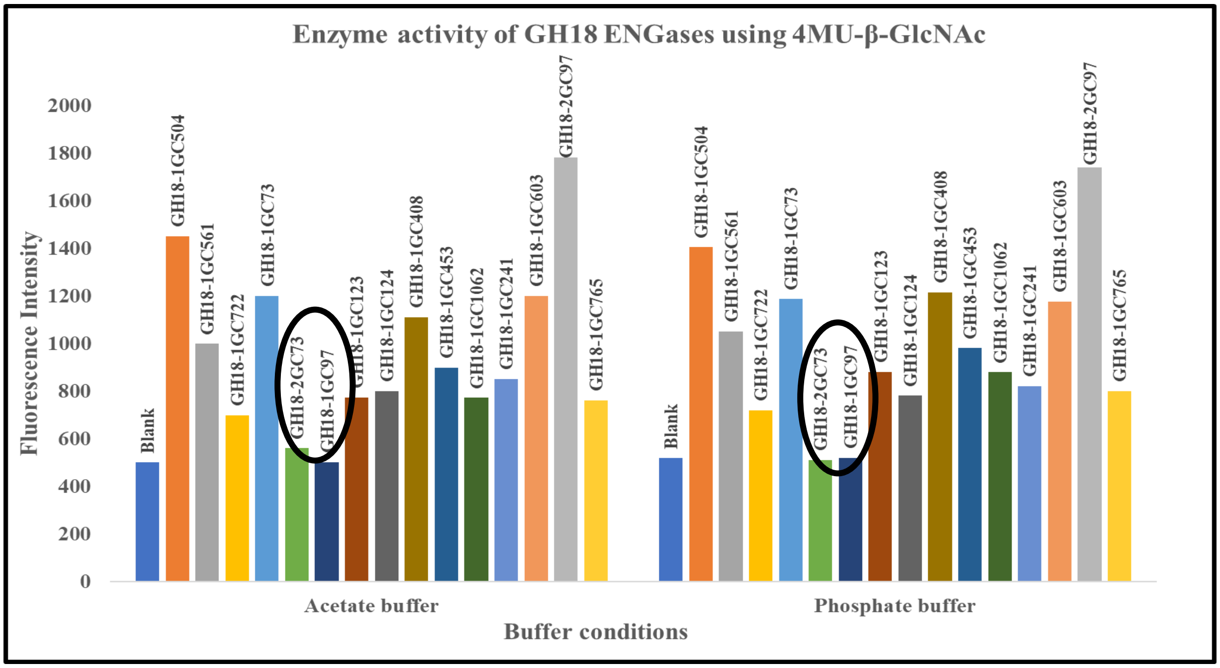
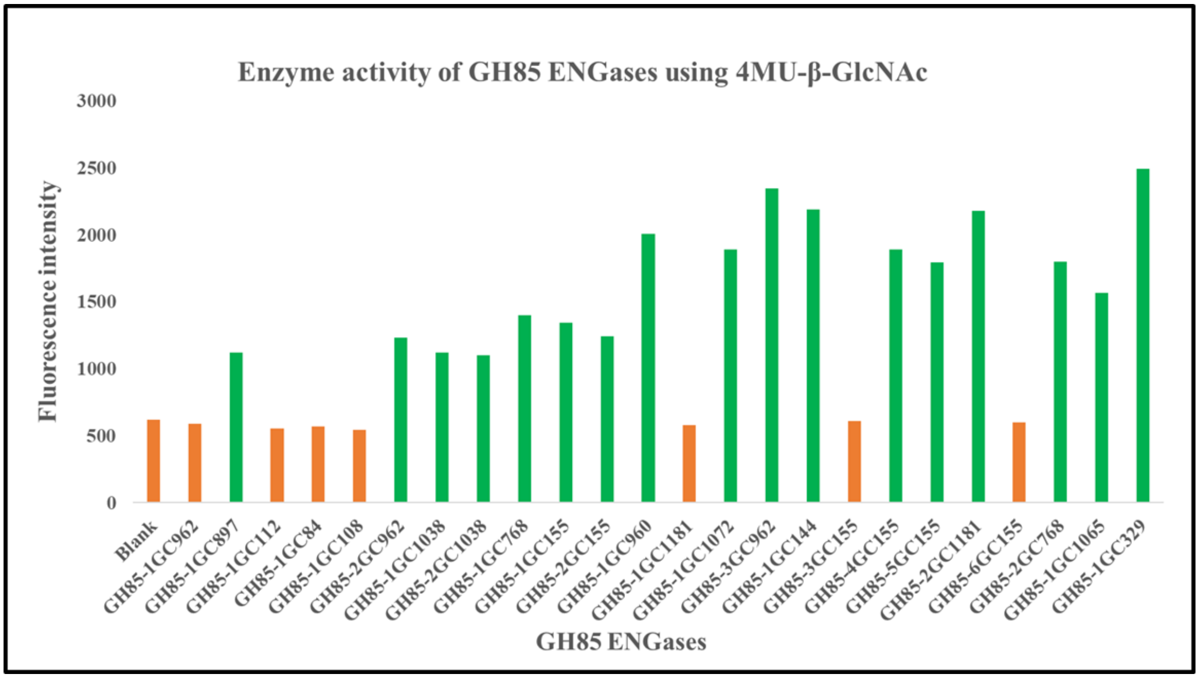


Figure 17 Enzyme activity assay of GH18 ENGases using small molecule substrate

An endpoint assay was performed using 4MU-β GlcNac as a substrate and recombinant putative ENGases as an enzyme. The assay was tested using two different buffer conditions – Sodium acetate buffer pH 5.2 and Sodium phosphate buffer pH 7.4. The empty vector without the inserts was used as the blank that was colored in blue bars. The other color bars represent fifteen GH18ENGases. The bars are labeled with protein IDs and the corresponding bacterial name are given on Table 5. The circled GH18 ENGases – GH18-2_{GC73} and GH18-1_{GC97} did not show any activity against 4MU-β GlcNac and are the same as that of the blank. The X-axis represents the buffer conditions, and the Y axis represents the fluorescence intensity. The protein ID with their taxonomic description is tabulated in Table 6.

A


Enzyme activity of GH85 ENGases using 4MU-β-GlcNAc
Fluorescence intensity
3000
2500
2000
1500
1000
500
0
Blank
GH85-1GC962
GH85-1GC897
GH85-1GC112
GH85-1GC84
GH85-1GC108
GH85-2GC962
GH85-1GC1038
GH85-2GC1038
GH85-1GC768
GH85-1GC155
GH85-2GC155
GH85-1GC960
GH85-1GC1181
GH85-1GC1072
GH85-3GC962
GH85-1GC144
GH85-3GC155
GH85-4GC155
GH85-5GC155
GH85-2GC1181
GH85-6GC155
GH85-2GC768
GH85-1GC1065
GH85-1GC329
GH85 ENGases

B

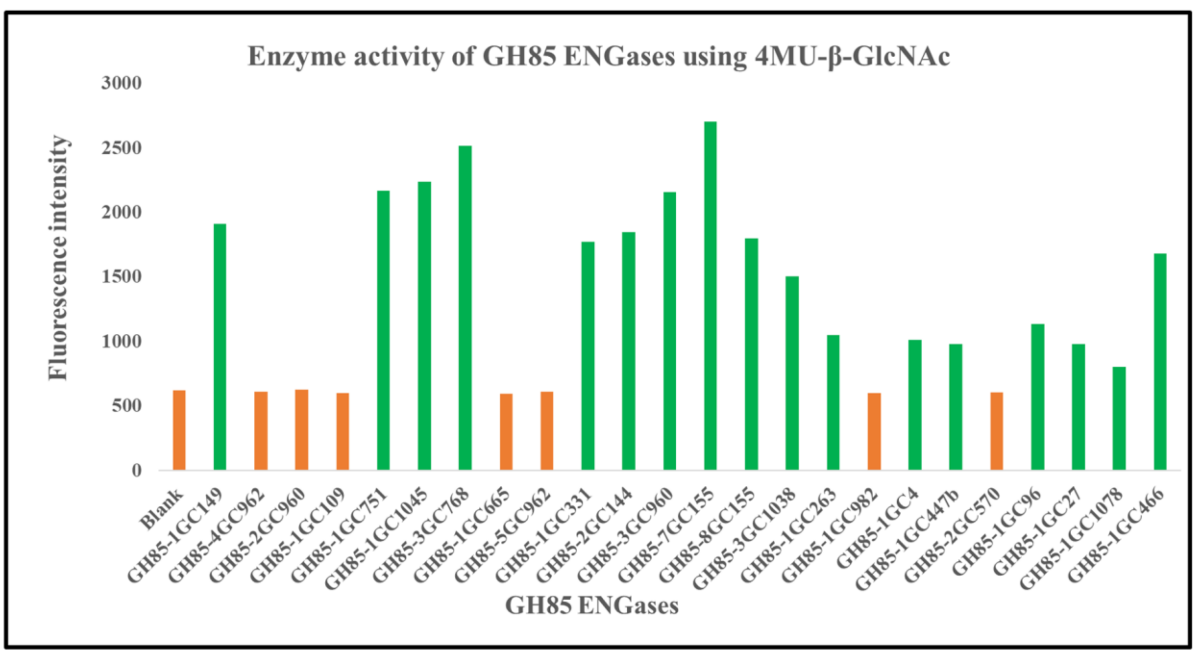
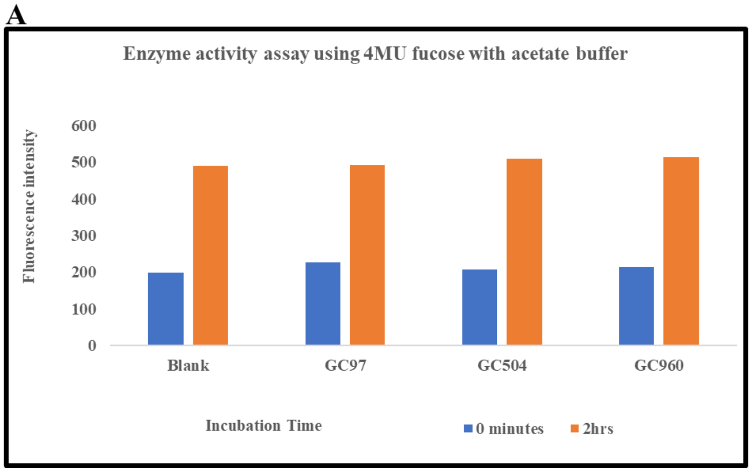

Enzyme activity of GH85 ENGases using 4MU-β-GlcNAc
Fluorescence intensity
3000
2500
2000
1500
1000
500
0
Blank
GH85-1GC149
GH85-4GC962
GH85-2GC960
GH85-1GC109
GH85-1GC751
GH85-1GC1045
GH85-3GC768
GH85-1GC665
GH85-5GC962
GH85-1GC331
GH85-2GC144
GH85-3GC960
GH85-7GC155
GH85-8GC155
GH85-3GC1038
GH85-1GC263
GH85-1GC982
GH85-1GC4
GH85-1GC447b
GH85-2GC570
GH85-1GC96
GH85-1GC27
GH85-1GC1078
GH85-1GC466
GH85 ENGases

Figure 18 A and B Enzyme activity assay of GH85 ENGases using a small molecule substrate

An endpoint assay was performed using 4MU-β GlcNac as a substrate and recombinant putative ENGases as an enzyme. The assay was tested using two different buffer conditions – Sodium acetate buffer pH 5.2 and Sodium phosphate buffer pH 7.4. The graph represents assay conditions tested using Sodium acetate buffer pH 5.2. Data is not shown for assays tested using Sodium phosphate buffer pH 7.4. The empty vector without the inserts was used as the blank that was colored in orange. All the active GH85 ENGases that show high fluorescence intensity are colored in green bars and those that are not active are the same as that of the blank and are labeled in orange bars. The X-axis represents the protein ID for all the individual GH85 ENGases and Y-axis represents the fluorescence intensity. The reactions were performed using both Sodium acetate buffer pH 5.2 and Sodium phosphate buffer pH 7.4. The data of the sodium phosphate buffer are shown in the figures. The protein ID with their taxonomic description are tabulated in Table 6.

A



B

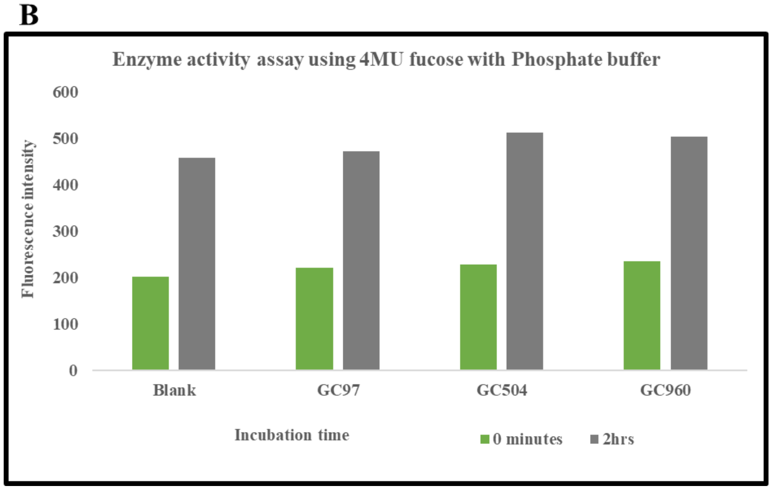


Figure 19 Enzyme activity assay of GH18 and GH85 ENGases using 4MU fucose

An endpoint assay was performed using 4MU-Fucose as a substrate and recombinant putative ENGases as an enzyme. The assay was tested using two different buffer conditions – Sodium acetate buffer pH 5.2 and Sodium phosphate buffer pH 7.4. The empty vector without the inserts was used as the blank. X- axis represents the incubation time - blue color bars in figure A and green color bars in figure B denotes intensity measurement at the start

(0 minutes); while orange color bars in figure A and grey color bars in figure B denotes intensity measurement at the end (2 hours). The Y axis represents the fluorescence intensity. The ENGases used in this assay were : GH18-2_{GC97} - *Enterococcus faecalis*, GH18-1_{GC504} - *Bacteroides uniformis* and GH85-2_{GC960} – *Gabonia massiliensis* labelled at the bottom in X axis. This assay was used as a negative control for the β- GlcNAc assay to show these enzymes specifically act on the bond between 4MU and β GlcNAc. None of the enzymes showed activity and all were same as that of the blank.

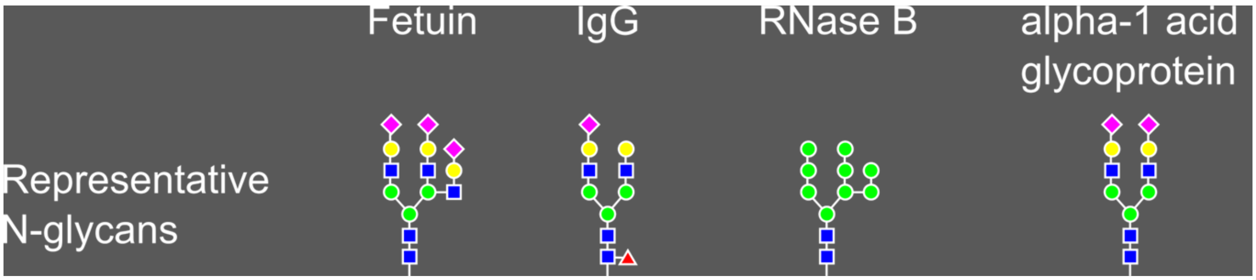
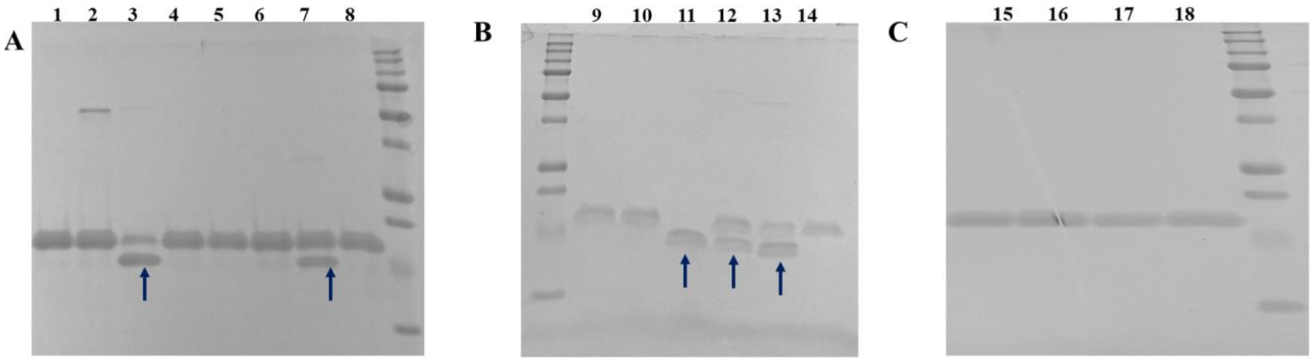


Figure 20 Structure of various glycoprotein substrates

Four different glycoprotein substrates carrying different N-glycan structures were used for enzyme activity assays. Fetuin is a 65 kDa glycoprotein with complex N-glycan structure. IgG is a complex N-glycan carrying glycoprotein having a 55kDa heavy chain that is glycosylated and a 27 kDa light chain. RNAse B, a 15 kDa high mannose N-glycan carrying glycoprotein and α1 acid glycoprotein a 41kDa glycoprotein with complex N-glycans.



Lanes 1,9,15 – RNAse B control	**Lane 10 – GH18-1$_{GC97}$**
Lane 2 – GH18-1$_{GC561}$	**Lane 11 – GH18-2$_{GC97}$**
Lane 3 – GH18-1$_{GC408}$	**Lane 12 – GH18-1$_{GC603}$**
Lane 4 – GH18-1$_{GC453}$	**Lane 13 – GH18-1$_{GC73}$**
Lane 5 – GH18-1$_{GC241}$	**Lane 14 – GH18-2$_{GC73}$**
Lane 6 – GH18-1$_{GC722}$	**Lane 16 – GH18-1$_{GC1062}$**
Lane 7 – GH18-1$_{GC504}$	**Lane 17 – GH18-1$_{GC123}$**
Lane 8 – GH18-1$_{GC765}$	**Lane 18 – GH18-1$_{GC124}$**

Figure 21 Enzyme activity assay of GH18 ENGases using RNAse B

Lanes were numbered from 1 to 18, corresponding to reactions of recombinant GH18 ENGases with RNAse B . Lanes 1, 9, and 15 are RNAse B control reactions without the enzymes. Five of GH18 ENGases were active against RNAse B whereas others that had no activity were similar to the control reaction. Blue arrows on lanes 3, 7, 11, 12, and 13 represent active GH18 ENGases which is visualized by a shift in the size of glycoprotein RNAse B because of the removal of glycans. The protein GH18-2$_{GC97}$ - *Enterococcus faecalis* on Lane 11 shows complete deglycosylation of RNAse B where a complete shift in the size of the RNAse B was visualized on the gel as a result of the removal of glycans. The other active GH18 ENGases : GH18-1$_{GC504}$ - *Bacteroides uniformis,* GH18-1$_{GC408}$ - *Bacteroides uniformis*, GH18-1$_{GC603}$ - *Bacteroides intestinalis*, and GH18-1$_{GC73}$ - *Bacteroides fragilis* shows partial deglycosylation of RNAse B where a partial shift in the size of the RNAse B was visualized on the gel. Corresponding protein IDs to lane numbers are labelled at the bottom of the gel pictures. (protein IDs with corresponding bacteria names are tabulated in Table 5).

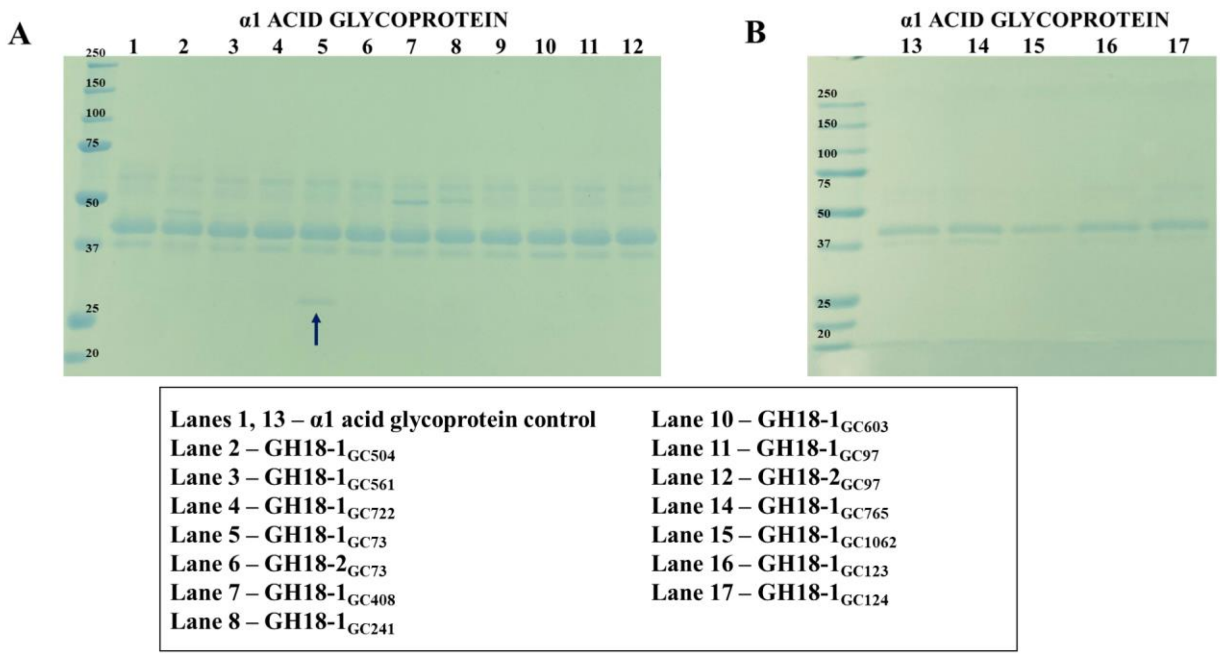


Figure 22 Enzyme activity assay of GH18 ENGases using α1 acid glycoprotein

Figures A and B correspond to SDS-PAGE gels of GH18 ENGase enzyme activity reactions using α1 acid glycoprotein. Lanes were numbered from 1 to 17 that correspond to reactions of putative GH18 ENGases with α1 acid glycoprotein. Lanes 1 and 13 are α1 acid glycoprotein control reactions without enzymes. GH18-1$_{GC73}$ - *Bacteroides fragilis* was active against α1 acid glycoprotein whereas others with no activity were similar to the control reaction. Blue arrow represents the activity of GH18 ENGase on α1 acid glycoprotein that was visualized as a thin band around 20kDa gel as a result of the removal of complex N-glycans from α1 acid glycoprotein, a shift in the size of glycoprotein. None of the other GH18 ENGases were active against α1 acid glycoprotein. Corresponding protein IDs to lane numbers are labelled at the bottom of the gel pictures. (protein IDs with corresponding bacteria names are tabulated in Table 5).

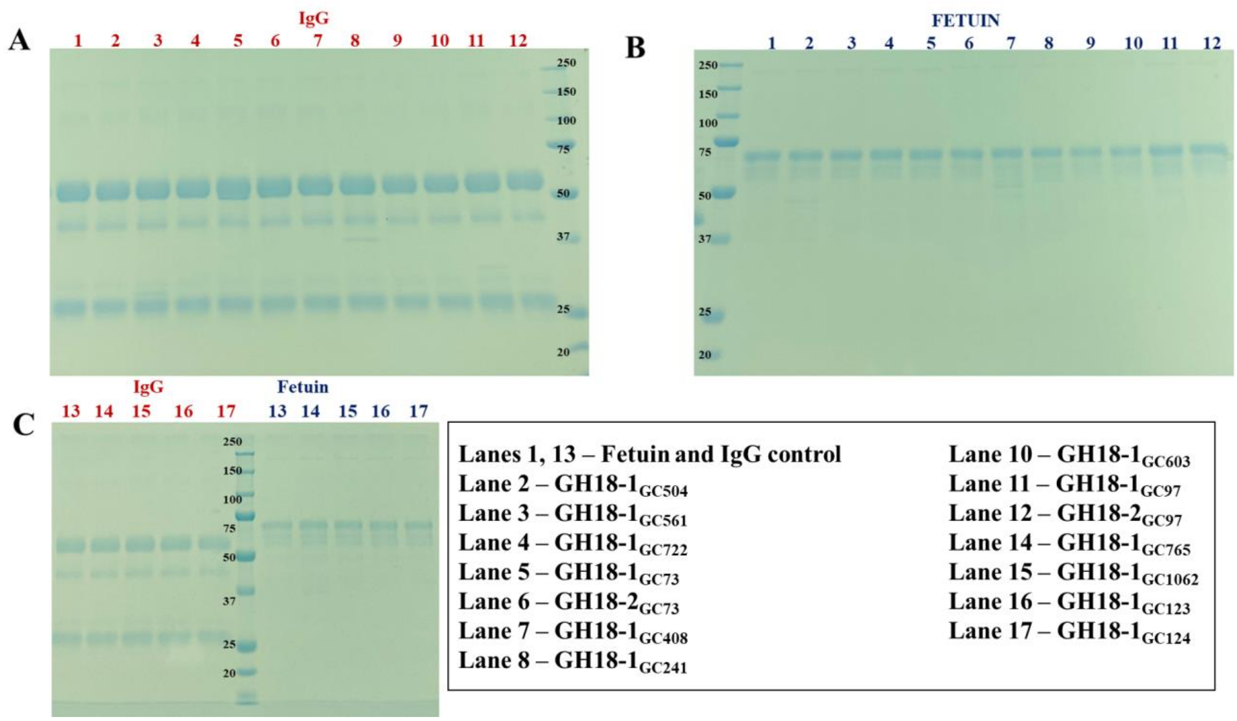
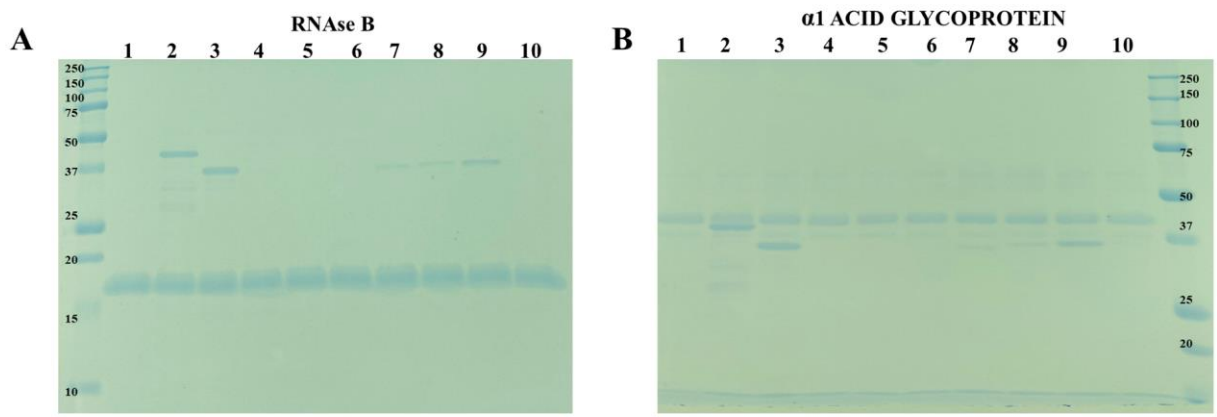


Figure 23 Enzyme activity assay of GH18 ENGases using Fetuin and IgG

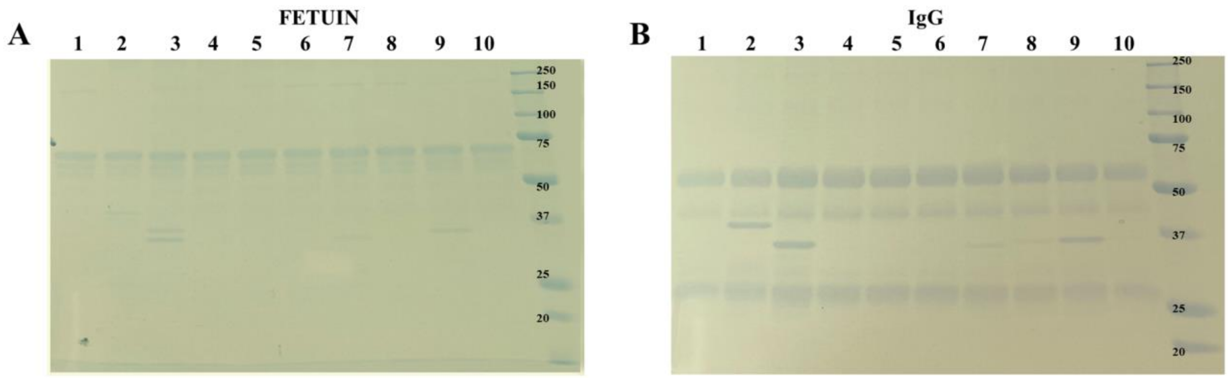
Figures A, B and C correspond to SDS-PAGE gels of GH18 ENGase enzyme activity reactions using IgG and Fetuin. Lanes were numbered from 1 to 17 that correspond to reactions of putative GH18 ENGases with IgG and Fetuin. Lanes 1 and 13 are IgG and Fetuin control reactions without the enzymes There was no change in the activity of GH18 ENGases with IgG and Fetuin and the reactions were similar to that of the controls. Fetuin, a 65 kDa glycoprotein visualized as band closer to 75kDa, and IgG having a 55kDa heavy chain that is glycosylated and a 27 kDa light chain. Corresponding protein IDs to lane numbers are labeled at the bottom of the gel pictures. (protein IDs with corresponding bacteria names are tabulated in Table 5).



Lane 1– Control RNAse B and α1 acid glycoprotein reactions
Lane 2 – GH85-1$_{GC1045}$
Lane 3 – GH85-3$_{GC768}$
Lane 4 – GH85-1$_{GC751}$
Lane 5 – GH85-1$_{GC331}$
Lane 6 – GH85-2$_{GC144}$
Lane 7 – GH85-3$_{GC960}$
Lane 8 – GH85-7$_{GC155}$
Lane 9 – GH85-8$_{GC155}$
Lane 10 – GH85-3$_{GC1038}$

Figure 24 Enzyme activity assay of GH85 ENGases using RNAse B and α1 acid glycoprotein

Figures A and B correspond to SDS-PAGE gels of GH85 ENGase enzyme activity reactions RNAse B and α1 acid glycoprotein, respectively. Lanes were numbered from 1 to 10 that correspond to reactions of putative GH85 ENGases with RNAse B and α1 acid glycoprotein. Lane 1 in figures A and B are the RNAse B and α1 acid glycoprotein control reactions without the enzymes. None of the GH85 ENGases show activity against RNAse B and α1 acid glycoprotein and the reactions were same as that of the controls. RNAse B is a 15kDa glycoprotein and α1 acid glycoprotein is a 41kDa glycoprotein. Corresponding protein IDs to lane numbers are labeled at the bottom of the gel pictures. Selected but not all the GH85 ENGase reactions are shown. (protein IDs with corresponding bacteria names are tabulated in Table 6).



Lane 1– Control Fetuin and IgG reactions
Lane 2 – GH85-1$_{GC1045}$
Lane 3 – GH85-3$_{GC768}$
Lane 4 – GH85-1$_{GC751}$
Lane 5 – GH85-1$_{GC331}$
Lane 6 – GH85-2$_{GC144}$
Lane 7 – GH85-3$_{GC960}$
Lane 8 – GH85-7$_{GC155}$
Lane 9 – GH85-8$_{GC155}$
Lane 10 – GH85-3$_{GC1038}$

Figure 25 Enzyme activity assay of GH85 ENGases using Fetuin and IgG

Figures A and B correspond to SDS-PAGE gels of GH85 ENGase enzyme activity reactions Fetuin and IgG, respectively. Lanes were numbered from 1 to 10 that correspond to reactions of putative GH85 ENGases with Fetuin and IgG. Lane 1 in figures A and B are the Fetuin and IgG control reactions without the enzymes. None of the GH85 ENGases show activity against Fetuin and IgG and the reactions were the same as that of the controls. Fetuin, a 65 kDa glycoprotein visualized as band closer to 75kDa, and IgG having a 55kDa heavy chain that is glycosylated and a 27 kDa light chain. Selected but not all the GH85 ENGase reactions are shown. (protein IDs with corresponding bacteria names are tabulated in Table 6).

A

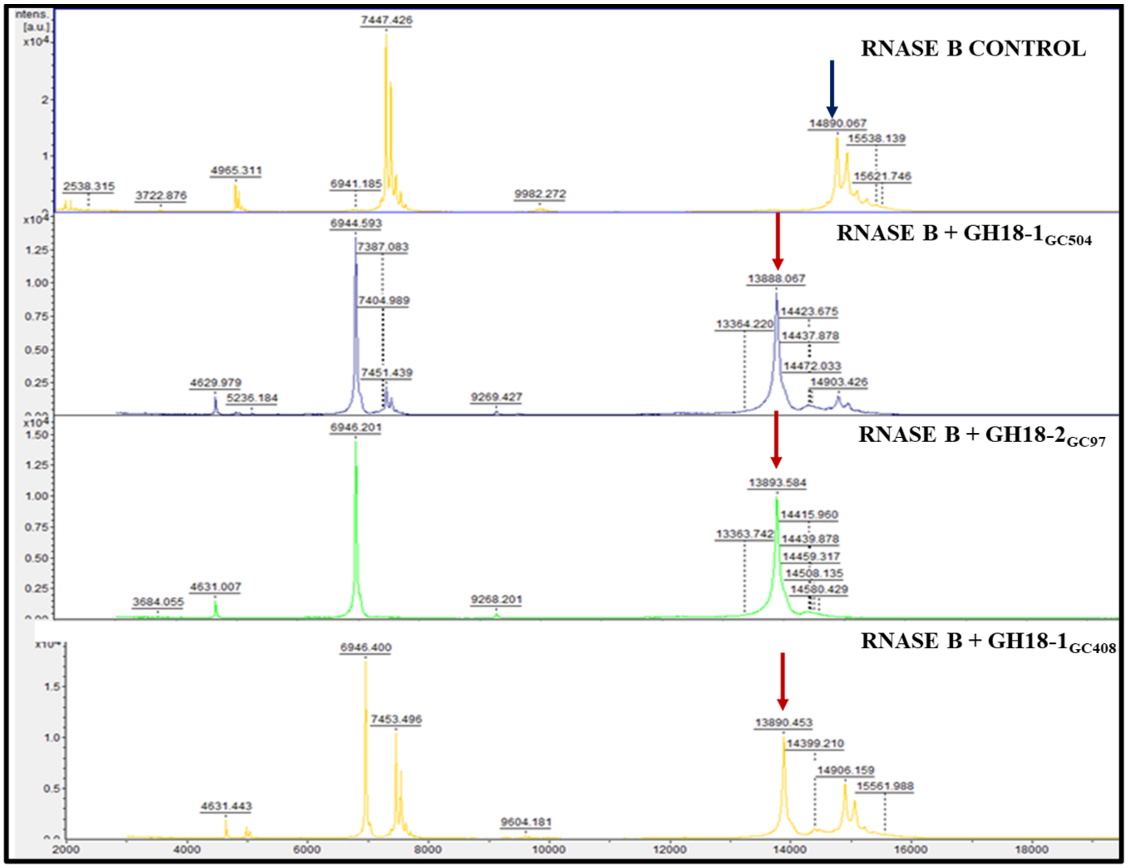


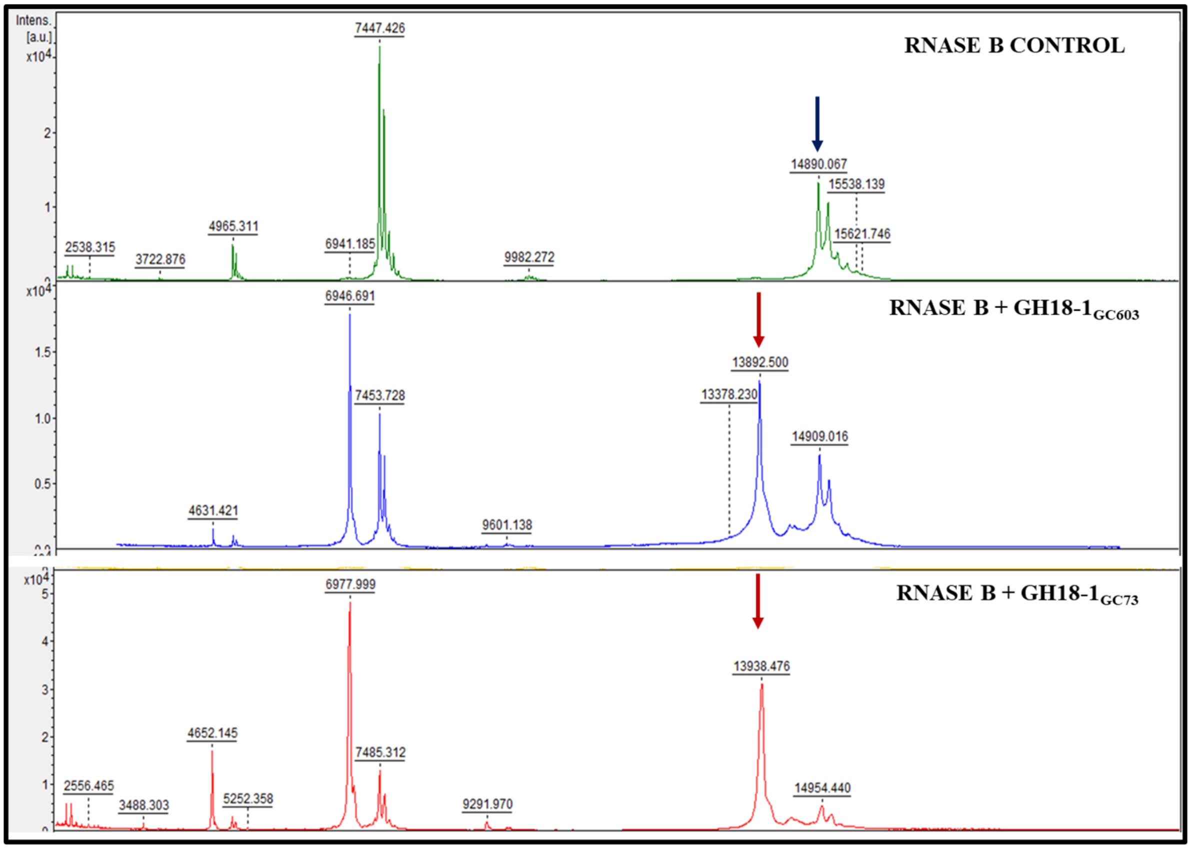
Figure 26 MALDI TOF assay for determining enzyme activity using RNAse B

The enzyme activity assay reaction mix was spotted onto MALDI target plate along with HCCA matrix. RNase B, glycoprotein carrying high mannose N-glycans in its core structure. The control RNAse B reaction contains the substrate glycoprotein, RNAse B without the enzyme. All the test reactions contain glycoprotein substrate RNAse B and respective putative ENGases. Those with enzyme activity show a spectrum different from the control reaction.

A This MALDI TOF spectrum represents the enzyme activity of three active putative GH18 ENGases. The enzyme activity of GH18-1$_{GC504}$ - *Bacteroides uniformis* and GH18-

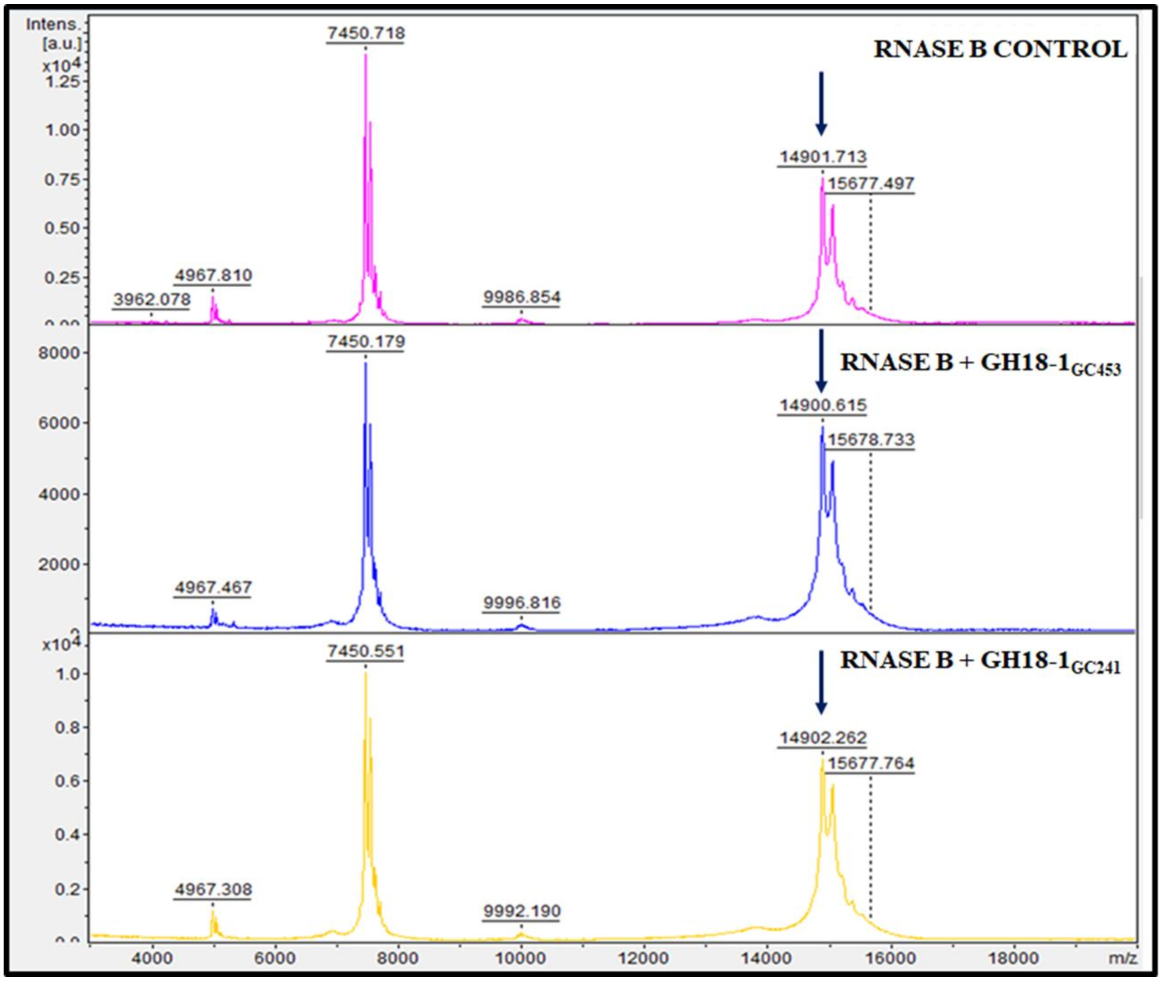
1_{GC408} - *Bacteroides uniformis* that showed partial deglycosylation of RNAse B on SDS-PAGE was visualized on MALDI TOF spectrum as a shift in the mass spectrum of RNAse B as a result of enzyme activity which was indicated by a red arrow. The blue arrow in the control reaction represents the RNAse B mass spectrum. The enzyme activity of GH18-2_{GC97} - *Enterococcus faecalis* that showed complete deglycosylation on SDS PAGE was visualized as a complete shift in the mass spectrum of the control peak that was indicated by a red arrow.

B



B This MALDI TOF spectrum represents the enzyme activity of other two active putative GH18 ENGases. The enzyme activity of GC603 - *Bacteroides intestinalis* and GH18-1$_{GC73}$ - *Bacteroides fragilis* that showed partial deglycosylation of RNAse B on SDS-PAGE was visualized on MALDI TOF spectrum as a shift in the mass spectrum of RNAse B as a result of enzyme activity which was indicated by a red arrow. The blue arrow in the control reaction represents the RNAse B mass spectrum.

C



C. This MALDI TOF spectrum represents the enzyme activity of two putative GH18 ENGases that does not show any activity on RNAse B. The putative GH18 ENGases of GH18-1$_{GC453}$ - *Cutibacterium acnes* and GH18-1$_{GC241}$ – *Bacteroides thetaiotaomicron* did not show any activity on RNASe B that was visualized on SDS PAGE similar to that of the control reaction. The MALDI TOF spectrum of these enzyme reactions were similar to those of control RNAse B reaction where no change in the mass spectrum was visualized that were represented by the blue arrow.

TABLE 2 List of GH18 ENGases used for cloning and protein expression with their Strain ID, phylum, and expected protein sizes of whole gene and catalytic domain

Protein ID	Organism	Phylum	Protein Size (Whole gene)kDa	Protein Size (Catalytic domain)kDa
GH18-1_{GC504}	*Bacteroides uniformis*	Bacteroidota	74.04	45.31
GH18-1_{GC561}	*Bacteroides vulgatus*	Bacteroidota	39.77	37.72
GH18-1_{GC722}	*Bacteroides ovatus*	Bacteroidota	60.66	35.39
GH18-1_{GC73}	*Bacteroides fragilis*	Bacteroidota	39.71	29.77
GH18-2_{GC73}	*Bacteroides fragilis*	Bacteroidota	39.48	33.85
GH18-1_{GC97}	*Enterococcus faecalis*	Firmicutes	94.06	52.22
GH18-1_{GC123}	*Alistipes shahii*	Bacteroidota	55.97	32.58
GH18-1_{GC124}	*Alistipes finegoldii*	Bacteroidota	60.19	37.65
GH18-1_{GC408}	*Bacteroides uniformis*	Bacteroidota	73.63	52.7
GH18-1_{GC453}	*Cutibacterium acnes*	Actinobacteria	32.93	29.36
GH18-1_{GC1062}	*Bacteroides xylanisolvens*	Bacteroidota	41.42	37.51
GH18-1_{GC241}	*Bacteroides thetaiotaomicron*	Bacteroidota	73.84	41.31
GH18-1_{GC603}	*Bacteroides intestinalis*	Bacteroidota	72.97	41.73
GH18-2_{GC97}	*Enterococcus faecalis*	Firmicutes	34.58	29.21
GH18-1_{GC765}	*Bacteroides thetaiotaomicron*	Bacteroidota	33.31	29.12

TABLE 3 List of GH85 ENGases used for cloning and protein expression with their Strain ID, phylum, and expected protein sizes of whole gene and catalytic domain

Protein ID	Organism	Phylum	Protein Size (Whole protein)kDa	Protein Size (Catalytic domain)kDa
GH85-1$_{GC962}$	*Coprobacter fastidiosus*	Bacteroidota	177.77	35.98
GH85-1$_{GC897}$	*Streptococcus parasanguinis*	Firmicutes	175.87	35.93
GH85-1$_{GC112}$	*Streptococcus cristatus*	Firmicutes	174.79	36
GH85-1$_{GC84}$	*Streptococcus gordonii*	Firmicutes	175.65	35.99
GH85-1$_{GC108}$	*Streptococcus dentisani*	Firmicutes	167.13	35.76
GH85-2$_{GC962}$	*Coprobacter fastidiosus*	Bacteroidota	162.40	35.2
GH85-1$_{GC1038}$	*Coprobacter fastidiosus*	Bacteroidota	160.63	35.4
GH85-2$_{GC1038}$	*Coprobacter fastidiosus*	Bacteroidota	161.82	34.92
GH85-1$_{GC768}$	*Bacteroides nordii*	Bacteroidota	157.51	35.64
GH85-1$_{GC155}$	*Prevotella nanceiensis*	Bacteroidota	157.25	34.96
GH85-2$_{GC155}$	*Prevotella nanceiensis*	Bacteroidota	158.60	35.52
GH85-1$_{GC960}$	*Gabonia massiliensis*	Bacteroidota	144.01	38.41
GH85-1$_{GC1181}$	*Barnesiella intestinihominis*	Bacteroidota	142.06	35.94
GH85-1$_{GC1072}$	*Bacteroides nordii W*	Bacteroidota	140.70	36.12
GH85-3$_{GC962}$	*Coprobacter fastidiosus*	Bacteroidota	141.10	36.11
GH85-1$_{GC144}$	*Rothia kristinae*	Actinobacteria	140.05	35.98

Protein ID	Organism	Phylum	Protein Size (Whole protein)kDa	Protein Size (Catalytic domain)kDa
GH85-3$_{GC155}$	*Prevotella nanceiensis*	Bacteroidota	140.85	36.17
GH85-4$_{GC155}$	*Prevotella nanceiensis*	Bacteroidota	139.92	36.38
GH85-5$_{GC155}$	*Prevotella nanceiensis*	Bacteroidota	138.59	36.27
GH85-2$_{GC1181}$	*Barnesiella intestinihominis*	Bacteroidota	137.56	35.73
GH85-6$_{GC155}$	*Prevotella nanceiensis*	Bacteroidota	139.78	36.24
GH85-2$_{GC768}$	*Bacteroides nordii*	Bacteroidota	138.95	35.43
GH85-1$_{GC1065}$	*Desulfovibrio desulfuricans*	Proteobacteria	138.26	35.94
GH85-1$_{GC329}$	*Prevotella melaninogenica*	Bacteroidota	137.77	35.92
GH85-1$_{GC149}$	*Prevotella melaninogenica*	Bacteroidota	137.22	35.87
GH85-4$_{GC962}$	*Coprobacter fastidiosus*	Bacteroidota	136.78	36.23
GH85-2$_{GC960}$	*Gabonia massiliensis*	Bacteroidota	134.54	36.33
GH85-1$_{GC109}$	*Streptococcus intermedius*	Firmicutes	129.53	35.99
GH85-1$_{GC751}$	*Bifidobacterium longum subsp. infantis*	Actinobacteria	122.25	40.05
GH85-1$_{GC1045}$	*Cutibacterium acnes*	Actinobacteria	117.23	40.34
GH85-3$_{GC768}$	*Bacteroides nordii*	Bacteroidota	113.33	35.62
GH85-1$_{GC665}$	*Barnesiella intestinihominis*	Bacteroidota	112.26	36.18
GH85-5$_{GC962}$	*Coprobacter fastidiosus*	Bacteroidota	112	35.19
GH85-1$_{GC331}$	*Prevotella jejuni*	Bacteroidota	111.92	36.64

Protein ID	**Organism**	**Phylum**	**Protein Size (Whole protein) kDa**	**Protein Size (Catalytic domain)kDa**
GH85-2$_{GC144}$	*Rothia kristinae*	Actinobacteria	111.75	36.34
GH85-3$_{GC960}$	*Gabonia massiliensis*	Bacteroidota	111.25	35.77
GH85-7$_{GC155}$	*Prevotella nanceiensis*	Bacteroidota	112.44	36.13
GH85-8$_{GC155}$	*Prevotella nanceiensis*	Bacteroidota	111.96	35.92
GH85-3$_{GC1038}$	*Coprobacter fastidiosus*	Bacteroidota	111.08	35.54
GH85-1$_{GC263}$	*Paraclostridium benzoelyticum*	Firmicutes	112.83	38.86
GH85-1$_{GC982}$	*Lactococcus lactis subsp. tructae*	Firmicutes	102.46	35.77
GH85-1$_{GC4}$	*Lactobacillus plantarum*	Firmicutes	104.39	38.04
GH85-1$_{GC447b}$	*Roseburia intestinalis*	Firmicutes	103.32	36.59
GH85-1$_{GC570}$	*Blautia wexlerae*	Firmicutes	104.46	36.64
GH85-1$_{GC96}$	*Bifidobacterium breve*	Actinobacteria	100.55	35.79
GH85-1$_{GC27}$	*Enterococcus durans*	Firmicutes	101.71	36.5
GH85-1$_{GC1078}$	*Bacillus kyonggiensis*	Firmicutes	99.51	35.25
GH85-1$_{GC466}$	*Butyricimonas paravirosa*	Bacteroidota	82.77	40.1

TABLE 4 List of protein expression and solubilization parameters

Expression and solubilization conditions	**Remarks**
Lowering expression temperature	Different temperatures such as 15°C, room temperature, 30°C and 37°C were used for protein expression to reduce protein aggregation
Varying inducer (IPTG) concentration	IPTG concentrations such as 0.05mM, 0.1mM, 0.2mM, 0.5mM, and 1mM
IPTG induction and growth condition	After growing the clones to absorbance of A600 of 0.6, IPTG induction was done for 2 hours, 4 hours, 8 hours, 12 hours, and 16 hours along with various temperature and inducer concentrations
Solubilization of inclusion body using classical denaturation and refolding protocol	Denaturants such as Sodium chloride, Triton X-100, and Guan HCL that help in the solubilization of proteins and simultaneous refolding using 6M Guan HCL in refolding buffer-50mM sodium phosphate buffer, pH 8.0

Solubilization of inclusion body using 0.2%N-lauroyl sarcosine in10mM Tris-HCl buffer, pH 8.0	Non-denaturing mild solubilization method. No need for refolding. Suitable for recovery of His-tag proteins using Ni-IMAC magnetic beads

TABLE 5 Details of recombinant GH18 ENGases used for activity assays

Protein ID	Bacteria with predicted ENGase domain (GH18)	Activity with 4MU-β-GlcNAc	Activity with glycoproteins			
			RNASE B	IgG	FETUIN	α1AGP
GH18-1$_{GC504}$	*Bacteroides uniformis*	✓	✓	X	X	X
GH18-1$_{GC561}$	*Bacteroides vulgatus*	✓	X	X	X	X
GH18-1$_{GC722}$	*Bacteroides ovatus*	✓	X	X	X	X
GH18-1$_{GC73}$	*Bacteroides fragilis*	✓	✓	X	X	✓
GH18-2$_{GC73}$	*Bacteroides fragilis*	X	X	X	X	X
GH18-1$_{GC97}$	*Enterococcus faecalis*	X	X	X	X	X
GH18-1$_{GC123}$	*Alistipes shahii*	✓	X	X	X	X
GH18-1$_{GC124}$	*Alistipes finegoldii*	✓	X	X	X	X
GH18-1$_{GC408}$	*Bacteroides uniformis*	✓	✓	X	X	X
GH18-1$_{GC453}$	*Cutibacterium acnes*	✓	X	X	X	X
GH18-1$_{GC1062}$	*Bacteroides xylanisolvens*	✓	X	X	X	X
GH18-1$_{GC241}$	*Bacteroides thetaiotaomicron*	✓	X	X	X	X
GH18-1$_{GC603}$	*Bacteroides intestinalis*	✓	✓	X	X	X
GH18-2$_{GC97}$	*Enterococcus faecalis*	✓	✓	X	X	X
GH18-1$_{GC765}$	*Bacteroides thetaiotaomicron*	✓	X	X	X	X

✓ = presence of activity, X = absence of activity. Those shaded in green are enzymes expressed as soluble form and in blue were enzymes active on both 4MU-β GlcNAc and RNAse B. The one shaded in orange was active on 4MU-β GlcNAc, RNAse B and α1 acid glycoprotein.

TABLE 6 Details of recombinant GH85 ENGases used for activity assays

Protein ID	Bacteria with predicted ENGase domain (GH85)	Activity with 4MU-β-GlcNAc	Activi/ty with glycoproteins			
			RNASE B	**IgG**	**FETUIN**	**α_1AGP**
GH85-1$_{GC962}$	*Coprobacter fastidiosus*	X	X	X	X	X
GH85-1$_{GC897}$	*Streptococcus parasanguinis*	✓	X	X	X	X
GH85-1$_{GC112}$	*Streptococcus cristatus*	X	X	X	X	X
GH85-1$_{GC84}$	*Streptococcus gordonii*	X	X	X	X	X
GH85-1$_{GC108}$	*Streptococcus dentisani*	X	X	X	X	X
GH85-2$_{GC962}$	*Coprobacter fastidiosus*	✓	X	X	X	X
GH85-1$_{GC1038}$	*Coprobacter fastidiosus*	✓	X	X	X	X
GH85-2$_{GC1038}$	*Coprobacter fastidiosus*	X	X	X	X	X
GH85-1$_{GC768}$	*Bacteroides nordii*	✓	X	X	X	X
GH85-1$_{GC155}$	*Prevotella nanceiensis*	✓	X	X	X	X
GH85-2$_{GC155}$	*Prevotella nanceiensis*	✓	X	X	X	X
GH85-1$_{GC960}$	*Gabonia massiliensis*	✓	X	X	X	X
GH85-1$_{GC1181}$	*Barnesiella intestinihominis*	X	X	X	X	X
GH85-1$_{GC1072}$	*Bacteroides nordii W*	✓	X	X	X	X
GH85-3$_{GC962}$	*Coprobacter fastidiosus*	✓	X	X	X	X
GH85-1$_{GC144}$	*Rothia kristinae*	✓	X	X	X	X
GH85-3$_{GC155}$	*Prevotella nanceiensis*	X	X	X	X	X

Protein ID	**Bacteria with predicted ENGase domain (GH85)**	**Activity with 4MU-β-GlcNAc**	**Activity with glycoproteins**			
			RNASE B	**IgG**	**FETUIN**	**α_1AGP**
GH85-4$_{GC155}$	*Prevotella nanceiensis*	✓	X	X	X	X
GH85-5$_{GC155}$	*Prevotella nanceiensis*	✓	X	X	X	X
GH85-2$_{GC1181}$	*Barnesiella intestinihominis*	✓	X	X	X	X
GH85-6$_{GC155}$	*Prevotella nanceiensis*	X	X	X	X	X
GH85-2$_{GC768}$	*Bacteroides nordii*	✓	X	X	X	X
GH85-1$_{GC1065}$	*Bacteroides nordii*	✓	X	X	X	X
GH85-1$_{GC329}$	*Prevotella melaninogenica*	✓	X	X	X	X
GH85-1$_{GC149}$	*Prevotella melaninogenica*	✓	X	X	X	X
GH85-4$_{GC962}$	*Coprobacter fastidiosus*	X	X	X	X	X
GH85-2$_{GC960}$	*Gabonia massiliensis*	X	X	X	X	X
GH85-1$_{GC109}$	*Streptococcus intermedius*	X	X	X	X	X
GH85-1$_{GC751}$	*Bifidobacterium longum subsp. infantis*	✓	X	X	X	X
GH85-1$_{GC1045}$	*Cutibacterium acnes*	✓	X	X	X	X
GH85-3$_{GC768}$	*Bacteroides nordii*	✓	X	X	X	X
GH85-1$_{GC665}$	*Barnesiella intestinihominis*	X	X	X	X	X
GH85-5$_{GC962}$	*Coprobacter fastidiosus*	X	X	X	X	X
GH85-1$_{GC331}$	*Prevotella jejuni*	✓	X	X	X	X
GH85-2$_{GC144}$	*Rothia kristinae*	✓	X	X	X	X
GH85-3$_{GC960}$	*Gabonia massiliensis*	✓	X	X	X	X

Protein ID	**Bacteria with predicted ENGase domain (GH85)**	**Activity with 4MU-β-GlcNAc**	**Activity with glycoproteins**			
			RNASE B	**IgG**	**FETUIN**	**α_1AGP**
GH85-7$_{GC155}$	*Prevotella nanceiensis*	✓	X	X	X	X
GH85-8$_{GC155}$	*Prevotella nanceiensis*	✓	X	X	X	X
GH85-3$_{GC1038}$	*Coprobacter fastidiosus*	✓	X	X	X	X
GH85-1$_{GC263}$	*Paraclostridium benzoelyticum*	✓	X	X	X	X
GH85-1$_{GC982}$	*Lactococcus lactis subsp. tructae*	X	X	X	X	X
GH85-1$_{GC4}$	*Lactobacillus plantarum*	✓	X	X	X	X
GH85-1$_{GC447b}$	*Roseburia intestinalis*	✓	X	X	X	X
GH85-1$_{GC570}$	*Blautia wexlerae*	X	X	X	X	X
GH85-1$_{GC96}$	*Bifibobacterium breve*	✓	X	X	X	X
GH85-1$_{GC27}$	*Enterococcus durans*	✓	X	X	X	X
GH85-1$_{GC1078}$	*Bacillus kyonggiensis*	✓	X	X	X	X
GH85-1$_{GC466}$	*Butyricimonas paravirosa*	✓	X	X	X	X

✓ = presence of activity, X = absence of activity. Those shaded in grey were enzymes active on 4MU-β GlcNAc. None of the GH85 ENGases showed activity against various glycoproteins.

CHAPTER FOUR - DISCUSSION

The overarching goal of this research program was to investigate how the activity of GHs from the human gut microbiome could modulate the host cell glycome. The bacterial glycoside hydrolases have the ability to alter host glycans by cleaving specific glycoside linkages removing carbohydrate moieties. However, despite the clear enzymatic specificity for specific glycosidic bonds for each GH family, the specificity of different enzymes for different macromolecular substrates is much less understood. The core focus of this thesis project is to establish a standard methodology for the rapid characterization that can be applied to various GH families of enzymes. In this study, I investigated two families of Endoglycosidases – the endo β N-acetyl glycosaminidases (ENGases) of the GH18 and GH85 families, that are predicted to cleave β 1,4 glycosidic bond between the two N-acetyl glucosamine residues present in the chitobiose core of N-glycans.

The background foundation of this thesis was the bioinformatic analysis of the Surette lab human microbiome collection carried out by Dr. Evan Mann from the Surette laboratory. In this, he identified uncharacterized GH18 and GH85 ENGases and generated sequence similarity networks (Figures 5 and 6) that categorized distinct clusters of related proteins within each GH family. We hypothesize that different ENGases from the same GH family differ in their substrate specificity and have specific substrate structural preferences at positions different from the point of cleavage, and that that each sequence similarity network cluster represents specificity for different macromolecular substrates. As a part of this larger research program, representatives from each cluster were selected for purification and functional characterization.

Sixty three putative ENGases from the sequence similarity network analysis were selected for cloning and overexpression. To use the proteins for the assays, they should be expressed in a soluble form. 61 of 63 of the expressed proteins were in inclusion bodies (insoluble protein aggregates common during recombinant protein expression). Different protocols for expression did not improve solubility and protocols using strong denaturants with refolding on a column were unsuccessful. A protocol for solubilization modified from Jevesar et al., 2008 was successful in recovering soluble protein from inclusion bodies. All the inclusion body fractions were solubilized using mild detergent such as N-Lauroyl sarcosine. However, due to the instability of these proteins after purifying them using Ni-IMAC magnetic beads, we decided to use the solubilized inclusion body fractions from all the insoluble expressed proteins for their enzyme activity assays.

Visualizing the sequence space of GH18 and 85 ENGases, we find many proteins from our collection falling into the same cluster as that of the already characterized and known ENGases (Figures 5 and 6). Of the fifteen GH18 ENGases and forty-eight GH85 ENGases (predicted by dbCAN analysis) assayed using a small molecule substrate (4MU β-GlcNAc), thirteen GH18 ENGases and thirty-three GH85 ENGases (Figures 17 and 18) show activity on this substrate confirming their ability to act on β-GlcNac.

The ENGases active on the small molecule substrates were then tested for their ability to act on various high molecular weight glycoproteins such as IgG, Fetuin, α1 acid glycoprotein, and RNAse B (Figure 20) that carry different types of N-glycans on their structure. These assays will help in identifying the structural preferences of these GH18 and 85 ENGases on various glycoproteins (Figure 4). From the thirteen GH18 ENGases with small molecule activity, five were active on RNAse B (GH18-1$_{GC504}$ - *Bacteroides uniformis*, GH18-1$_{GC408}$ - *Bacteroides uniformis*, GH18-1$_{GC603}$ - *Bacteroides intestinalis*, GH18-2$_{GC97}$ - *Enterococcus faecalis*, and GH18-1$_{GC73}$ - *Bacteroides fragilis*). This glycoprotein is a 15 kDa high mannose N-glycan carrying glycoprotein A MALD- TOF assay with RNAse B also confirmed the activity of these enzymes where a shift in the mass spectrum of glycoproteins corresponding to the loss of the high mannose glycan was observed. One of the GH18 ENGase (GH18-1$_{GC73}$ - *Bacteroides fragilis*) active on RNAseB also showed weak activity on the α1 acid glycoprotein (which has complex N-glycan (CNG)modifications). The activity was visualized on SDS-PAGE as a shift in the size of the glycoprotein because of removal of glycans (Figures 21 and 22). These results confirm one of my hypotheses that different proteins in the same family may have different or same substrate specificities. Of all the five active GH18 ENGases, four proteins belong to the same phylum Bacteriodetes and one belongs to the phylum Firmicutes. None of the GH85 ENGases were active on the glycoprotein substrates

Four of the GH18 ENGases - GH18-1$_{GC504}$, GH18-1$_{GC408}$, GH18-1$_{GC603}$ and GH18-2$_{GC97}$ that are active on high mannose N-glycan(HMNG) carrying substrate (RNAse B) fall into the same cluster as that of the already characterized known ENGases (visualized on SSN in Figure 27) whereas the GH18 ENGase - GH18-1$_{GC73}$ that was active on CNG and HMNG carrying substrates (α1 acid glycoprotein and RNAse B) fall in a separate cluster. The known characterized GH18 ENGases that are visualized on SSN in the same cluster include Endo H, Endo F1, Endo Fsp and ENGase BTVPI548, all of which act on HMNG and hybrid N-glycans of N-linked glycoproteins. This is consistent with my hypotheses that within

SSN, proteins from the same GH families fall into the same cluster and may have the same substrate specificities.

With respect to GH85 ENGases, though many of the predicted GH85 ENGases fall into the same cluster as that of the known characterized ones, we still did not see any activity against the different substrates tested (Figure 28). This could be because of the following conditions - they may require a different concentration of substrates and enzymes, need different types of complex substrates with different structural preferences of N-glycans (such as core fucosylation or presence of bisecting GlcNac), or may require removal of sialic acids from the glycoproteins so that they can recognize the substrates. These proteins may also be more sensitive to the solubilization detergent that was present in these assays.

4.1 CONCLUSION

The human gut cell surface, extracellular matrices, and cell surface proteins of innate immunity contain complex glycans and they form the major interface of host-microbiome interactions. Many human commensal and pathogenic microorganisms encode a variety of glycoside hydrolases that act on these host glycans and modulate this glycosylation. The main objective of this work is to characterize some of GH18 and GH85 ENGases active in N-glycan degradation. This microbial modification of cell surface glycans would essentially help in exploring various inflammatory disease-related immune responses. Understanding such host surface glycan modification by several other GH families would reveal interesting insights on enzyme-based therapeutics and remodeling glycans on various types of cells.

Though the identification of various enzymes of GHs can be established from genome sequences, the prediction of substrate specificity is currently not accurate. So far, a small fraction of GH enzymes have been characterized in detail. In this thesis, I set out to explore substrate specificities of a diverse set of putative ENGases of GH18 and 85 family from the gut microbiome that could reveal distinct specificities. Previous literature studies shows that various pathogens such as *Streptococcus pneumoniae* and *Enterococcus faecalis* use the ENGase activity for modifying surface glycans and utilize them for their attachment to surfaces to cause infection (Bohle et al., 2011 and Hobbs et al., 2018). In this study, we found that five predicted GH18 ENGases showed activity against high mannose N-glycans on the glycoproteins. Adding more enzymes and characterizing them could be important for understanding the physiological role of these enzymes which is the release of N-glycans from host glycoproteins, possibly modulating host immune response.

Also, in this study, we found more than 70% of GH85 ENGases tested were active on small molecule substrates but we were not able to detect any specific activity against different N-glycan-carrying glycoproteins that we tested. Exploration of different glycoproteins and other glycan substrate could be used to characterize these enzymes.

4.2 FUTURE DIRECTIONS

For future directions, exploring ways to maintain the stability of the purified proteins becomes an important part when we want to test them on *in vitro* cell culture assays to look at the modification of glycosylation. Though we were able to identify enzyme activity on solubilized inclusion body fractions, it may be difficult to use them on cell culture assays because of the possibility of interference of detergents in the assays.

The enzyme activity assays in the Aim 2 gave us interesting insights into characterization of many GH18 ENGases. However, we were not able to see the activity of GH85 ENGases with any of the macromolecular substrates. The use of other substrates such as secretory IgAs and hybrid N-glycan carrying glycoproteins in these enzymes assays and carrying out varying concentrations of substrates could pave the way for more characterization of these enzymes. These enzymes can further be used on TLR 2 and TLR4 assays to look at the modifications of N-glycans on these receptors.

The MALDI-TOF assay used in Aim 2 was more convenient, sensitive and less time-consuming for the analysis of enzyme activity. However, this method was not successful for higher molecular weight glycoproteins as HCCA matrix used in the assay were not suitable for proper ionization of these substrates. Hence, using a different matrix such as 2,5-dihydroxybenzoic acid (2,5-DHB) that are more suitable for complex glycoproteins.

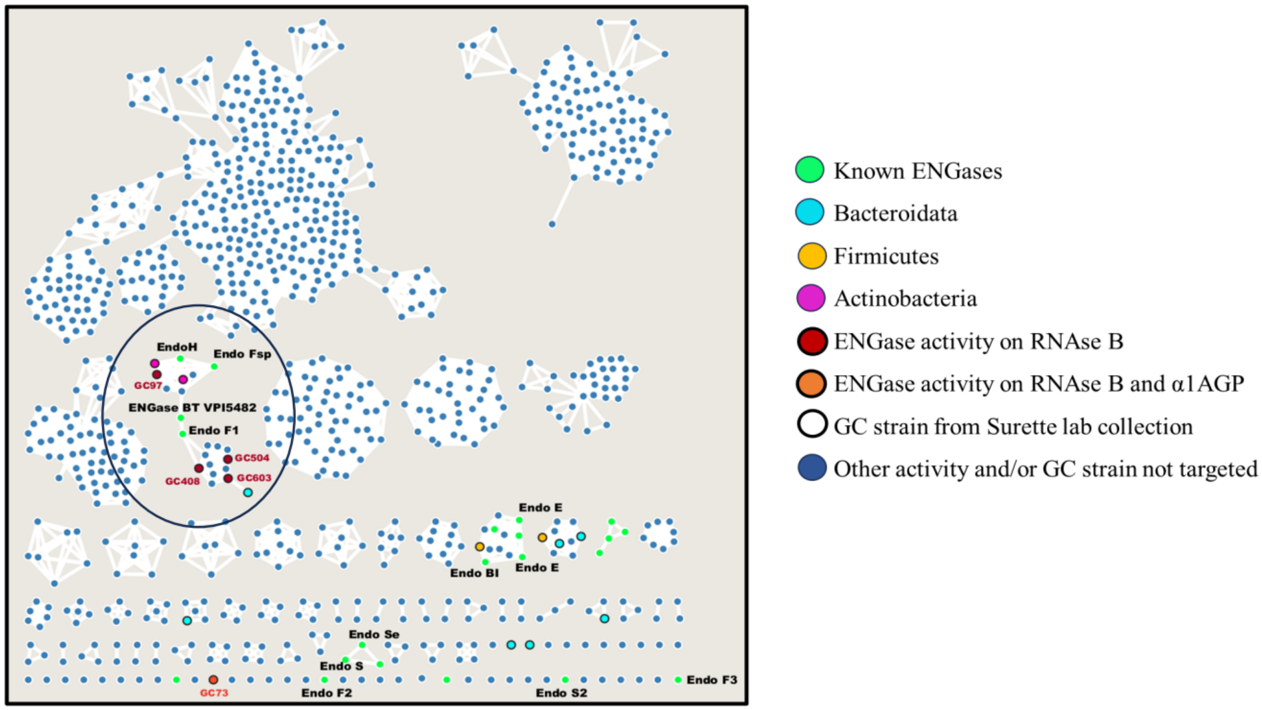


Figure 27 Sequence similarity network using GH18 ENGases

The available GH18 protein sequences (from a public database and Surette lab genome sequences) were clustered with cdhit to 95% identity and used to generate a sequence similarity network clustering at <45% edge identity threshold. Each circle represents a node (indicating protein sequence) and white lines connecting the nodes were the edges (indicating similarity in amino acid sequence). Proteins from different phyla were indicated by a specific color and putative GH18 ENGases used from the Surette lab collection were represented by a black circle around the node. GH18 ENGases active on RNAse B were labeled with their respective protein names and colored maroon. GH18 ENGase of GH18-1_{GC73} - *Bacteroides fragilis* that was active on RNAse B and α1 acid glycoprotein was colored orange. Known ENGase active proteins were labeled with their respective protein name and are colored green. Those proteins with other activity and/or GC strain not targeted were colored blue. The active GH18 ENGases - GH18-1_{GC504} - *Bacteroides uniformis*, GH18-1_{GC408} - *Bacteroides uniformis*, GH18-1_{GC603} - *Bacteroides intestinalis*, and GH18-2_{GC97} - *Enterococcus faecalis* were in the same cluster as that of the known characterized GH18 ENGases such as ENdoH, Endo Fsp, Endo F1 and ENGase of BT VPI5482. Data and figure were generated by Dr. Evan Mann.

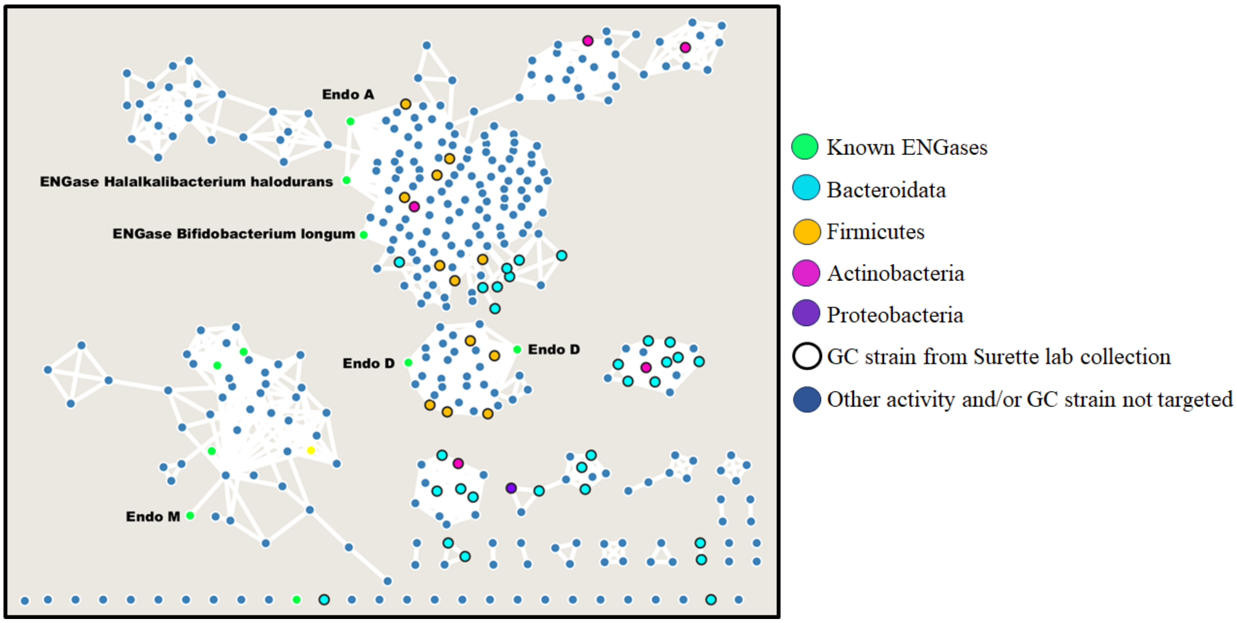


Figure 28 Sequence similarity network using GH85 ENGases

The available GH85 protein sequences (from a public database and Surette lab genome sequences) were clustered with cdhit to 95% identity and used to generate a sequence similarity network clustering at <45% edge identity threshold. Each circle represents a node (indicate protein sequence) and white lines connecting the nodes were the edges(indicating similarity in amino acid sequence). Proteins from different phyla were indicated by a specific color and putative GH85 ENGases used from the Surette lab collection were represented by a black circle around the node. Known ENGase active proteins were labeled with their respective protein name and are colored green. Many of the putative GH85 ENGases were in the same cluster as that of the known characterized ENGases but none of them were active on glycoproteins though some showed activity on fluorescence substrate.

Those proteins with other activity and/or GC strain not targeted were colored blue. Data and figure were generated by Dr. Evan Mann.

REFERENCES

Bell, A., & Juge, N. (2021). Mucosal glycan degradation of the host by the gut microbiota. *Glycobiology*, *31*(6), 691–696. https://doi.org/10.1093/glycob/cwaa097

Bennett, E. P., Mandel, U., Clausen, H., Gerken, T. A., Fritz, T. A., & Tabak, L. A. (2012). Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. *Glycobiology*, *22*(6), 736–756. https://doi.org/10.1093/glycob/cwr182

Berkhout, M. D., Plugge, C. M., & Belzer, C. (2022). How microbial glycosyl hydrolase activity in the gut mucosa initiates microbial cross-feeding. *Glycobiology*, *32*(3), 182–200. https://doi.org/10.1093/glycob/cwab105

Bøhle, L. A., Mathiesen, G., Vaaje-Kolstad, G., & Eijsink, V. G. (2011). An endo-β-N-acetylglucosaminidase from Enterococcus faecalis V583 responsible for the hydrolysis of high-mannose and hybrid-type N-linked glycans. *FEMS microbiology letters*, *325*(2), 123–129. https://doi.org/10.1111/j.1574-6968.2011.02419.x

Briggs, J. A., Grondin, J. M., & Brumer, H. (2021). Communal living: glycan utilization by the human gut microbiota. *Environmental microbiology*, *23*(1), 15–35. https://doi.org/10.1111/1462-2920.15317

Briliūtė, J., Urbanowicz, P. A., Luis, A. S., Baslé, A., Paterson, N., Rebello, O., Hendel, J., Ndeh, D. A., Lowe, E. C., Martens, E. C., Spencer, D. I. R., Bolam, D. N., & Crouch, L. I. (2019). Complex N-glycan breakdown by gut Bacteroides involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci. *Nature microbiology*, *4*(9), 1571–1581. https://doi.org/10.1038/s41564-019-0466-x

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic acids research*, *37*(Database issue), D233–D238. https://doi.org/10.1093/nar/gkn663

CAZypedia Consortium (2018). Ten years of CAZypedia: a living encyclopedia of carbohydrate-active enzymes. *Glycobiology*, *28*(1), 3–8. https://doi.org/10.1093/glycob/cwx089

Collin, M., & Olsén, A. (2001). EndoS, a novel secreted protein from Streptococcus pyogenes with endoglycosidase activity on human IgG. *The EMBO journal*, *20*(12), 3046–3055. https://doi.org/10.1093/emboj/20.12.3046

Cummings R. D. (2019). "Stuck on sugars - how carbohydrates regulate cell adhesion, recognition, and signaling". *Glycoconjugate journal*, *36*(4), 241–257. https://doi.org/10.1007/s10719-019-09876-0

Cuskin, F., Lowe, E. C., Temple, M. J., Zhu, Y., Cameron, E., Pudlo, N. A., Porter, N. T., Urs, K., Thompson, A. J., Cartmell, A., Rogowski, A., Hamilton, B. S., Chen, R., Tolbert, T. J., Piens, K., Bracke, D., Vervecken, W., Hakki, Z., Speciale, G., Munōz-Munōz, J. L., Gilbert, H. J. (2015). Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature*, *517*(7533), 165–169. https://doi.org/10.1038/nature13995

De, S., Kaus, K., Sinclair, S., Case, B. C., & Olson, R. (2018). Structural basis of mammalian glycan targeting by Vibrio cholerae cytolysin and biofilm proteins. *PLoS pathogens*, *14*(2), e1006841. https://doi.org/10.1371/journal.ppat.1006841

Drickamer K, Taylor ME. (2006). *Introduction to Glycobiology* (2nd ed.). Oxford University Press, USA.

Drula, E., Garron, M. L., Dogan, S., Lombard, V., Henrissat, B., & Terrapon, N. (2022). The carbohydrate-active enzyme database: functions and literature. *Nucleic acids research*, *50*(D1), D571–D577. https://doi.org/10.1093/nar/gkab1045

El Kaoutari, A., Armougom, F., Gordon, J. I., Raoult, D., & Henrissat, B. (2013). The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nature reviews. Microbiology*, *11*(7), 497–504. https://doi.org/10.1038/nrmicro3050

Fairbanks A. J. (2017). The ENGases: versatile biocatalysts for the production of homogeneous N-linked glycopeptides and glycoproteins. *Chemical Society reviews*, *46*(16), 5128–5146. https://doi.org/10.1039/c6cs00897f

Freedman, R. B., & Wetzel, R. (1992). Protein engineering. *Current opinion in biotechnology*, *3*(4), 323–325. https://doi.org/10.1016/0958-1669(92)90158-fs

Fujita, M., Shoda, S., Haneda, K., Inazu, T., Takegawa, K., & Yamamoto, K. (2001). A novel disaccharide substrate having 1,2-oxazoline moiety for detection of transglycosylating activity of endoglycosidases. Biochimica et Biophysica Acta (BBA) - General Subjects, 1528(1), 9–14. https://doi.org/10.1016/S0304-4165(01)00164-7

Gerbe, F., Legraverend, C., & Jay, P. (2012). The intestinal epithelium tuft cells: specification and function. *Cellular and molecular life sciences : CMLS*, *69*(17), 2907–2917. https://doi.org/10.1007/s00018-012-0984-7

Gloster, T. M., Turkenburg, J. P., Potts, J. R., Henrissat, B., & Davies, G. J. (2008). Divergence of Catalytic Mechanism within a Glycosidase Family Provides Insight into Evolution of Carbohydrate Metabolism by Human Gut Flora. Chemistry & Biology, 15(10), 1058–1067. https://doi.org/10.1016/j.chembiol.2008.09.005

Gornik, O., Pavić, T., & Lauc, G. (2012). Alternative glycosylation modulates function of IgG and other proteins - implications on evolution and disease. *Biochimica et biophysica acta*, *1820*(9), 1318–1326. https://doi.org/10.1016/j.bbagen.2011.12.004

Grondin, J. M., Tamura, K., Déjean, G., Abbott, D. W., & Brumer, H. (2017). Polysaccharide Utilization Loci: Fueling Microbial Communities. *Journal of bacteriology*, *199*(15), e00860-16. https://doi.org/10.1128/JB.00860-16

Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P., & Davies, G. (1995). Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(15), 7090–7094. https://doi.org/10.1073/pnas.92.15.7090

Henrissat, B., & Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. *The Biochemical journal*, *316 (Pt 2)*(Pt 2), 695–696. https://doi.org/10.1042/bj3160695

Higel, F., Seidl, A., Sörgel, F., & Friess, W. (2016). N-glycosylation heterogeneity and the influence on structure, function and pharmacokinetics of monoclonal antibodies and Fc fusion proteins. *European journal of pharmaceutics and biopharmaceutics : official journal of Arbeitsgemeinschaft fur Pharmazeutische Verfahrenstechnik e.V*, *100*, 94–100. https://doi.org/10.1016/j.ejpb.2016.01.005

Hobbs, J. K., Pluvinage, B., & Boraston, A. B. (2018). Glycan-metabolizing enzymes in microbe– host interactions: The Streptococcus pneumoniae paradigm. FEBS Letters, 592(23), 3865–3897. https://doi.org/10.1002/1873-3468.13045

Ioannou, A., Knol, J., & Belzer, C. (2021). Microbial glycoside hydrolases in the first year of life: An analysis review on their presence and importance in infant gut. *Frontiers in Microbiology*, *12*, 631282

Jevsevar, S., Gaberc-Porekar, V., Fonda, I., Podobnik, B., Grdadolnik, J., & Menart, V. (2005). Production of nonclassical inclusion bodies from which correctly folded protein can be extracted. *Biotechnology progress*, *21*(2), 632–639. https://doi.org/10.1021/bp0497839

Khosravi, A., & Mazmanian, S. K. (2013). Disruption of the gut microbiome as a risk factor for microbial infections. *Current opinion in microbiology*, *16*(2), 221–227. https://doi.org/10.1016/j.mib.2013.03.009

Koropatkin, N. M., Cameron, E. A., & Martens, E. C. (2012). How glycan metabolism shapes the human gut microbiota. *Nature reviews. Microbiology*, *10*(5), 323–335. https://doi.org/10.1038/nrmicro2746

Kudelka, M. R., Stowell, S. R., Cummings, R. D., & Neish, A. S. (2020). Intestinal epithelial glycosylation in homeostasis and gut microbiota interactions in IBD. *Nature reviews. Gastroenterology & hepatology*, *17*(10), 597–617. https://doi.org/10.1038/s41575-020-0331-7

Krautter, F., & Iqbal, A. J. (2021). Glycans and Glycan-Binding Proteins as Regulators and Potential Targets in Leukocyte Recruitment. Frontiers in cell and developmental biology, 9, 624082. https://doi.org/10.3389/fcell.2021.624082

Lapébie, P., Lombard, V., Drula, E., Terrapon, N., & Henrissat, B. (2019). Bacteroidetes use thousands of enzyme combinations to break down glycans. *Nature communications*, *10*(1), 2043. https://doi.org/10.1038/s41467-019-10068-5Marcobal, A., Southwick, A. M., Earle, K. A., & Sonnenburg, J. L. (2013). A refined palate: bacterial consumption of host glycans in the gut. *Glycobiology*, *23*(9), 1038–1046. https://doi.org/10.1093/glycob/cwt040

Ley, RE., Peterson, DA.,& Gordon, JI. (2006). Ecological and Evolutionary Forces Shaping Microbial Diversity in the Human Intestine. Cell, 124(4), 837-848. https://doi.org/10.1016/j.cell.2006.02.017

Li, F. Y., Wang, S. F., Bernardes, E. S., & Liu, F. T. (2020). Galectins in Host Defense Against Microbial Infections. *Advances in experimental medicine and biology*, *1204*, 141–167. https://doi.org/10.1007/978-981-15-1580-4_6

Maldonado, R. F., Sá-Correia, I., & Valvano, M. A. (2016). Lipopolysaccharide modification in Gram-negative bacteria during chronic infection. *FEMS microbiology reviews*, *40*(4), 480–493. https://doi.org/10.1093/femsre/fuw007

Mann, E., Shekarriz, S., & Surette, M. G. (2022). Human Gut Metagenomes Encode Diverse GH156 Sialidases. *Applied and environmental microbiology*, *88*(23) ,e0175522. https://doi.org/10.1128/aem.01755-22

Martens, E. C., Kelly, A. G., Tauzin, A. S., & Brumer, H. (2014). The devil lies in the details: how variations in polysaccharide fine-structure impact the physiology and evolution of gut microbes. *Journal of molecular biology*, *426*(23), 3851–3865. https://doi.org/10.1016/j.jmb.2014.06.022

Martens, E. C., Neumann, M., & Desai, M. S. (2018). Interactions of commensal and pathogenic microorganisms with the intestinal mucosal barrier. *Nature reviews. Microbiology*, *16*(8), 457–470. https://doi.org/10.1038/s41579-018-0036-x

McCarter, J. D., & Withers, S. G. (1994). Mechanisms of enzymatic glycoside hydrolysis. *Current opinion in structural biology*, *4*(6), 885–892. https://doi.org/10.1016/0959-440x(94)90271-2

O Sheridan, P., Martin, J. C., Lawley, T. D., Browne, H. P., Harris, H. M. B., Bernalier-Donadille, A., Duncan, S. H., O'Toole, P. W., P Scott, K., & J Flint, H. (2016). Polysaccharide utilization loci and nutritional specialization in a dominant group of butyrate-producing human colonic *Firmicutes*. *Microbial genomics*, *2*(2), e000043. https://doi.org/10.1099/mgen.0.000043

Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A., & Brown, P. O. (2007). Development of the human infant intestinal microbiota. *PLoS biology*, *5*(7), e177. https://doi.org/10.1371/journal.pbio.0050177

Prado Acosta, M., & Lepenies, B. (2019). Bacterial glycans and their interactions with lectins in the innate immune system. *Biochemical Society transactions*, *47*(6), 1569–1579. https://doi.org/10.1042/BST20170410

Raba, G., & Luis, A. S. (2023). Mucin utilization by gut microbiota: recent advances on characterization of key enzymes. *Essays in biochemistry*, *67*(3), 345–353. https://doi.org/10.1042/EBC20220121

Radovani, B., & Gudelj, I. (2022). N-Glycosylation and Inflammation; the Not-So-Sweet Relation. *Frontiers in immunology*, *13*, 893365. https://doi.org/10.3389/fimmu.2022.893365

Rinninella, E., Raoul, P., Cintoni, M., Franceschi, F., Miggiano, G. A. D., Gasbarrini, A., & Mele, M. C. (2019). What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases. *Microorganisms*, *7*(1), 14. https://doi.org/10.3390/microorganisms7010014

Robb, M., Hobbs, J. K., Woodiga, S. A., Shapiro-Ward, S., Suits, M. D., McGregor, N., Brumer, H., Yesilkaya, H., King, S. J., & Boraston, A. B. (2017). Molecular Characterization of N-glycan Degradation and Transport in Streptococcus pneumoniae and Its Contribution to Virulence. *PLoS pathogens*, *13*(1), e1006090. https://doi.org/10.1371/journal.ppat.1006090

Rudd, P. M., Elliott, T., Cresswell, P., Wilson, I. A., & Dwek, R. A. (2001). Glycosylation and the immune system. *Science (New York, N.Y.)*, *291*(5512), 2370–2376. https://doi.org/10.1126/science.291.5512.2370

Schjoldager, K. T., Narimatsu, Y., Joshi, H. J., & Clausen, H. (2020). Global view of human protein glycosylation pathways and functions. *Nature reviews. Molecular cell biology*, *21*(12), 729–749. https://doi.org/10.1038/s41580-020-00294-x

Singh, A., Upadhyay, V., Upadhyay, A. K., Singh, S. M., & Panda, A. K. (2015). Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. *Microbial cell factories*, 14, 41. https://doi.org/10.1186/s12934-015-0222-8

Sjögren, J., Struwe, W. B., Cosgrave, E. F., Rudd, P. M., Stervander, M., Allhorn, M., Hollands, A., Nizet, V., & Collin, M. (2013). EndoS2 is a unique and conserved enzyme of serotype M49 group A Streptococcus that hydrolyses N-linked glycans on IgG and α1-acid glycoprotein. *The Biochemical journal*, *455*(1), 107–118. https://doi.org/10.1042/BJ20130126

Stals, I., Karkehabadi, S., Kim, S., Ward, M., Van Landschoot, A., Devreese, B., & Sandgren, M. (2012). High resolution crystal structure of the endo-N-Acetyl-β-Dglucosaminidase responsible for the deglycosylation of Hypocrea jecorina cellulases. *PloS one*, *7*(7), e40854. https://doi.org/10.1371/journal.pone.0040854

Stütz, A. E., & Wrodnigg, T. M. (2011). Imino sugars and glycosyl hydrolases: historical context, current aspects, emerging trends. *Advances in carbohydrate chemistry and biochemistry*, *66*, 187–298. https://doi.org/10.1016/B978-0-12-385518-3.00004-3

Suwandi, A., Galeev, A., Riedel, R., Sharma, S., Seeger, K., Sterzenbach, T., García Pastor, L., Boyle, E. C., Gal-Mor, O., Hensel, M., Casadesús, J., Baines, J. F., & Grassl, G. A. (2019). Std fimbriae-fucose interaction increases Salmonella-induced intestinal inflammation and prolongs colonization. *PLoS pathogens*, *15*(7), e1007915. https://doi.org/10.1371/journal.ppat.1007915

Tailford, L. E., Crost, E. H., Kavanaugh, D., & Juge, N. (2015). Mucin glycan foraging in the human gut microbiome. Frontiers in Genetics, 6, 81. https://doi.org/10.3389/fgene.2015.00081

Tarentino, A. L., Quinones, G., Schrader, W. P., Changchien, L. M., & Plummer, T. H., Jr (1992). Multiple endoglycosidase (Endo) F activities expressed by Flavobacterium meningosepticum. Endo F1: molecular cloning, primary sequence, and structural relationship to Endo H. *The Journal of biological chemistry*, *267*(6), 3868–3872.

Tarentino, A. L., Quinones, G., Changchien, L. M., & Plummer, T. H., Jr (1993). Multiple endoglycosidase F activities expressed by Flavobacterium meningosepticum endoglycosidases F2 and F3. Molecular cloning, primary sequence, and enzyme expression. *The Journal of biological chemistry*, *268*(13), 9702–9708

Tomabechi, Y., Squire, M. A., & Fairbanks, A. J. (2014). Endo-β-N-acetylglucosaminidase catalysed glycosylation: tolerance of enzymes to structural variation of the glycosyl amino acid acceptor. *Organic & biomolecular chemistry*, *12*(6), 942–955. https://doi.org/10.1039/c3ob42104j

Tsuji, H., Nishimura, S., Inui, T., Kado, Y., Ishikawa, K., Nakamura, T., & Uegaki, K. (2010). Kinetic and crystallographic analyses of the catalytic domain of chitinase from Pyrococcus furiosus- the role of conserved residues in the active site. *The FEBS journal*, *277*(12), 2683–2695. https://doi.org/10.1111/j.1742-464X.2010.07685.x

Turner J. R. (2009). Intestinal mucosal barrier function in health and disease. *Nature reviews. Immunology*, *9*(11), 799–809. https://doi.org/10.1038/nri2653

Turner, R. D., Mesnage, S., Hobbs, J. K., & Foster, S. J. (2018). Molecular imaging of glycan chains couples cell-wall polysaccharide architecture to bacterial cell morphology. *Nature communications*, *9*(1), 1263. https://doi.org/10.1038/s41467-018-03551-y

Umekawa, M., Higashiyama, T., Koga, Y., Tanaka, T., Noguchi, M., Kobayashi, A., Shoda, S., Huang, W., Wang, L. X., Ashida, H., & Yamamoto, K. (2010). Efficient transfer of sialo-oligosaccharide onto proteins by combined use of a glycosynthase-like mutant of Mucor hiemalis endoglycosidase and synthetic sialo-complex-type sugar oxazoline. *Biochimica et biophysica acta*, *1800*(11), 1203–1209. https://doi.org/10.1016/j.bbagen.2010.07.003

Ursell, L. K., Haiser, H. J., Van Treuren, W., Garg, N., Reddivari, L., Vanamala, J., Dorrestein, P. C., Turnbaugh, P. J., & Knight, R. (2014). The intestinal metabolome: an intersection between microbiota and host. Gastroenterology, 146(6), 1470–1476. https://doi.org/10.1053/j.gastro.2014.03.001

Valdes, A. M., Walter, J., Segal, E., & Spector, T. D. (2018). Role of the gut microbiota in nutrition and health. *BMJ (Clinical research ed.)*, *361*, k2179. https://doi.org/10.1136/bmj.k2179

Varki, A., & Gagneux, P. (2015). Biological Functions of Glycans. In A. Varki, R. D. Cummings, J. D.Esko, P. Stanley, G. W. Hart, M. Aebi, A. G. Darvill, T. Kinoshita, N. H. Packer, J. H.

Prestegard, R. L. Schnaar, & P. H. Seeberger (Eds.), Essentials of Glycobiology (3rd ed.).

Cold Spring Harbor Laboratory Press. http://www.ncbi.nlm.nih.gov/books/NBK453034/

Varki, A., & Kornfeld, S. (2015). Historical Background and Overview. In A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, A. G. Darvill, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar, & P. H. Seeberger (Eds.), Essentials of Glycobiology (3rd ed.).

Cold Spring Harbor Laboratory Press. http://www.ncbi.nlm.nih.gov/books/NBK316258/

Varki A. (2017). Biological roles of glycans. *Glycobiology*, *27*(1), 3–49. https://doi.org/10.1093/glycob/cww086

Vasudevan, D., & Haltiwanger, R. S. (2014). Novel roles for O-linked glycans in protein folding. *Glycoconjugate journal*, *31*(6-7), 417–426. https://doi.org/10.1007/s10719-014-9556-4

Vuong, T. V., & Wilson, D. B. (2010). Glycoside hydrolases: Catalytic base/nucleophile diversity. Biotechnology and Bioengineering, 107(2), 195–205. https://doi.org/10.1002/bit.22838

Watson, M. E., Diepeveen, L. A., Stubbs, K. A., & Yeoh, G. C. (2015). Glycosylation-related Diagnostic and Therapeutic Drug Target Markers in Hepatocellular Carcinoma. *Journal of gastrointestinal and liver diseases : JGLD*, *24*(3), 349–357. https://doi.org/10.15403/jgld.2014.1121.243.mew

Wang L. X. (2008). Chemoenzymatic synthesis of glycopeptides and glycoproteins through endoglycosidase-catalyzed transglycosylation. *Carbohydrate research*, *343*(10-11), 1509–1522. https://doi.org/10.1016/j.carres.2008.03.025

Wang, B., Yao, M., Lv, L., Ling, Z., & Li, L. (2017). The Human Microbiota in Health and Disease. Engineering, 3(1), 71–82. https://doi.org/10.1016/J.ENG.2017.01.008

Yamamoto, K., Kadowaki, S., Fujisaki, M., Kumagai, H., & Tochikura, T. (1994). Novel specificities of Mucor hiemalis endo-beta-N-acetylglucosaminidase acting complex asparagine-linked oligosaccharides. *Bioscience, biotechnology, and biochemistry*, *58*(1), 72–77. https://doi.org/10.1271/bbb.58.72