In vitro characterization of polysaccharide utilization locus 25 from *Bacteroides caccae*

by

Bernadette Alvarez

B.Sc., Major in Biology, Vancouver Island University

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Biochemistry and Microbiology

We acknowledge and respect the Lək^{w}əŋən (Songhees and X^{w}sepsəm/ Esquimalt) Peoples on whose territory the university stands, and the Lək^{w}əŋən and W̱ SÁNEĆ Peoples whose historical relationships with the land continue to this day.

Supervisory Committee

In vitro characterization of polysaccharide utilization locus 25 from *Bacteroides caccae*

by

Bernadette Alvarez
BSc, Major in Biology, Vancouver Island University

Supervisory Committee

Alisdair B. Boraston, PhD
Supervisor, Department of Biochemistry and Microbiology

Lisa Reynolds, PhD
Department Member, Department of Biochemistry and Microbiology

John Burke, PhD
Department Member, Department of Biochemistry and Microbiology

Robert Chow, PhD
External Member, Department of Biology

Abstract

The human gut microbiota (HGM) plays a significant role in maintaining our overall health through its dynamic composition, ability to degrade recalcitrant nutrient sources, and production of metabolites that enable host-gut microbiota crosstalk. However, many aspects of HGM function remain poorly understood. Despite the highly diverse nature of the human diet, our understanding on how the HGM metabolises the glycan and peptide components of glycoproteins is limited. Here, we investigate how polysaccharide utilization locus 25 from *Bacteroides caccae* (BcPUL25) could target the glycosaminoglycan (GAG) and peptide components of aggrecan, a major dietary proteoglycan in animal cartilage. Through structure-function analyses of its encoded carbohydrate-active enzymes (CAZymes), we show that BcPUL25 targets desulfated chondroitin sulfate, or chondroitin, through a unique pathway that uses a carbohydrate dehydratase to prime saturated reducing end β-glucuronic acid residues for cleavage, allowing for more efficient GAG degradation by this PUL. Similarly, biochemical assessment of both BcM60_F and BcM60_G using mucin-specific FRET substrates revealed that BcM60_G can accommodate numerous glycan structures, whereas BcM60_F exhibited more stringent peptide specificity than its counterpart and could only bind to linear glycan moieties. All tested mucin *O*-glycan structures have been observed to decorate the aggrecan peptide backbone. Collectively, these findings establish *B. caccae* as a potential glycoprotein specialist and provides the molecular framework for exploring glycoprotein metabolism by the HGM. Overall, this research contributes toward our understanding of the molecular interplay within the trifecta of microbial composition, nutrient availability, and metabolite production. In the future, studies such as this will support, and eventually result in, the development of therapeutics that will alleviate HGM-influenced disease states.

Table of Contents

List of Figures

List of Supplementary Figures

List of Tables

List of Supplementary Tables

Acknowledgements

"To climb a ladder, one must begin at the bottom."

This was one hell of a ladder. I was lucky I had help. God knows, I needed it.

To name everyone would be another dissertation onto itself, so I will keep this short.

Firstly, I am perpetually grateful for Al Boraston for giving me a chance to climb this ladder. His knowledge, patience, and guidance allowed me to become the scientist I am today. I would not be here without him.

I would also like to profusely thank my committee, Lisa Reynolds, John Burke, and Bob Chow, for their continued support and encouragement, especially during the more difficult years of this project.

My everlasting affection goes to my lab mates, past and present, but most especially to Emily Knudson-Goerner, Olivia Canil, Liam Mihalynuk, Ashley Deventer, and Brendon Medley, for their surprisingly consistent faith when nothing was working and celebrating with me when stuff finally did. If food is love and love is food, you are always welcome at my table.

May the most tender blessings be showered upon the bloodlines of my dear friends, new and old, in and out of the UVic BCMB Department. It truly took a village to get here. If you ever doubt your role in my life, allow me to wax poetic about it over a meal. A very special shoutout to Katherine Lakusta Good, Cindy Kim, and Richard Yip, who kept my sanity stable and spirits high, despite everything, with their great kindness, wonderful cuddles, and (many) strong drink(s).

And lastly, I apologize to my Mama and Papa, Lulu and Gary Alvarez, as well as to the many felines and canines who have supported me through this and so much more. I may not have marriage prospects or job stability, but I'll have this degree and a lightsaber. I believe it makes for a good beginning ☺

Impossible dreams are easier to reach for when cradled in the brightness of your lives.

I am extremely blessed to have been born in these times that have you in it.

Dedication

To my parents, and the choices they made.
They led me here.

Dum spiro, spero.
Dum spero, amo.
Dum amo, vivo.

Chapter 1 : Introduction: Food, CAZymes, and the Phylum Bacteroidota

1.1. Glycans and the human gut microbiota (HGM)

The distal portion of the human gastrointestinal tract contains an incredibly dense and complex microbial ecosystem, commonly referred to as the human gut microbiota (HGM) (Marcobal et al., 2013). The extensive crosstalk between us and our gut microbiota has been well-established, as the HGM affects numerous processes in the body, including regulation of the immune system (Kunisawa et al., 2014; Z. Li et al., 2024; Mann et al., 2024); brain function (Hou et al., 2022; Silva et al., 2020; X. Wang et al., 2019); and gut maintenance (Desai et al., 2016; Shin et al., 2024). This extensive influence is primarily through two factors: the overall microbial composition of the HGM, and the degradation of normally recalcitrant substrates. The former has direct consequences on the host, especially if pathogenic microbes invade and cause infection during dysbiosis (Desai et al., 2016; Hou et al., 2022; Marcobal et al., 2013; Martens et al., 2018). The latter results in the release of bioactive products into the intestinal milieu. These include metabolites like short-chain fatty-acids (SCFAs) (L. W. Chia et al., 2020; Mann et al., 2024) and bioactive peptides (G. Liu et al., 2020; Wu et al., 2021), which have been repeatedly shown to benefit our health and improve numerous disease states (**Figure 1.1)**. Both SCFA and bioactive peptide production is dependent on the microbial degradation of complex carbohydrates and protein from host and dietary sources, respectively (Berkhout et al., 2024; Desai et al., 2016; Ghosh et al., 2020; Inokuma et al., 2023; G. Liu et al., 2020; Pan et al., 2021; Ross et al., 2024; Wu et al., 2021; C. Zhang et al., 2023). The microbial fermentation of glycans, and resulting SCFA production, is assisted by the activity of carbohydrate-active enzymes (CAZymes), such as glycoside hydrolases and polysaccharide lyases (Drula et al., 2022; Terrapon et al., 2018; Wardman et al., 2022). Understanding the molecular basis of how the HGM targets and depolymerize their glycan targets will better guide the research that studies the overall effects of HGM composition and their metabolites in the gut. Given this, as well as modern biotechnological advancements, studying the factors that govern enzyme substrate recognition and catalysis will eventually lead to microbially-informed therapeutics that will alleviate HGM-influenced disease states.

Within the HGM, microbial composition and their relative abundances are inextricably linked to nutrient availability and their ability to depolymerize existing substrates using CAZymes (Desai et al., 2016; Klassen et al., 2021; Marcobal et al., 2013; Soldán et al., 2024). Specifically, the consumption of host-specific and dietary glycans are among the most represented traits of this microbial community (Desai et al., 2016; Marcobal et al., 2013; Ravcheev & Thiele, 2017). Degrading glycan structures is complicated, as sugars may be modified through acetylation, sulfation, and phosphorylation at various positions (Muthana et al., 2012). Additionally, unlike peptides and nucleotides, there is greater variability in the possible glycosidic linkages between glycans, resulting in polysaccharide chains which can be linear or branched, further lending complexity to this already-elaborate molecule (Lebrilla et al., 2022; Seeberger, 2022). As a result, the HGM needs a consortium of enzymes to access and use sugars as a carbon and energy source. Metabolic flexibility (Desai et al., 2016; Martín et al., 2023; Nie et al., 2021; Shin et al., 2024; Ye et al., 2020), and/or mutualistic cooperation (Berkhout et al., 2022; Martín et al., 2023; Nie et al., 2021; Overbeeke et al., 2022; Ravcheev & Thiele, 2017), within this highly competitive community is required to break down complex carbohydrates and survive in the distal gut. Thus, it is not surprising that members of the HGM are highly adapted to efficiently catabolize these diverse glycan structures for downstream metabolic processing.

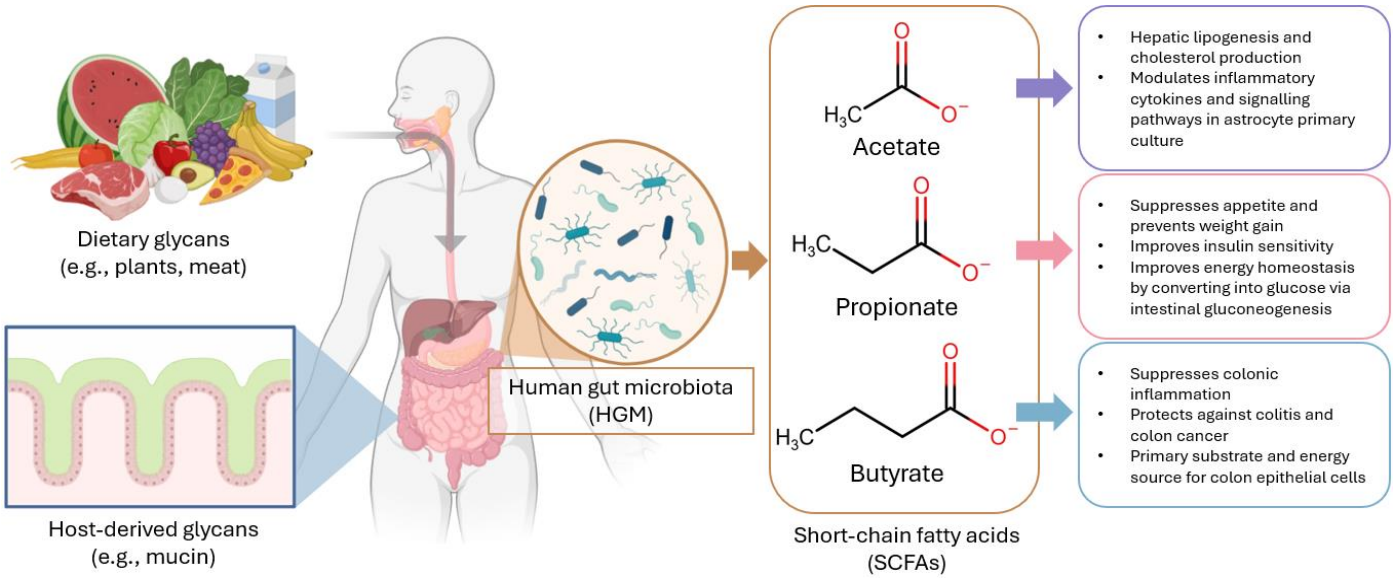


Figure 1.1: Glycan degradation by the human gut microbiota and its downstream effects.

Information acquired from the following sources (Harris et al., 2020; Mann et al., 2024; Silva et al., 2020). Images acquired from BioRender. Chemical drawings generated using Chemical Sketch Tool on RCSB PDB (ChemAxon, n.d.).

1.1.1. Mucin as a host glycan nutrient source

Maintaining mucosal integrity is essential for health (Desai et al., 2016; Hansson, 2020; Martens et al., 2018; Paone & Cani, 2020). Within the gastrointestinal tract, the biological role of mucus is twofold (Desai et al., 2016; Marcobal et al., 2013; Paone & Cani, 2020). It is a secretion that acts as an immunological barrier that physically separates the host epithelium from microbial communities, and a major source of nutrients for a dietarily deprived HGM, which can influence the composition of the resident microbes (Desai et al., 2016). Mucus is constitutively secreted by goblet cells, which are dispersed alongside other specialized cell types (Martens et al., 2018). In humans, mucus thickness, as well as goblet cell abundance, varies greatly along the gastrointestinal tract (Paone & Cani, 2020). For instance, in the large intestine, mucus has a densely packed inner layer that is impermeable to bacteria, and a loose, unattached, outer layer that acts as a habitat and nutrient source for commensal microbes. This is due to a higher abundance of goblet cells in the colon. The observed gradual increase in goblet cell-to-enterocyte ratio from small intestine to colon is likely in response to the proportional increase in number of microorganisms. Overall, these mucosal layers are primarily composed of water, lipids, electrolytes, and a family of heavily glycosylated proteins called mucins (Hansson, 2020; Marcobal et al., 2013; Paone & Cani, 2020; Ravcheev & Thiele, 2017).

Structurally, mucins are thought to resemble bottlebrushes. Their form comes from being extensively *O*-glycosylated on their PTS regions; up to 80% carbohydrate by mass (Paone & Cani, 2020; Tailford et al., 2015). There are two main types of mucins in the gut: transmembrane mucins and gel-forming mucins. The former is essential for properly establishing a protective glycocalyx around enterocytes and providing attachment points for the HGM, as well as playing important biological roles in cell signalling, cell-cell and cell-matrix interactions (Hansson, 2020; Tailford et al., 2015). The latter constitutes as the major structural and functional attributes of mucus (Paone & Cani, 2020; Tailford et al., 2015). Secreted mucins form a net-like structure with bicarbonate ions present in the intestinal milieu, expanding to 100-1000 times in volume by binding water (Paone & Cani, 2020). Of the thirteen mucins expressed in humans, four are differentially secreted along the gastrointestinal tract: mucin-2 (MUC2), MUC5A/C, MUC5B, and MUC6 (**Figure 1.2**). Of these mucins, MUC2 is the major mucin produced by goblet cells within the colonic mucosal layer (Marchix et al., 2018; Marcobal et al., 2013; Paone & Cani, 2020; Ravcheev & Thiele, 2017).

Notably, the abundance of sulfated mucin *O*-glycans increases significantly in the distal colon, wherein MUC2 is heavily sulfated up to ~10% by mass (Luis et al., 2023). Thus, accessing the mucin glycans, as well as peptide backbone, requires a concerted effort and specialized enzymes from the HGM.

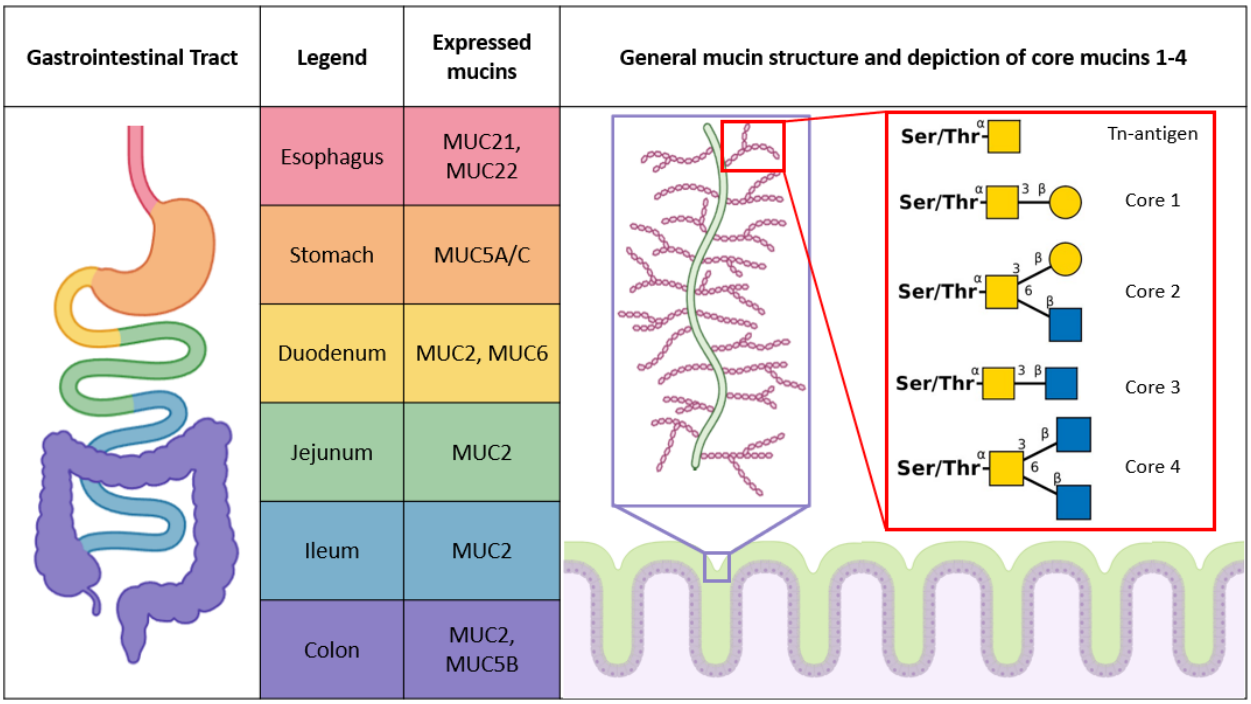


Figure 1.2: Mucin location, general structure, and glycosylation

Purple box (left) shows the general mucin bottlebrush structure from the mucosal layer secreted from the intestinal villi, while the red box (right) depicts core mucins structures. Apart from MUC21 and MUC22, all mucins mentioned are the predominant gel-forming mucins at each location (Hansson, 2020; Paone & Cani, 2020). Images acquired from BioRender. Glycan structures made using DrawGlycan-SNFG (Cheng et al., 2016). *Definitions*: yellow square: *N*-acetylgalactosamine (GalNAc), yellow circle: galactose (Gal), blue square: *N*-acetylglucosamine (GlcNAc).

Many commensal gut bacteria target mucins as a major nutrient source (Berkhout et al., 2024; Desai et al., 2016; Wardman et al., 2022). This includes bacterial genera *Bifidobacterium, Ruminococcus, Akkermansia,* and *Bacteroides*. Some notable species, such as *A. muciniphila*, are described as mucin specialists, wherein the microbe devotes most of their substrate-degrading genes towards mucin (Berkhout et al., 2024; Desai et al., 2016; Kostopoulos et al., 2021; Shon et al., 2020). In contrast, other mucin-degrading bacteria, such as *B. thetaiotaomicron* and *B. caccae*,

target a wider array of substrates in the human gut; these are considered generalists (Berkhout et al., 2024; Desai et al., 2016; Kostopoulos et al., 2021; Overbeeke et al., 2022). However, no single member of the HGM has been observed to fully degrade this complex glycoprotein by itself. A recent study hypothesized that mucin degradation occurs in an ecological network of gut microbes, wherein mucin degraders and cross-feeders collectively express a consortium of enzymes to completely catabolize mucin *O*-glycans (Berkhout et al., 2024). Using an *in vitro* synthetic gut microbiota with mucin as the main nutrient source, they discovered that this glycoprotein could sustain several mucin-degrading species and cross-feeders through the differential expression of CAZymes, peptidases, and other auxiliary proteins necessary to access the glycan backbone. Moreover, comparisons between concurrent mono- and co-culturing experiments sometimes showed different expression levels of the same enzyme. For instance, *B. thetaiotaomicron* expression of two peptidases, BT4244 and an M16 peptidase, were significantly upregulated in community compared to monoculture. This result suggests that *B. thetaiotaomicron* directly competes with other mucin-specialists, such as *A. muciniphila*, which is consistent with another study that focused on the close interactions between the two species (Kostopoulos et al., 2021).

1.1.2. Dietary glycan as a nutrient source

From birth to death, humans consume varied and dynamic diets composed of food groups with structurally distinctive glycan structures (Klassen et al., 2021) (**Figure 1.3**). Comprehensive reviews regarding the carbohydrates found in fungi (Abbott et al., 2015; Ali et al., 2024; Barreto-Bergter & Figueiredo, 2014; Routier et al., 2022), plants (O'Neill et al., 2022; Salimi et al., 2023; Yüksel et al., 2024), animal sources (S. Chia et al., 2022; Inokuma et al., 2023; Pan et al., 2021; J. Wang et al., 2022; C. Zhang et al., 2023), and seaweed (Abka-khajouei et al., 2022; C. Li et al., 2023; Pérez-Cruz et al., 2024) exist in the literature. Respectively, these include complex structures such as chitin and yeast mannan; pectin, arabinogalactan, and xyloglucan; glycosaminoglycans (GAGs) and human milk oligosaccharides (HMOs); as well as ulvan, fucoidan, and alginate. It should be noted that this is a mere fraction of the extraordinarily complex polysaccharides present in these nutrient sources. Degradation of these dietary carbohydrates not only modulate the HGM's microbial composition but impart abundant beneficial effects on the host through the production

of SCFAs (**Figure 1.1**) (Klassen et al., 2021a; Ross et al., 2024; Yang et al., 2024). As such, diet is an important factor in maintaining a healthy HGM.

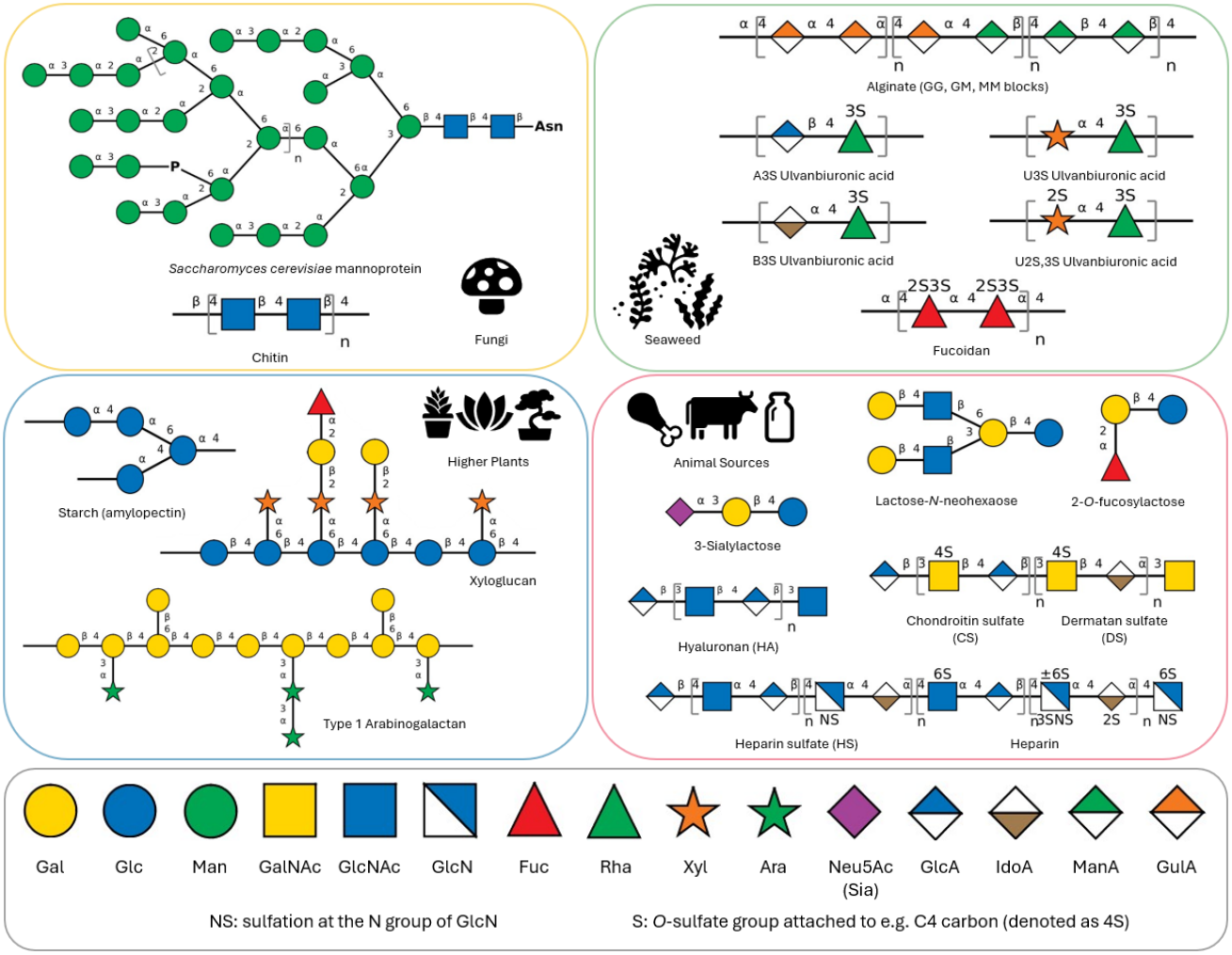


Figure 1.3: Representative carbohydrates in various human food sources.

Above figure based on Fig. 1 from (Klassen et al., 2021)**.** It should be noted that the structures shown are not an exhaustive list of the glycan diversity present in nature. Glycan structures made using DrawGlycan-SNFG (Cheng et al., 2016). *Abbreviations*: Gal: galactose, Glc: glucose, Man: mannose, GalNAc: *N*-acetylgalactosamine, GlcNAc: *N*-acetylglucosamine, GlcN: glucosamine, Fuc: fucose, Rha: rhamnose, Xyl: xylose, Ara: arabinose, Neu5Ac: *N*-acetylneuraminic acid, sialic acid (Sia), GlcA: glucuronic acid, IdoA: iduronic acid, ManA: mannuronic acid, GulA: guluronic acid.

Diet plays a major role in overall health state and preventing non-communicable diseases, including cancer, cardiovascular disease, as well as cognitive or metabolic impairments (**Figure 1.1**) (Golshany et al., 2024; Ross et al., 2024; Soldán et al., 2024). The diet of any individual is

heavily influenced by numerous factors, including culture, regional availability, socioeconomics, and consumer preference (Costlow et al., 2025; Ross et al., 2024). This has tremendous effects on HGM composition. Consistent dietary choices can also exert selective pressures to allow for unique trait acquisition by the HGM. For example, the ability of the gut microbiota to digest carrageenan came from the lateral transfer of an integrative conjugative element (ICE) from a marine species in the phylum Bacteroidota (Pudlo et al., 2022). This was selected for primarily because European and Chinese populations historically ate their local red algae, as far back as 400 BC, and carrageenan is now a widely used food additive. Overall, numerous studies on short- and long-term diets show altered HGMs with differing composition and functionality that greatly influences the host's health state at all stages of life (Al-Biltagi et al., 2022; Bradley & Haran, 2024; Desai et al., 2016; Ghosh et al., 2020; Gong et al., 2023; Ross et al., 2024; Saeed et al., 2022; Soldán et al., 2024). Therefore, it is no surprise that modern studies on the interplay between diet, HGM composition, and disease is a major area of research that has been ongoing for decades (**Figure 1.4**) (Hou et al., 2022).

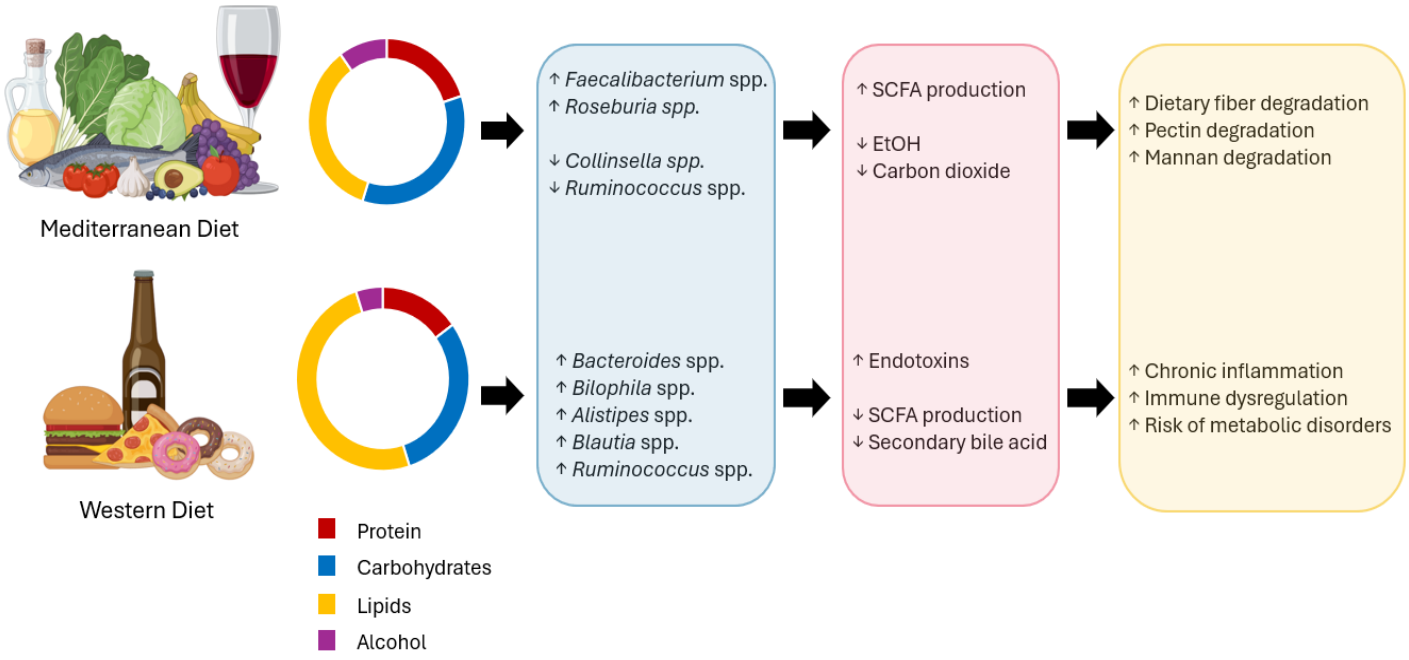


Figure 1.4: Macronutrient composition of whole diets, their effect on human gut microbiota composition, and subsequent host consequences.

(Agus et al., 2016; Meslier et al., 2020; Ross et al., 2024)

Comparisons between the Mediterranean and Western diet clearly display the impact of dietary choices on the HGM (**Figure 1.4**) (Agus et al., 2016; Ghosh et al., 2020; Ross et al., 2024). The Mediterranean diet is the gold standard of health maintenance and preventative medicine, which encourages the high consumption of olive oil, dairy products, and unprocessed whole plant food; moderate consumption of poultry and fish; and minimal amounts of red meat (Ross et al., 2024). A geriatric intervention study with this diet showed changes in the gut microbiota associated with improved cognitive function, reduced inflammatory status, and overall lower risk of frailty (Ghosh et al., 2020). This was due to increases in overall microbial diversity in the HGM and the abundances of *Faecalibacterium* spp. and *Roseburia* spp., which are bacterial genera associated with elevated production of SCFAs and other anti-inflammatory molecules, as well as increased microbial function for dietary fiber degradation (Ghosh et al., 2020; Martín et al., 2023; Nie et al., 2021; Ross et al., 2024). In contrast, the Western diet is known for its high calorie content, with an emphasis on saturated fats, simple sugars, animal protein, and ultra-processed foods, and deficient amounts of fiber, fruits, and vegetables (Agus et al., 2016; Ross et al., 2024). One murine model study demonstrated that this high fat and high sugar diet directly promotes an inflammatory environment in the gut, enriches growth of pro-inflammatory Proteobacteria, such as *Escherichia coli*, lowers microbial diversity of the HGM, and significantly decreases SCFA concentrations, leading to increased susceptibility to chemically induced colitis (Agus et al., 2016). Similarly, another murine study using a synthetic gut microbiota corroborated these findings and showed that these detrimental effects stemmed from a lack of dietary fiber, resulting in a microbially eroded mucus barrier due to the over-proliferation of mucus-degrading bacteria (Desai et al., 2016).

1.1.3. Glycosaminoglycans (GAG) as a nutrient source

Glycosaminoglycans (GAGs) are ubiquitously distributed, negatively charged, linear heteropolysaccharides found in the extracellular matrix (ECM) and on the surfaces of mammalian cells (Merry et al., 2022; Rawat et al., 2022). Membrane-bound GAGs found on the surfaces of intestinal epithelial cells contribute to the glycocalyx and regulate cell signalling, development, and homeostasis (Luis et al., 2023). Typically, GAGs in the ECM are found covalently attached to a core protein, wherein the resulting glycoconjugate are called proteoglycans (Merry et al., 2022). Hyaluronan is an exception to this, as this GAG interacts with specific proteoglycans via

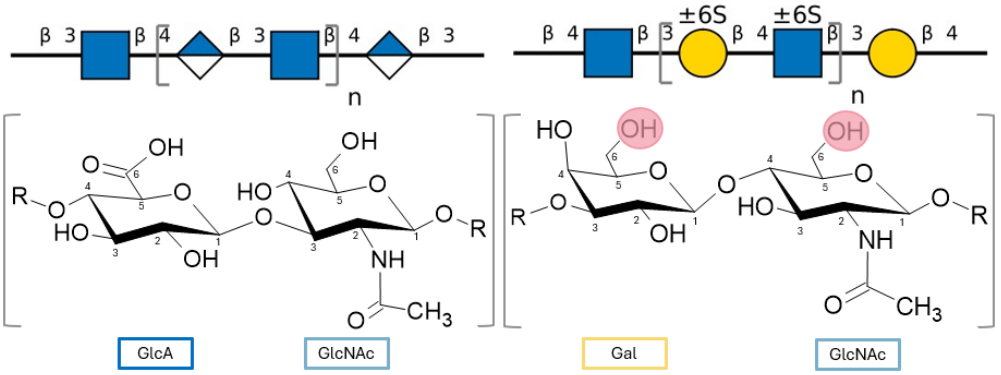
hyaluronan-binding motifs (Koch et al., 2020; Merry et al., 2022; Pan et al., 2021). Vertebrate proteoglycans are separated into five categories: intracellular granule, membrane-bound, secreted, non-secreted, and ECM proteoglycans (Merry et al., 2022). Although these have roles in tissue morphogenesis and signalling, proteoglycans primarily act as structural "shock absorbers" in the ECM (Farach-Carson et al., 2024; A. L. Gray et al., 2022). This is due to the highly electronegative nature of the attached GAGs. This negative charge attracts and retains large amounts of water, conferring tissues, such as articular cartilage, with elastic compressive properties.

There are four main GAG categories: hyaluronan (HA); heparin and heparan sulfate (HS); keratan sulfate (KS); and lastly, chondroitin sulfate (CS) and dermatan sulfate (DS) (Köwitsch et al., 2018). Each group has a characteristic repeating disaccharide unit of a uronic acid (or galactose) paired with an amino sugar. GAG distinctions are based on the monosaccharide composition, sulfation level, and type of glycosidic linkage present (**Figure 1.5**). For example, chondroitin (CH) has a repeating disaccharide of glucuronic acid (GlcA) that is β-1,3-linked to *N*-acetylgalactosamine (GalNAc); the repeating disaccharide is β-1,4-linked. Typically, chondroitin is sulfated to form CS. CS-A and CS-C are sulfated on the GalNAc residue at carbons 4 and 6, respectively. CS-D and CS-E comprise CS-C that is additionally sulfated at carbons 2 and 4, respectively of the GlcA (Mikami & Kitagawa, 2013).

The fermentation of GAGs by the HGM, especially by *Bacteroides* spp., is a well-studied topic, especially since its degradation is not widely prevalent in other gut phyla (Cartmell et al., 2017; D. Ndeh et al., 2018, 2020; Overbeeke et al., 2022; Rawat et al., 2022; Ulmer et al., 2014; L. Wei, Zou, et al., 2024). This process generates biologically active metabolites, such as SCFAs (Inokuma et al., 2023; Pan et al., 2021; C. Zhang et al., 2023), which beneficially affect our health state, as well as improve gastrointestinal (Abdul Rahim et al., 2019), immunological (Mann et al., 2024), and neuronal (Silva et al., 2020) disease states (**Figure 1.1** and **Figure 1.3**). For instance, HA was readily degraded in an *in vitro* gut microbiota system, with fecal samples taken from Chinese individuals (Pan et al., 2021). Subsequent addition of HA enhanced *Bacteroides*, *Bifobacterium*, *Dialister,* and *Faecalibacterium* spp. growth and generated a significant amount of acetate, propionate, and butyrate, although fermentation efficacy differed on an individual basis (Pan et al., 2021). Another *in vitro* study using the fecal samples of ten Japanese individuals

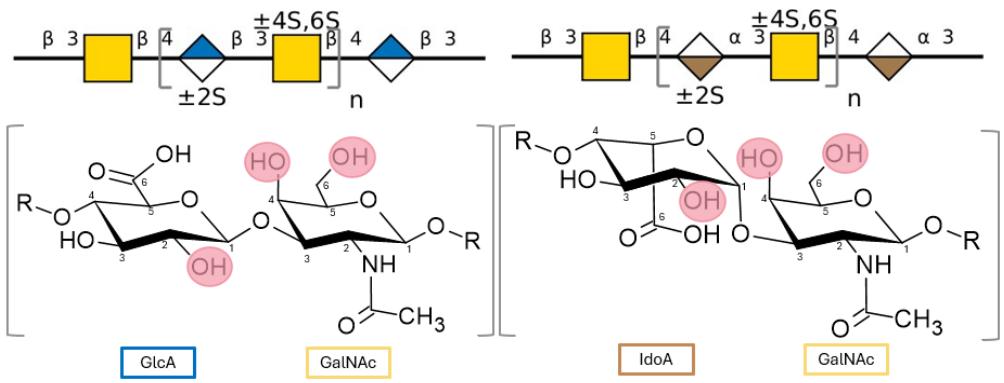
evaluated the effect of different types of CS on HGM composition and metabolite production (Inokuma et al., 2023). Similarly, the addition of CH and CS-E (*i.e.,* CS with sulfates on GalNAc's C4 and C6) (**Figure 1.5**), drastically changed the abundance and proportion of multiple *Bacteroides* spp., as well as the amount of acetate and propionate in the gut.

As a nutrient source for the HGM, GAGs can be acquired from dietary animal meat and cartilage, as well as found on the host mucosal layer and epithelial cells (**Figure 1.3**) (Klassen et al., 2021a; Overbeeke et al., 2022; Rawat et al., 2022). However, excessive foraging of the latter has been shown to induce inflammatory and opportunistic infections in the gut, further emphasizing that a healthy gut microbiota requires an assortment of diverse dietary glycans to modulate its microbial composition (Desai et al., 2016). As most dietary GAGs would be found as proteoglycan glycoconjugates (**Figure 1.3**), one potential dietary proteoglycan would be aggrecan (Kiani et al., 2002; Koch et al., 2020). Aggrecan is the major proteoglycan found in articular cartilage, and ~90% of its mass is mainly comprised of CS chains, although there is a region of KS polymers, as well as *N*- and *O*-linked oligosaccharides along the aggrecan core protein backbone (Kiani et al., 2002). Connective tissues, like cartilage and tendon, are widely used as a culinary ingredient by many cultures around the world, including Filipino, Japanese, Italian, Korean, and Chinese cuisine. Despite the absence of publications that directly address the molecular factors that affect dietary proteoglycan degradation by the HGM, it is likely that there are gut microbes that specialize in the catalysis of both glycan and peptide components of these glycoproteins.



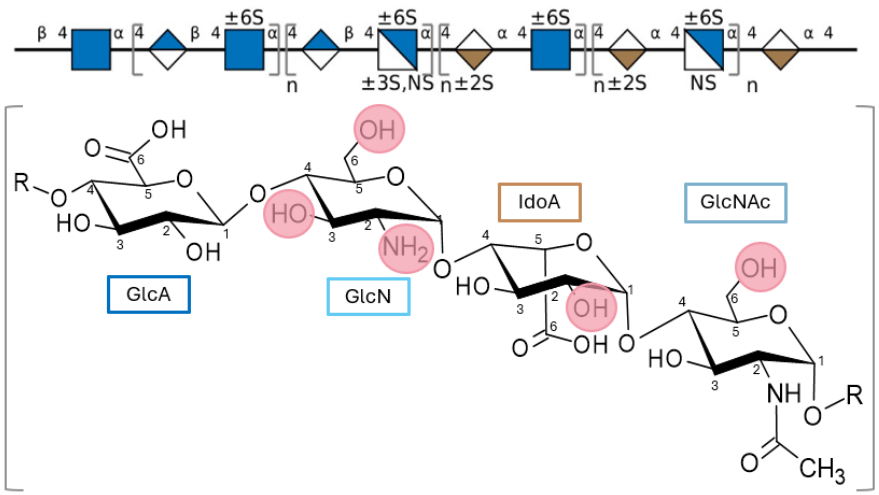
Hyaluronan (HA)

Keratan sulfate (KS)	
A unit	Gal-GlcNAc
B unit	Gal-GlcNAc(6S)
C unit	Gal(6S)-GlcNAc(6S)



Chondroitin sulfate (CS)	
O unit (chondroitin)	GlcA-GalNAc
A unit	GlcA-GalNAc (4S)
C unit	GlcA-GalNAc(6S)
D unit	GlcA(2S)-GalNAc(6S)
E unit	GlcA-GalNAc(4S,6S)

Dermatan sulfate (DS)	
iA unit	IdoA-GalNAc (4S)
iB unit	IdoA(2S)-GalNAc(4S)
iE unit	IdoA-GalNAc(4S,6S)



Heparan sulfate (HS)/Heparin

Figure 1.5: Chemical and symbolic depiction of disaccharide units of glycosaminoglycans.

Red shaded circles indicate potential sulfation positions. Glycan symbols above chair conformations made using DrawGlycan-SNFG according to nomenclature standards (Cheng et al., 2016). For clarification, heparan sulfate is differentially C5 epimerized and sulfated into heparin based on tissue type (Kjellén & Lindahl, 2024). Thus, this GAG does not have standardized disaccharide units.

Abbreviations:
GalNAc: *N*-acetylgalactosamine
GlcNAc: *N*-acetylglucosamine
GlcN: glucosamine
GlcA: glucuronic acid
IdoA: iduronic acid

1.2. Carbohydrate active enzymes (CAZymes)

Over thirty years have passed since the introduction of amino acid sequence-based classification of carbohydrate active enzymes (CAZymes) (Drula et al., 2022). Since then, these diverse, numerically organized, CAZyme families are maintained and curated in the carbohydrate-active enzyme (CAZy) database. Experimentally characterized members per family vary from one to several hundred, depending on research and/or biotechnological interest. Sequence-based classification also allows for the assignment of putative CAZymes using algorithmic methods. Biochemically untested proteins are annotated with the known functional information of the family they are associated with, but it is rare for any CAZyme family to have a single activity. There are often numerous substrate specificities within these sequence-based groupings, which can be shared between families. Recently, an additional search system called CAZyme activity descriptor (CAZac) was added to CAZy, which allows users to harness the functional information available on this database (Lombard et al., 2025)**.** Despite substrate variability, each family's respective catalytic mechanism is extremely conserved. Moreover, as CAZymes in the same family have similar amino acid sequences, members of the same family have the same protein fold. Thus, once an enzyme's mechanism of action is established, it is generally assumed to be shared within its family.

Besides sequence-based classification, there are four other ways to further categorize CAZymes: by mechanism, enzyme commission (EC) number, if the CAZyme is *endo*- or *exo*-acting, and if it has processive or non-processive properties (Drula et al., 2022). Respectively, *endo*- and *exo*-acting refers to the CAZyme's ability to act in the middle or on the ends of glycan chains. Processive versus non-processive describes an enzyme's ability to remain attached to their polymeric substrates and perform multiple rounds of catalysis before dissociating (Breyer & Matthews, 2001). Additionally, a polysaccharide's directionality is depicted by having the non-reducing end on the left and reducing end on the right, with the enzymatic sub-sites of CAZymes numbered relative to their cleavage point (G. J. Davies et al., 1997). Sub-sites closer to the non-reducing terminus has increasingly negative numbers (e.g., -1, -2, -3, etc.), while sites closer to the reducing terminus have increasingly positive numbers (e.g., +1, +2, +3, etc.) (**Figure 1.6**).

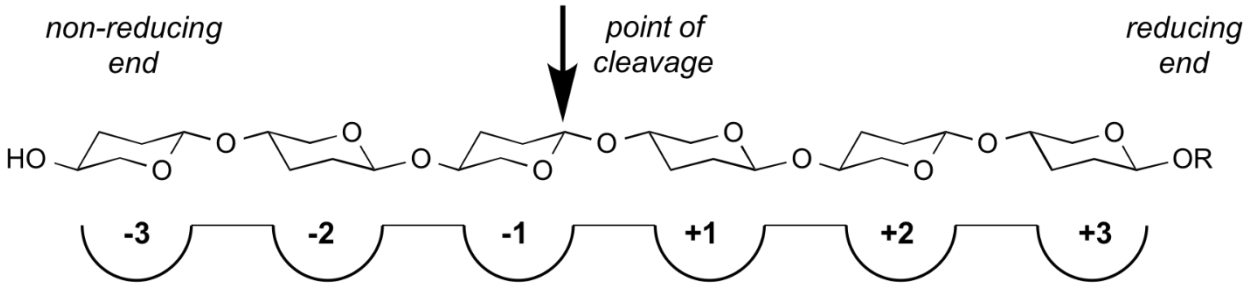


Figure 1.6: Depiction of CAZyme sub-site nomenclature and general directionality of polysaccharides.

In this example, the CAZyme would be described as *endo*-acting, as the point of cleavage is in the middle of a chain. Direct permission to use above figure (originally made by Dr. Spencer Williams) from CAZypedia was granted by primary curator, Dr. Harry Brumer (The CAZypedia Consortium, 2018). This figure is accessible at the following URL: https://www.cazypedia.org/index.php/Sub-site_nomenclature

Despite the importance of carbohydrates as a source of energy, humans have a startlingly minimal CAZyme repertoire in their genome (Bhattacharya et al., 2015; Klassen et al., 2021). There are only 17 human CAZymes that specifically digest lactose, starch, and sucrose; all relatively simple structures, especially when compared to the heavily modified host mucin glycans or dietary glycans from seaweed, amongst other nutrient sources (**Figure 1.3**) (Klassen et al., 2021). In contrast, members of the HGM dedicate remarkable proportions of their genomes to CAZymes (Wardman et al., 2022). For instance, the gut microbe *Bacteroides thetaiotaomicron* can encode hundreds of CAZymes (amounting to ~18% of their total genome) and has been extensively characterized to degrade numerous complex carbohydrates (Bhattacharya et al., 2015; Desai et al., 2016; Martens et al., 2008; D. Ndeh et al., 2020; Wardman et al., 2022; Ye et al., 2020). Hence, it is not surprising that the HGM is responsible for most carbohydrate degradation within the host. It is important to note that the HGM's microbial composition will determine its CAZyme profile. Different CAZyme profiles will result in differing metabolic capacities for glycan utilization, which will affect the production of bioactive metabolites, including SCFAs, that will influence the host's overall health state (**Figure 1.1**) (Inokuma et al., 2023; Pan et al., 2021; C. Zhang et al., 2023). As such, understanding the molecular factors that govern microbial CAZyme activity and substrate specificity serves as the foundation of research that focuses on the dietary and host glycan degradation effects on the host.

There are six main classes of CAZymes (Drula et al., 2022; The CAZypedia Consortium, 2018). Glycoside hydrolases (GH) catalyse hydrolysis of glycosidic linkages (G. Davies & Henrissat, 1995). Glycosyltransferases (GT) form glycosidic linkages. Polysaccharide lyases (PL) cleave glycosidic bonds using a β-elimination mechanism (Garron & Cygler, 2014). Carbohydrate esterases (CE) hydrolyse carbohydrate esters. Carbohydrate-binding modules (CBM) bind to specific sugars and are generally found within larger enzymes. And, lastly, auxiliary activities (AA), which refer to redox enzymes that act alongside other CAZymes in carbohydrate degradation.

1.2.1. Glycoside hydrolases (GHs)

Glycoside hydrolases (GHs) are found in all domains of life and is the largest and most-studied CAZyme class (Berlemont & Martiny, 2016; Hansen et al., 2020; Minic, 2008; Wardman et al., 2022). Currently, there are over one hundred GH families listed in the CAZy database (Drula et al., 2022). Of these, 140 GH families are curator-approved on CAZypedia, an accompanying open-access encyclopedic resource on CAZymes (The CAZypedia Consortium, 2018). Given the abundance of GHs in literature, groups of GH families have been further classified into clans, which describe GH families that share significant similarity in their catalytic residues, mechanism, as well as tertiary fold (G. J. Davies & Sinnot, 2008; The CAZypedia Consortium, 2018). Structurally, most GHs have a $(\alpha/\alpha)_n$ or a $(\beta/\alpha)_n$ fold, but other topologies, such as β-jelly roll or a Rossmann fold, are possible. Often, GHs also have carbohydrate-binding modules (CBMs) associated with their catalytic domain to enhance activity (Wardman et al., 2022).

In general, GHs must bind to specific monosaccharide units and are sensitive to the α- or β-glycosidic linkage present prior to cleavage. However, it is not unusual for a GH to prefer substrate targets containing additional functional groups, such as sulfates, prior to hydrolysis (Nakamichi et al., 2011).

GH families may be categorized mechanistically as retaining or inverting, which describes the end state of the released glycan's anomeric center (G. Davies & Henrissat, 1995). Both use a general acid and general base residue to catalyse hydrolysis and release a saturated hexose ring on the non-reducing end. The main difference between the two mechanisms is that a typical retaining

mechanism uses a two-step system, wherein a single residue acts as an acid/base, while another residue functions as a nucleophile to make the glycosyl-enzyme intermediate (**Figure 1.7**). Many variations on the retaining/inverting mechanisms exist (Nakamichi et al., 2014; The CAZypedia Consortium, 2018). The GH4, GH109, GH177, and GH179 families are a notable exception to these observed mechanisms, as they use a fundamentally different mechanism that requires a NAD^+ cofactor for catalysis (Teze et al., 2020; The CAZypedia Consortium, 2018).

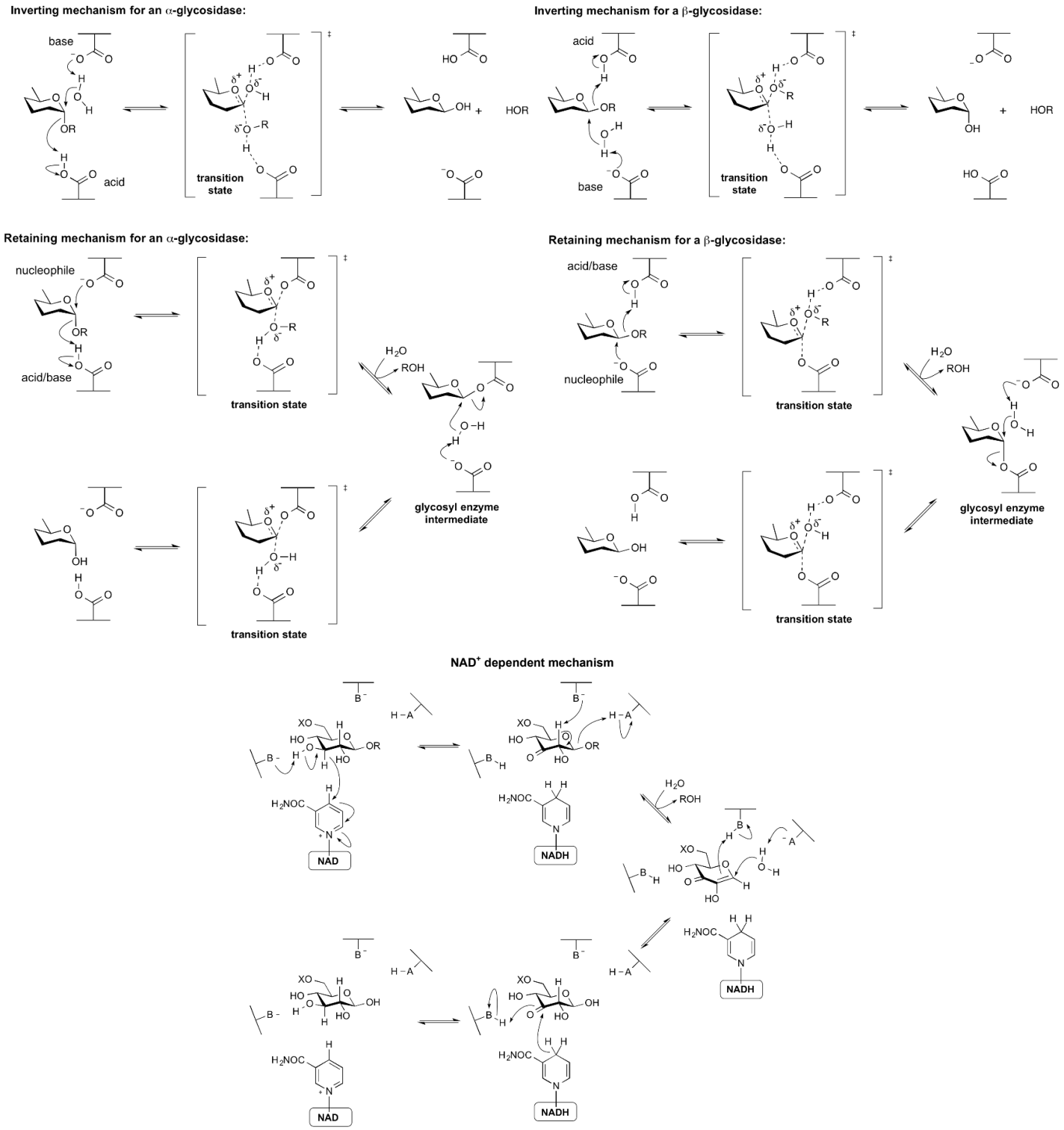


Figure 1.7: General depiction of retaining, inverting, and NAD$^+$ dependent mechanisms by glycoside hydrolases.

Direct permission to use above figures (originally made by Dr. Spencer Williams) from CAZypedia was granted by primary curator, Dr. Harry Brumer (The CAZypedia Consortium, 2018). These figures are accessible at the following URL: https://www.cazypedia.org/index.php/Glycoside_hydrolases.

1.2.2. Polysaccharide lyases (PLs)

Polysaccharide lyases (PLs) are CAZymes primarily expressed by bacteria and fungi, although there are some archaea and higher eukaryotes, such as plants and nudibranchs, noted to contain PLs in their genome (Drula et al., 2022; Qin et al., 2017). Presently, there are ~40 PL families listed in the CAZy database; significantly fewer than GH families (Drula et al., 2022). This could be indicative of two factors unique to PLs: a smaller substrate range and mechanism of action. These CAZymes are distinct for their specificity toward polysaccharides containing uronic acids, including pectin, the major structural component of plant primary cell wall (J. Li et al., 2024); GAGs, which are important signalling molecules found in the ECM and on the cellular surfaces of mammalian cells (D. Ndeh et al., 2020; L. Wei, Zou, et al., 2024); as well as alginate (Abka-khajouei et al., 2022) and ulvan (C. Li et al., 2023), which are complex structural carbohydrates found in brown and green algae, respectively (**Figure 1.3**).

As with GHs, PL families show a large variety of tertiary structures, implying that these lyases have emerged evolutionarily multiple times (Drula et al., 2022). However, there does seem to be a degree of structural parallel between families that degrade chemically similar substrates. For instance, the majority of structurally characterized heparin/HS lyases share a $(\alpha/\alpha)_n$ toroid and anti-parallel β-sheet fold. This is also seen among the CS lyases and HA lyases, although structures arranged in parallel β-helixes and triple strand β-helixes are also present (Drula et al., 2022; Loiodice et al., 2025; L. Wei, Cao, et al., 2024).

Generally, PLs use β-elimination mechanisms that generate a 4,5-unsaturated sugar at the new non-reducing end of the product (Garron & Cygler, 2010, 2014). This mechanism has been proposed to occur in three steps (**Figure 1.8**). Initially, the substrate's carboxyl group is neutralized, either by a metal cofactor or positively charged residue. This reduces the *p*Ka of a C5 proton, allowing for its abstraction and generating an enolate anion intermediate. Lastly, electron transfer occurs from the carboxyl group, forming an unsaturated bond between C4 and C5. Overall, PL mechanisms can be metal-assisted or rely on tyrosine/histidine (Tyr/His) residues for activity. Pectin-specific lyases predominantly use the former with Ca^{+2} as its cofactor, but PL2 is the exception, as it preferentially uses Mn^{+2} ions (Abbott et al., 2010; Garron & Cygler, 2014).

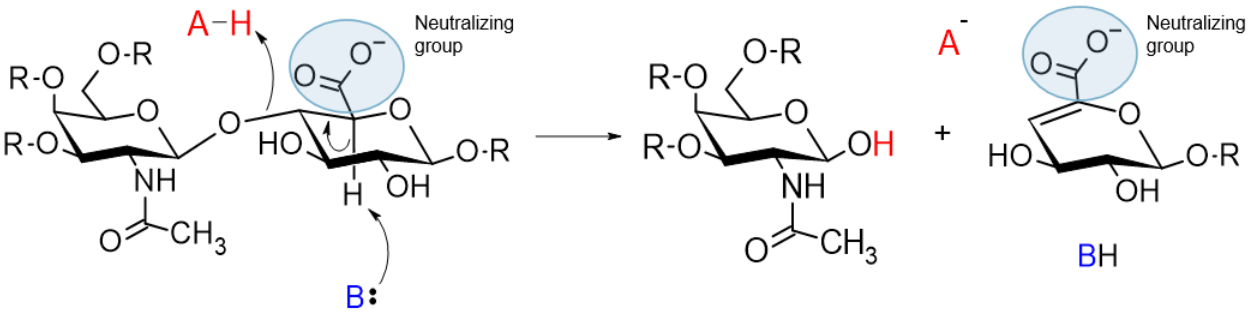


Figure 1.8: General β-elimination mechanism by polysaccharide lyases.

Shaded blue circle represents the neutralizing effect of a metal cofactor or a positively charged residue that would allow for the C5 proton to be removed, forming the unsaturated bond between C4 and C5 (Garron & Cygler, 2010).

1.3. Polysaccharide utilization loci (PUL)

Polysaccharide utilization loci (PULs) describe gene clusters that are co-localized and co-regulated in response to a single glycan, or chemically related group of glycans (Klassen et al., 2021) (**Figure 1.9**). The archetypal PUL is the starch-utilization system (Sus) from *B. thetaiotaomicron*, wherein eight genes were identified as part of a single cluster responsible for starch degradation (Foley et al., 2016; Grondin et al., 2017; D. Ndeh et al., 2020). Typically, PULs contain a complement of cell-surface glycan binding proteins, CAZymes, carbohydrate sensors and transcriptional regulators, as well as at least one sequential pair of genes encoding SusC and SusD homologs. These homologs are a hallmark of PULs (Bolam & Koropatkin, 2012; Brown & Koropatkin, 2020). They are found as a SusCD complex and specifically encode for an outer membrane TonB-dependent transporter (SusC) and an associated lipid-bound glycan binding protein (SusD).

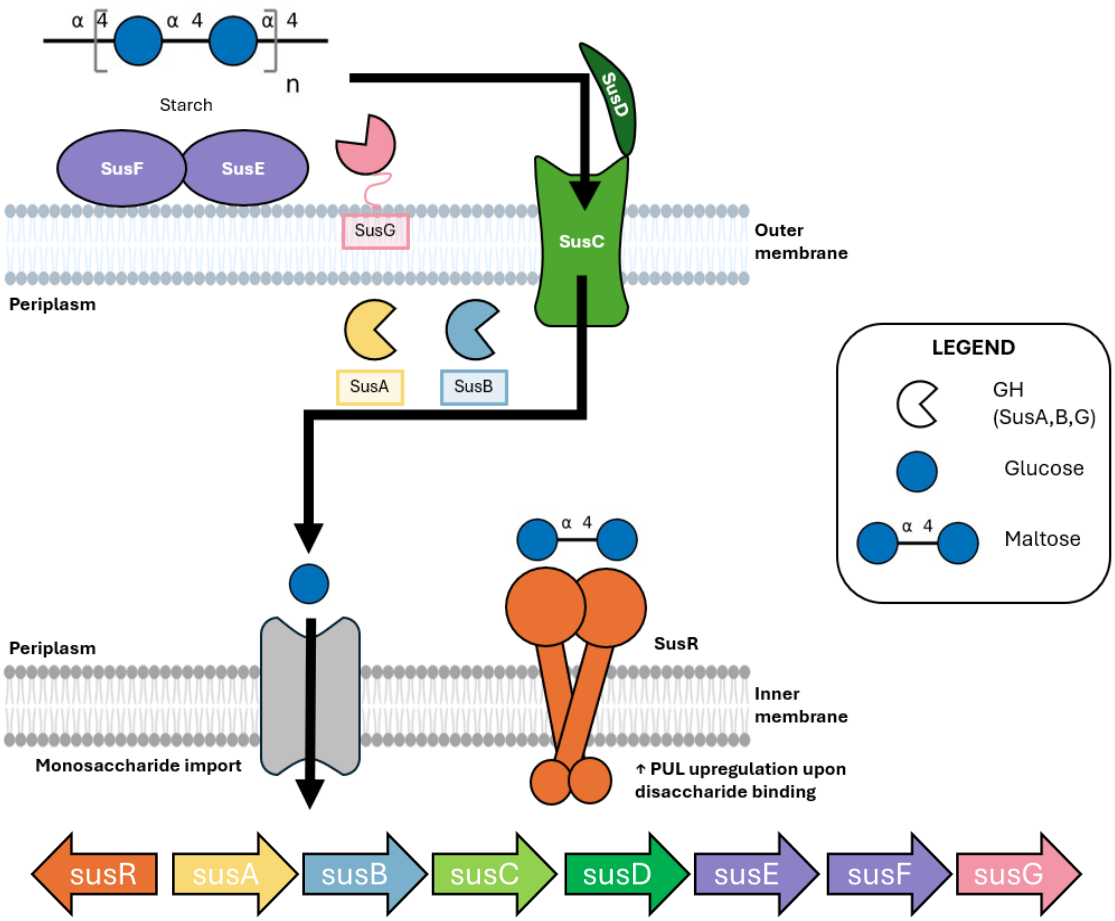


Figure 1.9: Depiction of the archetypal starch PUL in action.

Gene organization of this PUL follows the color coding shown in the diagram (Foley et al., 2016). Glycan symbols made using DrawGlycan-SNFG according to nomenclature standards (Cheng et al., 2016). Starch oligosaccharides generated by the outer membrane SusG interact with multidomain starch binding proteins, SusF and SusE. These oligosaccharides are transported into the periplasm via the SusCD complex. Disaccharides that bind to SusR result in the upregulation of the PUL. Monosaccharides are transported into the cellular cytosol by a PUL-independent monosaccharide importer. *Abbreviations*: GH: glycoside hydrolase.

Our understanding of *Bacteroides* spp. carbohydrate metabolism mainly stems from studies of how dietary plant polysaccharides, such as starch, are utilized using PULs (Grondin et al., 2017; D. Ndeh et al., 2020). This general system of carbohydrate utilization has been observed for other polysaccharides besides starch, leading to a general PUL system that is often described as "Sus-like". The emergence of multiple PULs that enable the use of different glycans reflects human dietary habits, as diets are often highly varied and contain structurally distinct carbohydrates from different food groups (**Figure 1.3**) (Klassen et al., 2021; Tuncil et al., 2017). Characterized and

putative PULs are stored and maintained in the CAZy PUL database (PULDB), wherein PULs are identified based on the identification of *susCD* genes in genome assemblies (Terrapon et al., 2018).

Overall, PULs are a major nutrient acquisition strategy for the phylum Bacteroidota, especially for *Bacteroides* spp. (Feng et al., 2022; D. Ndeh et al., 2020; Overbeeke et al., 2022; Terrapon et al., 2018). Given the high prevalence of this phylum in the HGM, the PUL system is intrinsically linked to the colonization of nutritional niches and subsequent establishment of microbial ecosystems in the gut (Grondin et al., 2017). Curiously, despite their numerous PUL complements, there are remarkably few homologous PULs shared between *Bacteroides* spp. Although different species can target the same glycans, the organization of similar PULs can differ in the encoded CAZymes present and overall regulation, allowing for each *Bacteroides* member to occupy a distinct glycan niche despite containing the capacity to degrade the same substrates.

1.3.1. Association of auxiliary proteins to PULs

Given the diversity of glycan structures found in nature, it is not uncommon for PULs to have other supporting enzymes associated with them, such as epimerases, proteases, and sulfatases (Terrapon et al., 2018). The lattermost enzyme class are relatively common additions to PULs, especially when a target polysaccharide is heavily sulfated (Klassen et al., 2021; D. Ndeh et al., 2020; Rawat et al., 2022; Terrapon et al., 2018). Sulfatases catalyse the hydrolysis of sulfate groups and are essential in allowing CAZymes to interact with the glycan backbone. They are widely distributed among prokaryotes and eukaryotes (Luis et al., 2023; Stam et al., 2023). Currently, there are four sulfatase families, designated S1-S4, that differ in amino acid sequence relatedness as well as catalytic mechanism. Putative and characterized sulfatases are added to the SulfAtlas database, which is dedicated to the sequence-based classification of sulfatases (Stam et al., 2023). The largest sulfatase family, S1, is ubiquitously distributed amongst all domains of life, and is the only family with characterized carbohydrate-specific members (Luis et al., 2023). It belongs to the alkaline phosphatase superfamily, which generally possess an α-/-β-/-α fold, and relies on a non-genome encoded formylglycine as the catalytic nucleophile, while a conserved lysine or histidine acts as a catalytic acid that protonates the leaving group sugar (Luis et al., 2023). This residue is generated using a sulfatase maturing enzyme that targets a conserved serine or cystine in the C/S-X-P/A-S/X-R motif (Luis et al., 2023; Ulmer et al., 2014).

Sulfatases may be found associated with PULs that target sulfated glycan structures, including mucins (Luis et al., 2021), GAGs (Ulmer et al., 2014), and seaweed glycans, such as fucoidan (Pérez-Cruz et al., 2024). Despite the abundance of sulfated host and dietary glycan substrates (**Figure 1.3**) (Cartmell et al., 2017; Klassen et al., 2021; Pérez-Cruz et al., 2024), as well as vast number of carbohydrate-degrading microbes known in literature (Cartmell et al., 2017; Overbeeke et al., 2022; Shin et al., 2024), only carbohydrate sulfatases found with *B. thetaiotaomicron* (Luis et al., 2021; D. Ndeh et al., 2020; Ulmer et al., 2014) and *Pedobacter heparinus* (Luis et al., 2023) have been characterized, with most work carried out with the former microbe. As desulfation is often the first step in degrading sulfated substrates, and there are several prominent members of the HGM that target sulfated glycans (Berkhout et al., 2024; Inokuma et al., 2023; Loiodice et al., 2025; Martín et al., 2023; Overbeeke et al., 2022; Pan et al., 2021; Shin et al., 2024; L. Wei, Zou, et al., 2024; C. Zhang et al., 2023), it is important to characterize the sulfatases used by these microbes in conjunction with their CAZymes.

1.4. M60-like peptidases

Through their ability to catalyze the hydrolysis of peptide bonds, peptidases are key enzymes in numerous biological processes across all taxonomic kingdoms (Noach et al., 2017; Rawlings et al., 2016; Rawlings & Bateman, 2021). Peptidases are organised into hierarchical classifications within the MEROPS database, wherein homologous sequences are clustered into families, and collections of families into clans (Rawlings et al., 2016, 2018). As there are billions of characterized and putative peptidases in this database, there is a distinct identifier of each at every level. For instance, at the family level, peptidases have a letter which describes their catalytic mechanism, or catalytic type, followed by a number. There are seven catalytic types: serine (S), cysteine (C), threonine (T), aspartic (A), glutamic (G), asparagine lyases (N), and metallo (M)**.** In the case of metallopeptidases, these enzymes specifically rely on a metal cofactor, such as zinc, to recruit a water molecule for hydrolysis (**Figure 1.10)** (Cerdà-Costa & Gomis-Rüth, 2014).

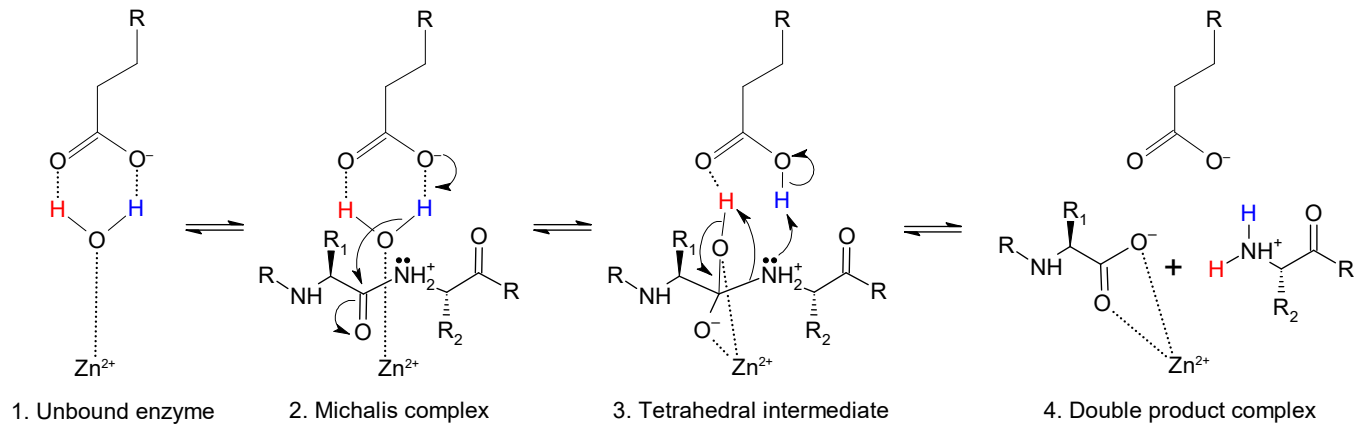


Figure 1.10: General zinc metallopeptidase mechanism.
(Cerdà-Costa & Gomis-Rüth, 2014).

Members of the M60-like superfamily are zinc metallopeptidases (clan MA, subclan MA(E)) that require binding to specific *O*-glycan moieties prior to peptide cleavage N-terminal to the glycan bearing residue (Noach et al., 2017). This superfamily is also annotated as Pfam family PF13402, enhancin, enhancin-like, and peptidase M60. There are three defining features in the M60-like superfamily: relatedness to viral enhancin zinc metallopeptidases, the presence of the gluzincin motif (HEXXHX(8,24)E), and multi-modularity with carbohydrate-binding modules (CBMs) (Nakjang et al., 2012; Pluvinage et al., 2021). The peptidase's eponymous catalytic domain contains the gluzincin binding motif necessary for catalysis and is conserved amongst host-associated bacterial and eukaryotic microbes (Nakjang et al., 2012). Additionally, members of this superfamily are called *O*-glycopeptidases, as these target *O*-glycosylated proteins, such as mucin (Noach et al., 2017; Pluvinage et al., 2021; Shon et al., 2020).

Many characterized members of the M60-like superfamily reside in taxa that commensally or pathogenically interact with animal hosts through degrading their mucosal layer (Nakjang et al., 2012; Pluvinage et al., 2021; Shon et al., 2020, 2022). These include glycopeptidases such as ZmpA, ZmpB, and ZmpC from *Clostridium perfingens* (Pluvinage et al., 2021); BT4244 from *B. thetaiotaomicron* (Noach et al., 2017; Taleb et al., 2022); and AMUC_0627, AMUC_0908, and AMUC_1514 from *A. muciniphila* (Medley et al., 2022; Shon et al., 2020; Taleb et al., 2022). However, not all sequenced strains of a given species contain these genes. Several bacterial pathogens, including the human pathogen *Vibrio cholera,* contain a M60-like peptidase gene

annotated as lipoprotein AcfD, whereas non-virulent strains of *V. cholera* do not (Nakjang et al., 2012). This implies that these glycopeptidase genes are part of a disposable pangenome that contributes to a microbe's specific niche adaptations. For instance, BT4244 from *B. thetaiotaomicron* is a well-characterized M60-like peptidase that contains a *Bacteroides* associated carbohydrate binding module often N-terminal (BACON) and a CBM32 domain (Berkhout et al., 2024; Noach et al., 2017; Pluvinage et al., 2021). This enzyme enables *B. thetaiotaomicron* to compete with *A. muciniphila* for mucin as a nutrient source, as it is upregulated in the presence of the substrate, as well as during co-culture in a microbial community (Berkhout et al., 2024; Desai et al., 2016; Kostopoulos et al., 2021). Seemingly in response, *A. muciniphila* upregulates its defense-related genes (Kostopoulos et al., 2021) along with multiple M60-like peptidases encoded in its genome (Desai et al., 2016) when co-cultured with *B. thetaiotaomicron* in the presence of mucin.

Clearly, the presence/absence of these glycopeptidases in a genome can provide insights to an organism's ecological role. The glycopeptidase redundancy present in *A. muciniphila* supports its role as a mucin-degrading specialist (Nakjang et al., 2012; Shon et al., 2020, 2022). Therefore, the fact that *B. caccae* (strain ATCC 48135) has no less than sixteen putative M60-like peptidases in its genome is particularly noteworthy (Nakjang et al., 2012; Rawlings & Bateman, 2021). Additionally, fifteen of these glycopeptidases are in PULs, implying enzymatic synergy between the encoded CAZymes and glycopeptidases against target glycoproteins. This has not yet been observed in any singular PUL.

1.5. *Bacteroides* spp. in the HGM

Within the HGM, the bacterial phylum Bacteroidota (formerly Bacteroidetes) is well-established to contain glycan generalists that can degrade a complex array of substrates (Briggs et al., 2021; Brown & Koropatkin, 2020; Klassen et al., 2021a; Overbeeke et al., 2022; Rawat et al., 2022; Shin et al., 2024). Members arrange their carbohydrate-metabolism-related genes in PULs, wherein genes are co-localized and co-regulated in response to a single glycan, or chemically related group of glycans (Brown & Koropatkin, 2020; D. Ndeh et al., 2020). Given the high prevalence of this phylum in the HGM, the PUL system is intrinsically linked to the colonization

of nutritional niches and establishment of microbial ecosystems in the gut (Feng et al., 2022; Grondin et al., 2017; Shin et al., 2024).

Bacteroides spp. exhibit preferences for different glycans, even when presented as a mixture in coculture growth experiments (Tuncil et al., 2017). In fact, some prevalent *Bacteroides* spp. prioritize GAG degradation over other glycans (Bolam & Koropatkin, 2012; Klassen et al., 2021; Tuncil et al., 2017). Through catabolism and the sharing of these recalcitrant glycan sources, *Bacteroides* spp. enhance the growth of other SCFA-producing gut microbes, leading to increased levels of SCFAs that regulate colonic health and act as substrates for key metabolic processes (Overbeeke et al., 2022; Pan et al., 2021; Rawat et al., 2022; Rios-Covian et al., 2017; C. Zhang et al., 2023). Thus, to appreciate how *Bacteroides* spp. interact with the host and other members of the HGM it is important to understand the molecular basis of how *Bacteroides* spp. interacts with GAGs as a nutrient source.

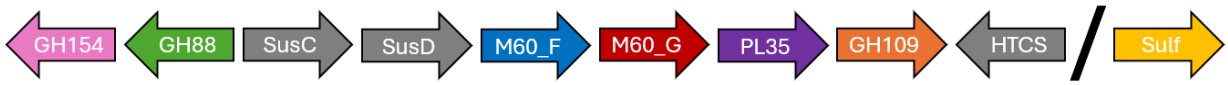
1.6. Overview of *Bacteroides caccae*

Bacteroides caccae is a relatively unstudied member of the genus whose abundance in the HGM is inextricably tied to diet, wherein it proliferates in high fat and high sugar diets (Johnson et al., 1987; McNulty et al., 2013; Terrapon et al., 2018). As with other members of its genus, *B. caccae* can degrade numerous nutrient sources, including pectin (Sirotek et al., 2004), GAGs (Overbeeke et al., 2022), and mucin (Desai et al., 2016). It is even capable of subsisting entirely on GAGs, including CS-A and CS-C, as its sole carbon source (D. Ndeh et al., 2020), and can enrich the growth of other *Bacteroides* spp. by sharing these glycan degradation products (Overbeeke et al., 2022).

Despite the noticeable impact of this species during various health states (Berkhout et al., 2024; Desai et al., 2016; B. Wei et al., 2001), *B. caccae* has no characterized PULs in the literature. Furthermore, of the 60 putative PULs it has in its genome, fifteen of these contain M60-like peptidases (Terrapon et al., 2018). There is presently no characterized PUL in any microbial species that describes the cooperation of both CAZyme and M60-like peptidase components. In other words: the ability of a singular PUL to degrade both glycan and peptide components of glycoproteins has not been observed. This marks *B. caccae* as a potential model for understanding

how members of the HGM completely degrade host and dietary glycoproteins, such as mucin and proteoglycans.

Here, this project focuses on PUL25 from *B. caccae* (**Figure 1.11**). PUL25 encodes nine proteins: the canonical SusC-like and SusD-like components of the SusCD-like complex, one regulatory hybrid two-component system (HTCS), two M60-like peptidases, and four CAZymes. To summarize the bioinformatic data available, the CAZyme families associated with PUL25 suggests that this PUL sequentially depolymerizes a GAG; specifically, chondroitin (*i.e.,* unsulfated state of chondroitin sulfate). This, along with association of two M60-like peptidases, indicates that PUL25 likely targets proteoglycans.



Locus Tag	Enzyme	Function/Activity
CGC64_04070	Glycoside hydrolase 154	Predicted uronic acid dehydratase
CGC64_04075	Glycoside hydrolase 88	4, 5 unsaturated uronic acid hydrolase
CGC64_04080	SusC-like	Ton-B dependent glycan transport
CGC64_04085	SusD-like	Glycan binding and acquisition
CGC64_04090	M60 peptidase (F)	*O*-glycopeptidase
CGC64_04095	M60 peptidase (G)	*O*-glycopeptidase
CGC64_04100	Polysaccharide lyase 35	Endo-processive surface lyase
CGC64_04105	Glycoside hydrolase 109	N-acetylgalactosamine hydrolase
CGC64_04110	Hybrid two-component system	BcPUL25 regulatory sensor
CGC64_04290	4-O-Sulfatase	Endo-acting 4-*O*-sulfatase

Figure 1.11: Schematic and summarized activities of PUL25 from *Bacteroides caccae*. Target proteins are colored. Genes not localized within the PUL are indicated with a slash, or with a line on the table.

1.7. Carbohydrate active enzyme (CAZyme) families associated with PUL25

1.7.1. PL35

Of the ~40 PL families listed on the CAZy database, ~25% of these contain members active on glycosaminoglycans (GAGs) (The CAZypedia Consortium, 2018; L. Wei, Zou, et al., 2024). GAG lyases (GAGases) are categorized based on substrate specificity, although lyases that are active on hyaluronan (HA) are often also active on chondroitin sulfate (CS). To summarize, PL12, PL13, PL15, and PL21 are heparin/heparin sulfate (HS) lyases, or hepases; and PL6, PL8, PL16, PL23, PL29, PL30, PL33, and PL37 families are HA lyases and chondroitin sulfate/dermatan

sulfate (CS/DS) lyases, or chondroitinases (Drula et al., 2022). Generally, the characterized GAGases in these lyase families are strictly specific for their substrates; they do not tolerate changes in glycosidic linkage, nor easily target different uronic acid moieties. Members of the PL35 family are an exception to this observation.

The PL35 family is a relatively new PL family, with putative and characterized members scattered amongst organisms that reside in numerous diverse environments, including marine, terrestrial, and within the human gut microbiota (HGM) (Drula et al., 2022). Its first functionally characterized member from *Victivallis vadensis* (VvPL35) was described as an *endo*-chondroitinase (Helbert et al., 2019). For clarification, chondroitin (CH) describes a unsulfated CS polymer. In the past year, more information about this family has been uncovered (L. Wei, Cao, et al., 2024; L. Wei, Zou, et al., 2024). The observed GAG activity of this family has been expanded to include HA, HS, CS, and CH (Lu et al., 2024; L. Wei, Zou, et al., 2024). One member was even observed to degrade alginate, which is a completely different uronic acid-containing polysaccharide found in brown algae (Abka-khajouei et al., 2022; L. Wei, Zou, et al., 2024). Additionally, the PL35 family now has two published apo-structures; it is described to have a N-terminal $(\alpha/\alpha)_7$ toroid domain and a C-terminal two-layered β-sheet domain (L. Wei, Cao, et al., 2024). However, the residues involved in the substrate recognition of this family remains a mystery, as there is no PL35-substrate complex yet.

It should be noted that, at the start of this project, VvPL35 was the sole characterized member of the family (Helbert et al., 2019). Thus, the PL35 member associated with *B. caccae*'s PUL25 (BcPL35) was proposed to be an *endo*-acting chondroitin lyase that it specifically targets CH, or unsulfated CS.

1.7.2. GH88

Often, PLs work with unsaturated glucuronyl hydrolases to degrade their substrates (Nakamichi et al., 2014). These include GH families such as GH5, GH88, and GH105 (Drula et al., 2022; Lombard et al., 2025). In the PUL context, this relationship is most exemplified by the GH88 family, as there are currently 31 characterized PULs with GH88 members that hydrolyze unsaturated glucuronic acid (ΔGlcA) moieties generated by PL activity (Terrapon et al., 2018).

Members of this family are broadly distributed in bacteria, archaea, and fungi, although only bacterial GH88s have been studied to date (Drula et al., 2022). Overall, the GH88 family acts on ΔGlcA moieties from xanthan and gellan, which are bacterial polysaccharides found in biofilms; as well as differentially sulfated GAGs, including CS, HA, and heparin/HS (Drula et al., 2022; Hashimoto et al., 1999; Itoh et al., 2006; Lombard et al., 2025).

Structurally, there are three characterized members of GH88: *Bacillus* sp. GL1 (BacUGL; BacGH88) (Itoh et al., 2006a), *Streptococcus agalactiae* (SagUGL; SaGH88) (Nakamichi et al., 2011), and *P. heparinus* (Phep_2830; PhGH88) (Nakamichi et al., 2014). These members share a $(\alpha/\alpha)_6$ barrel topology. Mutagenesis studies of BacUGL and SagUGL have revealed two conserved aspartic acid (Asp) residues that are essential for activity (Itoh et al., 2006b; Nakamichi et al., 2011). The unique catalytic mechanism of this family is also known (**Figure 1.12**). GH88-catalyzed hydrolysis relies on Asp, histidine (His), and glutamine (Gln), wherein hydration of the unsaturated alkene at C4 to C5 results in the release of a saturated 5-keto-4-deoxyuronate product (Itoh et al., 2006).

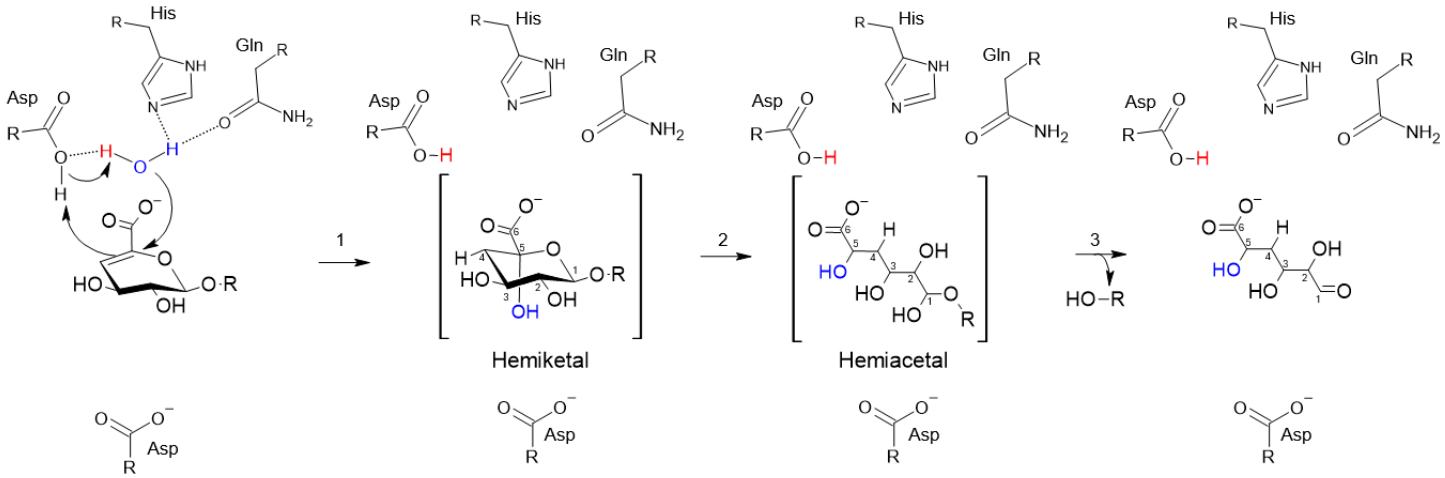


Figure 1.12: General GH88 catalytic mechanism.

An Asp acts as a general acid/base by **(1)** donating a proton to the C4 double bond on unsaturated hexuronate (ΔHexA), and then deprotonating a His/Gln coordinated water that nucleophilically attacks C5 of ΔHexA. **(2)** This resulting hemiketal is unstable and will convert into an α-keto acid (hemiacetal). **(3)** Lastly, because of the aldehyde-hemiacetal equilibrium, this hemiacetal will cleave off its saccharide leaving group (OH-R) to convert into an aldehyde product (5-keto-4-deoxyuronate) (Itoh et al., 2006).

Given that the GH88 member associated with *B. caccae*'s PUL25 (BcGH88) shares 46% sequence identity to PhGH88 (Nakamichi et al., 2014), which targets singly sulfated and unsulfated unsaturated heparin, it is expected that BcGH88 can tolerate singly sulfated CS (*i.e.,* CS-A and CS-C) and unsulfated CH oligosaccharides.

1.7.3. GH109

Initial interest in finding α-*N*-acetylgalactosaminidases and α-galactosidase was driven by the discovery of the ABO blood group system, and subsequent attempts to produce universal red blood cells for transfusions (Rahfeld & Withers, 2020). These efforts revealed the GH109 family, which are *exo*-*N*-acetylgalactosaminidases with members that can act on one or both α- and β-GalNAc (Q. P. Liu et al., 2007; Sulzenbacher et al., 2010; Teze et al., 2020). Members of this family are distributed among bacteria, with two previously characterized structures: NagA, from *Elizabethkingia meningosepticum* (Sulzenbacher et al., 2010), and AmGH109_A from *A. muciniphila* (Teze et al., 2020). Structural analysis of NagA showed this family's requirement for NAD^+, whereas analyses of the latter revealed a catalytic His in a conserved GGHGG motif. This NAD^+ requirement is an unusual trait only found in two other GH families (GH4 and GH97) (Teze et al., 2020).

As the closest homolog to the GH109 member associated with PUL25 from *B. caccae* (BcGH109) is AmGH109_A (36% sequence identity), BcGH109 is expected to target on α- and β GalNAc moieties, since AmGH109_A exhibits this activity (Teze et al., 2020).

1.7.4. GH154

Past the acquisition of its apo structure, the molecular factors underpinning substrate recognition of GH154 family has been elusive (Cartmell et al., 2018; Hameleers et al., 2024). Currently, there are two characterized GH154 members: BT3677, an *exo*-β-glucuronic acid (GlcA) hydrolase from *B. thetaiotaomicron* (BtGH154) (Cartmell et al., 2018), and an *exo*-β-galactose (Gal) hydrolase (BD-β-Gal) isolated from beaver droppings (Hameleers et al., 2024). The latter contributes the sole apo-structure of the family, showing a homo-tetramer with $(\alpha/\alpha)_6$ topology. However, analysis of the active site does not easily reveal the catalytic residues that contributes to activity. GH families typically use residues containing carboxylic acids (e.g., Glu, Asp) that are

positioned to act as general acid/bases or nucleophiles to allow for hydrolysis through a retaining or inverting mechanism (**Figure 1.7**) (G. Davies & Henrissat, 1995). Although the BD -β-Gal authors propose that Glu73 was a potential residue of interest, there were no other active site residues nearby that could serve as its catalytic counterpart. Additionally, mutagenic analysis of Glu73 did not result in the complete loss of activity, further confounding hypotheses on exactly how GH154 recognizes and acts on substrates, as well as emphasizing the need for an enzyme-substrate complex to elucidate this family's catalytic mechanism.

Given the predicted substrate target of PUL25 and characterized activity of BtGH154 (Cartmell et al., 2018), the GH154 member associated with *B. caccae*'s PUL25 (BcGH154) is expected to act on *exo*-β-GlcA.

1.8. Project hypothesis and objectives

By its composition, ability to degrade recalcitrant nutrient sources, and subsequent release of metabolites that enable crosstalk between the gut ecosystem and various host physiological processes, the HGM undeniably plays a significant role in our overall health (**Figure 1.1** and **Figure 1.4**) (Desai et al., 2016; Ross et al., 2024; Wardman et al., 2022). Although research in this field has been ongoing for decades, many interactions between the host and its gut microbiota remain poorly understood. This includes the latter's ability to target host and dietary glycoproteins. Despite the highly diverse nature of the human diet (Ross et al., 2024), there is a paucity of knowledge surrounding the HGM's ability to degrade nutrient sources that contain both glycan and protein components. To date, there is no study on any microbe that has directly examined how CAZymes and peptidases function together to catabolize these glycoproteins for downstream metabolic processes.

Here, I propose that *B. caccae* is a model of glycoprotein degradation by the HGM due to the sixteen putative M60-like peptidases associated with its genome (Nakjang et al., 2012; Noach et al., 2017; Terrapon et al., 2018). This is a remarkably high level of redundancy, especially given the observed transiency of these glycopeptidases in any genome and the metallopeptidase's requirement for specific *O*-glycans prior to cleavage (Nakjang et al., 2012). Additionally, fifteen of these putative glycopeptidase genes are associated with PULs (Terrapon et al., 2018). Put

together, this further corroborates *B. caccae*'s role as a glycoprotein specialist in the HGM. However, despite this and the repeated observation of the microbe's ability to proliferate and upregulate genes in the presence of GAGs and mucin (Berkhout et al., 2024; Desai et al., 2016; Overbeeke et al., 2022), it does not have any characterized PULs regarding either in literature.

Of the thirteen M60-like peptidase-containing PULs in *B. caccae* (ATCC 43185), this project focuses on the predicted enzymatic consortium associated with PUL25 (**Figure 1.11**) (Terrapon et al., 2018a). Given the bioinformatic analyses on the CAZyme families in this PUL, I hypothesize that PUL25-associated enzymes collectively target CS-containing proteoglycans. To address this hypothesis, I have the following objectives. I will demonstrate the glycan specificity and order by which the CAZyme consortium (BcPL35, BcGH88, BcGH109, and BcGH154) depolymerizes CS (Cartmell et al., 2018; Nakamichi et al., 2014; Teze et al., 2020; L. Wei, Zou, et al., 2024). Additionally, I will determine the *O*-glycan specificity of the two associated M60-like peptidases (BcM60_F and BcM60_G) that would cleave the proteoglycan peptide backbone (Noach et al., 2017; Pluvinage et al., 2021; Taleb et al., 2022). And, lastly, I will obtain structures of these enzymes to gain molecular insight on how these interact with their substrates.

Ultimately, this project will contribute toward our understanding of the molecular interplay within this trifecta of microbial composition, nutrient availability, and metabolite production. In the future, studies such as this will eventually result in the development of therapeutics that will alleviate HGM-influenced disease states.

Chapter 2 : Materials and Methods

2.1. Materials.

Chondroitin sulfate A (CS-A), CS-B, CS-C, hyaluronic acid (HA), and all tested *p*NP-sugars were acquired from Sigma. All unsaturated disaccharides (ΔHA, ΔHep, CSΔ0S, CSΔ4S, CSΔ6S), as well as chondrosine and methyl-β-D-glucuronide, were ordered from Biosynth. Saturated hyaluronic acid disaccharide (sHA) was obtained from Sussex Research. The mucin-*O*-glycan Core 2 attached to a threonine (Thr) residue (Core 2-Thr) was graciously gifted by Dr. Warren Wakarchuck (University of Alberta). Biotinylated *Vicia villosa* (VVA) lectin was acquired from Vector Laboratories for the detection of Tn-antigen on FRET substrates. All reagents, chemicals, and other materials were acquired from Sigma unless otherwise specified.

2.2. Protein production and purification.

All constructs were obtained as synthetic genes in pET28a from GenScript or TwistBiosciences. This includes plasmids containing BcGH154 (CGC64_04070; BcGDH; residues 24-423), BcGH88 (CGC64_04075; residues 31-433), BcM60_F (CGC64_04090; residues 437-959), BcM60_F_CBM (residues 245-959) BcM60_G (CGC64_04095; residues 76-510), BcPL35 (CGC64_04100; residues 27-613), BcGH109 (CGC64_04105; residues 19-514), the *endo*-sulfatases from *Bacteroides thetaiotamicron* (BT3349; BtSulf_S93C; residues 1-508) (Ulmer et al., 2014b), and its *B. caccae* ortholog (CGC64_04290; BcSulf_S91C; residues 20-515), as well as GH154 from *B. thetaiotamicron* (BT_3677; BtGH154; residues 30-673) (Drula et al., 2022; Hameleers et al., 2024; Terrapon et al., 2018). BcPL35 and BcGH154 mutant constructs were cloned using site-directed mutagenesis via polymerase chain reaction. BT4244 plasmid was acquired from a previous lab member. The BcM60_G _E327Q construct was acquired from Dr. Benjamin Pluvinage. All reading frames of the engineered genes incorporated an N-terminal 6-histidine tag separated from the target protein by a thrombin protease cleavage site. Oligonucleotide primers and DNA fragments used to generate mutants are given in **Table 2.1** and **Table 2.2**. The fidelity of all constructs was confirmed using whole plasmid sequencing.

Table 2.1: Oligonucleotide primer sequences used for gene amplification.

Construct	SEQUENCE (5' → 3')
PL35_H387A_FWD	GCGAACCGCGTCAAGCGGCCCAGCATTTGG
PL35_H387A_REV	CCGTCCAAATGCTGGGCCGCTTGACGCGG
PL35_W419A_FWD	GCCGATGATTACGCTTTGCCGGGTTTTGCCGATGGGCGCC
PL35_W419A_REV	CGTCGGGGCGCCCATCGGCAAAACCCGGC
PL35_R076A_Twist_FWD	GATAGAAACCTTTGCTGTCCATACATATTGCTAGCCATATG
PL35_R076A_Twist_REV	CCAACGAAGCATTGCTGTGGGTTTG
PL35_E228A_Y232A_Twist_FWD	CCTCTTTCGGGTAATCCTCGGCAATGCCGAGCG
PL35_E228A_Y232A_Twist_REV	CGGATGCTCAAAACAGCCCGGCGTTCTTCTTGTTCAG
PL35_Y419A_Twist_FWD	CTTGACGCGGTTCGCCGCCCTTTGCAATAAG
PL35_Y419A_Twist_REV	CCCTGTATAAAGACCAGGCGTCCAGCGTTTTTCGTACC
GH154_E074A_FWD	AAGGACATGACGCATCTGGCAGCCGTCGGTCGC
GH154_E074A_REV	CAACGTGCGACCGACGGCTGCCAGATGCGTCAT
GH154_R078A_FWD	GAGGCCGTCGGTGCTACGTTGGCTG
GH154_R078A_REV	GACGCCAGCCAACGTAGCACCGACG
GH154_N176A_FWD	GATCGTACCGGCGCATACGCCAACTGGCTG
GH154_N176A__REV	GCCGGTAAATAACAGCCAGTTGGCGTATGCGCC
GH154_N177A_FWD	ACCGGCGCATACAACGCCTGGCTGTTATTTAC
GH154_N177A _REV	CAGGCCGGTAAATAACAGCCAGGCGTTGTATGC
GH154_W178A_FWD	GGCGCATACAACAACGCACTGTTATTTACC
GH154_W178A _REV	TCAGGCCGGTAAATAACAGTGCGTTGTTGTATG
GH154_F181A_FWD	AACAACTGGCTGTTAGCTACCGGCCTG
GH154_F181A _REV	CGACTCGGTCAGGCCGGTAGCTAACAG
GH154_Y229A_FWD	TTCAGCATGGATAATGCTAACGCGTAC
GH154_Y229A _REV	CATCACGTACGCGTTAGCATTATCCATG
GH154_R285A_FWD	ACGTACCCGGCGTTTGGTGCTTCCGTGACC
GH154_R285A _REV	ACGATAGGTCACGGAAGCACCAAACGCCGG
GH154_Y289A_FWD	GGTCGTTCCGTGACCGCTCGTACTGCTG
GH154_Y289A _REV	GAAAGCAGCAGTACGAGCGGTCACGGAAC
GH154_Y354A_FWD	GAGTGCGCTGACGGCGCTACTTCGACC
GH154_Y354A _REV	GCTCCCGGTCGAAGTAGCGCCGTCAGC

Table 2.2: List of BcPL35 Twist fragments used for mutagenesis.

Name	SEQUENCE (5' → 3')
R076A	AATATGTATGGACAGCAAAGGTTTCTATCAAACCACCCGAGACTGCTGTT TACCGGTGCTGAGGAAGCGGCGGTCAAACAATTAATCCAGAACAACCGC CTGGCGGGCGAGCTGGCGGAGTTCCTGAAGGCGAAGGCCGACACCCTGG TCATTACCCCGCAAAAACCGTATCTTAAAGACAAGTACGGCAATATTCTG TGGACGAGCGCCTCGTATGTTAATCGCCTGGGTACCTTGGCCCTGGCCTA CCGGTTGTACGGTGAGCGTAAATACCTGGACGCTGCCAACGAAGCATTG CTG
E228A	CATTGCCGAGGATTACCCGAAAGAGGCTGCGGTTATCCTGGACAACGCT GCGAAGTACATGCCGAATTGTCTGAAGCACTTTGCACCGGATGGTGTTTG TTATGCCGGTCCGGCGTACTGGGGTTATACCACATCGTACTTGACCCTGT ACTTAAAGGCGGTGGCCGATAACGACAACGGTAAGGGCGGGATCGCCCA GTTACCGGGTCTGGAGCGTACTGCGCTCTACCAGAAGCGCACCCTGACTC CGAGCGGTCGCTTATTCAACTTTGGTAACGCAGGCGCGGATGCTCAAAAC AG
Y232A	CATTGCCGAGGATTACCCGAAAGAGGCTGCGGTTATCCTGGACAACGCT GCGAAGTACATGCCGAATTGTCTGAAGCACTTTGCACCGGATGGTGTTTG TTATGAGGGTCCGGCGGCCTGGGGTTATACCACATCGTACTTGACCCTGT ACTTAAAGGCGGTGGCCGATAACGACAACGGTAAGGGCGGGATCGCCCA GTTACCGGGTCTGGAGCGTACTGCGCTCTACCAGAAGCGCACCCTGACTC CGAGCGGTCGCTTATTCAACTTTGGTAACGCAGGCGCGGATGCTCAAAAC AG
Y413A	GCGGCGAACCGCGTCAAGCGCACCAGCATTTGGACGGCGGCACCTTCAT CGTGGAAAGCAATGGCGTTTGCTGGACCGAAGATCTGGGTGCCGATGAT GCCGCTTTGCCGGGTTTTTGGGATGGGCGCCCCGACGGCCAGCGTTGGAA ATATTTCCGCAACAATAACTTTAGCCATAACACGTTGTCCATCGATCATA AGATCCAATACGCGAATGGCGAGGCGTTTGTGTGCGAGGAACACACCGA TGCGAAGCAACCGAGCGTTAAGCTGGACATGACCACCCTGTATAAAGAC CAGG

Proteins were produced in *Escherichia coli* strain BL21 (DE3). KdgF, KduI, and KduD were produced and purified as described previously (Hobbs et al., 2016) (**Table 2.3**). For assaying preliminary activity against *para*-nitrophenyl-containing (*p*NP) sugars, BcGH154 was initially expressed in *E. coli* strain Tuner (DE3) to account for the endogenous β-galactosidase in BL21 strains. After confirmation of the CAZyme's true activity, it was grown in *E. coli* strain BL21 (DE3) for further functional characterization. In brief, cultures were grown at 37°C, with shaking at 170 rpm, until an OD_{600} of ~0.6 was reached. These growths were moved to 16°C for at least 30 minutes prior to addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) (0.5 mM final concentration) to induce overnight expression of target protein.

Table 2.3: Properties and expression conditions of all target proteins used in project.
Color-coded proteins indicate direct association with PUL25 in *B. caccae*. *Abbreviations*: Kan: kanamycin (final concentration: 50 µg/mL), Cam: chloramphenicol (final concentration: 25 µg/mL), Amp: ampicillin (final concentration: 100 µg/mL)

Locus Tag	Construct	Molecular weight (kDa)	Extinction coefficient (M^{-1} cm^{-1})	Expression strain	Growth/expression temperatures (°C)	Selective antibiotics
CGC64_04070	BcGH154_WT	45.5	92,820	BL21 (DE3)/Tuner	37/16	Kan
	BcGH154_E074A	45.5	92,820	BL21 (DE3)	37/16	Kan
	BcGH154_R078A	45.5	92,820	BL21 (DE3)	37/16	Kan
	BcGH154_N176A	45.5	92,820	BL21 (DE3)	37/16	Kan
	BcGH154_N177A	45.5	92,820	BL21 (DE3)	37/16	Kan
	BcGH154_W178A	45.5	87,320	BL21 (DE3)	37/16	Kan
	BcGH154_F181A	45.5	92,820	BL21 (DE3)	37/16	Kan
	BcGH154_Y229A	45.5	91,330	BL21 (DE3)	37/16	Kan
	BcGH154_R285A	45.5	92,820	BL21 (DE3)	37/16	Kan
	BcGH154_Y289A	45.5	91,330	BL21 (DE3)	37/16	Kan
	BcGH154_Y354A	45.5	91,330	BL21 (DE3)	37/16	Kan
CGC64_04075	BcGH88	46.0	81,250	BL21 (DE3)	37/16	Kan
CGC64_04090	BcM60_F	59.4	87,920	BL21 (DE3)	37/16	Kan
	BcM60_F_CBM	82.0	125,030	BL21 (DE3)	37/16	Kan
CGC64_04095	BcM60_G_WT	49.7	84,355	BL21 (DE3)	37/16	Kan
	BcM60_G_E327Q	49.7	84,355	BL21 (DE3)	37/16	Kan
CGC64_04100	BcPL35_WT	66.6	122,160	BL21 (DE3)	37/16	Kan
	BcPL35_R076A	66.6	122,160	BL21 (DE3)	37/16	Kan
	BcPL35_E228A	66.6	122,160	BL21 (DE3)	37/16	Kan
	BcPL35_Y232A	66.6	120,670	BL21 (DE3)	37/16	Kan
	BcPL35_H387A	66.6	122,160	BL21 (DE3)	37/16	Kan
	BcPL35_Y413A	66.6	120,670	BL21 (DE3)	37/16	Kan
	BcPL35_W419A	66.6	116,660	BL21 (DE3)	37/16	Kan
CGC64_04105	BcGH109	55.8	91,790	BL21 (DE3)	37/16	Kan
BT3677	BtGH154	72.9	155,620	BL21 (DE3)	37/16	Kan
CGC64_04290	BcSulf	56.7	70,750	BL21 (DE3)	37/16	Kan, Amp
BT3349	BtSulf	56.7	70,750	BL21 (DE3)	37/16	Kan, Amp
YE1889	Ye_KdgF	12.5	13,980	BL21 (DE3)	37/16	Kan
YE1888	Ye_Kdul	31.2	34,380	BL21 (DE3)	37/16	Kan
YE1887	Ye_KduD	27.3	29,450	BL21 (DE3)	37/16	Kan
	NKD-FRET	~55	77,240	BL21 (DE3)	37/16	Kan
	Tn-FRET	~55	77,240	Origami2 (DE3) with pOGO42	37/18	Kan/Cam

Mutant sulfatases, where the catalytic serine was mutated to cysteine to enable maturation in *E. coli,* were co-expressed with a sulfatase maturing enzyme (*Mycobacterium tuberculosis* formylglycine generating enzyme; MT_FGE) as described previously (Carrico et al., 2007). Upon reaching an OD_{600} of ~0.6, sulfatase cultures were moved to 16°C prior to addition of 250 mg L-arabinose to induce production of MT_FGE. After three hours, IPTG was added to induce production of target sulfatases. All proteins were left to express overnight. Maturation of sulfatases was confirmed via colormetric assay with *p*NP-SO_4, wherein increased absorbance at 405 nm indicated activity.

Target proteins were purified from clarified cell extracts via Ni^{+2}-NTA immobilized metal affinity chromatography, wherein target protein was eluted using a gradient of 20-500 mM imidazole in binding buffer (20 mM Tris pH 8.0, 500 mM NaCl). Fractions were assessed for purity using SDS-PAGE. Chosen fractions were pooled and dialyzed overnight at 4°C. Isolated BcPL35 constructs was either dialyzed into crystallography buffer (20 mM Tris pH 8.0, 500 mM NaCl, 10 mM dithiothreitol) or into activity buffer (20 mM Bis-Tris pH 6.0, 500 mM NaCl, 10% glycerol). For activity assays, BcGH88, BcGH109, BcGH154 constructs, and all sulfatases were dialysed into binding buffer, while BcM60_F, BcM60_F_CBM, BcM60_G, KdgF, Kdul, and KduD were desalted into 20 mM Tris pH 7.4 buffer. For crystallography, BcGH154 was desalted into 20 mM Tris pH 8.0 (PD MidiTrap G-25; Cytiva) prior to overnight thrombin cleavage of the His-tag.

Additional purification of protein for crystallography was done using size exclusion chromatography (SEC) using a HiPrep 16/60 Sephacryl S-200 HR column (GE Healthcare) equilibrated in an appropriate buffer. All proteins were concentrated using a stirred ultrafiltration unit (Amicon) with a 10 kDa molecular weight cut-off membrane (EMD Millipore).

Protein concentration was determined using a UV absorbances at 280 nm. Extinction coefficients were calculated using Expasy ProtParam (**Table 2.3**). Concentrated proteins were stored at 4°C, or flash frozen in liquid nitrogen and stored at -80°C, prior to use.

2.2.1. Production, modification, and isolation of FRET substrates.

For clarification: Forster resonance energy transfer (FRET) substrates are composed of two fluorescent proteins (N-terminal mNeonGreen and C-terminal mScarlet) linked by a peptide sequence that can be *O*-glycosylated. These linker sequences are specific for sections of glycoprotein targets, such as MUC2 or MUC5A/C. Hence, FRET substrates are generally designated by the origin of their linker sequence (**Table 2.4**).

FRET substrates were cloned by Dr. Alisdair Boraston. Stocks of these substrates were produced and maintained primarily by Olivia Canil and Julien Case. Additional glycotransferase-containing plasmid (pOGO42) was provided by Dr. Warren Wakarchuck (University of Alberta) (Wardman et al., 2023). All reading frames of the engineered genes incorporated an N-terminal 6-histidine tag separated from the target protein by a thrombin protease cleavage site. The fidelity of all constructs was confirmed using whole plasmid sequencing.

Escherichia coli strain Origami2 (DE3) were doubly transformed with an appropriate recombinant FRET plasmid and pOGO-42 using a heat shock protocol (**Table 2.4**). These FRET cultures were moved to 18°C for at least 30 minutes prior to adding IPTG and 0.2% *N*-acetylglucosamine (**Table 2.3**). All proteins were left at their final temperature to express overnight.

Target proteins were purified from clarified cell extracts via Ni^{+2}-NTA immobilized metal affinity chromatography, wherein the supernatant was washed with three column volumes of binding buffer with 20 mM imidazole prior to elution with 500 mM imidazole in binding buffer to isolate the His-tagged FRET substrates. Isolated protein was desalted into 50 mM HEPES pH 7.0 buffer (PD MidiTrap G-25; Cytiva) and assessed for purity using SDS-PAGE. FRET substrate concentration was determined using a UV absorbances at 280 nm **(Table 2.3)** and glycosylation status was confirmed through screening with BT4244. Extinction coefficients were calculated using Expasy ProtParam. After were stored at 4°C, or flash frozen in liquid nitrogen and stored at -80°C, prior to use.

Table 2.4: FRET substrate linker origins and designations.
Identity of each peptide linker is listed on the left, wherein linkers found on the same glycoprotein are differentiated by their repeats (R1, R2) and/or their sequence (S1, S2). *Abbreviations*: IgA: immunoglobulin A, ASF: asialofetuin, GPR: glycoprotein Iba.

Linker Origin		Linker sequence	FRET #
MUC1	MUC1_S1	PAPGSTAPPAH	A1
	MUC1_S2	AHGVTSAPDTR	A2
MUC2	MUC2_R1	TTVTPTPTPTG	A3
	MUC2_R2_S1	VPPTTTPSPPP	A4
	MUC2_R2_S2	SPPPTSTTTLP	A5
MUC5A/C	MUC5AC_R1	PPTTSTTSAPP	A6
	MUC5AC_R2	PVPATTVGPVP	A7
MUC5B	MUC5B_S1	GSTATPSSTPG	A8
	MUC5B_S2	VLTTTATTPTP	A9
IgA	IgA1_HP_S2	SPSTPPTPSPS	A10
MUC4	MUC4	STGHATPLPVT	A11
MUC3	MUC3_A_B	TSSITTTETTS	A12
MUC7	MUC7_S1	TTAAPPTPSAT	B1
	MUC7_S2	PSATTPAPPSS	B2
MUC8	MUC8_S1	PLQEGTPGSRV	B3
	MUC8_S2	VHELPTSSPGG	B4
MUC10	MUC10	TTDSTTPAPTT	B5
MUC11/12	MUC11/12	SSPGSTHTTLS	B6
MUC16	MUC16	SVPTTSTPGTS	B7
MUC19	MUC19	STTVAPGSTTV	B8
ASF	ASF	AEAPTAVPDKG	B9
GPR	GPR	PAPSPTTPEPT	B10
MUC6	MUC6_S1	TTTYPTPSHPQ	B11
MUC17	MUC17_S1	SPTNSSPTTAE	C3

2.3. Generation of chondroitin.

Chondroitin was made via chemical desulfation of chondroitin sulfate A (CS-A) using previously described methods (Nagasawa et al., 1979). Briefly, CS-A-pyridine salt was prepared by dissolving 1.40 g of sodium CS-A in 40 mL dH_2O. This was passed through a cation exchange column (Dowex 5W, X8, 50-100 mesh). After washing the column with three column volumes of dH_2O, the washes were pooled and neutralized to pH 7 using pyridine prior to lyophilization. The final reaction used 0.53 g of CS-A-pyridine dissolved in 50 mL DMSO in 10% methanol and

incubated at 80°C for at least 5 hours. This solution was diluted in an equivalent volume of dH_2O, pH adjusted to between pH 9.0 to 9.5, and dialyzed in dH_2O overnight, with multiple dH_2O changes per hour the following day. Afterwards, the chondroitin solution is filtered, lyophilized, and stored at -20°C. The substrate was resuspended in dH_2O prior to use.

2.4. Fluorophore-assisted carbohydrate electrophoresis (FACE).

All enzymatic reactions for analysis by FACE were done at 37°C for two hours, with 3 mg/mL GAG substrate, 5 μM per enzyme, 100 mM Tris pH 7.5, 1 mM $MgCl_2$, and 10 μM sulfatase (when used). 2-aminoacridone (AMAC) were used to label reactions. Reaction progression was halted by adding 95% ethanol. Samples were dried in a speed vacuum for at least two hours, or until samples had a jelly-like consistency. These were resuspended in 0.02M AMAC dissolved in diluted acetic acid (3:17, acetic acid: dH_2O) and 0.1M sodium cyanoborohydride in DMSO ($NaBH_3CN$; Sigma) prior to overnight incubation at 37°C, wrapped in foil. AMAC-tagged reactions were dried in a speed vacuum for at least two hours and resuspended in FACE loading buffer (62 mM Tris pH 6.8, 0.1% glycerol). All reactions were run on 30% polyacrylamide gels at 100V for 30 minutes, and 200V for 120 minutes after a 15-minute rest time, in an icebox (4°C) prior to imaging.

2.5. Screening glycoside hydrolase activity.

Initial screening with synthetic aryl-glycosides was done to explore potential hydrolase activity of BcGH154 and BcGH109. All reactions were done in triplicate using Costar® 96-well microplates (Corning Incorporated) and incubated for 1 hour at 37°C. Reactions comprised 100 mM Tris pH 7.4, 1 mM aryl-glycoside, and 2 μM enzyme. Reaction progress was stopped by adding 100 μL of 100 mM NaOH, and endpoint results by absorbance at 405 nm were read with a SpectraMax M5 plate reader (Molecular Devices).

2.6. Coupled assay for 5-keto-4-deoxyuronate detection.

The activity of BcGH88 on unsaturated chondroitin disaccharide (CSΔ0S) was indirectly confirmed using a coupled assay with KduI and KduD for the detection of the BcGH88 product 5-keto-4-deoxyuronate. KduD activity consumes NADH, providing a readout at 340 nm. All reactions were done in triplicate in UV-Star® 96-well microplates (Greiner Bio-One) and read at 340 nm for 1 hour at 25°C. Reactions included 100 mM Tris pH 7.4, 0.200 mM NADH, 0.500

mM CSΔ0S disaccharide, and 0.500 μM per enzyme. Substrate was equilibrated in plate reader for 15 minutes prior to enzyme addition. Reaction progression was monitored with a SpectraMax M5 plate reader (Molecular Devices).

2.6.1. Coupled assay for β-glucuronidase detection.

β-glucuronidase activity on chondrosine was assessed using a coupled assay based for free glucuronic acid detection (Megazyme). The assay is based on oxidation of glucuronic acid by uronate dehydrogenase to generate NADH from NAD^+. Reactions were performed in triplicate according to manufacturer's instructions in UV-Star® 96-well microplates (Greiner Bio-One) (Megazyme, 2019). Reaction progress was monitored at 37°C for 30 min using a SpectraMax M5 plate reader (Molecular Devices), wherein reads were taken at 340 nm (Megazyme, 2019). Reaction progress was monitored at 37°C for 30 min using a SpectraMax M5 plate reader (Molecular Devices), wherein reads were taken at 340 nm. Final reaction conditions include 50 mM Tris pH 7.4, 1.5 mg/mL chondrosine or glucuronic acid standard, and 5 μM BcGH154.

2.7. UV absorption assays for activity and assessment of CAZyme kinetic parameters.

BcPL35 activity on various GAGs was initially assessed by absorbance at 232 nm following the generation of the Δ4,5 bond at the new non-reducing end. All reactions were done in triplicate in UV-Star® 96-well microplates (Greiner Bio-One) and read at 232 nm for 1 hour at 37°C. Samples that displayed activity were left overnight at room temperature to reach equilibrium. BcGH88 was then added to these reactions and read under the same parameters as above. Reaction progression was monitored with a SpectraMax M5 plate reader (Molecular Devices). Final reactions comprised 100 mM Tris pH 7.4, 1 mg/mL chondroitin, and 1 μM per enzyme.

Kinetic analysis of chondroitin cleavage by BcPL35 at 37°C was done through the continuous monitoring of double bond formation at 232 nm using substrate concentration up to 7 mg/mL. Initial assessment of all BcPL35 mutants were done with 1 μM enzyme, 3 mg/mL of chondroitin, and 100 mM Tris pH 7.5, in triplicate at 37°C. Upon confirmation of activity, the kinetics of mutants with measurable activity were determined as for wild-type. Reactions were done in UV-Star® 96-well microplate (Greiner Bio-One) and read in a SpectraMax M5 plate reader (Molecular

Devices). Saturation could not be reached, so k_{cat}/K_M values were determined by linear fits to the plots of rate vs chondroitin concentration.

Cleavage of Δ4,5-unsaturated GAG derived disaccharides by BcGH88 was monitored by loss of absorbance at 232 nm. All reactions were done in triplicate in UV-Star® 96-well microplates (Greiner Bio-One) and incubated for 1 hour at 37°C. Reaction progression was monitored with a SpectraMax M5 plate reader (Molecular Devices). Reaction components include: 100 mM Tris pH 7.4, 0.500 mM GAG disaccharide, and 0.025 µM or 1 µM BcGH88. Kinetic assays were performed by continuous assay of A_{232} nm depletion at substrate concentrations up to 1.5 mM.

The activity of BcGH154 on chondrosine was followed by increased absorbance at 232 nm. As with BcPL35, these reactions were allowed to plateau followed by addition of BcGH88 and continued monitoring at 232 nm. All reactions were done in triplicate in UV-Star® 96-well microplates (Greiner Bio-One) at 25°C in a SpectraMax M5 plate reader (Molecular Devices). BcGH154 and BcGH88 were used at 1 µM final concentration in 100 mM Tris pH 7.4. BcGH154 mutants were qualitatively assessed for activity using 10 mM chondrosine in a continuous A_{232} nm assay. Kinetic assays were performed by continuous assay of A_{232} nm at substrate concentrations up to 6 mM. Kinetic analysis of BtGH154 on 1-*O*-methyl-β-D-glucuronide was done under identical conditions using substrate up to 10 mM concentration.

In these kinetic assays, the absorbance was converted to concentration using the extinction coefficient of 5,200 M^{-1} cm^{-1}. Time windows of 10 minutes at the beginning of the kinetic runs were used to determine the initial rates. In all cases, saturation could not be reached, so k_{cat}/K_M values were determined by linear fits to the plots of rate vs substrate concentration.

2.7.1. Protein stability of BcGH154 constructs.

The stability of BcGH154 mutant proteins was assessed by differential scanning fluorimetry as reported previously (Hobbs et al., 2018). Protein was used at a concentration of 0.1 mg ml^{-1} in 20 mM Tris (pH 7.5) and 8× SYPRO Orange (Invitrogen, Carlsbad, CA). Reaction mixtures were

incubated in an Applied Biosystems 7500 Fast-Real Time PCR machine at 25 to 95°C with a ramp rate of 1°C min^{-1}. Data was analyzed using the accompanying StepOne software (version 2.3).

2.8. Mass spectrometry.

Mass spectrometry was used to further clarify the final products of BcPL35 and BcGH154. Samples were sent to Dr. Wade Abbott (University of Lethbridge, Agri-Food Canada) and Dr. Kristin Low (Agri-Food Canada) for mass spectrometry analysis. Reactions were run overnight at 37°C, flash frozen with liquid nitrogen, and stored at -80°C prior to use. Final reactions were in 100 mM Tris pH 7.4 with 1 µM enzyme and 200 µg substrate. Chondroitin and chondrosine enzyme reaction products (100 µg) were run through a Supelclean ENVI-Carb SPE column (Sigma) whereby columns were preconditioned with 2 mL of 80% (v/v) acetonitrile 0.1% (v/v) TFA, and 6 mL of water. Reaction products were then added, washed with 10 mL water followed by 2 mL of 5% (v/v) acetonitrile 0.1% (v/v) TFA, and eluted using 8 mL of 80% (v/v) acetronitrile 0.1% (v/v) TFA and evaporated to dryness under nitrogen gas. Each sample was then resuspended in 150 µL of water and filtered (0.20 µm) prior to injection and separation. Liquid chromatography was performed on a Vanquish ultra-high performance liquid chromatography (UHPLC) system (Thermo Scientific). Separation of the chondroitin and chondrosine enzyme products was achieved using an Acquity UPLC BEH Amide (HILIC) Column, 130Å, 1.7 µm, 2.1 mm x 150 mm (Waters) at a flow rate of 400 µL/min at 30 °C, using a gradient as shown in **Table 2.5**.

Enzyme products were injected in a volume of 10 µL for electrospray ionization mass spectrometry (ESI-MS) on an Orbitrap Fusion Tribrid system (Thermo Scientific) in negative ion mode. Mass spectra parameters are shown in **Table 2.6**. To select ions for MS2 experiments, a data-dependent MS2 strategy was employed using expected m/z values as listed in **Table S*3.2***, and a dynamic exclusion filter was used after 1 time for 2.5 s with default mass tolerance values. Higher-energy collisional dissociation (HCD) was employed to generate fragments. MS spectra were analyzed using Xcalibur and Freestyle software packages (Thermo Scientific), and product ions were identified based on manual interpretation.

Table 2.5: Gradient conditions for separation of chondroitin and chondrosine enzyme products.

Time (min)	A (%) 10 mM ammonium formate 50 mM formic acid 80% acetonitrile 20% water	B (%) 10 mM ammonium formate 50 mM formic acid 20% acetonitrile 80% water
0	100	0
1	70	30
10	62.5	37.5
10.1	0	100
12	0	100
12.1	100	0
22	100	0

Table 2.6: Parameters for ESI-MSn on the Orbitrap Fusion Tribrid

Parameter (units)	Value
ESI	
Spray voltage: negative ion (V)	2500
Sheath Gas (Arb)	45
Aux Gas (Arb)	10
Sweep Gas (Arb)	1
Ion Transfer Tube Temp (°C)	325
Vaporizer Temp (°C)	250
MS	
Detector Type	Orbitrap
Orbitrap Resolution	120K
Mass Range	Normal
Scan Range (m/z)	150-2000
RF Lens (%)	50
Data-dependent MS2	
Collision Energy Type	Normalized
Isolation Mode	Quadrupole
Activation Type	HCD
Collision Energy Mode	Stepped
Collision Energies (%)	15,30,45,60,80
Detector Type	Orbitrap
Orbitrap Resolution	30K

2.9. Forster resonance energy transfer (FRET) assay.

Three initial stocks were made per set of reactions: enzyme (50 mM HEPES pH 7.0, 200 μM $ZnCl_2$) and substrate stocks (50 mM HEPES pH 7.0) were prepared at 1.0 μM, and 0.4 mg/mL, respectively. A negative control (substrate only; no enzyme) stock was made using buffer instead of enzyme to make up the loss in volume. Reactions were prepared in varying concentration gradients to a final volume of 10 μL and were set up in 384-well plates (Greiner Bio-One) in a 1:1 ratio (enzyme: FRET substrate) in duplicate. Each well had a final concentration of 0.500 μM enzyme, and 0.2 mg/mL FRET substrate in 100 μM $ZnCl_2$, 50 mM HEPES pH 7.0 buffer. Prior to enzyme addition, substrates were equilibrated and monitored at 25°C for 30 minutes. After enzyme (and negative control stock) addition, FRET substrates were excited at 430 nm and measured at 518 nm and 590 nm using a SpectraMax M5 plate reader (Molecular Devices). Reads were taken

before and immediately after enzyme addition, as well as at 1, 2, 4, 24, 48, 72, and 96 hours. Plates were kept inside the plate reader between measurements, or wrapped in foil, stored in a drawer, at room temperature.

Raw fluorescence readings were taken at 518 nm and 590 nm per timepoint. Per reaction, the 518:590 nm ratio was determined for the substrate only controls and averaged to calculate the background fluorescence. The relative change in FRET (ΔFRET) was calculated by dividing each sample's respective 518:590 ratio by this background ratio and subtracting 1. Sample values were averaged between duplicates. The resulting heat maps were generated via conditional formatting of cells from black to yellow, with a value range of 0.2 to 1, wherein higher values indicate activity.

2.9.1. Assessment of M60 peptidase kinetic parameters using DynaFit.

Quantifying glycopeptidase kinetic parameters followed similar methods, as described in Section 2.9. Three initial stock solutions were made per set of reactions. Enzyme stocks were prepared so that, once combined at a 1:1 ratio with substrate, reactions had a final concentration of 0.5 μM or 1.0 μM. Similarly, substrate stocks were made in a concentration gradient, with a final concentration of 0.00, 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 2.00, 2.50, 3.00, 4.00, and 6.00 μM. Negative control (substrate only; no enzyme) stock was made using dH_2O instead of enzyme to make up the loss in volume. Prior to enzyme addition, reaction wells containing substrate were equilibrated and monitored at 25°C for 30 minutes. Excitation and emission wavelengths proceeded as described above, with measurements taken every 85 seconds or 10 minutes for 24 hours.

Raw fluorescence readings taken at 518 nm and 590 nm were processed differently than as described in Section 2.9. Per timepoint, each glycopeptidase-treated sample replicate value was subtracted by the average substrate-only value at 518 nm and 590 nm. Then, each substrate-normalized dataset at 518 nm and 590 nm were subtracted against each other (518 nm – 590 nm). The resulting background-corrected datasets were fit by global analysis using DynaFit 4.11.073 software (Kuzmič, 1996, 2009).

2.10. Crystallography.

Crystallization conditions for BcPL35, BcGH154, and BcM60_G_E327Q were initially screened using MCSG-1 to 4 (Anatrace) crystal screen kits and the sitting drop method of vapor diffusion in 96-well plates. Initial screening with BcM60_G_E327Q was done by Dr. Benjamin Pluvinage, while BcGH154 R285A was screened and optimized by Olivia Canil. Respectively, x-ray data collection of BcGH154 and BcGH154 R285A was done by Dr. Andrew Hettle and Dr. Alisdair Boraston.

Optimized BcPL35 native crystals were obtained at 18°C via hanging drop vapor diffusion from protein concentrated to 8.6 g/L mixed at a 1:1 ratio with well solution containing 100 mM MES:NaOH pH 6.5, 300 mM $MgCl_2$, and 12.5% PEG 4000. Similarly, BcGH154 crystals were obtained at 18°C from protein concentrated to 26 g/L mixed at a 1:1 ratio with a crystallization solution containing 220 mM KCl and 24% PEG 3350. BcGH154 R285A at 18 g/L crystallized in 0.25M sodium iodide, 5-13% glycerol, and 8-13% PEG 3350. Lastly, BcM60_G_E327Q crystals were obtained from at 18°C from protein concentrated to 20 g/L mixed at a 1:1 ratio in a well solution containing 0.2 M NH_4OAc, 0.1 M Tris pH 9.0, and 24% PEG 3350.

All crystals were cryo-protected in crystallization solution supplemented with 20% ethylene glycol prior mounting directly in a nitrogen cryo-stream at 100 K for data collection. The cryo-protecting solution for BcGH154 R285A also contained 40 mM chondrosine, while the cryo solution for BcM60_G_E327Q has Core 2-Thr in excess. X-ray diffraction data was collected in-house on a system comprising a Rigaku MM-007HF X-ray generator with a VariMaxTM-HF ArcSec Confocal Optical System and an Oxford Cryostream 800.

All datasets were processed using HKL3000. Structures were solved using AlphaFold2 predicted models for molecular replacement using PhaserMR. All models were manually correct by building in COOT with refinement using REFMAC and/or Phenix. Waters were added in COOT with FINDWATERS and manually checked after refinement. In all datasets, refinement procedures were monitored by flagging 5% of all observations as “free”. Model validation was performed with MOLPROBITY. All data processing and model refinement statistics and PDB ID accession codes are shown in **Table S3.1** and **Table S4.1**.

Chapter 3 : Analysis of chondroitin degradation by components of a *Bacteroides caccae* polysaccharide utilization locus

Bernadette Alvarez[1], Olivia F. Canil[1], Kristin E. Low[2], Andrew G. Hettle[1], D. Wade Abbott[2,3], and Alistair B. Boraston[1].

[1]Department of Biochemistry and Microbiology, University of Victoria, PO Box 1700 STN CSC, Victoria, British Columbia, V8W 2Y2, Canada

[2]Lethbridge Research and Development Centre, Agriculture and Agri-Food Canada, Lethbridge, AB, Canada

[3]Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, AB, Canada

Adapted from: Analysis of chondroitin degradation by components of a Bacteroides caccae polysaccharide utilization locus. *JBC*. 2025 June 7; 10.1016/j.jbc.2025.110354

Contributed to all aspects of research shown here, except for data collection of the BcGH154 native structure (**PDB:** 9O4U; Dr. Andrew Hettle); generation and data collection of BcGH154 complex (**PDF:** 9NWF; Olivia Canil and Dr. Alistair Boraston); and the acquisition and analysis of mass spectrometry data (**Figure 3.5** and **Figure S3.4**) (Dr. Kristin Low and Dr. Wade Abbott).

3.1. Introduction

Bacteria of the phylum Bacteroidota (formerly Bacteroidetes) are well-established in literature as glycan generalists, capable of degrading a complex array of substrates within the distal human gut (Briggs et al., 2021; Brown & Koropatkin, 2020; Klassen et al., 2021; Overbeeke et al., 2022; Rawat et al., 2022). Members are known to arrange their carbohydrate-metabolism-related genes in polysaccharide utilization loci (PULs), wherein genes are co-localized and co-regulated in response to a single glycan, or chemically related group of glycans (Brown & Koropatkin, 2020; D. Ndeh et al., 2020). In brief, PULs contain a complement of cell-surface glycan binding proteins, CAZymes, carbohydrate sensors and transcriptional regulators, as well as at least one sequential pair of genes encoding SusC and SusD homologs (**Figure 1.9**). PULs are a major nutrient acquisition strategy for *Bacteroides* spp. (Cartmell et al., 2018; Feng et al., 2022; Grondin et al., 2017; Tuncil et al., 2017). Given the high prevalence of this genus in the HGM, the PUL system is intrinsically linked to the colonization of nutritional niches and establishment of microbial ecosystems in the gut (Feng et al., 2022; Grondin et al., 2017).

Bacteroides spp. exhibit preferences for different glycans even when presented as a mixture in coculture growth experiments (Tuncil et al., 2017). Some prevalent species of this genus even prioritize glycosaminoglycans (GAGs) degradation over other glycans (D. Ndeh et al., 2020; Tuncil et al., 2017). GAGs are ubiquitously distributed, linear heteropolysaccharides found in the extracellular matrix and on the cellular surfaces of mammalian cells (**Figure 1.5**) (Köwitsch et al., 2018; Mikami & Kitagawa, 2013). The microbial degradation of GAGs generates biologically active microbial metabolites, such as short chain fatty-acids (SCFAs), that influence the host's overall health state (P. Liu et al., 2024; Overbeeke et al., 2022) and modulate the microbial composition of the HGM (**Figure 1.1** and **Figure 1.4**) (Chen et al., 2017; F. Liu et al., 2017; P. Liu et al., 2024). Members of *Bacteroides* genus can catabolize these recalcitrant glycan sources, allowing for the growth enrichment of other SCFA-producing gut microbes, which leads to increased levels of SCFAs that regulate colonic health and act as substrates for key metabolic processes (Overbeeke et al., 2022; Pan et al., 2021; Rawat et al., 2022; Rios-Covian et al., 2017; C. Zhang et al., 2023). Thus, to appreciate how *Bacteroides* spp. interact with the host and other members of the HGM it is important to understand the molecular basis of how *Bacteroides* spp. interacts with GAGs as a nutrient source.

GAGs are separated into four main categories: hyaluronan (HA); heparin and heparan sulfate (HS); keratan sulfate (KS); and lastly, chondroitin sulfate (CS) and dermatan sulfate (DS) (**Figure 1.5**) (Köwitsch et al., 2018). Each group has a characteristic repeating disaccharide unit of a uronic acid (or galactose) paired with an amino sugar. GAG distinctions are based on the monosaccharide composition, sulfation level, and type of glycosidic linkage present. For example, chondroitin (CH) has a repeating disaccharide of glucuronic acid (GlcA) that is β-1,3-linked to N-acetylgalactosamine (GalNAc); the repeating disaccharide is β-1,4-linked. Chondroitin is typically sulfated to form CS. CS-A refers and CS-C are sulfated on the GalNAc residue at carbons 4 and 6, respectively. CS-D and CS-E comprise CS-C that is additionally sulfated at carbons 2 and 4, respectively of the GlcA (Mikami & Kitagawa, 2013).

Bacteroides thetaiotamicron is a widely used model organism to study microbe-host interactions, with extensively studied PULs that illustrate how it can target complex substrates, including GAGs (Cartmell et al., 2017; D. Ndeh et al., 2020; Overbeeke et al., 2022; Rawat et al.,

2022). However, less is known about how other *Bacteroides* spp. target these same substrates. *B. caccae* is a less studied member of the genus, with 60 predicted PULs in its genome (Johnson et al., 1987; Terrapon et al., 2018). Its abundance in the HGM is inextricably tied to diet, wherein it proliferates in high fat and high sugar diets (McNulty et al., 2013). As with other members of its genus, *B. caccae* can degrade numerous glycan sources, including pectin (Sirotek et al., 2004), GAGs (Overbeeke et al., 2022), and mucin (Desai et al., 2016). It is even capable of subsisting entirely on GAGs, including CS-A and CS-C, as its sole carbon source (D. Ndeh et al., 2020), and can enrich the growth of other *Bacteroides* spp. by sharing these glycan degradation products (Overbeeke et al., 2022).

The content of the PUL in the *B. caccae* (strain ATCC 43185) genome that is annotated as "PUL25" is unusual in its gene content (**Figure 1.11**). The predicted functions of the gene products include peptidases and CAZymes, including predicted members of PL35, GH88, GH109, and GH154 (**Figure 3.1A**). None of these proteins have been functionally characterized. Other members of PL35 show lyase activity on the GAGs HA, HS, and CS (**Figure 1.5**) (Lu et al., 2024; L. Wei, Zou, et al., 2024), while members of GH88, GH109, and GH154 families are reported to have unsaturated uronyl hydrolase (Itoh et al., 2006; Jongkees et al., 2014; Maruyama et al., 2009; Nakamichi et al., 2014), hexosaminidase (Teze et al., 2020), and β-glucuronidase (Cartmell et al., 2018) activities. Hence, we hypothesize that the CAZymes encoded by PUL25 targets a glycosaminoglycan. Through structure-function analyses of these enzymes we show that the target substrate is most likely CH (*i.e.,* chondroitin lacking sulfation). Unexpectedly, we found the predicted GH154 to lack glycoside hydrolase activity but rather it has a carbohydrate dehydratase activity that primes saturated non-reducing end β-glucuronic acid residues for cleavage by BcGH88, the unsaturated uronyl hydrolase. We also show that the founding member of the GH154 family, BT_3677 (Hameleers et al., 2024), also has dehydratase activity, leading us to propose that the entire family has this activity and are not glycoside hydrolases. Together, these results give new insight into GAG metabolism by bacteria, particularly members of the gut microbiota, and reclassifies GH154 as a novel family of dehydratases.

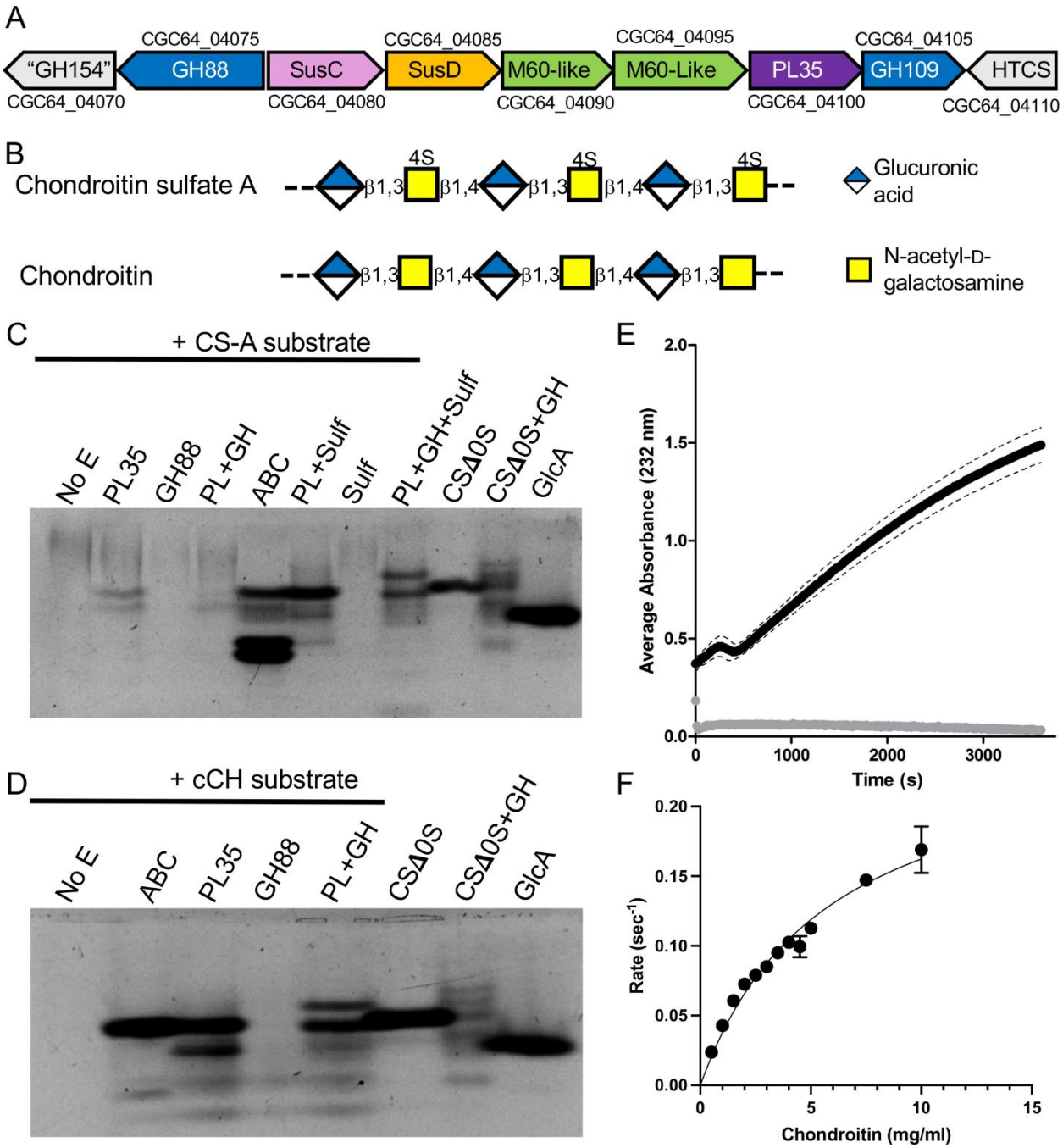


Figure 3.1: Structure of BcPUL25 and activity of the keystone enzyme.

A) Schematic of *B. caccae* PUL25 with initial annotations (centered) and locus tags (above and below). **B)** Schematics of chondroitin sulfate and chondroitin. **C)** Activity analysis by FACE of BcPL35 (PL35 or PL), BcGH88 (GH88 or GH), BcSulf (Sulf), and combinations thereof on chondroitin sulfate, CS-A. ABC lyase is included as a positive control. CSΔ0S, CSΔ0S with BcGH88, and GlcA are included for reference. **D)** Progress curve of cCH (black) and CS-A (grey) cleavage by BcPL35. Dashed lines represent the standard deviation of triplicate samples. **E)** Kinetic analysis of cCH cleavage by BcPL35 (see also **Figure S3.3**). Error bars indicate the standard deviation of triplicate samples.

3.2. Results

3.2.1. BcPL35 from PUL25 prefers chondroitin.

The founding member of the PL35 family was originally characterized to have *endo*-chondroitinase activity (i.e. on unsulfated chondroitin) (Helbert et al., 2019), though additional members have been recently identified with broad GAG-lyase activity (Lu et al., 2024; L. Wei, Zou, et al., 2024). We screened recombinant BcPL35 against CS-A, CS-B (also known as dermatan), CS-C, and HA as possible substrates. Activity was only detected on HA in this initial screen (**Figure S3.1A** and **Figure S3.1C**). Given the chondroitinase activity in the family, we sought to test the activity of BcPL35 on this substrate. BT_3349 from *B. thetaiotamicron* is a previously characterized chondroitin active *endo* 4-O-sulfatase (Ulmer et al., 2014). Its *B. caccae* ortholog is CGC64_04290 (BcSulf) from CAZyme cluster 1. As BcSulf has 92% amino acid sequence identity with BT_3349 and 100% conservation of the active site, we used recombinant BcSulf to determine if this putative chondroitin active *endo* 4-O-sulfatase would potentiate the activity of BcPL35 on CS-A. BcSulf was active on *p*NP-SO_4, confirming its sulfatase activity (**Figure S3.1D**). Treatment of CS-A with BcSulf rendered this otherwise recalcitrant polysaccharide a substrate for BcPL35 (**Figure 3.1B** and **Figure S3.2A**). The product produced by BcPL53 on BcSulf treated CS-A had similar mobility by fluorophore assisted carbohydrate electrophoresis (FACE) as an unsaturated, unsulfated chondroitin disaccharide standard (CSΔ0S) and with one of the three main products of CS-A digestion by *Proteus vulgaris* ABC lyase, which is known to produce primarily unsaturated disaccharide products (Z. Zhang et al., 2009). With this result in hand, we also used chemically desulfated CS-A (denoted as cCH) prepared using a previously described method (Nagasawa et al., 1979). Again, the product produced by BcPL53 had similar mobility by FACE as CSD0S and the main product of CS-A digestion by ABC lyase (**Figure 3.1C** and **Figure S3.2B**).

The cleavage of cCH was also monitored by UV absorption at 232 nm to follow the generation of the 4,5-unsaturated bond typical of polysaccharide lyase activity (**Figure 3.1D**). Consistent with the FACE results, BcPL35 displayed a time dependent increase in product formation on cCH but not CS-A. A kinetic analysis using absorbance to quantify product formation allowed us to determine parameters of 15.6 (± 1.8) min^{-1} for k_{cat} and 5.8 (± 1.1) mg mL^{-1} for K_M (errors represent 95% CI). k_{cat}/K_M was 2.7 ml mg^{-1} min^{-1} (**Figure 3.1E** and **Figure S3.3A)**. However, due to

solubility issues with the substrate we were unable to assay saturating substrate concentrations, therefore, we corroborated the k_{cat}/K_M by three independent determinations of k_{cat}/K_M at low concentrations of cCH (< 1/5 K_M), which gave an average value of 2.8 (± 0.5) ml mg^{-1} min^{-1} (error indicates standard deviation of the independent replicates).

The FACE results suggested that the BcPL35 product was most likely predominantly CSΔ0S. To confirm this, we used mass spectrometry to analyze the products of cCH digestion by BcPL35. This confirmed that the major product from this reaction was CSΔ0S along with relatively lower amounts of unsaturated and unsulfated tetra- and hexasaccharide (**Figure S3.4** and **Table S3.1**). Sulfated oligosaccharide species were not readily detectable, consistent with either complete chemical desulfation of the substrate and/or strict specificity of the enzyme for non-sulfate stretches of polysaccharide.

3.2.2. Structure and mutagenesis of BcPL35.

BcPL35 crystallized in the spacegroup $P2_12_12_1$ and we solved its structure to a resolution of 1.75 Å resolution (**Table S3.1)**. The overall fold is a bi-domain structure of a $(a/a)_6$ toroid fused to an anti-parallel b-sheet domain (**Figure 3.2A**), thus generally similar to other PL families 8, 12, 15, 17, 21, 23 and 39 (Drula et al., 2022). The interface of the two domains is characterized by a deep groove that contains the putative catalytic machinery (**Figure 3.2A**). The groove architecture of the active site that spans the full protein suggests *endo*-recognition of the substrate, which is consistent with the production of oligos with a degree of polymerization from two to six. The closest structural homologs of BcPL35 are GAGase II and GAGase VII, which are *endo*-acting PL35 enzymes from *Spirosoma fluviale* and *Bacteroides intestinalis* DSM 17393, respectively (L. Wei, Cao, et al., 2024). The structures of both GAGase II (**PDB:** 8KHV) and GAGaseVII (**PDB:** 8KHW) have a root mean square deviation (RMSD) of 2.2 Å with BcPL35. The relatively high RMSD values appear to reflect that BcPL35 adopts a conformation with more widely separated N- and C-terminal domains (**Figure 3.2B**). If the domains are separately overlapped with the respective domains in either GAGase II or VI the RMSD values drop to ~1 Å, more accurately reflecting the high structural identity of the proteins. This did suggest, however, that there may be flexibility between the domains of BcPL35, which we interrogated by normal mode analysis using El Nemo (Suhre & Sanejouand, 2004). The primary predicted mode of flexibility involved a hinge

motion at the separation between the domains resulting in ~10 Å of movement between the domains with the crystal structure approximating an intermediate conformation (**Figure 3.2C**). The most closed predicted conformation showed an RMSD of 1.8 Å with both GAGase II and GAGaseVII, respectively.

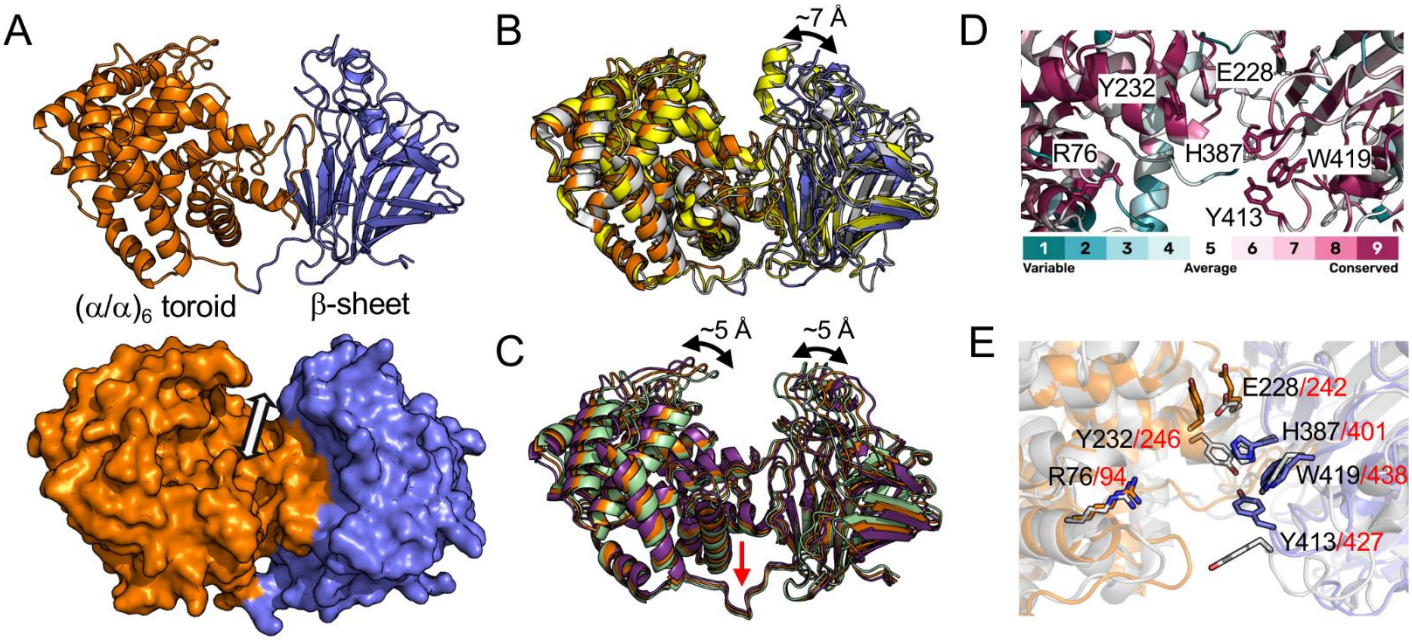


Figure 3.2: Structural analysis of BcPL35.

A) Cartoon and surface representations of the 1.75 Å resolution structure of BcPL35. The two domains are colored orange and blue. **B)** overlap of BcPL35 (orange and blue) with GAGase II (grey, **PDB:** 8KHV) and GAGaseVII (yellow, **PDB:** 8KHW). **C)** Cartoon representation of BcPL35 conformations determined by normal mode analysis with El Nemo comparted with the crystal structure (orange). The predicted conformational extremes are shown in green and purple with the hinge point indicated by a red arrow. **D)** Consurf analysis of BcPL35. Residue conservation is color ramped teal to purple based on 150 randomly sampled homologs out of ~1700 found with amino acid sequence identities between 25%-95%. Residues chosen for mutagenesis are shown as sticks. **E)** Conservation of mutated residues with GAGase II (grey, **PDB:** 8KHV).

An analysis of conserved residues in the putative BcPL35 active site cleft allowed us to select six highly conserved residues for mutation to alanine: R76, E228, Y232, H387, Y413, and W419 (**Figure 3.2D**). This includes conservation of the proposed catalytic tyrosine and histidine in GAGase II (Y246, H401) (**Figure 3.2D**) and GAGaseVII (Y241, H397), which are Y232 and H387 in BcPL35. Using cCH as a substrate, the activity for the R76A, E228A, H387A, and Y413A mutants was below detectable levels. The Y232A mutant showed a roughly 3-fold reduction in

activity relative to wild-type with a k_{cat}/K_M of 0.8 ± 0.1 ml mg^{-1} min^{-1} (**Figure S3.3A**). The W419A mutant had slightly reduced activity of ~60% wild-type with a k_{cat}/K_M of 1.7 ± 0.1 ml mg^{-1} min^{-1} (**Figure S3.3A**). This largely agrees with the dramatically reduced or absent activity of analogous mutants of GAGase II (L. Wei, Cao, et al., 2024). The exception was W419 in BcPL35, which had minimal impact on activity whereas the corresponding W438A mutation in GAGase II eliminated its activity. This may reflect the use of HA as a substrate for the GAGase II assays vs cCH in our BcPL35 assays and potential differences in the modes of substrate recognition of these polysaccharides.

3.2.3. BcGH88 works in tandem with BcPL35.

The *exo*-uronyl hydrolase activity of the GH88 family targets Δ4,5-unsaturated glucuronyl or galacturonyl residues and through their unusual catalytic mechanism release the saturated 5-keto-4-deoxyuronate product (Itoh et al., 2006). These enzymes typically act after polysaccharide lyase activity to hydrolyze the newly formed unsaturated monosaccharide at the non-reducing end (Itoh et al., 2006; Maruyama et al., 2009; Nakamichi et al., 2011). We therefore tested BcGH88 on the products generated by BcPL35 substrate cleavage. Reactions of CS-A with BcPL35 and BcSulf or cCH with just BcPL35 were supplemented with BcGH88 and analyzed by FACE. In both cases, the major products formed by BcPL35 were completely converted to two products by BcGH88, indicating the activity of BcGH88 on products produced by BcPL35 (**Figure 3.1B** and **Figure 3.1C**). We followed up on this by using absorbance at 232 nm to track the formation of BcPL35 products followed by the addition of BcGH88 and the loss of absorbance owing to the removal of the terminal Δ4,5-unsaturated glucuronyl sugar and conversion to a non-absorbing species. As anticipated, after BcPL35 reactions were allowed to come to equilibrium, the addition of BcGH88 resulted in the rapid depletion of absorbance at 232 nm (**Figure 3.3A**). We further pursued this by directly treating the disaccharide CSΔ0S, which has a Δ4,5-unsaturated glucuronyl residue, with BcGH88 and monitoring the reaction progress by loss of absorbance indicating conversion of the Δ4,5-unsaturated glucuronyl residue (**Figure S3.5A**). To provide additional evidence that BcGH88 is producing the expected 5-keto-4-deoxyuronate product we used a coupled assay utilizing KduI (5-keto-4-deoxyuronate isomerase) and KduU (2-dehydro-3-deoxy-*D*-gluconate dehydratase) (Hobbs et al., 2016). Qualitatively, the oxidation of NADH observed when CSΔ0S is cleaved by BcGH88 in the presence of KduI and KduU indicates production of the expected 5-keto-4-

deoxyuronate product by BcGH88 (**Figure S3.5B**). Notably, 5-keto-4-deoxyuronate has a free aldehyde and keto group, both of which can be labeled by the reductive amination used to introduce the AMAC fluorophore for FACE. This likely explains the multiple labeled species observed by FACE when BcGH88 releases 5-keto-4-deoxyuronate (**Figure 3.1B** and **Figure 3.1C**).

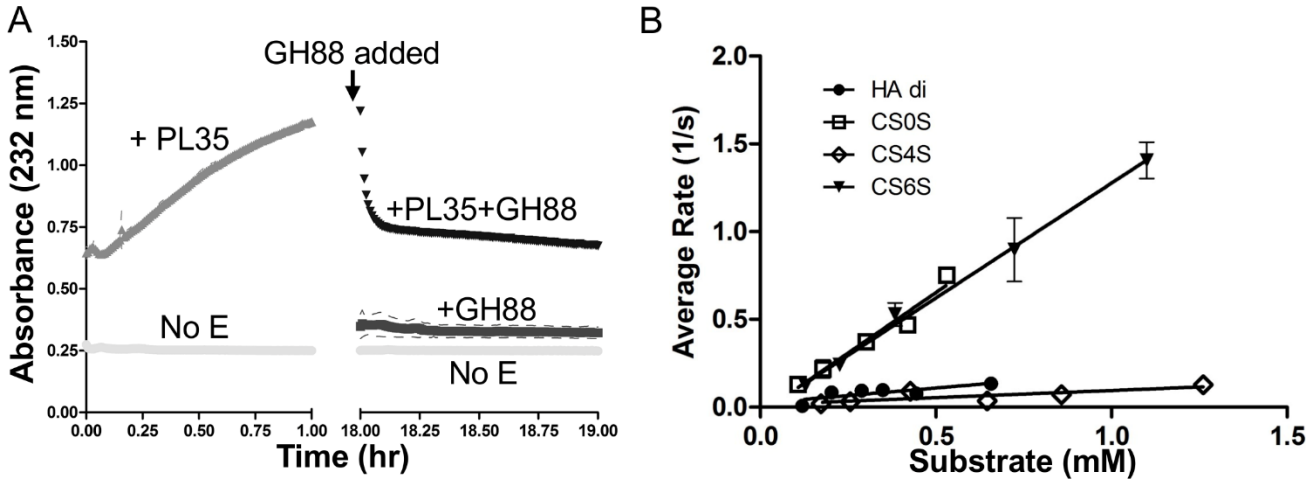


Figure 3.3: Activity of BcGH88.

A) BcGH88 works on the products of BcPL35. cCH was treated with BcPL35 until equilibrium was reached, as judged by a plateau in absorbance. The reaction was then spiked with BcGH88. A substrate-only control was included; this was also spiked with BcGH88 at the same time. **B)** Kinetic analysis of BcGH88 on different substrates, as indicated (see also **Figure S3.5C**). Error bars indicate the standard deviation of triplicate samples.

Using the assay for loss of absorbance at 232 nm, we assessed the kinetics of BcGH88 activity on a variety of Δ4,5-unsaturated disaccharides (**Figure 3.3B** and **Figure S3.5C).** The activity on Δ4,5-unsaturated heparin disaccharide (no sulfation) was too low to quantify. BcGH88 had the best activity on CSΔ0S and CSΔ6S with k_{cat}/K_M values of 1.38 ± 0.11 s^{-1} mM^{-1} and $1.31 \pm 0.11 s^{-1}$ mM^{-1}, respectively, and therefore no significant preference for either. The enzyme had roughly 10-fold lower efficiency on ΔHA and CSΔ4S with k_{cat}/K_M values of 0.17 ± 0.04 s^{-1} mM^{-1} and 0.08 ± 0.02 s^{-1} mM^{-1}, respectively.

3.2.4. BcGH109 is an α/β-N-acetylgalactosaminidase.

Recombinant BcGH109 was screened against a panel of aryl-glycosides to identify its substrate repertoire. Significant activity was only observed on *p*NP-α-GalNAc and *p*NP-β-GalNAc (**Figure 3.4**), which is consistent with all other prior examples of the family being classified as *N*-

acetylgalactosaminidases, often with activity on both α- and β-*N*-acetylgalactosamine moieties (Q. P. Liu et al., 2007; D. A. Ndeh et al., 2024; Sulzenbacher et al., 2010; Teze et al., 2020). This suggested that the role of BcGH109 may be to cleave the terminal β-1,3-linked GalNAc left after BcGH88 activity on tetra- or hexasaccharide products of BcPL35. We attempted to examine this using FACE to separate reactions of CS-A with BcPL35, BcSulf, and BcGH88 or cCH with BcPL35 and BcGH88 that were supplemented with BcGH109. However, this was inconclusive, likely owing to the relatively small amounts of longer oligosaccharides produced by BcPL35.

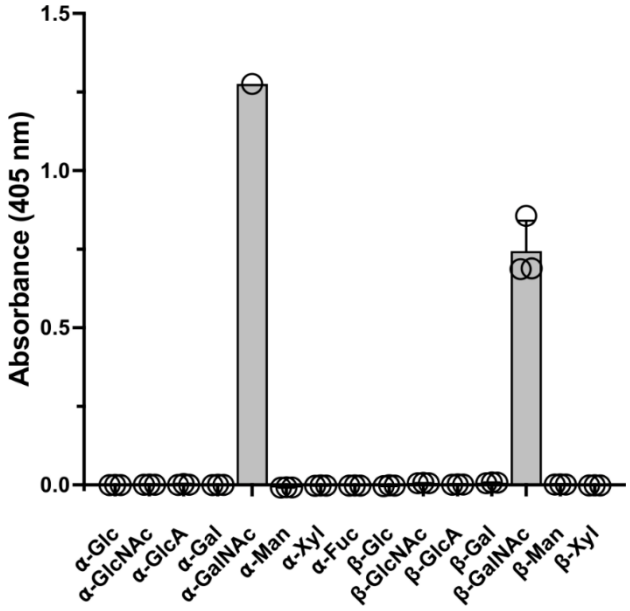


Figure 3.4: Activity of BcGH109 on 4-nitrophenyl glycosides

3.2.5. Identification of a novel dehydratase.

To identify the activity of a recombinant version of the putative GH154 protein, we screened it against a panel of aryl-glycosides on the premise it would have β-glucuronidase or β-galactosidase activity like other reported members of the family (Cartmell et al., 2018; Hameleers et al., 2024). IMAC purified preparations of the protein produced from *E. coli* BL21 did display glycoside hydrolase activity on some aryl-glycosides, most notably pNP-β-D-galactopyranoside (**Figure S3.6A**). However, the activities were inconsistent between multiple individual preparations, leading us to believe that the activity was from contaminating proteins, such as LacZ. We therefore used a purification protocol that used the *E. coli* Tuner strain, in which LacZ is absent, and an additional size exclusion chromatography purification step after IMAC. We were

unable to find any activity of these preparations on aryl glycosides (**Figure S3.6A**). Glycoside hydrolases occasionally have strict specificity for the sugar residue preceding the hydrolyzed glycosidic bond (*i.e.,* in the -1 subsite or so-called "aglycon" specificity). We then used a linked assay to detect glucuronic acid released from chondrosine (GlcAβ-1,3-GalN), a saturated chondroitin disaccharide mimic, after treatment with the enzyme but failed to detect the monosaccharide (**Figure S3.6B**).

Seeking an alternative hypothesis for the activity of the putative GH154 protein, we noted that analysis of the interaction network of the GH154 family by STRING (Szklarczyk et al., 2023) revealed frequent co-occurrence and co-localization of GH154 encoding genes nearby genes encoding unsaturated uronyl hydrolases in the GH88 and GH105 families. A prior study revealed the role of a unique dehydratase, P29_PDnc, in generating a new Δ4,5-unsaturated non-reducing end, thus creating a substrate for an accompanying GH105 in ulvan depolymerization (Bäumgen et al., 2021). Though GH154 is not related at the primary structure level to P29_PDnc, we hypothesized that our protein may be a dehydratase with a similar activity relationship to unsaturated uronyl hydrolases. We tested the activity of our protein on chondrosine by monitoring the absorbance at 232 nm to detect the formation of the predicted Δ4,5 double bond. The absorbance increased in a time dependent manner, with the rate of increase dependent upon the chondrosine concentration (**Figure 3.5A**). The efficiency of BcGH154 catalyzed generation of the absorbing species was determined to have a k_{cat}/K_M of 0.54 (± 0.02) min^{-1} mM^{-1} for chondrosine (**Figure 3.5B** and **Figure S3.7A**). Reactions of BcGH154 with chondrosine were allowed to go to completion then spiked with either buffer or BcGH88. Only reactions spiked with BcGH88 showed a rapid decrease in absorbance, consistent with cleavage of the glycosidic bond by BcGH88 and indicating formation of a Δ4,5-unsaturated glucuronyl residue on chondrosine (**Figure 3.5A**). To confirm this, we used LC-ESI MS to examine the structure of the reaction product. The mass of the reaction product, 336 Da (**Figure 3.5C**) (**Table S3.2**), was consistent with formation of the dehydrated product while MS/MS analysis supported the structural assignment of the product as containing the Δ4,5 bond of dehydrated chondrosine (**Figure 3.5D**). This protein, which was previously classified as a GH154, shows properties most consistent with dehydratase activity. Thus, we have renamed it BcGDH (GDH for glucuronic acid dehydratase).

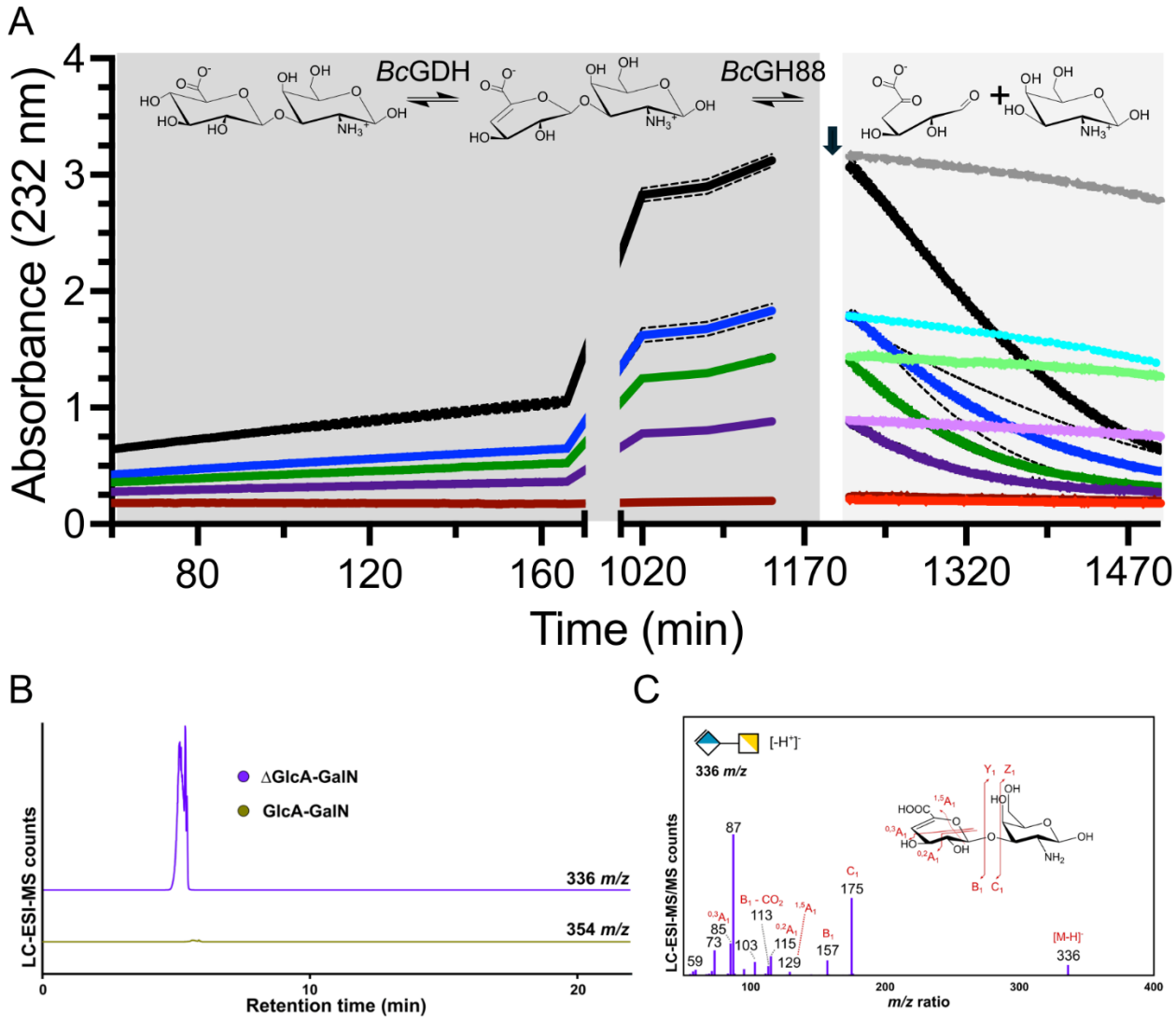


Figure 3.5: Activity of BcGDH (GH154) on chondrosine.

A) BcGDH is active on chondrosine and BcGH88 is active on the products. Chondrosine at 0 mM (dark red), 0.1 mM (dark purple), 0.5 mM (dark green), 1.5 mM (dark blue), and 3.0 mM (black) was treated with BcGDH until equilibrium was reached, as judged by a plateau in absorbance. Reactions were in triplicate and dashed lines indicate the standard deviation of these readings. Two of each triplicate was then spiked with BcGH88 (indicated by an arrow), this data is shown in the corresponding dark color. Buffer was added to one of the triplicate samples, this data is shown in a light version of the corresponding color. The proposed reactions and structures are shown above. **B)** Enzyme reaction products of chondrosine with BcGDH was analyzed by LC-ESI-MS with extracted ion chromatograms are shown for ions of interest with ion counts scaled relative to the most intense peak. **C)** ESI-MS/MS with HCD was performed to identify the extracted ion for the primary catalytic product in each.

Based on this result, we also tested the possible dehydratase activity of BT_3677, the founding member of GH154. BT_3677 did not display activity on chondrosine but had activity on 1-*O*-methyl-β-D-glucuronate with a k_{cat}/K_M of 0.35 (± 0.01) min^{-1} mM^{-1} (**Figure S3.7A** and **Figure S3.7B**). BcGDH did not have activity on 1-*O*-methyl-β-D-glucuronate, suggesting this family of proteins may have different specificity for the monosaccharide and/or linkage preceding the terminal glucuronic acid residue.

3.2.6. Structural analysis of BcGDH.

We solved the X-ray crystal structure of BcGDH and refined it in the spacegroup $P2_1$ to 2.25 Å resolution. The resulting structure comprised a well-organized tetramer in the asymmetric unit that is predicted by PISA analysis to be stable in solution (**Figure 3.6A**). We also obtained slightly lower resolution and lower quality diffraction data sets of BcGDH in P1, two different $P2_12_12_1$ crystal forms, and a second $P2_1$ crystal form. We chose not to refine these but note that the same tetramer was present in all acquired crystal forms, supporting this as a stable quaternary structure.

The structure shows a pronounced slot on the surface of each monomer that is slightly tilted towards the center of the tetramer (**Figure 3.6A, circled region) (Table S3.1)**. An analysis of the surface residues of BcGDH for conservation in the family revealed the residues in the slot to be highly conserved and likely comprising the active site **(Figure 3.6B** and **Figure 3.6C)**. Residues on surfaces making up the quaternary structure interfaces were not conserved, possibly indicating different or lacking quaternary structures amongst other family members. Based on this analysis, we chose to make alanine mutations of ten residues in the putative active site (E74A, R78A, N176A, N177A, W178A, F181A, Y229A, R285A, Y289A, Y354A) assay them for activity. None of the mutants had activity when using up to 10 mM chondrosine, despite displaying thermal stability similar to the unmutated enzyme (**Table S3.3**). We note, however, that chondrosine is a relatively poor substrate, likely by virtue of the substitution of what should be an acetamido group on *N*-acetylgalactosamine for the amino group on galactosamine. Any small perturbations in the enzyme substrate interaction that even subtly decrease activity may render it below levels of reliable detection.

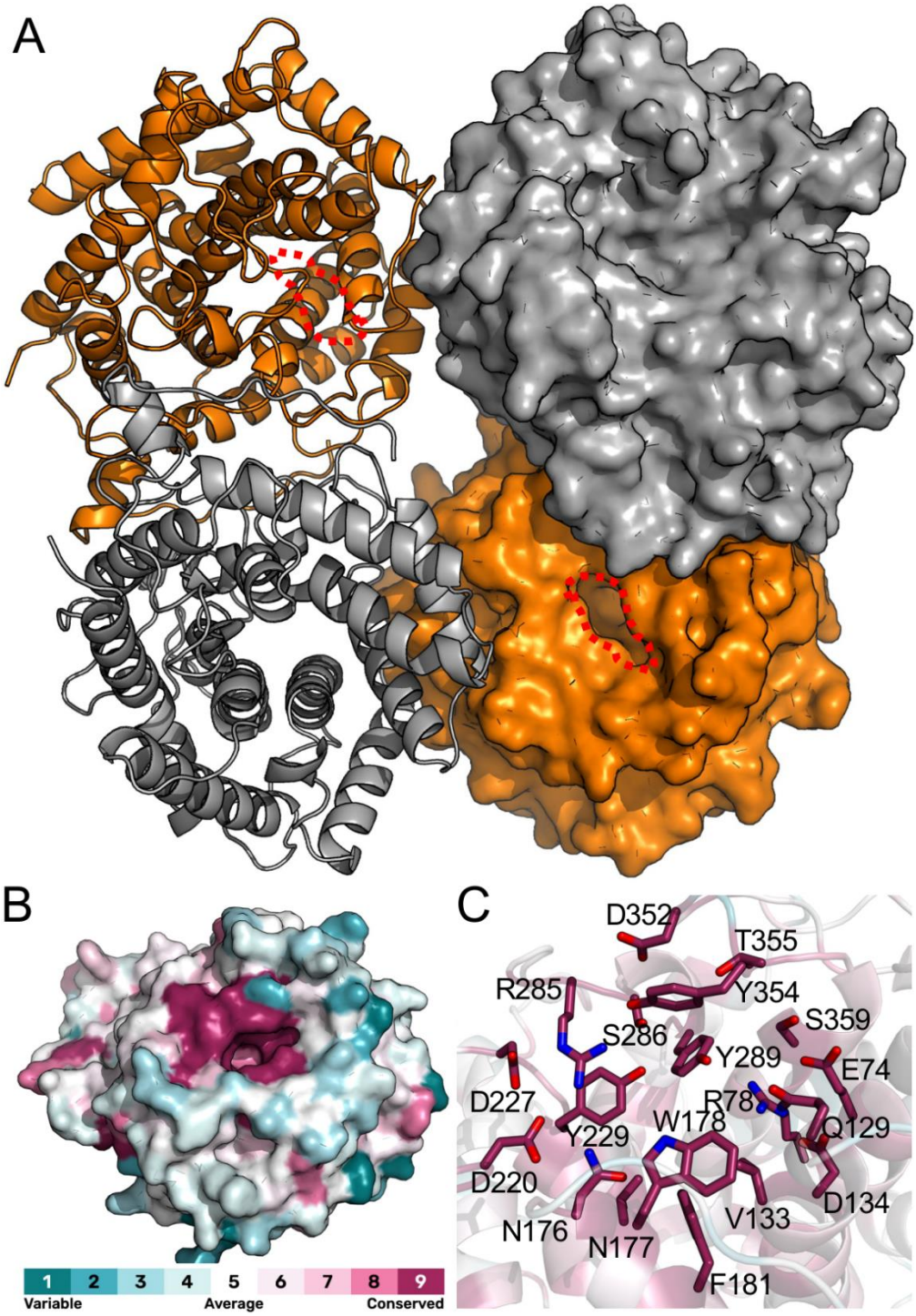


Figure 3.6: Structural analysis of BcGDH.

A) Structure of the BcGDH tetramer determined at 1.75 Å resolution. Two of the monomers are shown as cartoons while the other two are shown in surface representation. The active site is indicated with a red dashed line. **B)** Surface representation of a BcGDH monomer colored by Consurf analysis. Residue conservation is color ramped teal to purple based on 150 randomly sampled homologs out of ~1800 found with amino acid sequence identities between 35%-95%. **C)** Active site conservation, colored as in panel B, with conserved residues shown as sticks.

To trap a substrate complex, we crystallized the inactive R285A mutant and soaked crystals in chondrosine **(Table S3.1)**. The structure comprising a BcGDH tetramer in the asymmetric unit was refined in the spacegroup $P2_12_12_1$ to 2.6 Å resolution. Electron density in the active site of one monomer was clear and allowed modelling of the disaccharide (**Figure 3.7A** and **Figure 3.7B**). A consortium of amino acid side chains in the active site, consistent with those predicted by conservation, were involved in interactions with the substrate (**Figure 3.7C**). In this case, a tyrosine residue that is not conserved in the family is donated by a separate monomer in the tetramer and may help define the opening into the active site (**Figure 3.7C**). The loop containing Y175 was modelled in two conformations. In one conformation the tyrosine sidechain is oriented out of the active site and makes interactions with another monomer in the tetramer (**Figure 3.7C**). This is the conformation observed in the unliganded structure. The other conformation is one where the sidechain of Y175 moves over the active site to interact with the galactosamine residue (**Figure 3.7C**).

The structure of trapped substrate on the inactive BcGDH R285A mutant overlaid with the unliganded structure to estimate the position of R285 provides unique insight into the mechanism of this enzyme. Y289 is positioned in proximity to O4 of the glucuronic acid and may act as a catalytic acid (**Figure 3.7D**). The catalytic base is slightly more ambiguous as Y229 or R285 may perform the role of abstracting a proton from C5 of the glucuronic acid (**Figure 3.7D**). However, Y229 is nearer the C5 proton and thus we propose is more likely to perform the role. The role of R289, a mutant of which was completely inactive, may be two-fold. First, the proximity and positive charge on the guanidino group may assist with keeping the hydroxyl of Y229 in a deprotonated state for a role as catalytic base. Second, there are no other obvious candidates to neutralize the negative charge on the glucuronic acid carboxylate, but R285 is in sufficient proximity to this group to assist in this respect, though it is not in the typical geometry relative to the substrate carboxylate. Taken together we proposed a two-step lyase catalytic mechanism of dehydration that by using two catalytic tyrosine sidechains resembles the alginate lyase mechanism proposed to be used by Alg17C from *Saccharophagus degradans* (**Figure 3.7E**) (H. T. Kim et al., 2012; Park et al., 2014).

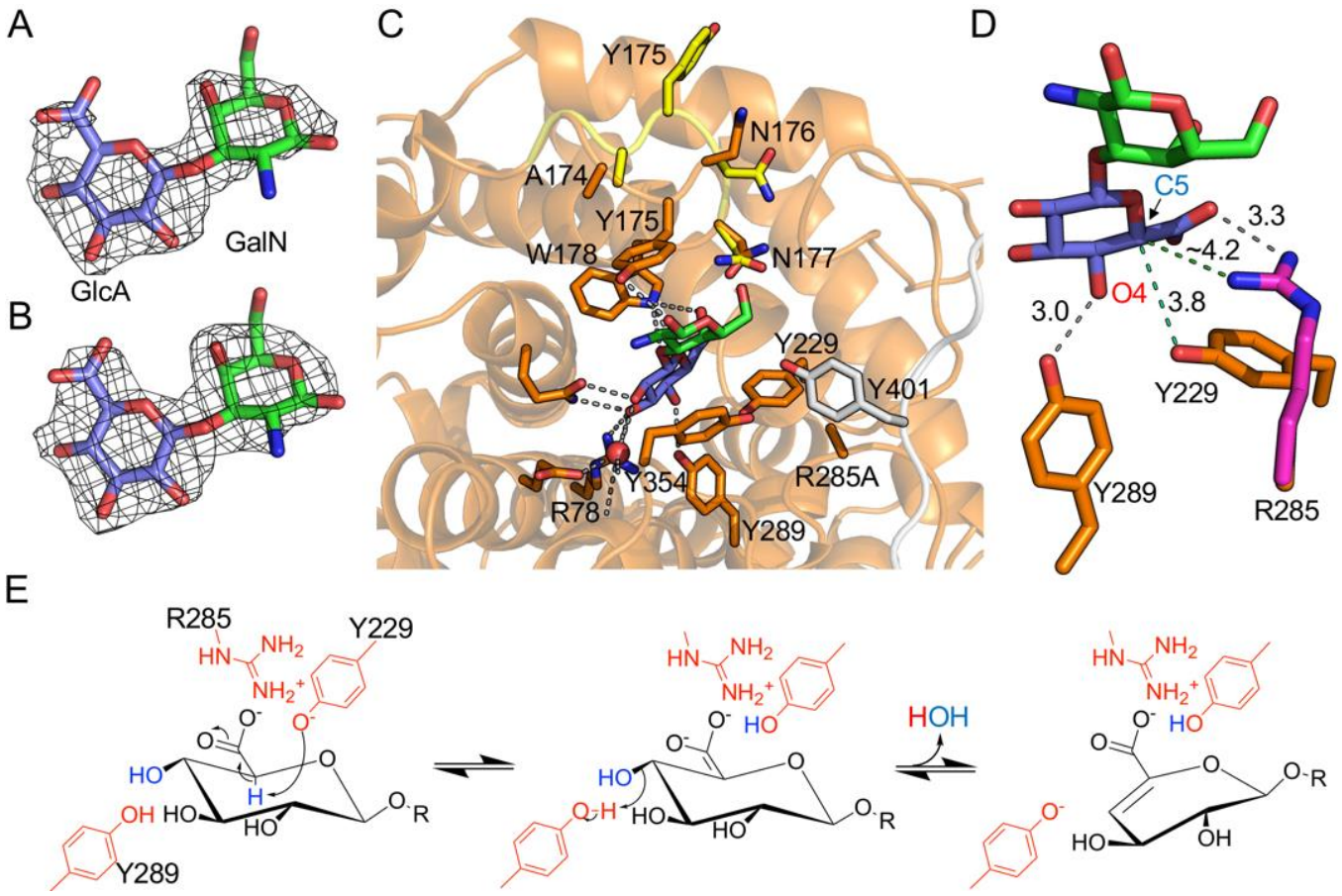


Figure 3.7: Structure of BcGDH R285A mutant in complex with chondrosine.

A) Difference electron density map (F_o-F_c) contoured at 2.6σ determined by refinement in the absence of the ligand coordinates. **B)** Electron density map ($2F_o$-F_c) contoured at 0.9σ determined by refinement with the ligand coordinates. **C)** Interactions within the active site. The mobile loop is shown in both conformations with the yellow-colored loop indicating the disengaged conformation. The grey loop containing Y401 is donated by a neighbouring monomer in the tetramer. **D)** Proposed key catalytic residues and distances in the active site. R285 (magenta) was placed via overlap with the unliganded structure. **E)** Proposed catalytic mechanism.

3.3. Discussion

GAGs are important molecules with profound biological impact (Köwitsch et al., 2018; Rawat et al., 2022). Except for HA, GAGs are generally found attached to a protein core, forming proteoglycans. In the human gut, altered abundance of specific proteoglycans and surface-bound GAGs play critical roles in inflammatory bowel disorders (Overbeeke et al., 2022; Rawat et al., 2022). *Bacteroides* spp., including *B. caccae,* can break down numerous glycan sources (Desai et al., 2016), including GAGs, which it can use as a sole carbon source (D. Ndeh et al., 2020). Given the prevalence of the genus *Bacteroides* in the HGM (Feng et al., 2022; Grondin et al., 2017); its preference for GAGs as a nutrient source (D. Ndeh et al., 2020; Tuncil et al., 2017); and its ability to generate SCFAs from GAGs (Desai et al., 2016; Pan et al., 2021; C. Zhang et al., 2023); it is important to understand the molecular basis of how *Bacteroides* spp. targets GAGs for catalysis. Through biochemical analysis of the CAZymes from *Bacteroides caccae* ATCC 43185 (new assembly), we have uncovered support for an unusual biochemical cycle to depolymerize the GAG chondroitin sulfate.

The PL35 family was originally annotated as an *endo*-chondroitinase family, as its founding member from *Victivallis vadensis* (VvPL35) showed this activity (Helbert et al., 2019). Like VvPL35, BcPL35 also displayed a preference for chondroitin, in this case produced by enzymatic (*i.e.,* by BcSulf) or chemical desulfation. BcPL35 was also active on HA, which lacks sulfation and only differs from chondroitin by the substitution of GalNAc by GlcNAc. However, this appeared to make a relatively poor substrate, as judged by attempts to monitor reaction progress by UV absorbance, while the kinetic parameters on chemically desulfated chondroitin were similar to those observed for more recently PL35 enzymes on their preferred substrates (L. Wei, Zou, et al., 2024). Consistent with other members of the GH88 family, BcGH88 displayed *exo*-uronyl hydrolase activity. Furthermore, the observation that BcGH88 had activity on the products of BcPL35 but had poor activity on unsaturated disaccharides other than those derived from chondroitin, including the unsaturated disaccharide from HA, support chondroitin as main component of the substrate targeted by BcPUL25.

BcPL35 produces mainly unsaturated disaccharide and tetrasaccharide from chondroitin, which, *in vivo,* would most likely be produced by BcPL35 after enzymatic desulfation (**Figure**

S3.4). However, it is possible the enzyme targets regions of chondroitin sulfate that are naturally low in sulfation. The BcGH88 readily hydrolyzes the disaccharide product to the components 5-keto-4-deoxyuronate and GalNAc, which can feed into additional metabolic pathways (D. A. Ndeh et al., 2024; Wargaki et al., 2012). The activity of BcGH88 on the tetrasaccharide would result in a GalNAcβ-1,3-GlcAβ-1,4-GalNAc trisaccharide. We postulate that the demonstrated β-*N*-acetylgalactosaminidase activity BcGH109 would translate to removal of the terminal β-1,4-linked GalNAc, thus yielding the saturated chondroitin disaccharide repeating unit of GlcAβ-1,3-GalNAc. However, we were unable to confirm this due to the lack of substrate, or even substrate mimics.

The first gene in BcPUL25 encodes a protein that, by its high sequence identity to members of glycoside hydrolase family 154, is annotated as a putative β-glucuronidase (Cartmell et al., 2018). This hypothesised activity was entirely consistent with the potential need to hydrolyze terminal saturated β-1,4-linked GlcA residues, as suggested would be revealed in the reactions above, but we were unable to find any glycoside hydrolase activity for the recombinant protein. Instead, the evidence using chondrosine as a mimic of the natural substrate clearly supports that this enzyme, which we refer to as BcGDH, is a β-D-glucuronic acid dehydratase that converts the terminal β-1,4-linked GlcA into a terminal Δ4,5-unsaturated uronic and, therefore, renders the sugar a substrate for BcGH88 (**Figure 3.5A**). This would close the loop and allow the cyclical pathway as proposed to depolymerize the BcPL35 cleavage products of any length, or indeed any chondroitin fragment with either a non-reducing end GalNAc or GlcA, with the latter being either saturated or unsaturated (**Figure 3.8**).

The two other characterized members of the GH154 family, BT3677 and BD-β-Gal, were reported to have *exo*-β-1,6-glucuronidase and *exo*-β-1,6-galactosidase activity, respectively (Cartmell et al., 2018; Hameleers et al., 2024). BcGDH has 22% and 50% amino acid sequence identity with BT3677 and BD-β-Gal, respectively. The uncomplexed structure of BD-β-Gal has been determined to 1.76 Å resolution (**PDB:** 8OI4). It displays the same tetrameric organization as BcGDH while monomers of each protein have a root mean square deviation of < 0.8 Å (**Figure S3.8**). The residues in BcGHD that interact with the terminal GlcA are highly conserved across the family (**Figure 3.6B** and **Figure 3.6C**) and essentially invariant with BD-β-Gal and BT3677,

as judged by overlays with the crystal structure of BD-β-Gal and an AlphaFold 3 model of BT3677 (**Figure S3.8B** and **Figure S3.8C**). None of the structures display an active site architecture consistent with typical glycoside hydrolase activity (*i.e.*, potential acid/base or nucleophilic side chains in appropriate proximity to the substrate glycosidic bond). This strongly supports the contention that the "GH154" family is actually a family of carbohydrate dehydratases. This is additionally substantiated by the observation that BT3677 also displays dehydratase activity on 1-*O*-methyl-β-D-glucuronate. The predicted structure of BT3677, however, reveals a second domain that contributes sidechains to the vicinity of the active site that are absent in BcGDH (**Figure S3.8C**). Likewise, BD-β-Gal has a loop that contributes additional sidechains on the rim of the active site and oriented into the catalytic pocket (**Figure S3.8B**). Such accessorizing of the entrance to the active site may help dictate substrate selectivity. For example, BT3677 was active on 1-*O*-methyl-β-D-glucuronate but not active on chondrosine, while BcGDH showed a reversed pattern of selectivity.

The pathway of chondroitin degradation, therefore, utilizes an unusual cycle that pairs a dehydratase with an *exo*-uronyl hydrolase, in this case a GH88, to remove a non-reducing end uronate monosaccharide (**Figure 3.8**). This has only been observed before in ulvan degradation where P29_PDnc, a dehydratase that is unrelated to BcGDH at the amino acid sequence level, is paired with a GH105 *exo*-uronyl hydrolase (Bäumgen et al., 2021). The protein family to which BcGDH belongs contains over 6000 members (Drula et al., 2022), suggesting this model of uronate sugar removal may be widespread. Indeed, genes encoding these dehydratases frequently co-occur and co-localize with genes encoding GH88 or GH105 *exo*-uronyl hydrolases. In the case of BcPUL25, the pairing of chondroitin specific BcPL35 with BcGH88, BcGDH, and BcGH109 creates a logical cascade to completely depolymerize chondroitin and bypasses the need for a dedicated β-glucuronidase. BcPUL25, however, lacks a gene encoding a chondroitin sulfate specific sulfatase while it somewhat cryptically contains genes encoding M60-like proteins, which are putative *O*-glycopeptidases (Noach et al., 2017). It is possible that BcPUL25 cooperates with the enzymes produced from other co-transcriptionally controlled genes, with the gene encoding BcSulf (CGC64_04290) being a logical candidate that would render at least some species of chondroitin sulfate (*e.g.*, CS-A) a substrate for BcPUL25. However, the presence of putative *O*-glycopeptidases generally suggests a more complex target for BcPUL25. Chondroitin sulfate is

typically a component of proteoglycans where the chondroitin sulfate is covalently attached a protein backbone, such as in the lecticans, of which aggrecan from cartilaginous tissue is a prominent and abundant member (Kiani et al., 2002; Koch et al., 2020). In addition to chondroitin sulfate modifications, the protein backbone of aggrecan also bears mucin-like *O*-glycans (Kiani et al., 2002), which may be the target of the putative *O*-glycopeptidases. Thus, the unique GAG degradation pathway encoded by BcPUL may be coupled with a greater capacity to target proteoglycans, again invoking the concept of the incredible metabolic diversity of the HGM.

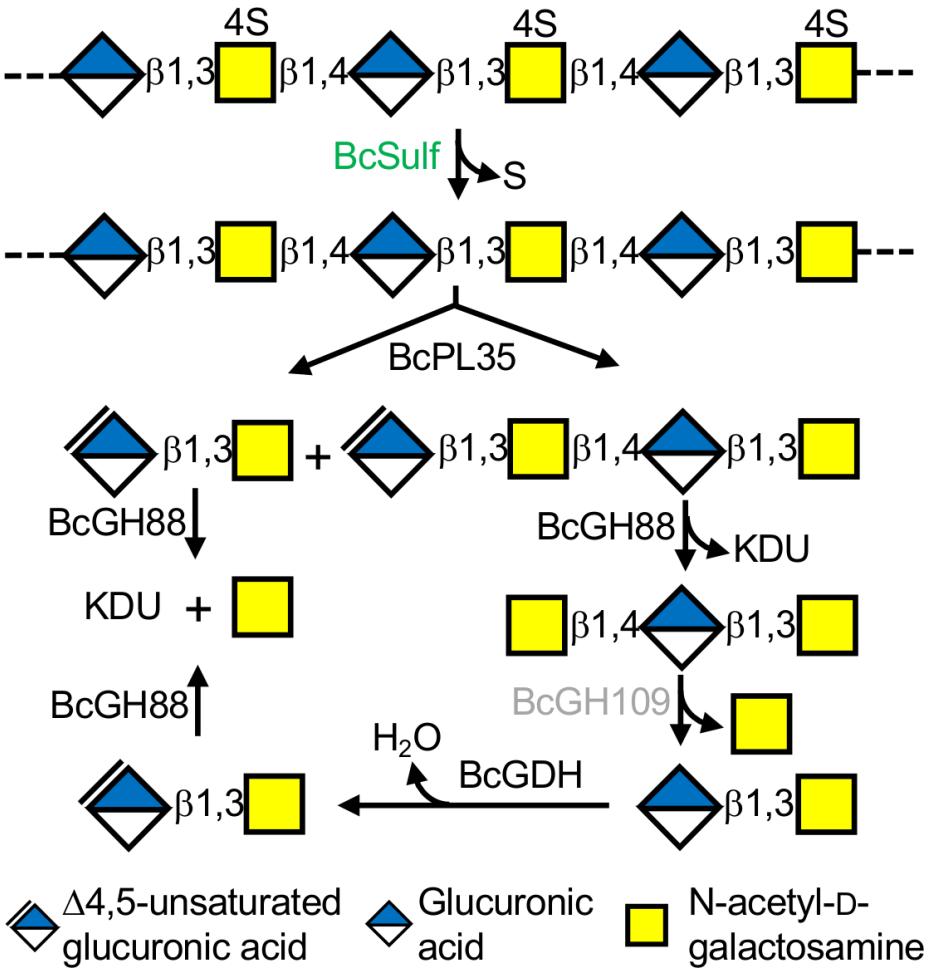


Figure 3.8: Schematic of the chondroitin depolymerization pathway.

Pathway begins with the desulfation of CS-A into chondroitin. Steps not encoded by BcPUL25 are shown in green. BcGH109 has confirmed β-*N*-acetylgalactosaminindase activity but activity on chondroitin or a mimic has not been demonstrated and so it is colored grey.

3.4. Supplementary Figures and Tables

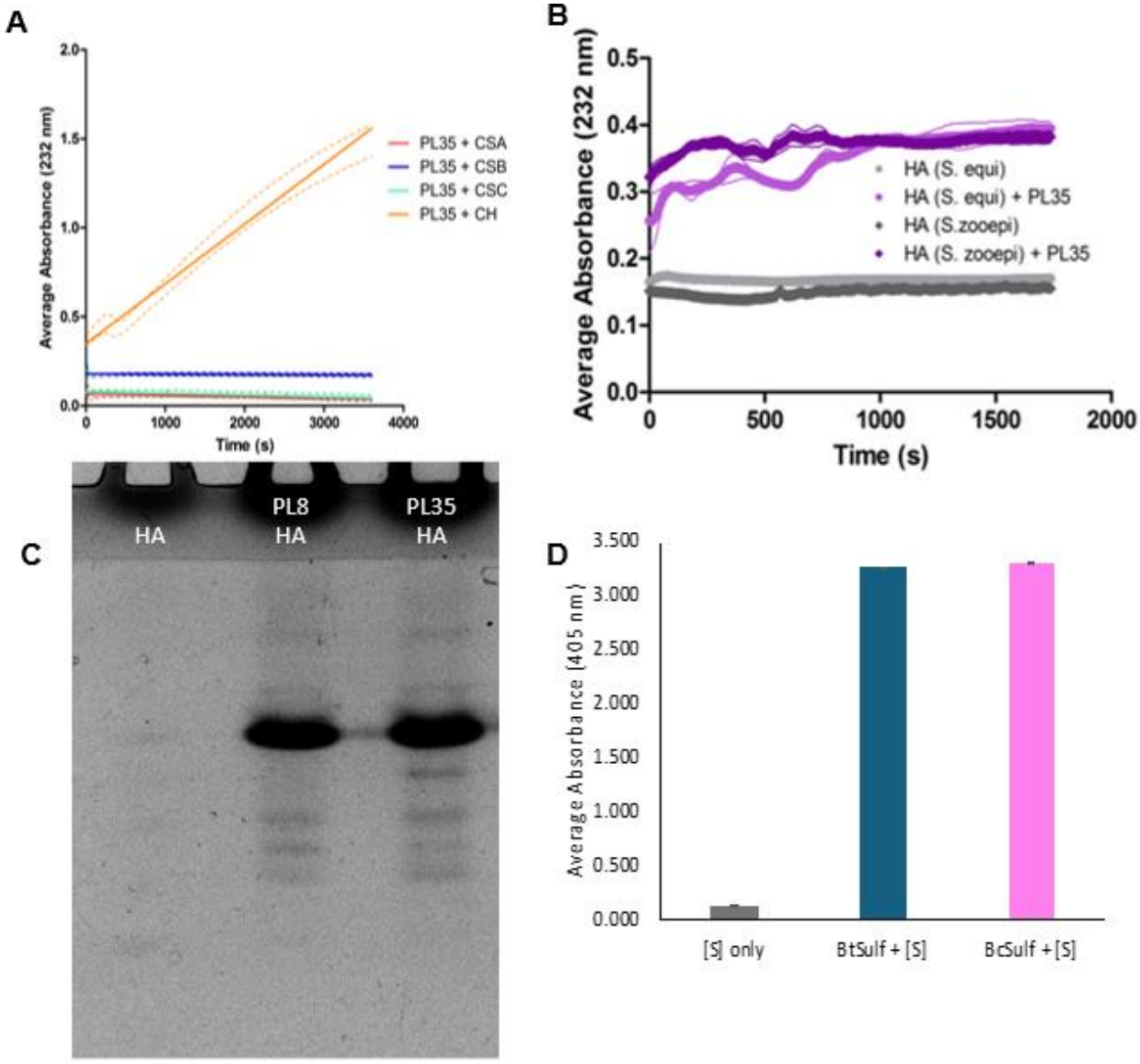


Figure S3.1: Screening BcPL35 activity.

A) Tracking formation of unsaturated products by UV absorbance. CH refers to cCH, which was included as a positive control. **B)** As in panel A but tested two forms of hyaluronic acid. **C)** FACE analysis of HA degradation products by BcPL35 (PL35) and the *Streptococcus pnuemoniae* hyaluronate lyase (PL8) as a positive control. **D)** Activity of sulfatases on *p*NP-sulfate.

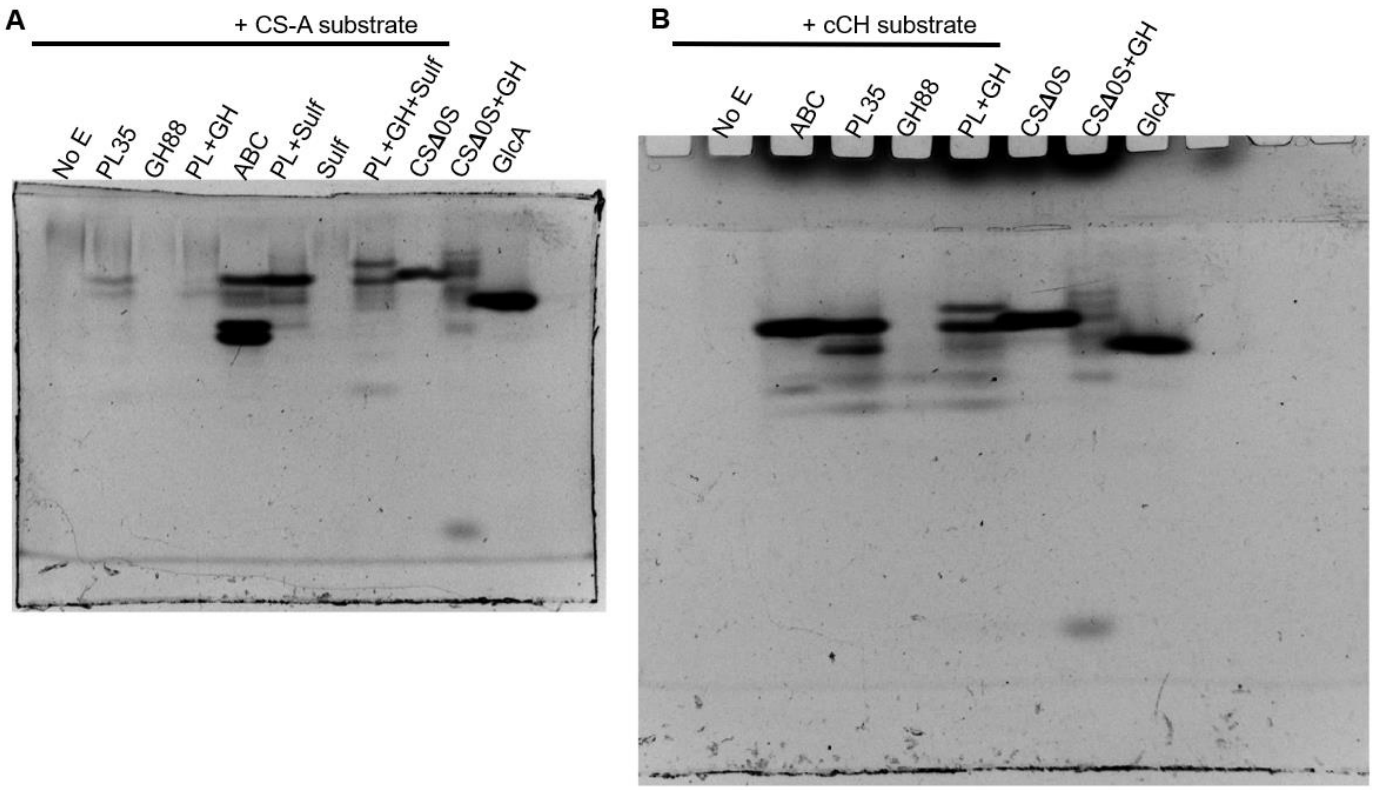


Figure S3.2: Full BcPL35 vs CSA and cCH FACE gels.

From **Figure 3.1B** and **Figure 3.1C** of the main text.

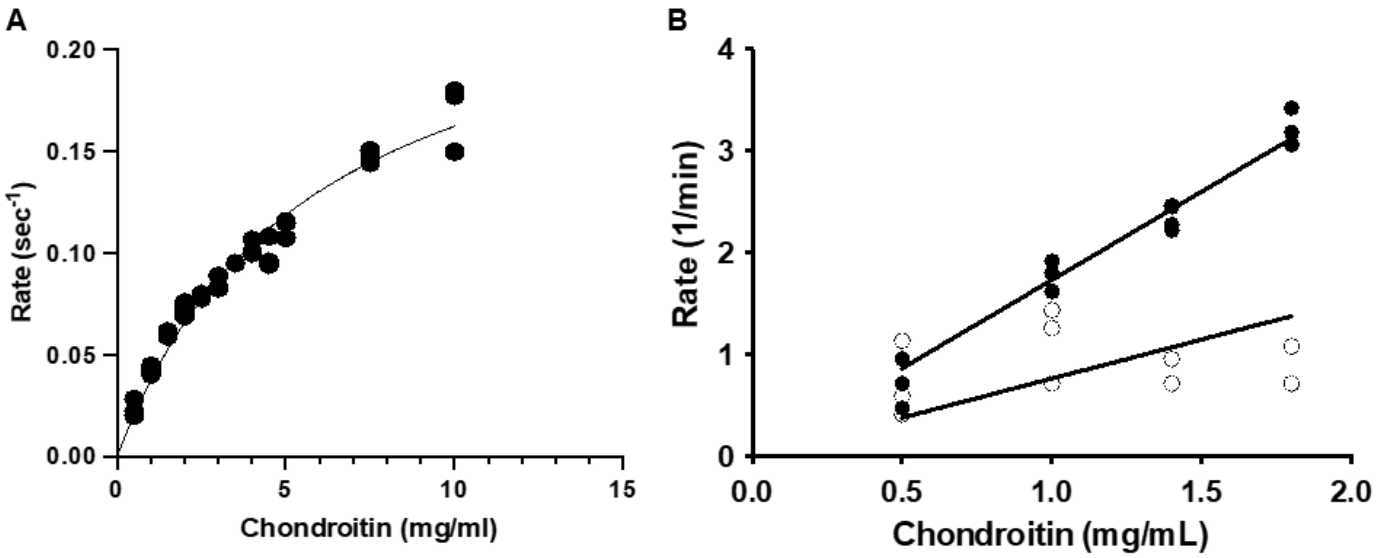


Figure S3.3: BcPL35 kinetics on chondroitin.

A) Kinetic analysis of BcPL35 on cCH showing individual replicates. The solid line shows the best-fit line to the Michaelis-Menton model. **B)** Kinetic analysis of BcPL35 W419A (solid circles) and Y232A (open circles) on cCH showing individual replicates. The solid line shows the best-fit line from linear regression.

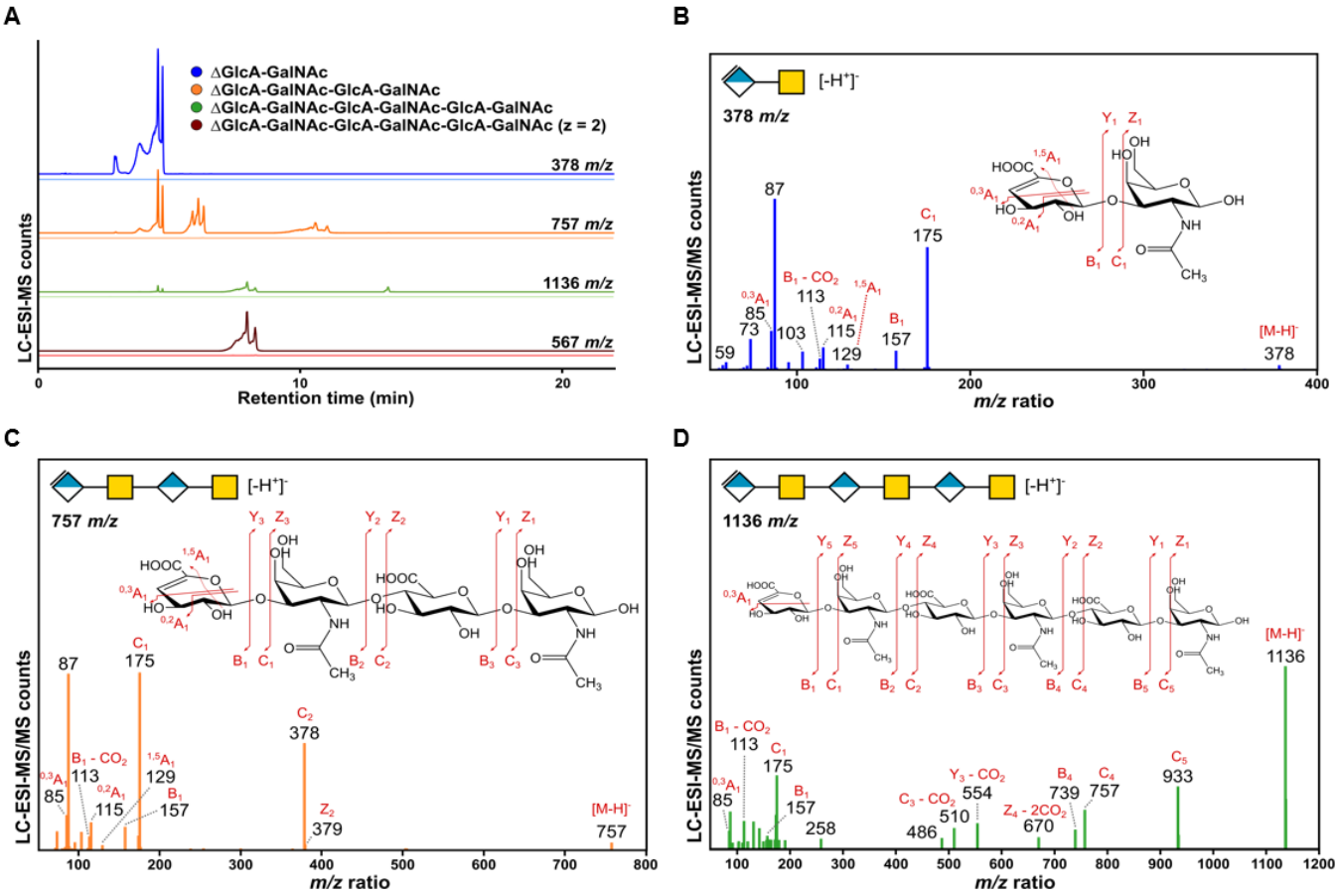


Figure S3.4: Enzyme reaction products from cCH digested with BcPL35 analyzed by LC-ESI-MS.

A) Extracted ion chromatograms are shown for ions of interest with ion counts scaled relative to the most intense peak for all chromatograms. Chromatograms are for extracted ions in the enzyme reaction (dark colours) or control/no enzyme reactions (light colours). **B-D)** ESI-MS/MS with HCD was performed to identify the extracted ion for the primary catalytic products in each.

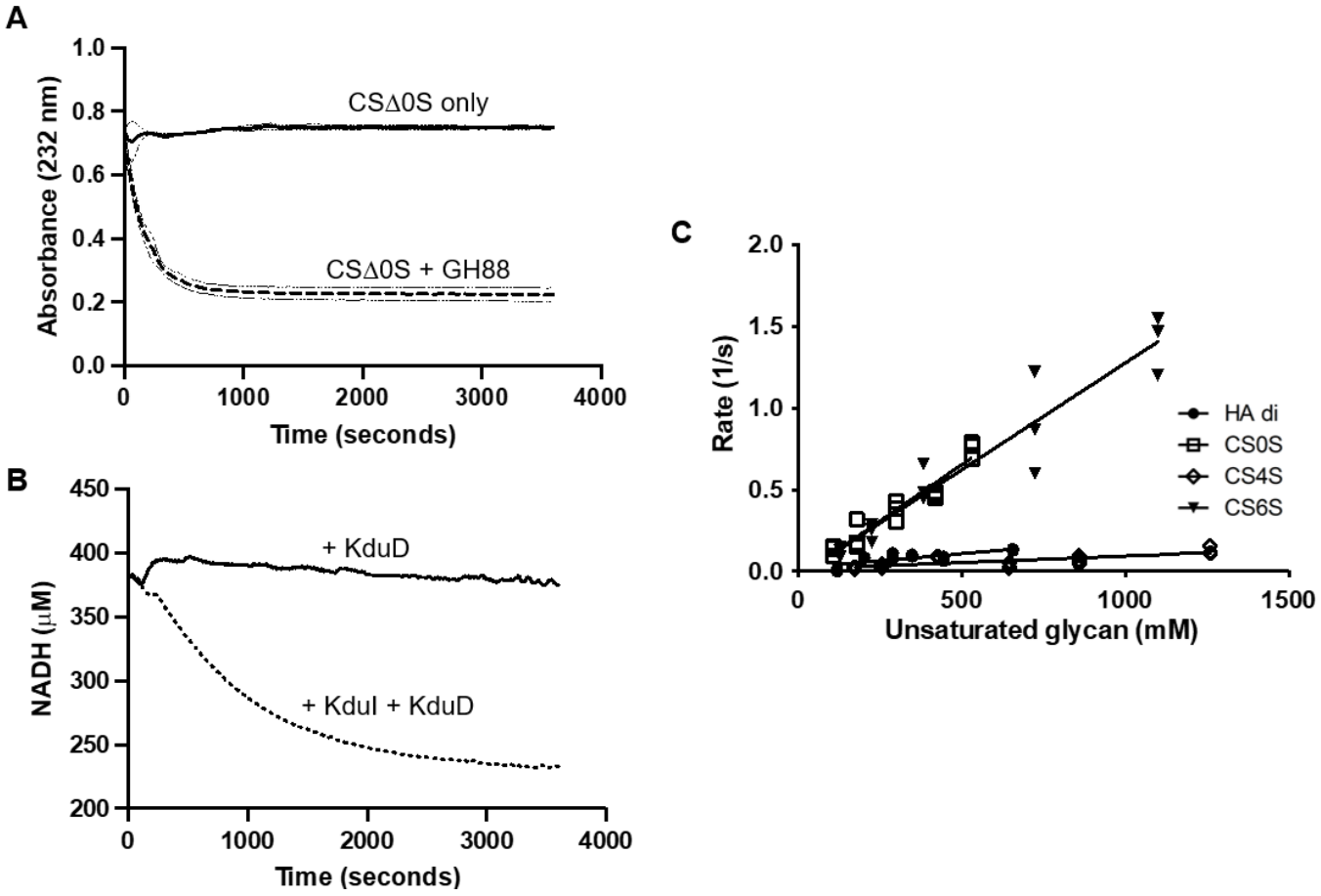


Figure S3.5: Activity of BcGH88.

A) Activity of BcGH88 on CSΔ0S. **B)** Oxidation of NADH by KduD and KduI activity on the 5-keto-4-deoxyuronate produced by BcGH88 from CSΔ0S. **C)** Kinetics of BcGDH showing individual replicates. Solid lines show best fits from linear regression.

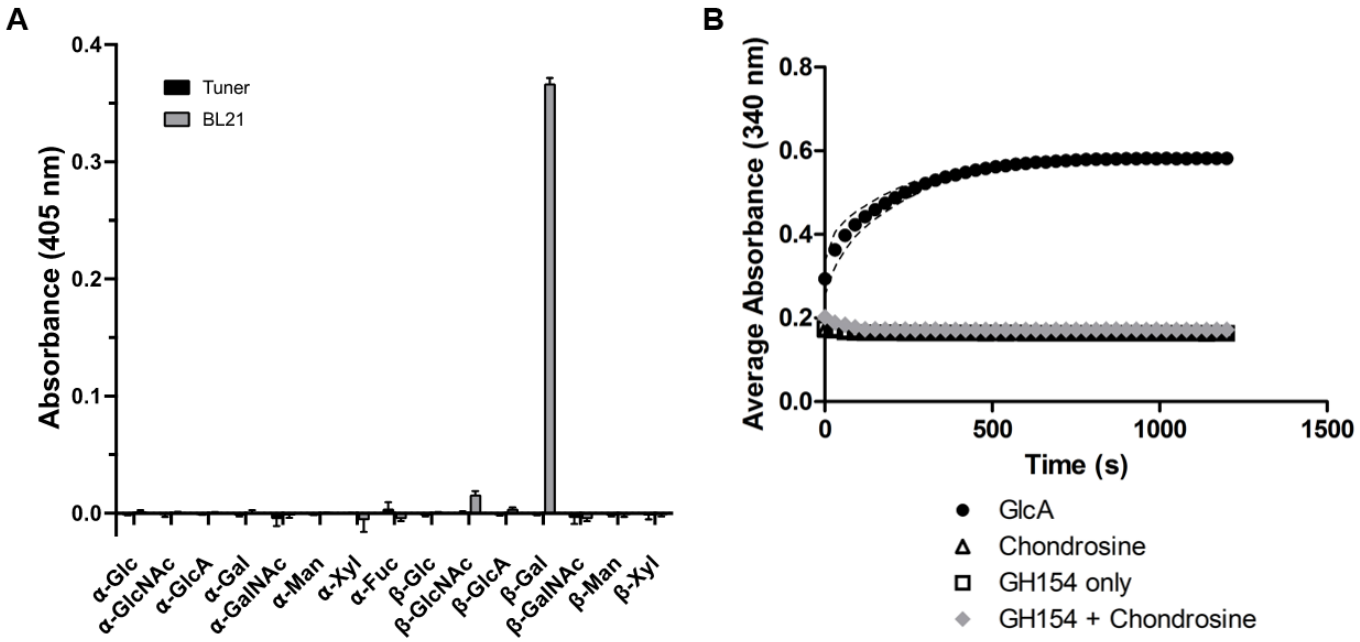


Figure S3.6: Activity of BcGDH on glycosides.

For clarification, BcGDH refers to BcGH154. **A)** Screen of BcGDH purified from *E. coli* BL21 (grey bars) or Tuner (black bars) on aryl-glycosides. Error bars represent the standard deviation of triplicate samples. **B)** Glucuronic acid detection assay for activity on chondrosine.

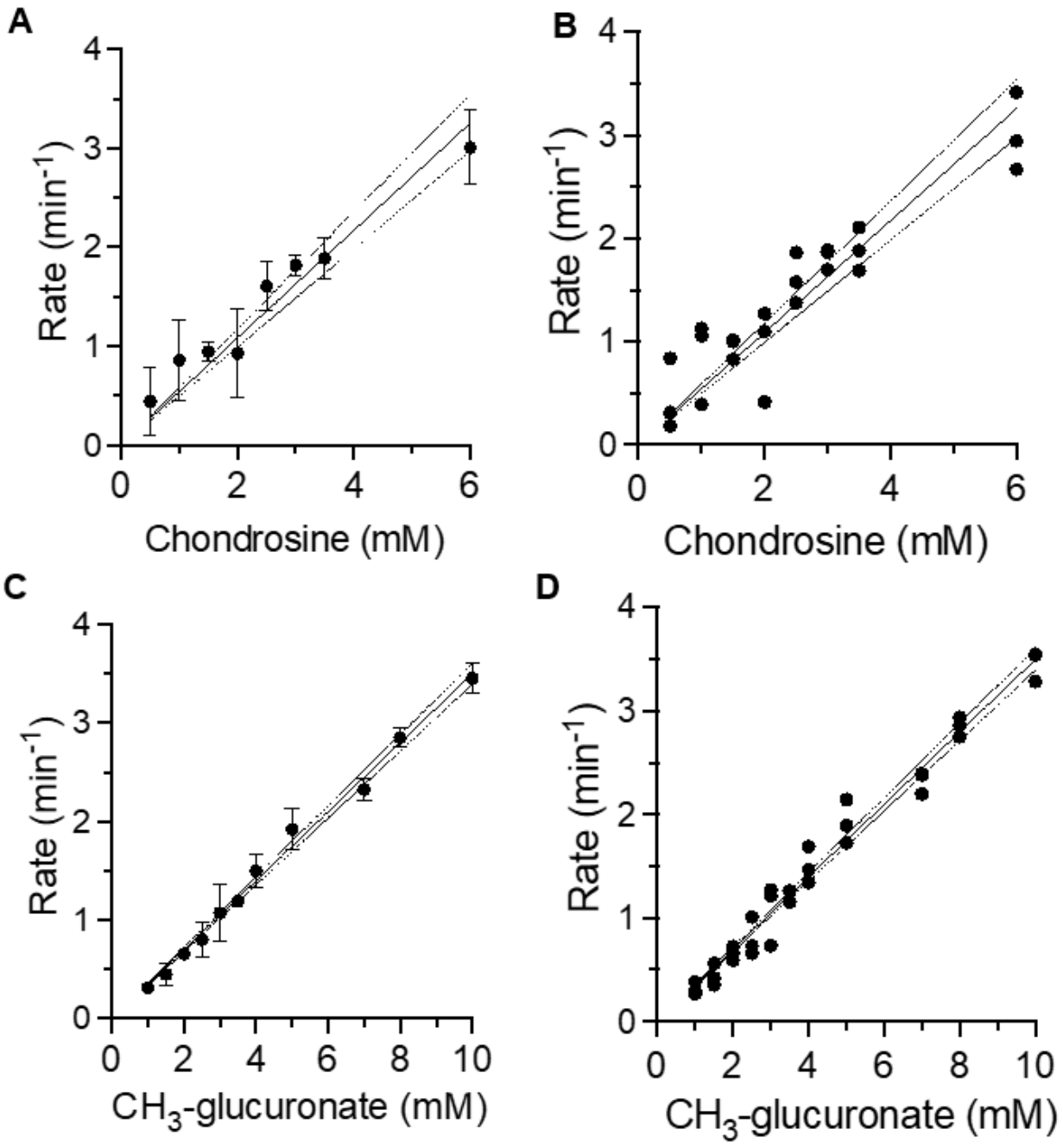


Figure S3.7: Kinetics of β-glucuronate dehydrogenases.

A) Kinetics of BcGDH (i.e. BcGH154) on 1-*O*-methyl-β-D-glucuronate. Error bars represent the standard deviation of triplicate samples. **B)** Kinetics of BcGDH on chondrosine showing individual replicates. **C)** Kinetics of BT_3677 (*i.e.,* BtGH154) on 1-*O*-methyl-β-D-glucuronate. Error bars represent the standard deviation of triplicate samples. **D)** Individual replicates of data in panel B. The solid lines in all panels represent the best fit lines by linear regression; dashed lines indicate 95% confidence limits.

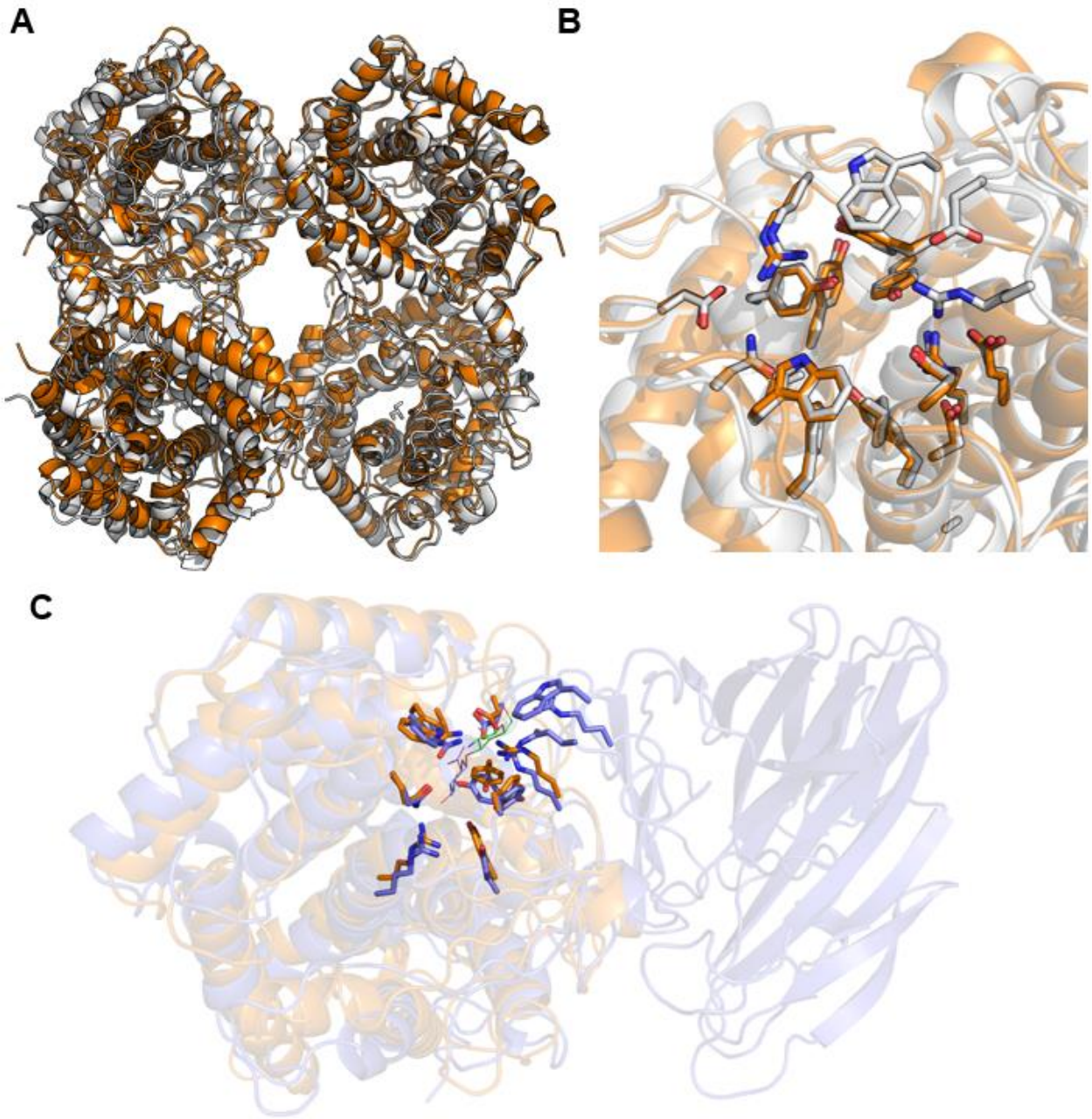


Figure S3.8: Structural comparisons of BcGDH.

A) The BcGDH tetramer (orange) overlaid with the BD-β-Gal tetramer (grey; **PDB:** 8OI4). **B)** The conserved active site of BcGDH with BD-β-Gal with important residues shown as sticks. **C)** The BcGDH monomer (orange) overlaid with an AlphaFold model of BT3677 (blue). Important active site residues are shown as sticks.

Table S3.1: BcPL35, BcGDH, and BcGDH_R285A x-ray data collection statistics.
For clarification, BcGDH refers to BcGH154.

	BcPL35	**BcGDH**	**BcGDH_R285A**
			+ chondrosine
Data Collection			
Wavelength	1.541	1.541	1.541
Space Group	$P2_12_12_1$	$P2_1$	$P2_12_12_1$
Cell Dimensions			
a, b, c (Å)	69.284, 84.043, 112.870	53.535, 161.37, 88.61 (β=95.70)	99.55, 105.86, 179.93
Resolution (Å)	30.00-1.75 (1.78-1.75)	20.00-2.20 (2.24-2.20)	20.00-2.60 (2.65-2.60)
R_{meas}	0.083 (0.649)	0.083 (0.325)	0.191 (0.994)
R_{pim}	0.031 (0.357)	0.046 (0.199)	0.092 (0.500)
CC1/2	0.998 (0.924)	0.968 (0.997)	0.995 (0.844)
$\langle I/\sigma I \rangle$	20.6 (1.8)	13.2 (3.0)	7.5 (1.3)
Completeness (%)	99.8 (99.5)	94.1 (80.7)	98.7 (98.9)
Redundancy	4.9 (3.0)	2.7 (2.0)	4.0 (3.9)
No. of Reflections	329,566	191,567	231,408
No. Unique	67,073	71,486	58,307
Refinement			
Resolution (Å)	1.75	2.20	2.60
R_{work}/R_{free}	0.185/0.217	0.164/0.216	0.22/0.26
No. of Atoms	5,311	13,277	12,912
Protein	4,631	12,270	12,377
Ligand	2	n/a	96
Water	678	956	412
B-factors	21.88	32.04	46.78
Protein	20.68	31.79	46.58
Ligand	16.63	n/a	67.28
Water	30.08	35.18	43.31
r.m.s.d.			
Bond Lengths (Å)	0.007	0.006	0.002
Bond Angles (°)	0.929	0.804	0.583
Ramachandran (%)			
Preferred	97.8	97.5	97.0
Allowed	2.24	2.48	2.9
Disallowed	0.0	0.0	0.0
PDB ID	**9O3Q**	**9O4U**	**9NWF**

Table S3.2: Enzyme products as observed by LC-ESI-MS.

Species	Ion	Expected m/z	Observed m/z	Retention Time (peak)(min)
GlcA-GalNAc	[M-H]-	396.1147	396.1124	5.13
ΔGlcA-GalNAc	[M-H]-	378.1042	378.1020	4.58
GlcA-GalNAc-GlcA-GalNAc	[M-H]-	775.2262	775.2227	6.88
ΔGlcA-GalNAc-GlcA-GalNAc	[M-H]-	757.2156	757.2120	6.12
GlcA-GalNAc-GlcA-GalNAc-GlcA-GalNAc	[M-H]-	1154.3377	1154.3315	9.17
ΔGlcA-GalNAc-GlcA-GalNAc-GlcA-GalNAc	[M-H]-	1136.3271	1136.3210	7.98
ΔGlcA-GalNAc-GlcA-GalNAc-GlcA-GalNAc	$[M-2H]^{2-}$	567.6599	567.6570	7.99
GlcA-GalN	[M-H]-	354.1042	354.1018	5.62
ΔGlcA-GalN	[M-H]-	336.0936	336.0916	5.14

Table S3.3: Differential scanning fluorimetry melts of BcGDH mutants.

Protein	Melting temperature ^{o}C (± SD of triplicates)
Wild type	50.8 (± 0.1)
E74A	49.9 (± 0.2)
R78A	48.3 (± 0.1)
N176A	47.8 (± 0.7)
N177A	52.5 (± 0.1)
W178A	53.6 (± 0.6)
F181A	47.9 (± 0.1)
Y229A	50.0 (± 0.1)
R285A	49.5 (± 0.1)
Y289A	48.9 (± 0.3)
Y354A	50.1 (± 0.2)

Chapter 4 : Exploring the specificity of two PUL-associated M60-like peptidases from *Bacteroides caccae*

4.1. Introduction

In normal or pathogenic processes, protein turnover requires the hydrolysis of peptide bonds, which is catalysed through the activity of peptidases (Pluvinage et al., 2021; Rawlings et al., 2016; Rawlings & Bateman, 2021). These enzymes are organised into hierarchical classifications within the MEROPS database, wherein homologous sequences are clustered into into families, and collections of families into clans (Rawlings et al., 2016; Rawlings & Bateman, 2021). At each level, peptidases are assigned an identifier that describes their catalytic mechanism, or catalytic type. There are seven catalytic types: serine (S), cysteine (C), threonine (T), aspartic (A), glutamic (G), asparagine lyases (N), and metallo (M). However, the protein substrates of peptidases are often post-translationally modified, therefore peptidases must also account for these modifications. It has been estimated that half of all proteins in nature bears covalently linked glycans. This indicates that glycosylation is the most frequent post-translational modification that prevents the core protein from chemical and enzymatic degradation (Apweiler et al., 1999; Noach et al., 2017). Consequently, glycosylation is challenging for peptidases, as these large glycans prevent access to the peptide backbone of their glycoprotein substrates.

Members of the M60-like peptidase super-family (Pfam PF13402; enhancin; enhancin-like; peptidase_M60) are known to cleave peptide bonds of highly glycosylated proteins and use *O*-glycan modifications as a specific substrate recognition determinant (Rawlings & Bateman, 2021). This unique glycan requirement was linked to distinct glycan-binding subsites in the M60-like peptidase active sites that confer specificity for protein substrates bearing specific *O*-linked glycans. Consequently, these peptidases are also called *O*-glycopeptidases (**Figure 4.1**). The M60-like peptidase super-family is defined by three traits: relatedness to viral enhancin metallopeptidases, a conserved gluzincin metallopeptidase motif (HEXXHX(8,24)E) and reoccurring multi-modularity with carbohydrate binding modules (CBMs) (Boraston et al., 2004; Nakjang et al., 2012). These CBMs can be located at either the N- or C-terminal side of the catalytic M60 domain, but the CBM positioning relative to the M60 domain appears to be conserved, implying that CBM configuration is key for enzyme functionality. Overall, this peptidase

superfamily is widely dispersed amongst bacteria, plants, fungi, and animals (Nakjang et al., 2012; Noach et al., 2017; Rawlings & Bateman, 2021).

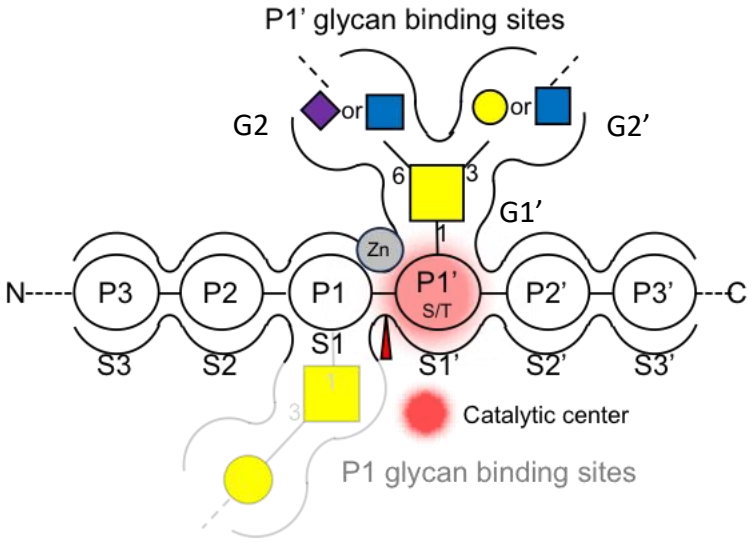


Figure 4.1: Depiction of M60 peptidase active site.

Red triangle indicates cleavage point on target peptide after glycan recognition. Peptide-binding residues (S) are labelled relative to the cleavage point. Glycan-binding subsites (G) are designated relative to catalytic center, wherein glycan-binding residues interacting with GalNAc noted as G1', while glycan-binding residues interacting with sugars attached to C6 and/or C3 of the GalNAc are labelled G2 and G2', respectively (Noach et al., 2017).

Most characterized members of the M60-like peptidase superfamily reside in taxa that have transient or long-term interactions with animal hosts (Nakjang et al., 2012; Pluvinage et al., 2021; Shon et al., 2020, 2022). These include glycopeptidases such as IMPa from *Pseudomonas aeruginosa* (Noach et al., 2017); ZmpA, ZmpB, and ZmpC from *Clostridium perfingens* (Pluvinage et al., 2021); BT4244 from *Bacteroides thetaiotamicron* (Noach et al., 2017; Taleb et al., 2022); and AMUC_0627, AMUC_0908, and AMUC_1514 from *Akkermansia muciniphila* (Medley et al., 2022; Shon et al., 2020; Taleb et al., 2022). These bacteria are members of the human gut microbiota (HGM) that commonly target the peptide backbone of mucins; densely *O*-glycosylated proteins in the mucosal layer, which are 60-70% carbohydrate by mass (**Figure 1.2**). Although M60-like peptidase genes are found in many bacterial species, not all sequenced strains of specific species contain these genes (Nakjang et al., 2012). This implies that these glycopeptidase genes are part of a disposable pangenome that

contributes to a microbe's specific niche adaptations. In other words, the presence/absence of these peptidases in a genome can provide insight into its host's ecological role. For instance, the glycopeptidase redundancy present in *A muciniphila* supports its role as a mucin-degrading specialist (Nakjang et al., 2012; Shon et al., 2020, 2022). Thus, the observation that *B. caccae* (strain ATCC 48135) has no less than sixteen putative M60-like peptidases in its genome is particularly noteworthy (Nakjang et al., 2012; Rawlings & Bateman, 2021).

Bacteroides caccae is a less studied, diet-sensitive member of the HGM (Desai et al., 2016). As with other members of its genus, *B. caccae* can degrade numerous glycan sources, including GAGs, which are generally found as proteoglycans (*i.e.,* GAGs attached to a protein core) (Overbeeke et al., 2022); and mucin (Desai et al., 2016). Of the fifteen putative M60-like sequences this microbe has in its genome, twelve contain CBMs, while eight contain the structurally distinct, but functionally similar, *Bacteroides*-associated carbohydrate binding often N-terminal (BACON) domains (Blum et al., 2021; Nakjang et al., 2012) (**Figure 4.2)**. Preliminary alignment of these sequences showed the conserved gluzincin motif, characteristic of this family (**Figure 4.3**). Interestingly, all these glycopeptidases are found in polysaccharide utilization loci (PULs) (Terrapon et al., 2018). Classically, PULs are defined as co-localized and co-regulated gene clusters that target a single glycan, or chemically related group of glycans. Given this, the presence of M60 peptidases in *B. caccae* PULs imply enzymatic synergy between the encoded carbohydrate active enzymes (CAZymes) and glycopeptidases. Here, we focus on structurally and functionally characterizing two *B. caccae* M60-like peptidases: BcM60_F and BcM60_G (Nakjang et al., 2012; Terrapon et al., 2018). Both glycopeptidases are found in BcPUL25, whose CAZyme consortium was determined to target glycosaminoglycans (GAGs) (**Figure 1.11**) (Alvarez et al., 2025). Given this, we hypothesized that BcM60_F and BcM60_G can target the *O*-linked oligosaccharides found on proteoglycan peptide backbones, such as aggrecan (Kiani et al., 2002; Koch et al., 2020).

Overall, this study seeks to understand the molecular features that confer glycopeptidase specificity using two putative M60-like peptidases from *B. caccae.* Given that glycopeptidases have outstanding utility as tools to study glycoprotein structure by mass spectrometry (Malaker et al., 2016, 2019), expanding the repertoire of glycopeptidases with different glycan specificities will improve their utility as biotechnological tools. Additionally, this work will contribute to the

growing literature surrounding *B. caccae* and its role in the HGM, as this species has been implicated in exacerbating various health states (Berkhout et al., 2024; Desai et al., 2016; B. Wei et al., 2001).

Figure 4.2: Multiple sequence alignment of BcM60 peptidases' catalytic domains.

Red boxes indicate shared residues between glycopeptidases. Yellow boxes indicate similar residues between glycopeptidases. Sequences were aligned using ClustalOmega (Madeira et al., 2024).

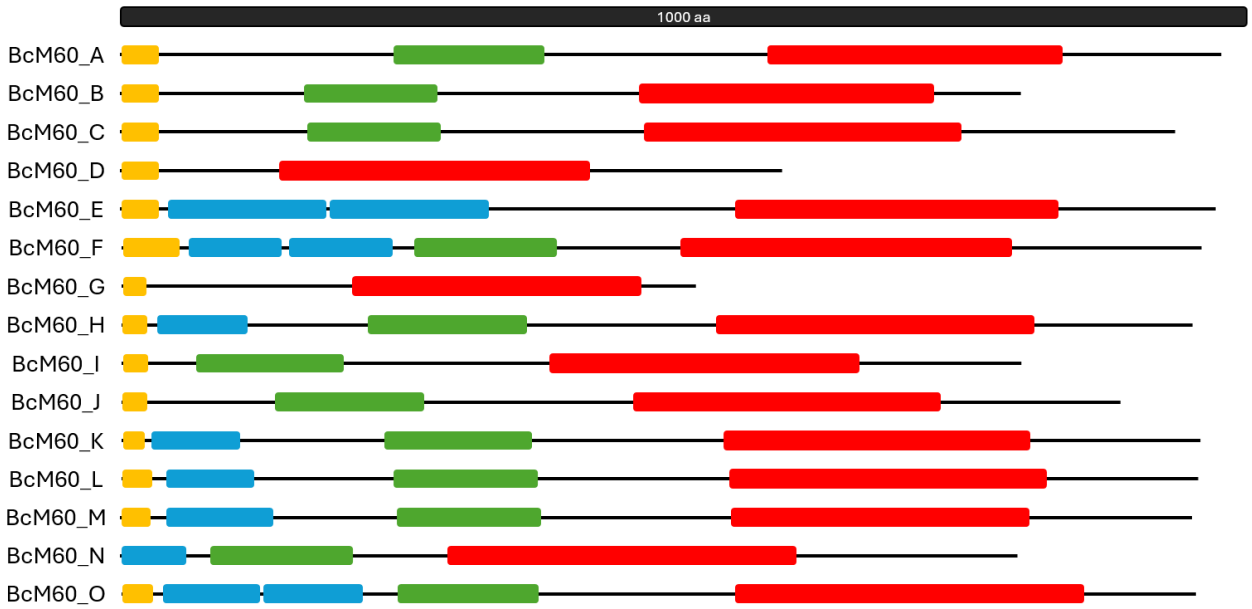


Figure 4.3: InterProScan predictions on the co-occurrence of CBM and/or BACON domains with *B. caccae* M60-like peptidases.

Domain boundaries and position determined using InterProScan and SignalP (6.0) (Blum et al., 2021; Teufel et al., 2022). *Color legend*: yellow, signal peptide; blue, BACON domain; green, CBM domain; red, peptidase_M60 domain (Pfam 13402). For scale, the black box above these putative genes represents 1000 amino acids.

4.2. Results

4.2.1. BcM60_F and BcM60_G are *O*-glycopeptidases with different glycan specificities.

Previous transcriptional analyses of these M60-like peptidases in *B. caccae* (BcM60) show upregulation under conditions that promote degradation of the host mucosal layer (Desai et al., 2016). Given this, we assessed the glycopeptidase activity of the BcM60_F and BcM60_G catalytic modules to determine whether these are selective for *O*-glycosylation present on mucin (**Figure 4.4A** and **Figure 4.4B)**. This was tested using 24 mucin-specific FRET substrates with one of the following glycosylations: Tn (GalNAcα-Ser/Thr), Core 1 (Galβ1-3GalNAcα-Ser/Thr), 3SC1 (Neu5Acα2-3Galβ1-3GalNAcα-Ser/Thr), Core 2 (GlcNAcβ1-6(Galβ1-3)GalNAcα-Ser/Thr), and 6SC1 (Neu5Acα2-6(Galβ1-3)GalNAcα-Ser/Thr) (Wardman et al., 2023). Activity is detected through cleavage of the *O*-glycosylated linker, as this will result in a decrease in FRET efficiency, leading to detectable changes in fluorescent signal. A previously characterized M60-like peptidase from *B. thetaiotamicron*, BT4244, was used here as a positive control, as it can cleave numerous peptide sequences with Tn and Core 1 glycan moieties (**Figure 4.4C)** (Noach et al., 2017; Shon et al., 2020; Taleb et al., 2022).

By itself, the catalytic domain of BcM60_F was inactive on the available mucin-specific FRET constructs (**Figure S4.2**). Previous studies have observed that removing or mutating the glycan binding domains of enzymes significantly decreases, if not abolishes, their activity (Arai et al., 2003; Boraston et al., 2004). Given this, BcM60_F likely relies on its CBM and two BACON domains associated with its eponymous module to recognize and position glycoprotein substrates for peptide hydrolysis (**Figure 4.3**) (Boraston et al., 2004). This idea was reinforced by the activity shown by BcM60_F after restoring the CBM domain (BcM60_F_CBM) (**Figure 4.4A**). To summarize, BcM60_F_CBM appears to recognize Tn, Core 1, and 3SC1 moieties on MUC5A/C_R2 (FRET ID: A7); Tn and Core 1 glycans on MUC5A/C_R1 (FRET ID: A6); Core 1 on MUC5B (FRET ID: A9) and MUC8 (FRET ID: B4). Notably, all of these are linear glycan structures; this glycopeptidase does not appear to tolerate branched sugars on any tested linker peptide.

In contrast, it was immediately apparent that BcM60_G has a broader substrate range than BcM60_F_CBM (**Figure 4.4B)**, as it recognizes all tested glycoforms on multiple different linker

peptides. Additionally, unlike BcM60_F_CBM, this glycopeptidase can interact with linear and branched sugars prior to peptide cleavage. To summarize, BcM60_G targets MUC2_R2_S1 (FRET ID: A4), MUC2_R2_S2 (FRET ID: A5), MUC5AC_R1 (FRET ID: A6), MUC5AC_R2 (FRET ID: A7), MUC5B_S1 (FRET ID: A8), MUC5B_S2 (FRET ID: A9), IgA1_HP_S2 (FRET ID: A10), MUC7_S1 (FRET ID: B1), MUC16 (FRET ID: B7), ASF (FRET ID: B9), and MUC17_S1 (FRET ID: C3) at varying degrees of glycosylation.

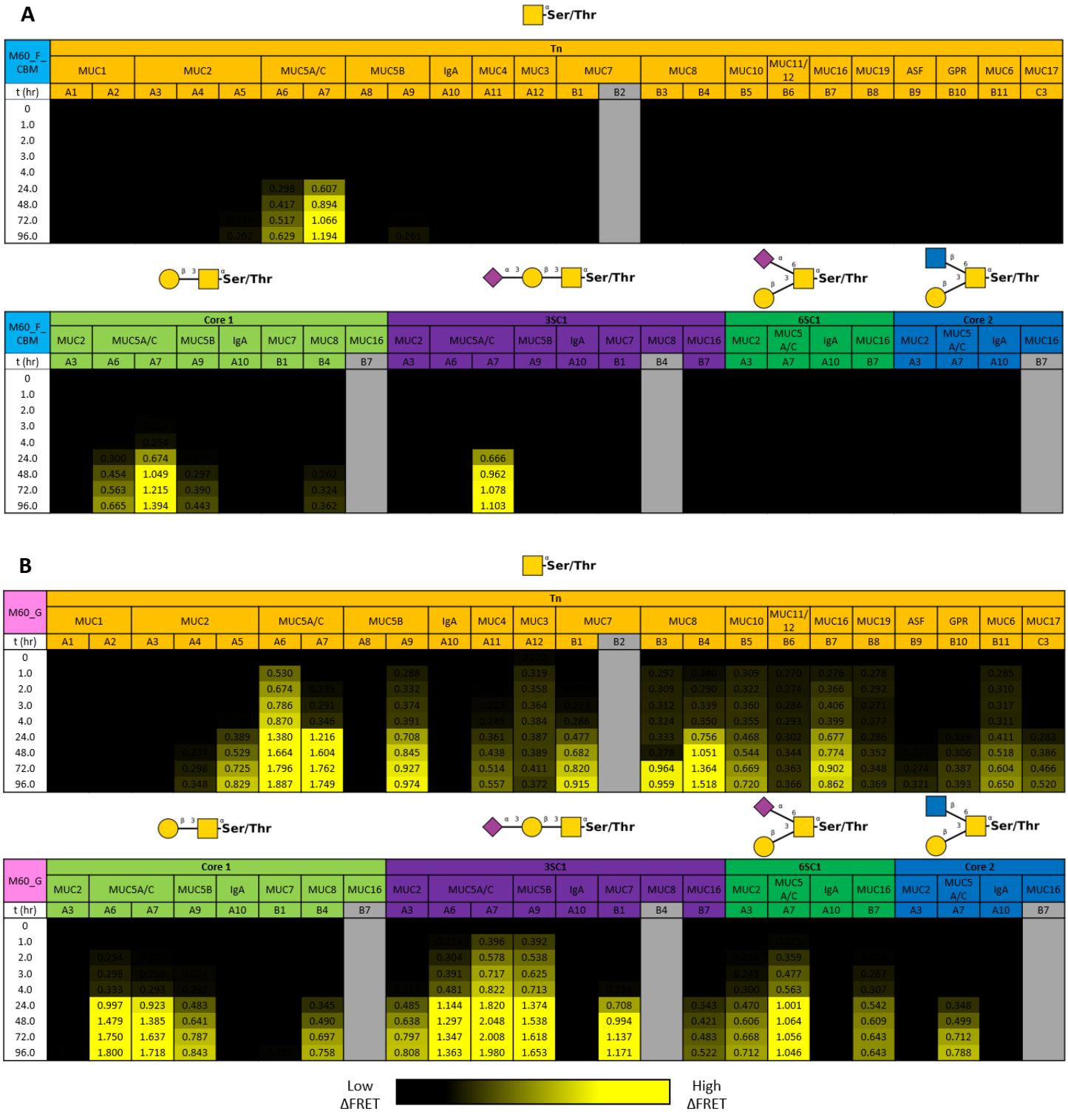
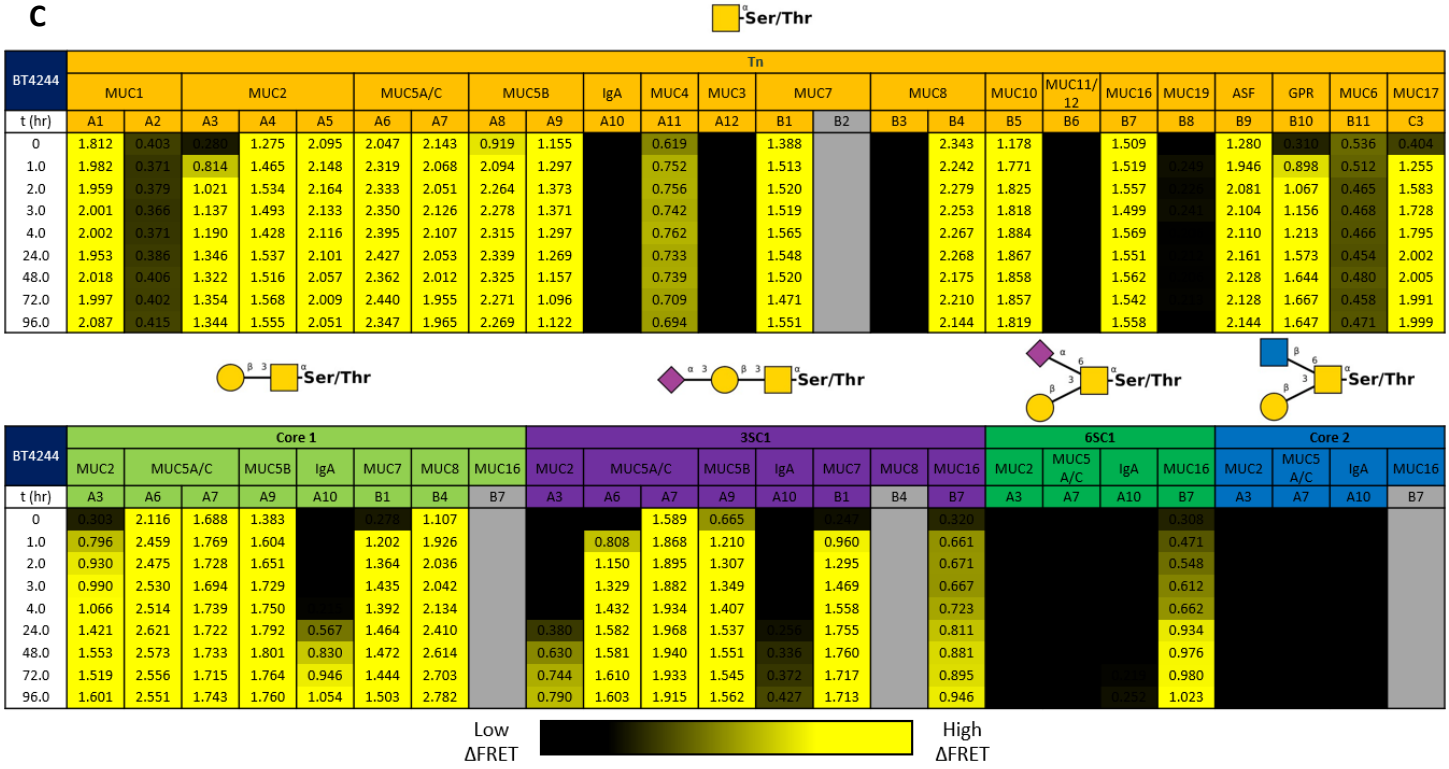


Figure 4.4: Qualitative comparison of BcM60-like peptidases to BT4244 glycan recognition.

All reactions done in duplicate alongside substrate only controls (**Figure S4.1**). Activity is measured by the difference in FRET signal (ΔFRET), wherein greater ΔFRET indicates higher activity. Greyed out boxes indicate unavailable substrate. **A)** BcM60_F_CBM. **B)** BcM60_G.

C



BT4244	Tn																							
	MUC1		MUC2			MUC5A/C		MUC5B		IgA	MUC4	MUC3	MUC7		MUC8		MUC10	MUC11/12	MUC16	MUC19	ASF	GPR	MUC6	MUC17
t (hr)	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	C3
0	1.812	0.403	0.280	1.275	2.095	2.047	2.143	0.919	1.155		0.619		1.388			2.343	1.178		1.509		1.280	0.310	0.536	0.404
1.0	1.982	0.371	0.814	1.465	2.148	2.319	2.068	2.094	1.297		0.752		1.513			2.242	1.771		1.519	0.249	1.946	0.898	0.512	1.255
2.0	1.959	0.379	1.021	1.534	2.164	2.333	2.051	2.264	1.373		0.756		1.520			2.279	1.825		1.557	0.276	2.081	1.067	0.465	1.583
3.0	2.001	0.366	1.137	1.493	2.133	2.350	2.126	2.278	1.371		0.742		1.519			2.253	1.818		1.499	0.241	2.104	1.156	0.468	1.728
4.0	2.002	0.371	1.190	1.428	2.116	2.395	2.107	2.315	1.297		0.762		1.565			2.267	1.884		1.569		2.110	1.213	0.466	1.795
24.0	1.953	0.386	1.346	1.537	2.101	2.427	2.053	2.339	1.269		0.733		1.548			2.268	1.867		1.551		2.161	1.573	0.454	2.002
48.0	2.018	0.406	1.322	1.516	2.057	2.362	2.012	2.325	1.157		0.739		1.520			2.175	1.858		1.562		2.128	1.644	0.480	2.005
72.0	1.997	0.402	1.354	1.568	2.009	2.440	1.955	2.271	1.096		0.709		1.471			2.210	1.857		1.542		2.128	1.667	0.458	1.991
96.0	2.087	0.415	1.344	1.555	2.051	2.347	1.965	2.269	1.122		0.694		1.551			2.144	1.819		1.558		2.144	1.647	0.471	1.999

BT4244	Core 1								3SC1								6SC1				Core 2			
	MUC2	MUC5A/C		MUC5B	IgA	MUC7	MUC8	MUC16	MUC2	MUC5A/C		MUC5B	IgA	MUC7	MUC8	MUC16	MUC2	MUC5 A/C	IgA	MUC16	MUC2	MUC5 A/C	IgA	MUC16
t (hr)	A3	A6	A7	A9	A10	B1	B4	B7	A3	A6	A7	A9	A10	B1	B4	B7	A3	A7	A10	B7	A3	A7	A10	B7
0	0.303	2.116	1.688	1.383		0.278	1.107				1.589	0.665		0.242		0.320				0.308				
1.0	0.796	2.459	1.769	1.604		1.202	1.926			0.808	1.868	1.210		0.960		0.661				0.471				
2.0	0.930	2.475	1.728	1.651		1.364	2.036			1.150	1.895	1.307		1.295		0.671				0.548				
3.0	0.990	2.530	1.694	1.729		1.435	2.042			1.329	1.882	1.349		1.469		0.667				0.612				
4.0	1.066	2.514	1.739	1.750		1.392	2.134			1.432	1.934	1.407		1.558		0.723				0.662				
24.0	1.421	2.621	1.722	1.792	0.567	1.464	2.410		0.380	1.582	1.968	1.537	0.256	1.755		0.811				0.934				
48.0	1.553	2.573	1.733	1.801	0.830	1.472	2.614		0.630	1.581	1.940	1.551	0.336	1.760		0.881				0.976				
72.0	1.519	2.556	1.715	1.764	0.946	1.444	2.703		0.744	1.610	1.933	1.545	0.372	1.717		0.895				0.980				
96.0	1.601	2.551	1.743	1.760	1.054	1.503	2.782		0.790	1.603	1.915	1.562	0.427	1.713		0.946				1.023				

Figure 4.4: (Continued): Qualitative comparison of BcM60-like peptidases to BT4244 glycan recognition.

All reactions done in duplicate alongside substrate only controls (**Figure S4.1**). Activity is measured by the difference in FRET signal (ΔFRET), wherein greater ΔFRET indicates higher activity. Greyed out boxes indicate unavailable substrate. **C)** BT4244.

4.2.2. Kinetics

After qualitative substrate specificity analysis with the mucin and aggrecan FRET screen, the same assay was adapted towards quantitative kinetic analysis. Given that both glycopeptidases were previously active on multiple glycoforms of MUC5AC_R2 (FRET ID: A7) (**Figure 4.4**), this substrate was tested against BcM60_F_CBM and BcM60_G. Initial analysis using the Van Slyke Cullen method did not yield reproducible values at the two enzyme concentrations tested (0.5 µM and 1.0 µM) (Kuzmič, 2009). Thus, a simplified enzyme mechanism was used for modelling eleven progress curves at different substrate concentrations (Kuzmič et al., 2010), wherein the following was used to solve for k_{cat}/K_M:

$$E + S \rightarrow E + P$$

Global analysis of these datasets using this simplified enzyme mechanism via DynaFit was reliable, as resulting residual plots showed a random pattern (**Figure S4.3**). Additionally, k_{cat}/K_M values for both glycopeptidases and two enzyme concentrations tested were reproducible for Tn, Core 1, Core 2, and 6SC1 (± 0.2 min^{-1} mM^{-1}). In contrast, 3SC1 varied by ± 0.4 min^{-1} mM^{-1} between datasets (**Table S4.2**). Overall, BcM60_F_CBM and BcM60_G had relatively similar catalytic efficiency (k_{cat}/K_M) across linear MUC5A/C_R2 glycoforms (**Table 4.1**). BcM60_G can tolerate branched glycans, although its efficiency was generally lower compared to its activity on linear glycosylations.

Table 4.1: Summarized catalytic efficiency of tested BcM60-like peptidases.
Reactions were done in technical duplicates. Values acquired from global analysis using DynaFit4 software (Kuzmič, 1996). Residual plots to validate analysis per glycoform per M60-like peptidase shown in **Figure S4.3**. Notation below represent an average ± standard error.

Construct	Linker ID	Glycosylation	k_{cat}/K_M (min^{-1} μM^{-1}) (x 10^{-5})	CV (%)	95% CI (low, high) (x 10^{-5})
BcM60_F_CBM	MUC5A/C	Tn	0.64 ± 0.01	6.8	0.56, 0.73
		Core 1	0.56 ± 0.01	4.0	0.51, 0.60
		3SC1	0.73 ± 0.01	2.2	0.69, 0.77
		Core 2	n/a		
		6SC1	n/a		
BcM60_G	MUC5A/C	Tn	0.93 ± 0.01	4.1	0.85, 1.00
		Core 1	0.67 ± 0.01	3.3	0.63, 0.71
		3SC1	0.75 ± 0.01	3.2	0.70, 0.80
		Core 2	0.57 ± 0.02	8.9	0.47, 0.68
		6SC1	0.59 ± 0.01	6.5	0.39, 0.66

4.2.4.: Structural analysis of BcM60_G complexed to Core 2-Thr

The structure of BcM60_G has previously been determined, unliganded and in complex with Core 1 (Deventer, 2021). However, to gain insight into how BcM60_G can tolerate branched glycoforms, we pursued and successfully acquired its structure in complex with Core 2 using X-ray crystallography (**Figure 4.5**) (**Table S4.1**). A catalytically inactive BcM60_G construct (BcM60_G_E327Q) was expressed, purified, and set into crystal trays. Then, these crystals were soaked with Core 2-Thr. After collecting a dataset in the space group C_2, this structure was solved by molecular replacement using an AlphaFold model to a resolution of 1.85 Å (Jumper et al.,

2021). The final refined model was composed of a single monomer containing the Core 2-Thr in the active site and a zinc cofactor positioned at the gluzincin motif, coordinating a nucleophilic water located ~3 Å away from where the N-terminal scissile peptide bond would be located (**Figure 4.5D**) (Nakjang et al., 2012).

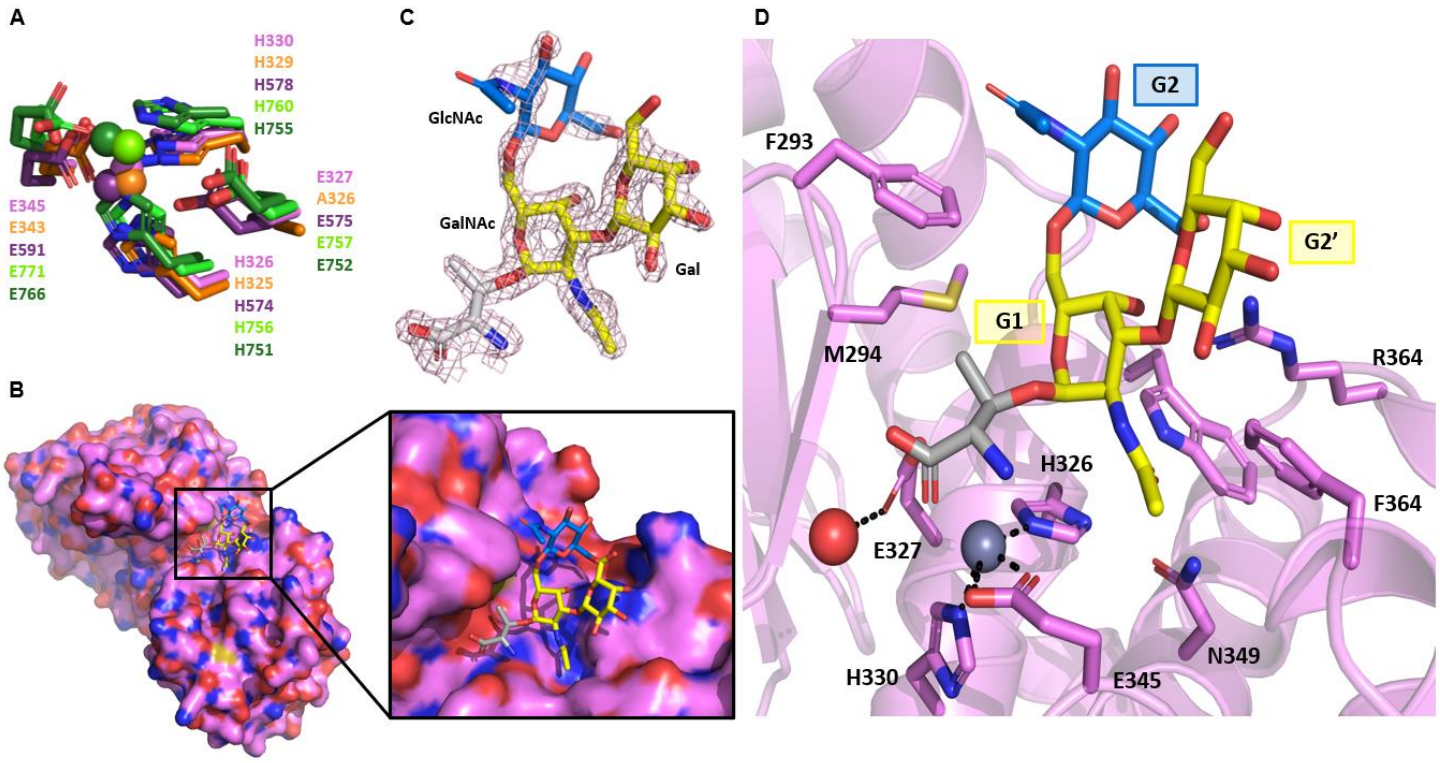


Figure 4.5: BcM60_G in complex with Core 2 glycan.

A) Structural conservation of gluzincin motif between BcM60_G (pink) and AMUC_0627 (orange) (**PDB:** 7YX8) (RMSD: 1.33 Å), BT4244 (purple) (**PDB:** 5KD8) (RMSD: 2.15 Å), and ZmpB/C (light/dark green) (**PDB:** 5KDU/6XT1) (RMSD: 2.31 Å/ 2.47 Å) (Noach et al., 2017; Pluvinage et al., 2021; Taleb et al., 2022). Zinc ions of each structure are shown as spheres, wherein each cofactor was color coded based on structure identity. **B)** Electrostatic surface view of BcM60_G with Core 2. Black box zooms into the ligand in the binding pocket. **C)** Ligand surrounded by light pink mesh representing $2F_O$-F_C density map contoured at 1.0 σ. **D)** Active site of BcM60_G complexed with Core 2. Glycopeptidase side chains shown in pink, substrate amino acid in grey, and Core 2 glycans shown in blue (GlcNAc) and yellow (Gal, GalNAc) sticks according to the glycan symbol nomenclature conventions. Zinc cofactor shown in grey, with the recruited water in red.

As expected for a member of this superfamily, alignment with other structurally characterized M60-like peptidases showed high spatial conservation of the BcM60_G gluzincin motif (**Figure 4.2** and **Figure 4.5A**). Electrostatic surface view of the monomer shows the Core 2 ligand fitting into the binding pocket (**Figure 4.5B**). Although there is only partial density of the GlcNAc moiety

of the Core 2, the electron density map indicated that the GalNAc and Gal glycans interacting in the G1 and G2' subsites of BcM60_G were in stable chair conformations (**Figure 4.5C**). It has been noted previously that structurally characterized M60-like peptidases share relatively conserved residues (Trp/Asn/Arg) near the GalNAc moiety at the G1 position (Taleb et al., 2022). BcM60_G recapitulated this observation, as Trp322/Asn349/Arg364 interacted with the GalNAc in the same way. Notably, BcM60_G positions and shares the same residues surrounding G1 (*i.e.,* GalNAc) and G2 (*i.e.,* GlcNAc) as AMUC_0627, BT4244, and ZmpB/C (**Figure 4.6A**) (Noach et al., 2017). Its G2 subsite has a Met ($Met294^{BcM60\_G}$/$Met728^{ZmpB}$/$Met723^{ZmpC}$) and Phe ($Phe293^{BcM60\_G}$/$Phe727^{ZmpB}$/$Phe723^{ZmpC}$) that is found in ZmpB/C (**Figure 4.6B**). This is notable as ZmpB/C has been shown to accommodate branched glycans (Noach et al., 2017; Pluvinage et al., 2021). Of these two residues, Phe293 is positioned close enough to form CH-π interactions with GlcNAc (~3.4 Å), while Met294 may contribute to substrate orientation prior to cleavage (4.8 Å). Overall, based on this BcM60_G-Core 2 complex, glycan recognition is primarily reliant on the Trp/Asn/Arg residues associated with the G1 and G2' subsites, as well as the Phe residue at the G2 subsite.

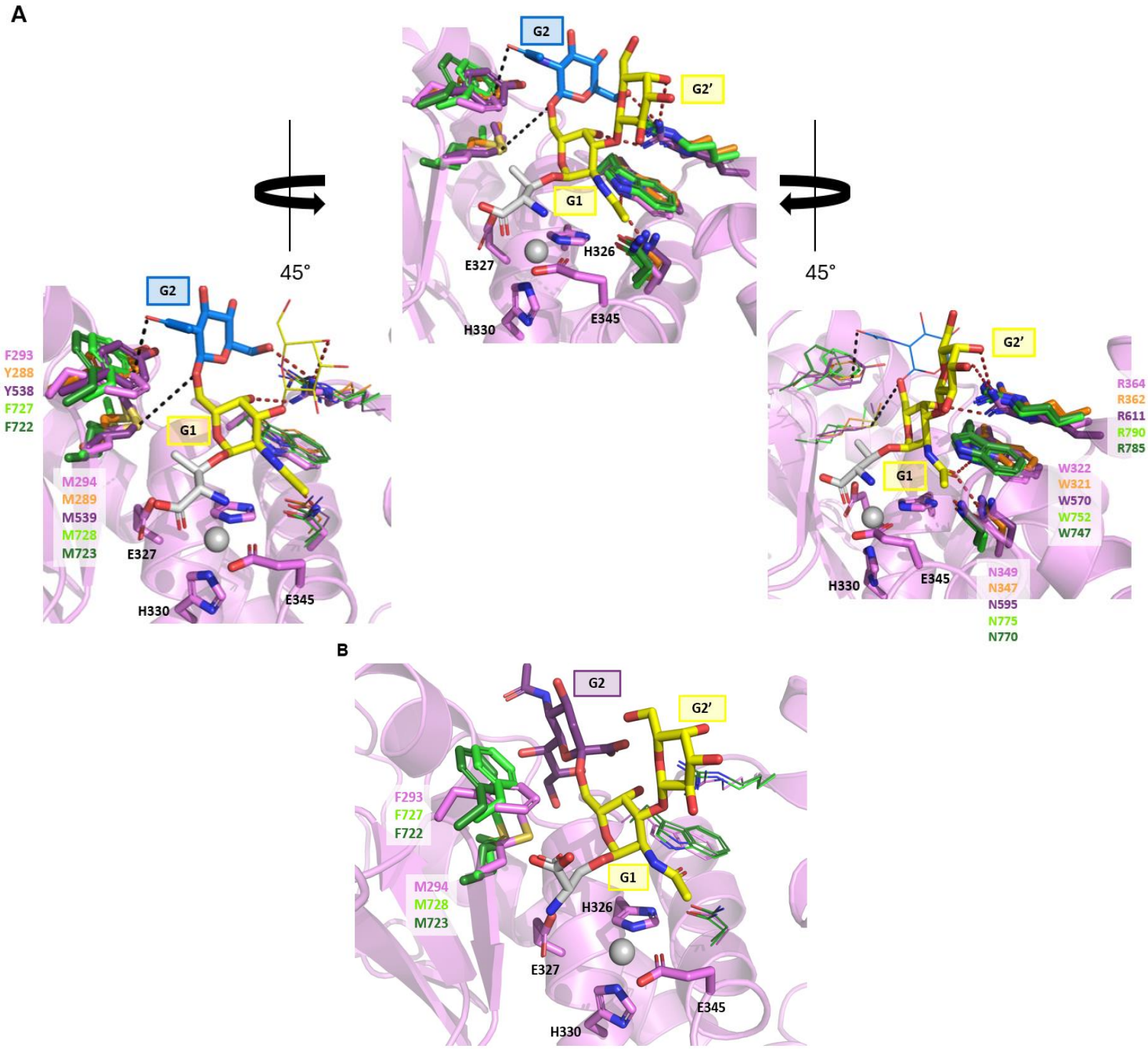


Figure 4.6: BcM60_G shares key residues with AMUC_0627, BT4244, and ZmpB/C.

Residues of interest shown as sticks, while notable residues are shown in line form. Glycans that are not the focus of the panel are also shown in line form. Dashes denote distances between BcM60_G and ligand. Black dashes signify bonds that are within 5 Å, with Phe293 being 3.4 Å away, while red bonds indicate hydrogen bonds. BcM60_G gluzincin motif labeled in black and shown for orientation. Glycans shown in blue (GlcNAc), purple (Neu5Ac), and yellow (Gal, GalNAc) sticks according to the glycan symbol nomenclature conventions. **A)** Core 2 interactions and shared residues between BcM60_G (pink) and AMUC_0627 (orange) (**PDB:** 7YX8), BT4244 (purple) (**PDB:** 5KD8), and ZmpB/C (light/dark green) (**PDB:** 5KDU/6XT1). **Right:** Phe/Met residues near G2. **Left**: Conservation of Trp/Asn/Arg residues at G1 and G2' subsite. **B)** Overlay with 6SC1 from ZmpB.

4.3. Discussion

Bacteroides caccae possesses sixteen putative M60-like peptidases in its genome, fifteen of which are associated with thirteen polysaccharide utilization loci (PULs) (Terrapon et al., 2018). Although the PUL system is a major nutrient acquisition strategy for this genus (Feng et al., 2022; Grondin et al., 2017; D. Ndeh et al., 2020; Overbeeke et al., 2022; Terrapon et al., 2018), there are remarkably few homologous PULs shared between *Bacteroides* spp., suggesting that each member occupies a distinct glycan niche. As *B. caccae* has several M60-like peptidases in its genome (**Figure 4.2**), this suggests that this microbe specializes in degrading glycoproteins, such as mucin and aggrecan. Both glycoproteins contain distinct glycan structures that protect against hydrolysis of the peptide backbone. Mucin is heavily *O*-glycosylated with four main core structures that extend off a Tn-antigen (**Figure 1.2**) (Brockhousen et al., 2022). In contrast, proteoglycans, such as aggrecan, contain *O*-linked oligosaccharides throughout its peptide backbone alongside and GAG attachments; negatively charged, linear heteropolysaccharides that are typically heavily sulfated (**Figure 1.5**) (Kiani et al., 2002; Köwitsch et al., 2018). Although the ability of the HGM to utilize complex carbohydrates on glycoproteins is a major research area that has been ongoing for decades (Desai et al., 2016; Hou et al., 2022; Klassen et al., 2021; Marcobal et al., 2013; Soldán et al., 2024; Wardman et al., 2022), its ability to target the core protein component of these glycoproteins remains relatively understudied in comparison. Thus, to understand how CAZymes and peptidases work together to degrade glycoproteins, we must investigate the molecular mechanisms by which PUL-associated M60-like peptidases recognize and cleave their substrates.

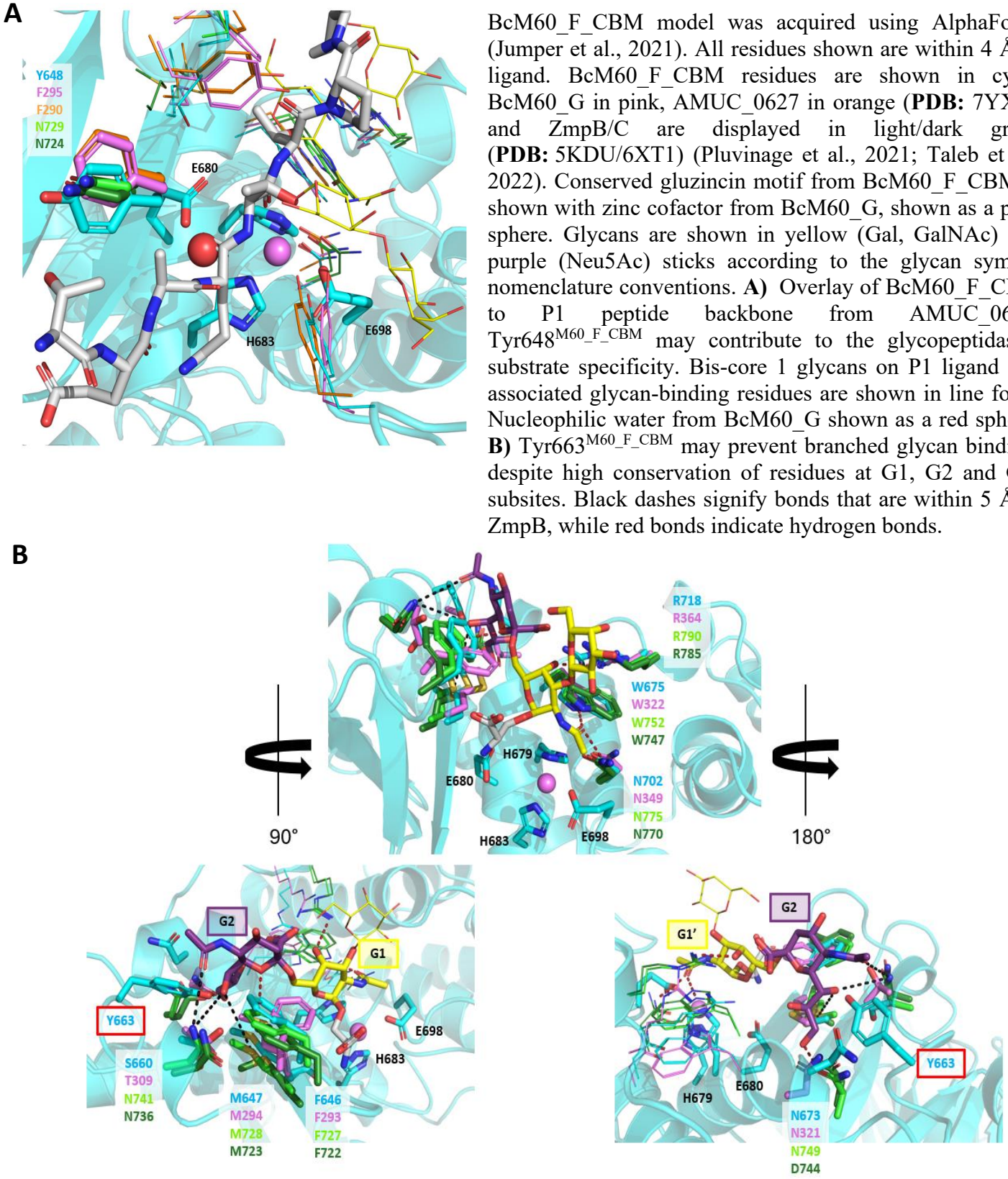
Initial work on the catalytic unit of BcM60_F and BcM60_G demonstrated that both are glycopeptidases that are active on bovine submaxillary mucin (BSM), wherein peptide cleavage occurs upon recognition of specific *O*-glycans immediately N-terminal adjacent to the scissile bond (**Figure 4.1**) (Deventer, 2021). However, due to the heterogeneity of mucin oligosaccharides (**Figure 1.2**), it can be challenging to identify the specific glycan structure/s recognized by each enzyme (J. Kim et al., 2020). The development and use of FRET substrates with defined glycosylated linkers have greatly enhanced sensitivity in detecting glycopeptidase activity, especially as BcM60_F displayed measurable activity on the available substrates only after inclusion of the CBM domains (**Figure 4.4** and **Figure S4.2**) (Arai et al., 2003; Boraston et al., 2004; Wardman et al., 2023). However, it should be noted that the FRET substrates used here do

not represent the full diversity of glycosylated peptides found in nature. Given its activity on BSM, the catalytic domain of BcM60_F is clearly able to function independently of its CBM and BACON domains (Deventer, 2021). However, we were unable to replicate this observation since BSM-specific FRET substrates do not exist yet. Future work should focus on incorporating a broader range of glycoforms across more peptide linkers, as this will help further clarify the substrate specificities of putative and previously characterized glycopeptidases. Nonetheless, the findings presented here, especially in conjunction with structural analyses, offer valuable insight into the substrate specificities of BcM60_F and BcM60_G, and adds to our understanding of both glycopeptidases.

Curiously, of the 44 mucin-specific FRET substrates, BcM60_F_CBM was primarily active on MUC5A/C containing linear glycosylation, with minimal activity on MUC5B and MUC8 (**Figure 4.4**). However, MUC5A/C, MUC5B, and MUC8 are all found along the respiratory tract, with MUC5A/C also being reported in the stomach (Hansson, 2020; Paone & Cani, 2020). Although *Bacteroides* spp. have been found associated with the oropharynx, it is not considered a dominant taxon, and the anaerobic *B. caccae* would be unlikely to survive in this environment (Johnson et al., 1987; Natalini et al., 2023). Consequently, this result suggests that BcM60_F_CBM shows more stringent peptide specificity, and/or requires different glycan structures than the ones tested to achieve activity on gut-specific mucin peptide linkers. The former has been previously observed in ZmpB and ZmpC, from *Clostridium perfringens* (Noach et al., 2017; Pluvinage et al., 2021), which share identical G' subsites but differing S' subsite residues, which are responsible for interacting with the peptide backbone. This resulted in different kinetic properties between the two glycopeptidases for the same substrates. To date, specific S residues have not been identified in a M60-like peptidase. However, superposition of the BcM60_F_CBM Alphafold model on AMUC_0627 complexed with P-selectin glycoprotein ligand 1 (PSGL-1)-like bis-Core 1 glycopeptide (hereafter termed P1; TEAQT**T**PPPA in which ** denotes a Galβ1-3GalNAc disaccharide) (**PDB**: 7YX8) shows a Tyr unique to BcM60_F_CBM ($Tyr648^{BcM60\_F\_CBM}$/$Phe295^{BcM60\_G}$/$Phe290^{AMUC\_0627}$/$Asn729^{ZmpB}$/$Asn724^{ZmpC}$) near the substrate peptide backbone and the gluzincin motif (**Figure 4.7A)**. Depending on conformational changes that occur upon glycan recognition, this residue may contribute to BcM60_F's more stringent peptide specificity.

In contrast with its counterpart, BcM60_G recognizes numerous glycoforms on multiple peptide linkers, including Core 2 (GlcNAcβ1-6(Galβ1-3)GalNAcα-Ser/Thr) and *N*-acetylneuraminic acid (Neu5Ac)-modified substrates (**Figure 4.6**) (**Table 4.1**). Previous structural work with BcM60_G has successfully yielded an apo-form and a complex bound to Core 1-Thr, recapitulating the conserved interactions between the GalNAc and the Trp/Asn/Arg residues in the G1 subsite seen in other M60-like peptidases (Deventer, 2021; Noach et al., 2017; Pluvinage et al., 2021; Taleb et al., 2022). In addition to supporting these observations, this BcM60_G-Core 2 complex revealed Phe293 at the G2 subsite that is close enough to form CH-π bonds with the GlcNAc (**Figure 4.5**). Overlaying BcM60_G onto ZmpB/C complexed with 6SC1 (**PDB**: 5KDU/6XT1) (RMSD: 2.31/2.47 Å) revealed this same Phe residue at this subsite, corroborating BcM60_G activity on 6SC1 (**Figure 4.6B**).

Figure 4.7: BcM60_F_CBM *in silico* analysis with other glycopeptidases.



BcM60_F_CBM model was acquired using AlphaFold2 (Jumper et al., 2021). All residues shown are within 4 Å of ligand. BcM60_F_CBM residues are shown in cyan, BcM60_G in pink, AMUC_0627 in orange (**PDB:** 7YX8), and ZmpB/C are displayed in light/dark green (**PDB:** 5KDU/6XT1) (Pluvinage et al., 2021; Taleb et al., 2022). Conserved gluzincin motif from BcM60_F_CBM is shown with zinc cofactor from BcM60_G, shown as a pink sphere. Glycans are shown in yellow (Gal, GalNAc) and purple (Neu5Ac) sticks according to the glycan symbol nomenclature conventions. **A)** Overlay of BcM60_F_CBM to P1 peptide backbone from AMUC_0627. $Tyr648^{M60\_F\_CBM}$ may contribute to the glycopeptidase's substrate specificity. Bis-core 1 glycans on P1 ligand and associated glycan-binding residues are shown in line form. Nucleophilic water from BcM60_G shown as a red sphere. **B)** $Tyr663^{M60\_F\_CBM}$ may prevent branched glycan binding, despite high conservation of residues at G1, G2 and G2' subsites. Black dashes signify bonds that are within 5 Å of ZmpB, while red bonds indicate hydrogen bonds.

Sequence analysis of BcM60_G using BLAST (UniProt) revealed significant sequence identity (50.6%) to AMUC_0627, an extensively characterized M60 peptidase from *A. muciniphila* (Shon et al., 2020, 2022; Taleb et al., 2022). Superposition of BcM60_G onto an AMUC_0627-P1 complex (**PDB**: 7YX8) revealed an active site that closely resembles each other (RMSD: 1.33 Å) (**Figure 4.8**) (Taleb et al., 2022). Besides the gluzincin motif and the G1 subsite residues (Trp/Asn/Arg) shared by all characterized M60 peptidases to date (**Figure 4.5A, Figure 4.6A, Figure 4.7B**), BcM60_G contains many of the side chains used by AMUC_0627 to orient its ligand in the active site. $Phe293^{BcM60\_G}$ is an exception to this, as its orientation more closely resembles the corresponding residues in ZmpB/C ($Phe727^{ZmpB}$/$Phe722^{ZmpC}$), which establishes CH-π interactions with the Neu5Ac moiety of 6SC1 (**Figure 4.6**) (Noach et al., 2017; Pluvinage et al., 2021; Taleb et al., 2022). The lack of an equivalent residue in AMUC_0627 and BT4244 was previously proposed by another study to prevent their activity on 6SC1-substrates (Taleb et al., 2022).

Curiously, the impact of Phe versus Tyr on 6SC1 binding was not recapitulated with BcM60_F_CBM (**Figure 4.7B**). Despite having an equivalent Phe residue at the same position that allowed BcM60_G and ZmpB/C to bind the Neu5Ac moiety of the 6SC1 glycan structure ($Phe646^{BcM60\_F\_CBM}$/$Phe293^{BcM60\_G}$/$Phe727^{ZmpB}$/$Phe722^{ZmpC}$), BcM60_F_CBM was not active on any tested 6SC1-containing substrates (**Table 4.1**). This result may reflect the glycopeptidase's proposed peptide specificity and the current limitations of the available FRET substrates. However, this observation may also be explained by a Tyr residue in BcM60_F (Tyr663), which is absent in the G2 subsite of BcM60_G and ZmpB/C (**Figure 4.7**). This Tyr's positioning may sterically hinder binding at this subsite entirely, potentially accounting for BcM60_F_CBM's lack of activity towards branched glycan substrates, such as 6SC1 and Core 2.

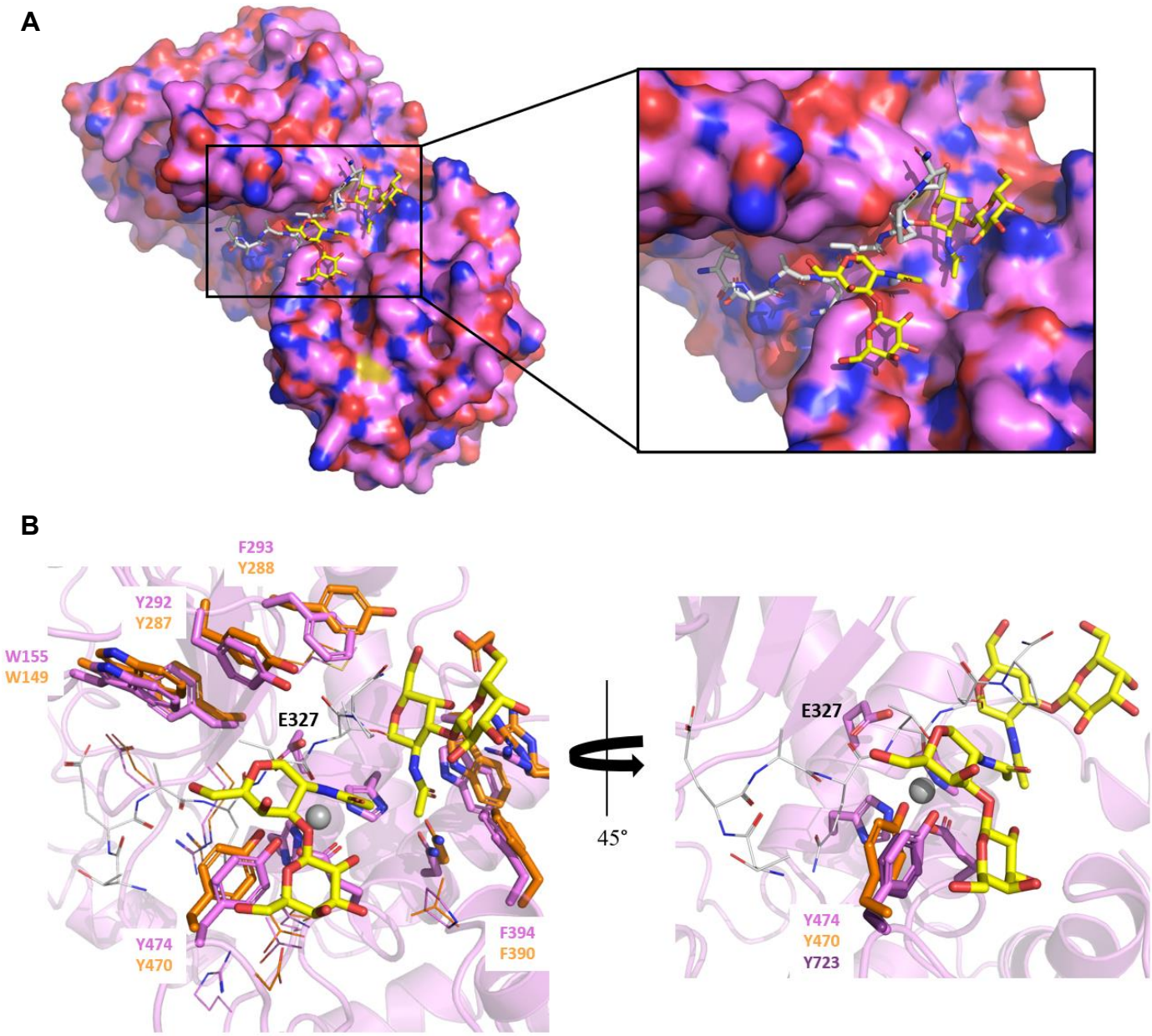


Figure 4.8: BcM60_G may act on bis-glycosylated peptide substrates.

BcM60_G residues are shown in pink, while AMUC_0627 (**PDB:** 7YX8) (RMSD: 1.33 Å) and BT4244 (**PDB:** 5KD8) (RMSD: 2.15 Å) are in orange and purple, respectively (Noach et al., 2017; Taleb et al., 2022). Full view of active site with BcM60_G, AMUC_0627, and BT4244 alignment shown in **Figure S4.4.** Glycans are shown in yellow (Gal, GalNAc) sticks according to the glycan symbol nomenclature conventions, while substrate peptides are in grey. **A)** Electrostatic surface view of BcM60_G aligned to P1 ligand from AMUC_0627. Black box zooms into the ligand in the binding pocket. **B)** Active site similarities between BcM60_G and AMUC_0627. Residues within 4 Å of ligand are visible, wherein aromatic side chains and glycan binding residues are labelled and shown in stick form. Similarly, the ligand peptide backbone, as well as predicted peptide-binding residues, are shown in line form. BcM60_G gluzincin motif shown and partially labelled for orientation.

BcM60_G's activity on 3SC1-modified FRET substrates may be attributed to a neutral charge on the surface close to the G2' subsite, where Neu5Ac is positioned in 3SC1 (**Figure 4.9B**) (**Table 4.1**). In contrast, AMUC_0627's inability to accommodate sialyated substrates was attributed to a negative surface charge that would repel this moiety (**Figure 4.9C**) (Taleb et al., 2022). Interestingly, AlphaFold modelling of BcM60_F_CBM's showed that it has an overall positive surface charge in this region, bordered with electronegative protrusions and seemingly less spatial availability (**Figure 4.9A**). While this observation is based on an untested *in silico* model rather than an experimentally derived structure, the positive charge in this region compared to BcM60_G may explain BcM60_F_CBM's relatively high catalytic efficiency on 3SC1, despite containing protrusions at the Neu5Ac position that are not seen in the other glycopeptidases (**Table 4.1**).

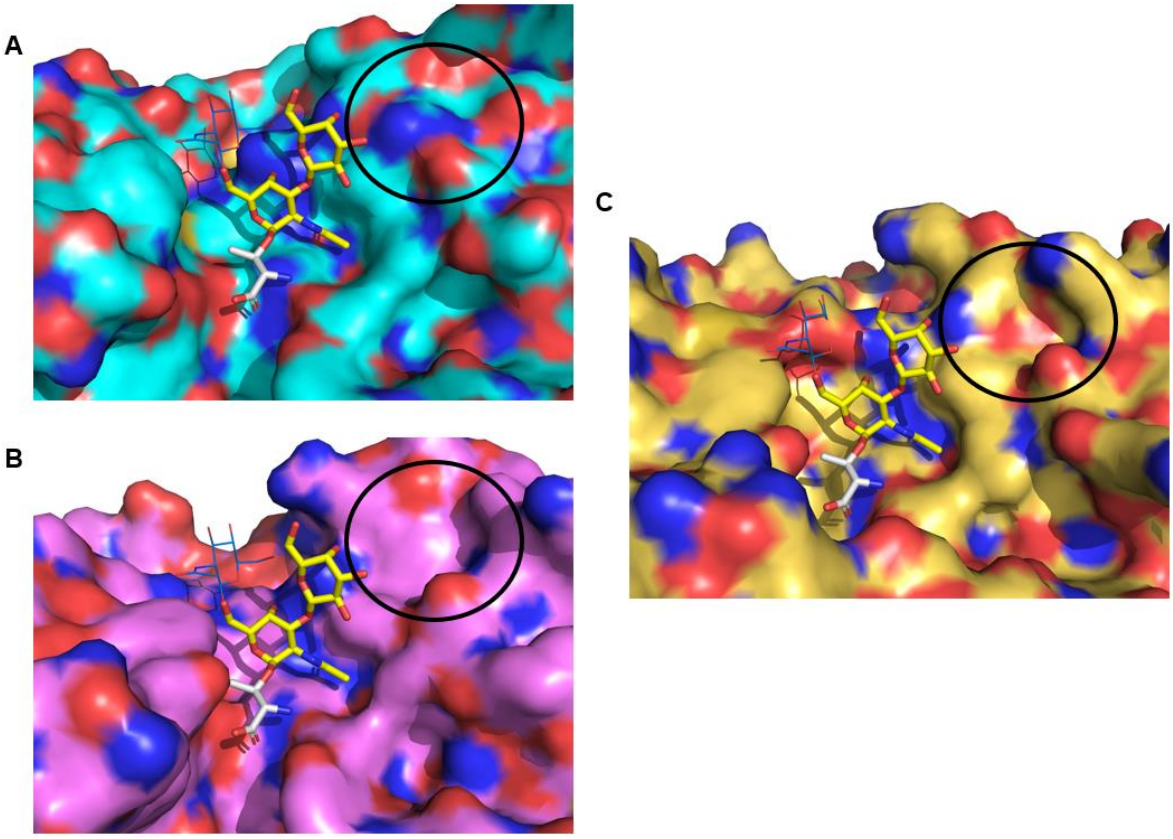


Figure 4.9: Overall surface charge near the G2' subsite may determine 3SC1-binding ability.

BcM60_F_CBM residues are shown in cyan (**A**), BcM60_G in pink (**B**), while AMUC_0627 (**PDB:** 7YX8) is in gold (**C**). Core 2 glycan was colored blue (GlcNAc) and yellow (Gal, GalNAc) according to the glycan symbol nomenclature conventions, while substrate amino acid is in grey. GlcNAc moiety is presented in line form, as it is not the focus of this figure; the GalNAc and Gal are shown in stick form. Black circle indicates the area where Neu5Ac would be located in a 3SC1 moiety.

Functional and structural data have supported AMUC_0627 recognition of adjacently glycosylated Core 1 (bis-Core 1) and bis-Tn peptides, with a preference for the former glycoform (Shon et al., 2020; Taleb et al., 2022). Clustered *O*-glycan stretches are commonly found on mucin and mucin-like domains but are especially prominent on MUC2 and MUC5A/C, often with stretches of 3-6 adjacent *O*-glycans (Schjoldager et al., 2020; Taleb et al., 2022). This distinctive glycan recognition specificity of AMUC_0627 was confirmed via mass spectrometry and structural analysis with P1 (Shon et al., 2020; Taleb et al., 2022). Remarkably, not only is BcM60_G able to accommodate the full P1 ligand (**Figure 4.8A**), but it also shares an equivalent Tyr474 ($Tyr470^{AMUC\_0627}$/$Tyr723^{BT4244}$) that was shown to be critical for activity on bis-Core 1 glycopeptides. This residue interacts with both GalNAc and Gal moieties on the Core 1 moiety N-terminal to the scissile bond; its mutation to Ala was shown in AMUC_0627 to abolish activity (**Figure 4.8B**) (Taleb et al., 2022). Given this, it is likely that BcM60_G can cleave bis-glycosylated peptides, although this will require validation via activity-based assays, mass spectrometry, and/or complexed structure acquisition with bis-glycosylated substrates.

Overall, the results presented substantially extend the previous work done with BcM60_F and BcM60_G (Deventer, 2021). It has expanded our knowledge surrounding their respective substrate specificities, as well as substantiated the foundation on how *B. caccae* interacts with its glycoprotein substrates. Functional glycopeptidase activity screening using FRET revealed that BcM60_G can accommodate numerous glycan structures, whereas BcM60_F exhibited more stringent peptide specificity than its counterpart and could only bind to linear glycan moieties. Structural analysis of BcM60_G complexed with Core 2 revealed that its G2 subsite contains a Phe residue shared by ZmpB/C, which have demonstrated activity on 6SC1. This suggests that this residue is necessary for accommodating branched glycans. However, there are exceptions to this observation. Despite having an equivalent Phe in this position, a unique Tyr663 from BcM60_F_CBM may obstruct binding at G2 subsite. Superposition of BcM60_G to AMUC_0627 complexed with P1 revealed remarkably similar active site architectures, with key exceptions that allow BcM60_G to tolerate sialylated glycans better than AMUC_0627, as well as the potential for this glycopeptidase to target bis-glycosylated peptides (Taleb et al., 2022). Likewise, structural comparisons with BcM60_F highlighted Tyr648 and Tyr663 as notable residues that potentially contributes to its peptide specificity and inability to tolerate branched glycans, respectively

(Pluvinage et al., 2021; Taleb et al., 2022). Although these insights were based on an *in silico* AlphaFold model (Jumper et al., 2021), they offer a plausible explanation for the observed kinetic differences between the two BcM60 peptidases against the same substrates. Future analyses of BcM60_F, especially through the mutagenesis of Tyr648 and Tyr663 would help clarify the factors that influence its substrate specificity. Additionally, mass spectrometric analysis of both BcM60-like peptidase activity on glycoproteins would provide valuable insight into uncovering their cleavage motifs and clarify both their glycan and peptide specificity (Shon et al., 2020).

4.4. Supplementary Figures and Tables

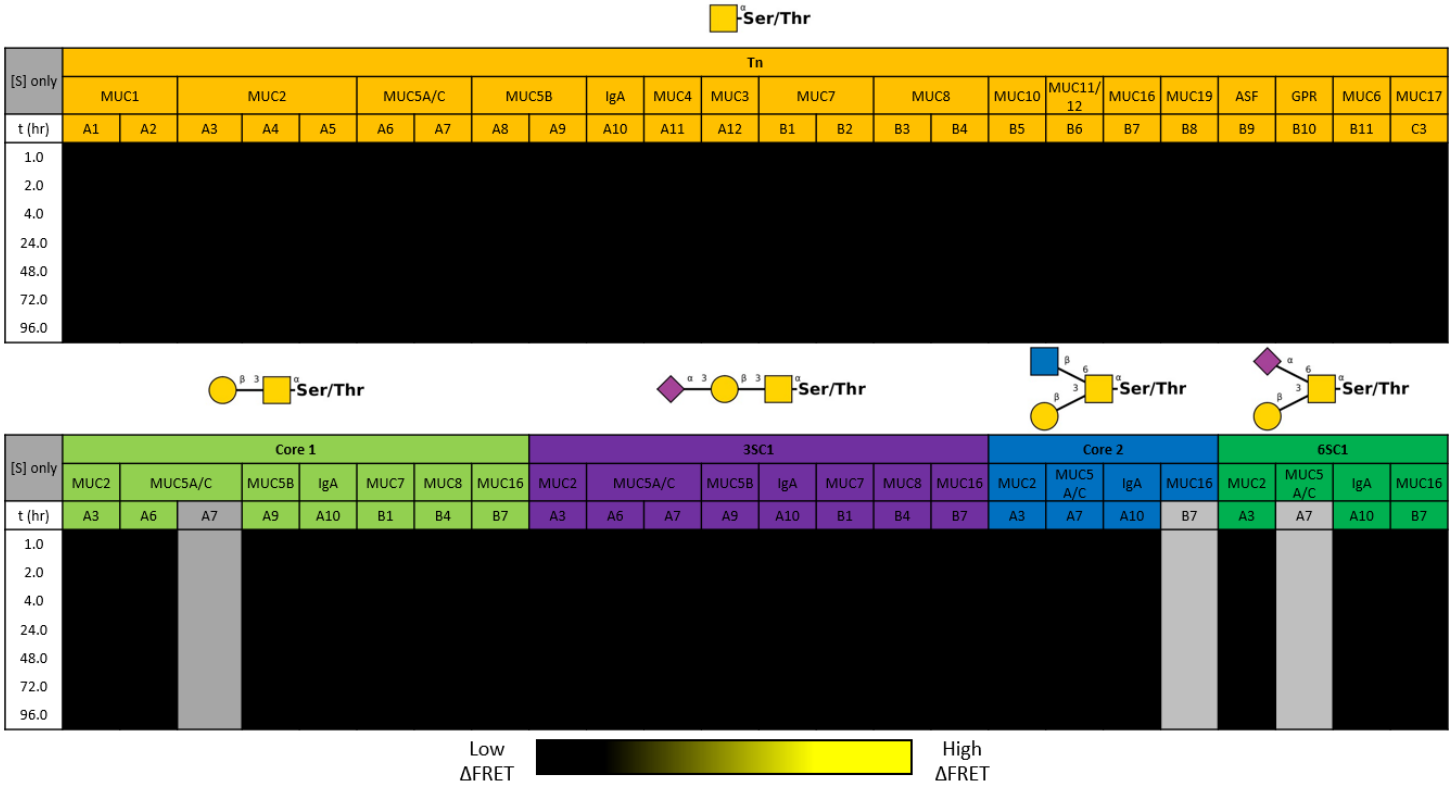


Figure S4.1: Substrate-only control for glycopeptidase FRET results in Figure 4.5.

All reactions done in duplicate alongside BT4244, BcM60_F_CBM, and BcM60_G-treated samples. Greyed out boxes indicate unavailable substrate.

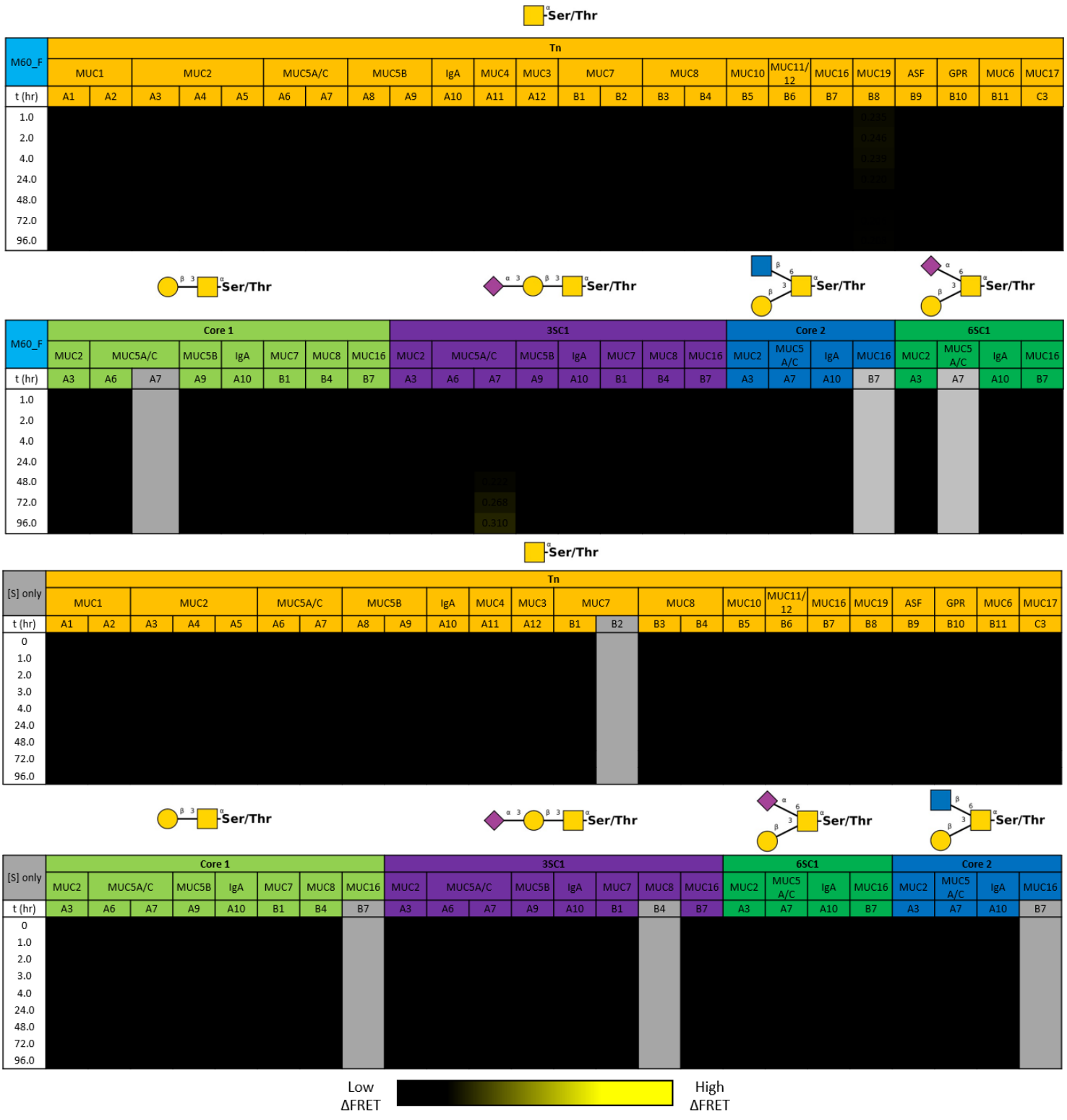


Figure S4.2: Qualitative assessment of BcM60_F glycan recognition.

All reactions done in duplicate. Greyed out boxes indicate unavailable substrate.

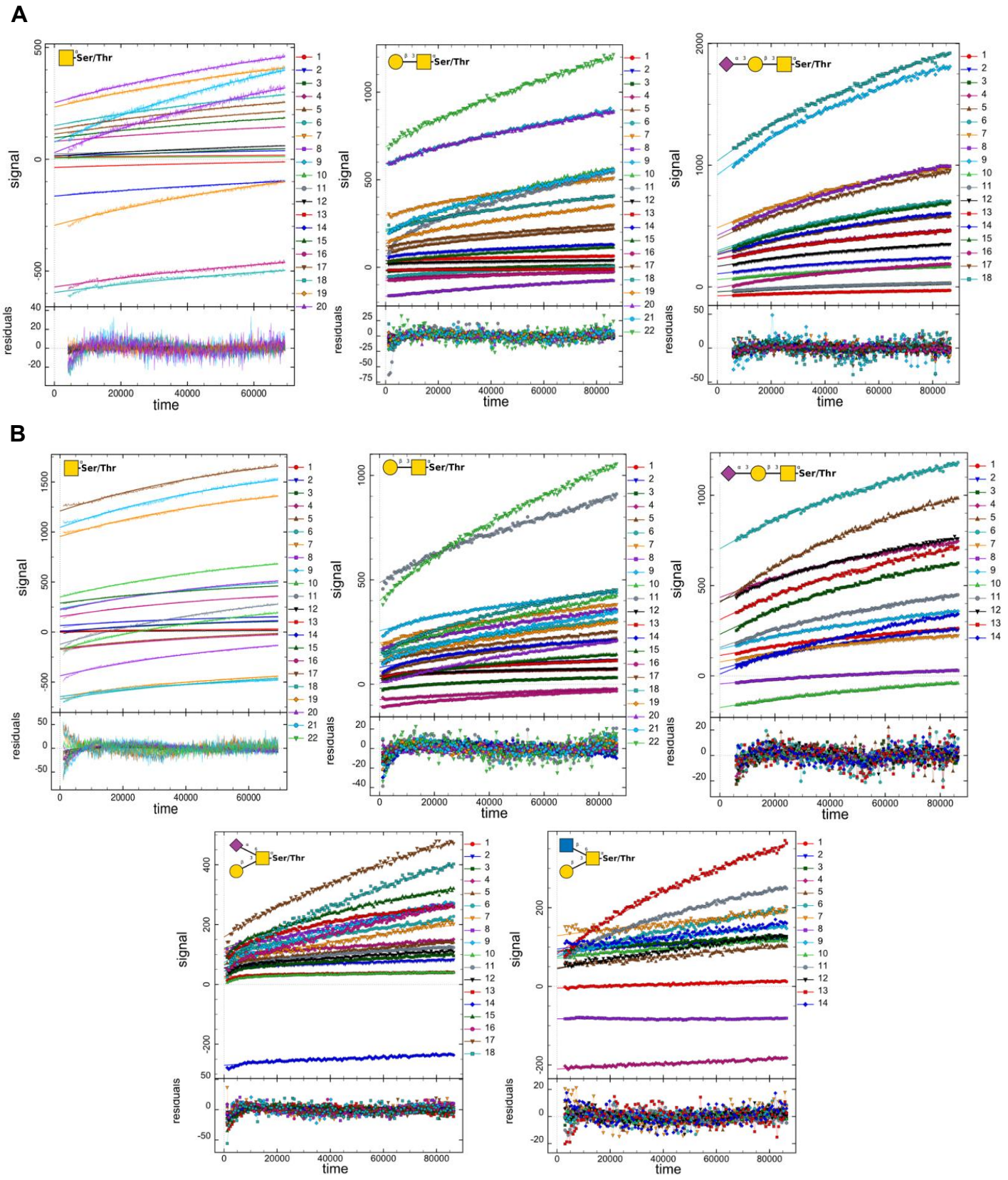


Figure S4.3: DynaFit kinetic curves and residual plots.

Reactions were done in duplicate. A simplified Michalis-Menton mechanism was used to model eleven progress curves at different substrate concentrations for global analysis (Kuzmič et al., 2010). **A)** BcM60_F_CBM. **B)** BcM60_G.

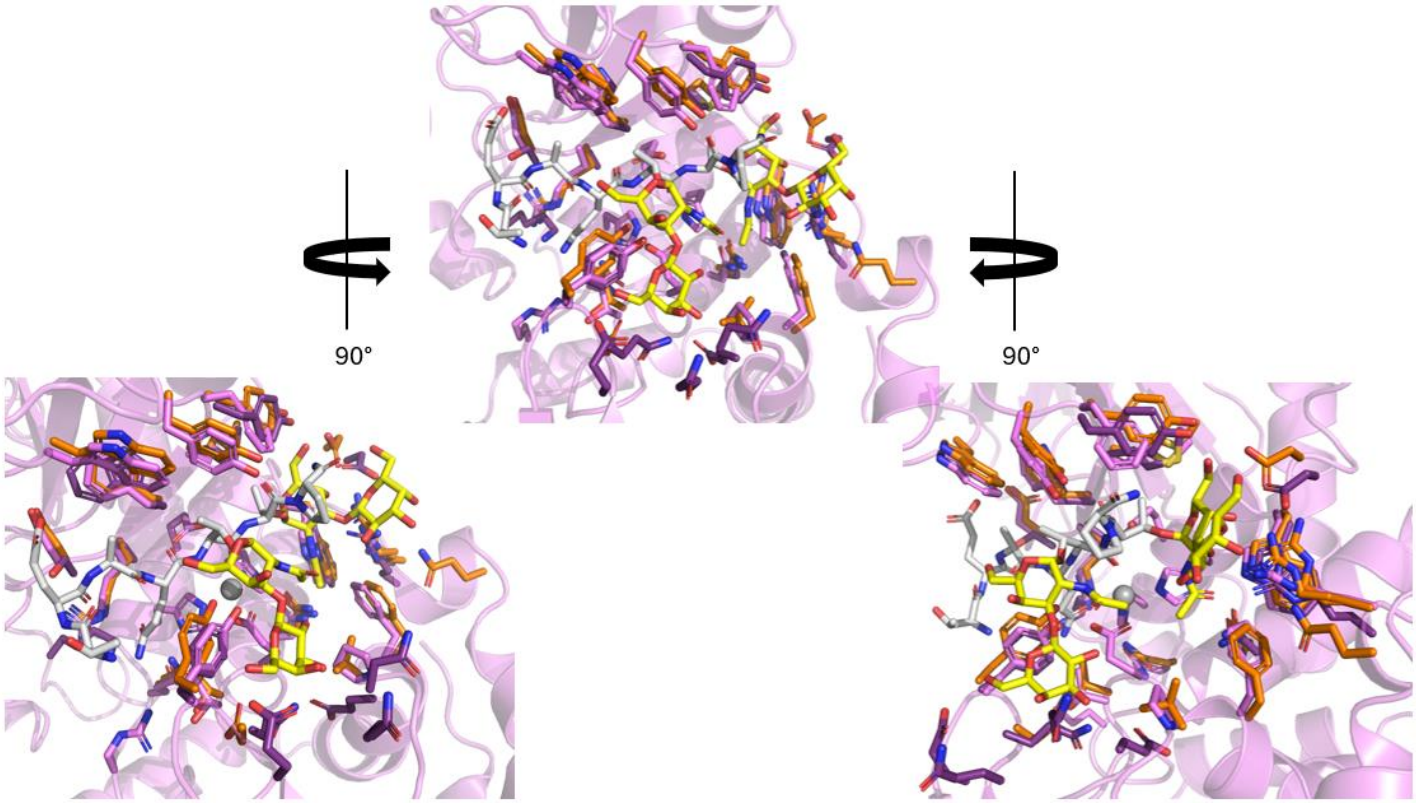


Figure S4.4: Full stick view of BcM60_G alignment with AMUC_0627 with P1 ligand.

BcM60_G residues are shown in pink, while AMUC_0627 (**PDB:** 7YX8) (RMSD: 1.33 Å). All residues shown are within 4 Å of P1 ligand.

Table S4.1: BcM60_G_E327Q x-ray collection statistics.

	BcM60_G_ E327Q
	+ Core 2 - Thr
Data Collection	
Wavelength	1.541
Space Group	C_2
Cell Dimensions	
a, b, c (Å)	174.553, 39.386, 71.085 (β=103.95)
Resolution (Å)	20.00-1.85 (1.88-1.85)
R_{meas}	0.073 (0.536)
R_{pim}	0.032 (0.335)
CC1/2	0.999 (0.928)
<I/σI>	19.6 (2.0)
Completeness (%)	97.0 (76.4)
Redundancy	4.2 (2.2)
No. of Reflections	172,297
No. Unique	40,593
Refinement	
Resolution (Å)	1.85
R_{work}/R_{free}	0.163/0.194
No. of Atoms	3,856
Protein	3,466
Ligand	46
Water	341
B-factors	22.11
Protein	21.05
Ligand	37.82
Water	30.87
r.m.s.d.	
Bond Lengths (Å)	0.007
Bond Angles (°)	0.826
Ramachandran (%)	
Preferred	98.4
Allowed	1.62
Disallowed	0.0
PDB ID	n/a

Table S4.2: Summarized catalytic efficiency of BcM60_F_CBM and BcM60_G at 1.0 µM and 0.5 µM.

Construct	[E] (µM)	Linker ID	Glycosylation	k_{cat}/K_M (min^{-1} mM^{-1}) (x 10^{-5})	CV (%)	95% CI (low, high)
BcM60_F_CBM	1.0	MUC5A/C	Tn	0.64 ± 0.01	6.8	0.56, 0.73
	1.0		Core 1	0.56 ± 0.01	4.0	0.51, 0.60
	1.0		3SC1	0.73 ± 0.01	2.2	0.69, 0.77
			Core 2	n/a	n/a	n/a
			6SC1	n/a	n/a	n/a
BcM60_F_CBM	0.5	MUC5A/C	Tn	0.73 ± 0.01	11.2	0.57, 0.89
	0.5		Core 1	0.78 ± 0.01	5.0	0.71, 0.86
	0.5		3SC1	1.10 ± 0.01	4.6	1.00, 1.19
			Core 2	n/a	n/a	n/a
			6SC1	n/a	n/a	n/a
BcM60_G	1.0	MUC5A/C	Tn	0.93 ± 0.01	4.1	0.85, 1.00
	1.0		Core 1	0.67 ± 0.01	3.3	0.63, 0.71
	1.0		3SC1	0.75 ± 0.01	3.2	0.70, 0.80
	1.0		Core 2	0.57 ± 0.02	8.9	0.47, 0.68
	1.0		6SC1	0.59 ± 0.01	6.5	0.39, 0.66
BcM60_G	0.5	MUC5A/C	Tn	0.80 ± 0.03	19.9	0.48, 1.12
	0.5		Core 1	0.68 ± 0.04	14.9	0.47, 0.88
	0.5		3SC1	1.10 ± 0.03	6.9	0.93, 1.23
	0.5		Core 2	0.86 ± 0.09	28.4	0.38, 1.36
	0.5		6SC1	0.60 ± 0.01	25.7	0.29, 0.92

Chapter 5 : Final Discussion

5.1. BcPUL25 targets CS-containing proteoglycans.

The functional and structural analyses presented in Chapter 3 revealed that the CAZymes associated with PUL25 from *Bacteroides caccae* (BcPUL25) primarily target desulfated CS, or chondroitin, through a unique pathway that pairs a dehydratase (BcGDH, or BcGH154) with an *exo*-uronyl hydrolase (BcGH88), to remove a non-reducing end uronate monosaccharide (**Figure 3.8**). This was observed only once before in ulvan degradation, wherein P29_PDnc, a dehydratase unrelated to BcGDH, functions in tandem with a GH105 *exo*-uronyl hydrolase (Bäumgen et al., 2021). Additionally, we demonstrated that the two M60-like peptidases associated with BcPUL25 differentially recognize *O*-glycans as a requisite to peptide hydrolysis (**Table 4.1**). Using a glycopeptidase-specific FRET screen, this work deepened our understanding on their respective substrate specificities (**Figure 4.4**) and contributes to the foundational knowledge on how *B. caccae* interacts with its glycoprotein substrates (**Figure 4.5** and **Figure 4.7**) (Deventer, 2021; Wardman et al., 2023). Given that PUL-associated enzymes are co-regulated and co-localized to collaboratively depolymerize the same substrates, BcPUL25 is hypothesized to collectively target CS-containing proteoglycans (**Figure 1.11**).

Unlike GAGs, which can be found in the mucosal layer or embedded on the cellular surface of epithelial cells, any host-associated proteoglycan in a healthy human gut would be inaccessible for the HGM to degrade, as these glycoproteins are located within the intestinal ECM rather than secreted into the intestinal lumen (Merry et al., 2022; Overbeeke et al., 2022; Rawat et al., 2022). Thus, HGM degradation of proteoglycans would primarily be from dietary sources, such as animal meat (**Figure 1.3** and **Figure 1.5**). Connective tissues, such as cartilage and tendon, are widely used as a culinary ingredient in numerous cuisines around the world. As aggrecan is the major proteoglycan found in articular cartilage, and ~90% of its mass is mainly comprised of CS chains, we propose that dietary aggrecan is the primary target of BcPUL25 (Kiani et al., 2002).

In cartilage, aggrecan is secreted by chondrocytes (Kiani et al., 2002). The proteoglycan is typically found in multimolecular aggregates that are non-covalently bound to a central filament of HA, with each aggrecan-HA interaction stabilized with a link protein (**Figure 5.1**). It has three globular domains (G1, G2, G3) and three extended domains: interglobular domain (IGD), KS, and

CS-rich domain (Kiani et al., 2002; Merry et al., 2022; Plaas et al., 2023). The N-terminal G1, IGD, and G2 domain sequences are conserved in vertebrates, with the two GAG domains varying in length between species (Flannery et al., 1998; Kiani et al., 2002; Roughley & Mort, 2014). The C-terminal G3 contains a variable structure due to alternative splicing of its own domains.

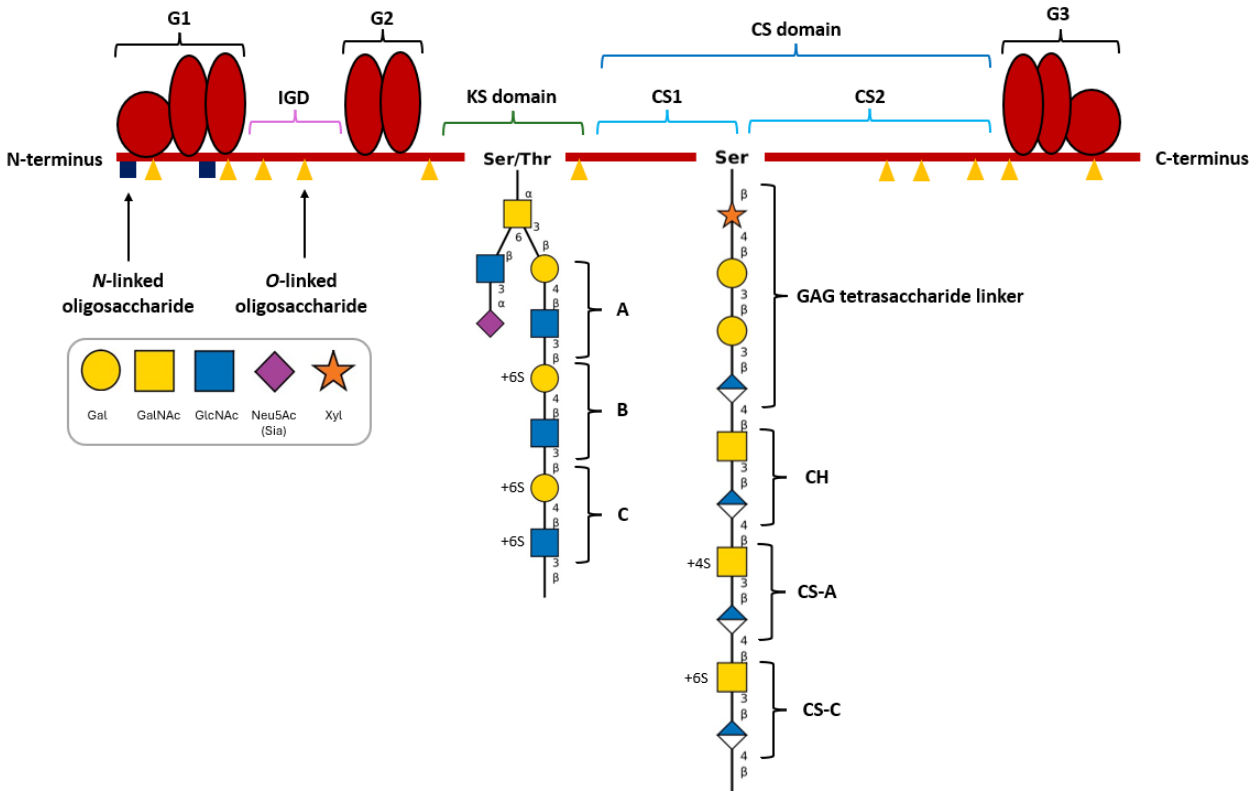


Figure 5.1: Structure of aggrecan.

General aggrecan structure acquired from (Kiani et al., 2002; Klein et al., 2018; Plaas et al., 2023) KS and CS disaccharide unit designations determined by sulfation level present (see **Figure 1.5**). Glycan structures made using DrawGlycan-SNFG (Cheng et al., 2016). *Abbreviations*: Gal: galactose, GalNAc: *N*-acetylgalactosamine, GlcNAc: *N*-acetylglucosamine, Neu5Ac: *N*-acetylneuraminic acid, Xyl: xylose.

5.1.1. Degradation of aggrecan CS chains.

Of the two GAG regions in aggrecan, the CS domain is the largest, with up to 100 CS chains attached that are ~20 kDa each (Roughley & Mort, 2014). It is subdivided into two domains (CS1 and CS2) and contains multiple repeats of a nineteen amino acid sequence containing Ser-Gly dipeptides flanked by an acidic and a hydrophobic residue (Kiani et al., 2002; Rodriguez et al.,

2006). The recognition sequence for the initial *O*-xylose (Xyl) attachment by xylose glycotransferase 1 and 2 (XYLT-1/2) is a-a-a-a-G-S-G-a-(A/G)-a, wherein "a" refers to an Asp or Glu residue. These CS chains on aggrecan are attached to the core protein backbone via a GAG tetrasaccharide linker (GlcAβ1–3Galβ1–3Galβ1–4Xylβ1-*O*-Ser) (**Figure 5.1**) (Götting et al., 2007; Schjoldager et al., 2020). The type and degree of sulfation on these CS chains is dependent on host age, species, tissue location, and attachment site (Plaas et al., 2023; Rodriguez et al., 2006). Typically, the CS chains in terrestrial animals decreases up to 50% in length with increasing age and transition into predominantly CS-C units, as opposed to maintaining relatively equal abundance of CS-A and CS-C (**Figure 1.5**) (Plaas et al., 2023).

Human diets include a wide range of animal protein sources, with the most commonly consumed being pork (*Sus scrofa*), beef (*Bos taurus*), poultry (*Gallus gallus*), and varying species of fish. Despite major taxonomic differences, the cartilage in these species have the same predominant CS units, differing only in their disaccharide ratio (**Figure 1.5**) (Shionoya et al., 2022). The major CS disaccharide units in terrestrial animal cartilage are CS-A (50-70%) and CS-C (20-40%), with the ratio between the two units (CS-A:CS-C; 4S:6S) at 4.5-7.00 in pigs, 1.50-2.00 in cows, and 3.00-4.00 in chickens. Similarly, CS isolated from bony fish (class Osteichthyes) or cartilaginous fish (clade Chondrichthyes) have CS-A (13-29%) and CS-C (31-71%), but at lower unit ratios, with 0.45 in salmon, 0.36-0.69 in sturgeon, and 0.45-0.70 in sharks.

The fermentation of GAGs, such as CS/DS, by *Bacteroides* spp. is a well-studied topic, especially since its degradation is not widely prevalent in other gut phyla (Alvarez et al., 2025; Cartmell et al., 2017; D. Ndeh et al., 2018, 2020; Overbeeke et al., 2022; Rawat et al., 2022; Ulmer et al., 2014; L. Wei, Zou, et al., 2024). One seminal paper describes a versatile PUL from *B. thetaiotaomicron* that can metabolize CS, DS, and HA ($BtPUL_{CS/DS/HA}$) (**Figure 1.5**) (D. Ndeh et al., 2020). The components of $BtPUL_{CS/DS/HA}$ were able to independently target CS-A, CS-C, CS-E, and DS. However, the association of a non-PUL associated sulfatase (2-*O*-sulfatase^{BT1696}) and DS/HA-specific lyase ($PL33^{BT4410}$) enabled the degradation of CS-D and HA, as well as more efficient DS degradation. While this presented a model of GAG metabolism that is broadly applicable to other members of the HGM, it also highlighted that its PUL components are widely conserved in the *Bacteroides* genus. This was observed with the CAZyme components of

BcPUL25 (Alvarez et al., 2025). Analysis of BcPL35 activity revealed its desulfation requirement, despite the absence of sulfatases directly encoded with BcPUL25. To date, 4-*O*-sulfatase^{BT3349} (BtSulf) from $BtPUL_{CS/DS/HA}$ is the only characterized *endo*-acting sulfatase active on CS-A (D. Ndeh et al., 2020; Ulmer et al., 2014). *Bacteroides caccae* contains its ortholog (CGC64_04290; BcSulf) in CAZyme cluster 1, which shares 92% amino acid sequence identity and 100% active site conservation with BtSulf (Terrapon et al., 2018). The activity of these sulfatases was shown to be necessary for BcPL35 activity on CS-A (**Figure 3.1** and **Figure S3.1D)**. A BLAST search of *B. caccae* (ATCC 43185 new assembly) identified 25 putative sulfatases in its genome (Sayers et al., 2025). Therefore, despite the lack of known *endo*-acting sulfatases specific for other CS variants, it is likely that *B. caccae* has appropriate sulfatases that will allow BcPUL25 to act on a broader range of CS substrates. Further characterization of sulfatases involved in GAG degradation will clarify the molecular mechanisms *Bacteroides* spp. and other HGM members use to recognize and process these complex glycans.

5.1.2. Degradation of aggrecan core protein.

Throughout the aggrecan core protein, there are a high variable number of mucin *O*-linked oligosaccharides, estimated at ~120 *O*-linked oligosaccharides per proteoglycan (**Figure 5.1**) (Kiani et al., 2002; Nilsson et al., 1982). These *O*-glycosylated sites are generated by GalNAc glycosyltransferases (GALNT 1-20), which recognize structural, isoform-specific motifs rather than an amino acid sequence (Bennett et al., 2012; Schjoldager et al., 2020). Although little is known on the overall KS and CS synthesis on aggrecan, it is thought that that chondrocyte GALNT and XYLT-1/2 glycotransferases are in competition with each other (Brockhousen et al., 2022; Schjoldager et al., 2020). This was observed in a murine study that overexpressed Galnt-3 in chondrocytes, resulting in dwarfism due to the overabundance of mucin *O*-glycans and significant reduction of GAGs on aggrecan (Yoshida et al., 2014). In healthy cartilage, these *O*-glycans eventually become sites of KS elongation.

Unlike other GAGs, KS can be *N*- or *O*-linked to the core protein backbone, with the former occurring on an Asn with a β-GlcNAc (KSI, or corneal KS) and the latter occurring on a Ser/Thr with a α-GalNAc (KSII, or skeletal KS) (**Figure 1.5** and **Figure 5.1)** (Merry et al., 2022). Of the two types of glycosylation, the latter is more common on aggrecan (~6 *N*-linked KS vs ~30-60 *O*-

linked KS) (Kiani et al., 2002; Merry et al., 2022; Plaas et al., 2023; Roughley, 2006). In contrast to its CS counterpart, there are fewer KS chains on aggrecan, with up to 60 KS chains attached at ~5-15 kDa each (Kiani et al., 2002; Plaas et al., 2023). These chains are primarily in the KS-rich domain of aggrecan, though KS may be present within the G1, IGD, and G2 region (Roughley, 2006). In humans, the consensus sequence for KS attachment is proposed to be E-(E/K)-P-F-P-S or E-E-P-(S/F)-P-S, with a T-T-A-P motif found in the CS2 region being considered a potential *O*-glycosylation site (Kiani et al., 2002). Curiously, initial synthesis of KS (*i.e.,* the generation of mucin *O*-glycans) does not guarantee extension of the chain. As individuals age, KS chains become longer, and sites previously occupied by mucin *O*-glycans are substituted with KS containing higher levels of sulfation on its Gal residues (Roughley & Mort, 2014).

Although aggrecan has long been recognized to bear mucin *O*-linked oligosaccharides, no study has considered the potential involvement of *O*-glycopeptidases in targeting the proteoglycan peptide backbone. Moreover, despite four decades of aggrecan centered research, only two publications have explored the potential structures of *O*-linked oligosaccharides attached to the core protein (**Figure 5.2**). One study from 1982 reported three *O*-linked glycans on aggrecan from murine chondrosarcomas (Nilsson et al., 1982). Since then, a recent LC-MS paper has expanded this list with the site-specific identification of seven mucin *O*-glycans on bovine aggrecan (Klein et al., 2018). These identified *O*-glycopeptides were distributed throughout the proteoglycan, and contained Hex, HexNAc, and Neu5Ac, with some sequences featuring multiple glycosylation sites. While the full structures of the glycans on each glycopeptide was not elucidated in this study, it still provides valuable information on the types of mucin *O*-glycans present, as these suggest the presence of Core 1, Core 2, 3SC1, and 6SC1 moieties on aggrecan. Notably, both BcM60_F and BcM60_G exhibited differential activity on mucin-specific FRET substrates bearing these glycans (**Table 4.1**). Further investigation through development of aggrecan-specific FRET substrates with these glycoforms would corroborate the ability of these BcM60-like peptidases to target proteoglycans.

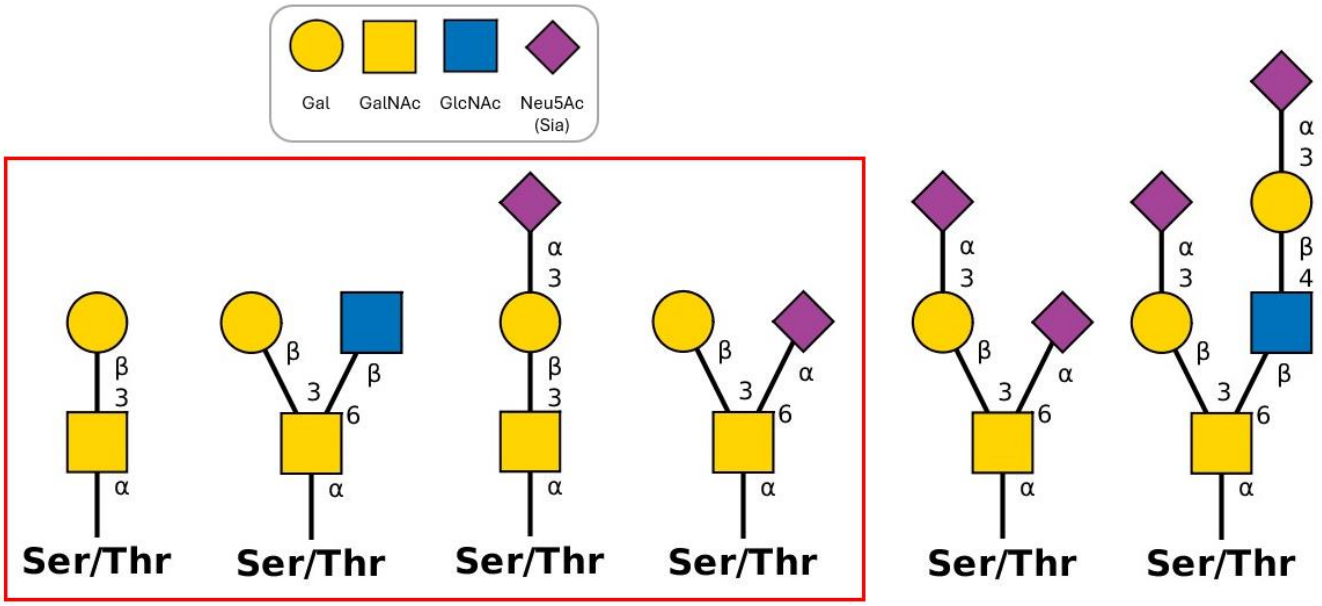


Figure 5.2: Potential mucin O-linked oligosaccharides on aggrecan.

Structures for potential aggrecan *O*-linked oligosaccharides acquired from supplementary data from (Klein et al., 2018) or main text of (Nilsson et al., 1982). The glycans highlighted in the red box indicates glycoforms that have been tested against BcM60_F and BcM60_G from BcPUL25.

5.1.3. Complete degradation of aggrecan by *Bacteroides caccae*.

Complete aggrecan degradation by BcPUL25 will require the coordinated activity of BcPUL-associated enzymes and CAZymes acting independently of this locus (Alvarez et al., 2025; Nakjang et al., 2012; D. Ndeh et al., 2020). As M60-like peptidases do not appear to require removal of mucin *O*-glycan chains prior to activity (Noach et al., 2017), the initial step of the BcPUL25 aggrecan degrading pathway will likely depend on the cellular location of each PUL-associated enzyme. SignalP 6.0 analysis of the CAZymes and M60-like peptidases in BcPUL25 predict lipoprotein signal peptides (Sec/SPII) on BcM60_F and BcGH109, while BcM60_G, BcGH88, and BcPL35 contain signal peptide for general secretion into the periplasm (Sec/SPI) (**Figure 5.3**) (Kaushik et al., 2022; Teufel et al., 2022). Based on these results; distribution of potential mucin *O*-glycans along the aggrecan core protein (**Figure 5.2**); and characterized activity of the CAZyme components of BcPUL25 (**Figure 3.8)**; we propose that BcM60_F initiates aggrecan cleavage by BcPUL25. The capacity of SusCD transporters to import glycopeptides into the periplasm has not been explored. However, as SusCD systems are highly structurally conserved, it was previously proposed that these complexes have an upper size limit for oligosaccharide intake of approximately ~5 kDa (D. A. Gray et al., 2021). Therefore, if the

predicted localization of BcPL35 into the periplasm is correct, this implies the existence of another CS lyase positioned on the outer membrane to assist with uptake. This additional PL would be necessary to ensure import of aggrecan into the cell, as CS chains on aggrecan are typically ~20 kDa in size (Roughley & Mort, 2014). Following successful uptake and desulfation of CS-glycopeptides, the degradation of chondroitin by BcPUL25 would proceed as previously described (**Figure 3.8)** (Alvarez et al., 2025). During this process, BcM60_G likely facilitates access to chondroitin through cleavage of the aggrecan core protein, increasing the number of exposed CS chains for BcPUL25-associated CAZymes to act on.

It is important to note that, while PULs can be studied independently *in vitro*, the complete degradation of structurally diverse glycans *in vivo* likely requires the coordinated action of multiple PULs and/or CAZymes not encoded within a given PUL. As shown here, BcPUL25 activity on GAGs requires the action of a PUL-independent sulfatase (Alvarez et al., 2025). Depending on the complexity of the target substrate, full depolymerization can even involve enzymes from different microbes. A recent publication that examined the mucin-driven ecological interactions by the HGM emphasises this point (Berkhout et al., 2024). Using a synthetic gut microbiota with mucin as the only carbon and energy source, the authors demonstrated that the complete mucin degradation requires activity from both mucin generalists and specialists, which enabled the stable co-existence of secondary degraders and hydrogen consumers that would otherwise be unable to survive under these conditions. In the context of aggrecan degradation by *B. caccae*, BcPUL25 can depolymerize the desulfated CS chains and peptide backbone (**Figure 5.3**). However, it lacks the CAZymes necessary to depolymerize the KS chains and the GAG tetrasaccharide linker on this proteoglycan. The molecular mechanisms underlying KS degradation by the HGM are relatively understudied compared to those involved in the breakdown of other GAGs, such as heparin and CS. Aggrecan's main KS disaccharide unit is KSII, wherein an elongated KS chain is made of highly sulfated *N*-acetyllactosamine (GlcNAcβ1-3Galβ1-4) (Brockhousen et al., 2022; Merry et al., 2022; Schjoldager et al., 2020). Extended mucin chains are also composed of the same base disaccharide unit, although the linkage and modifications on these may vary (Schjoldager et al., 2020). Thus, it is likely that HGM members that have been shown to degrade mucin, such as *B. caccae*, can also contribute to KS degradation.

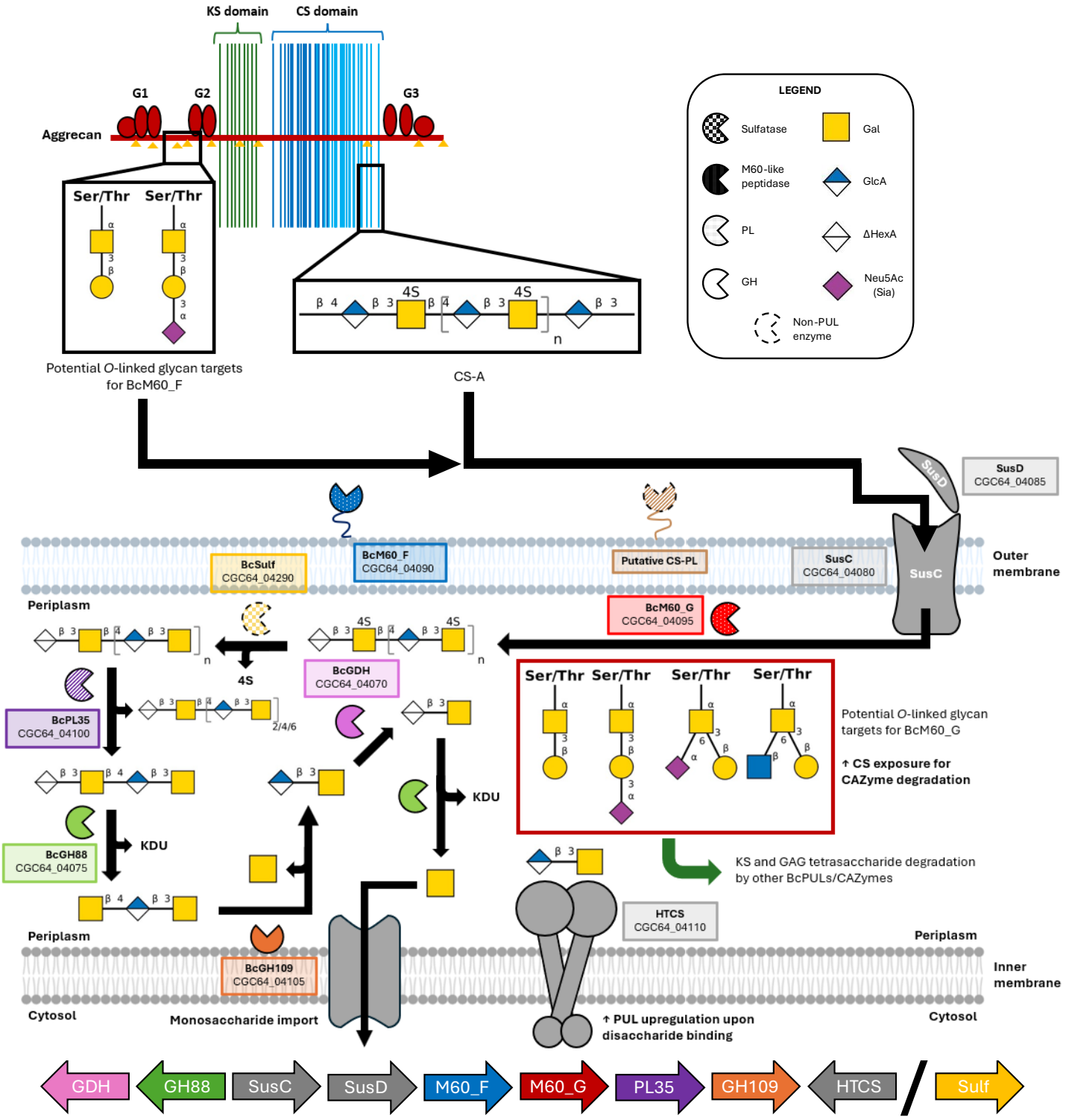


Figure 5.3: Model of CS degradation by BcPUL25.

Enzymes of interest color are coded according to PUL schematic at the bottom of figure (Alvarez et al., 2025). Greyed out components indicate proteins that were not investigated. For clarity, although BcPL35 releases di, tetra, and hexasaccharide oligomers, only the degradation of the tetrasaccharide is depicted here. *Abbreviations*: KDU: 5-keto-4-deoxyuronate, HTCS: hybrid two component system.

Of the putative PULs in *B. caccae*, the one designated as BcPUL16 is particularly notable (Terrapon et al., 2018). Besides its SusCD complex and regulatory HTCS, this PUL contains one GH105-PL33, one GH31, two GH2 enzymes, and one GH29. Bioinformatic assessment of these putative BcPUL16-annotated CAZymes suggest that it may contribute to the breakdown of CS and the GAG tetrasaccharide linker (Garron & Henrissat, 2019; Schjoldager et al., 2020; Terrapon et al., 2018). Additionally, the BcPUL16 GH29 member is annotated as an α-fucosidase. Mucin *O*-glycans are often capped with α-fucose, implying that this CAZyme may work with mucin degrading PULs. Additionally, it may also allow access to fucosylated GAGs. Fucosylated *N*-linked oligosaccharides and KS chains have been identified on bovine aggrecan (Klein et al., 2018; Plaas et al., 2023). Similarly, this would allow access to uniquely fucosylated GAGs found in marine animals, such as mollusks and sea cucumbers, which have been reported to have fucosylated CS and HS (Myron et al., 2014; Onishi et al., 2023). Thus, if these annotations are supported by functional testing of each CAZyme, BcPUL16 activity would complement chondroitin degradation by BcPUL25 by contributing to the breakdown of the CS-tetrasaccharide linker and removal of α-fucose from fucosylated GAGs.

5.2. *Bacteroides caccae*: A Potential Glycoprotein Connoisseur

Among the potential glycoprotein sources available to the HGM, mucin degradation remains the most extensively studied largely due to its strong association with host health and disease (Berkhout et al., 2024; Desai et al., 2016; Glover et al., 2022; Hansson, 2020; Paone & Cani, 2020; Ravcheev & Thiele, 2017). Numerous mucin-degrading bacteria have been identified, with the best studied being *A. muciniphila* and *Bacteroides* spp. (Berkhout et al., 2024; Glover et al., 2022). However, even though these bacteria are used as a model for mucin degradation, the molecular basis underlying complete mucin degradation (*i.e.,* degradation of both glycan and peptide components) remains fragmented and poorly integrated.

Bacteroides caccae has constantly been identified as a prominent mucin degrader in gut microbiota studies (Berkhout et al., 2024; Desai et al., 2016; Glover et al., 2022). While the specific PULs involved in its mucin utilization remain unknown, the enzymatic expression profile of *B. caccae* in response to host mucin and dietary fiber has been documented (Berkhout et al., 2024; Desai et al., 2016; Glover et al., 2022; Hansson, 2020; Paone & Cani, 2020; Ravcheev & Thiele,

2017). One seminal murine study demonstrated that dietary fiber deprivation leads to erosion of the gut's mucosal barrier and heightened pathogen susceptibility (Desai et al., 2016). In this context, *B. caccae* showed broad transcriptional changes when exposed to a fiber-free (*i.e.,* a mucus-degrading) diet compared to a fiber-rich diet, wherein the i*n vivo* expression of 230 genes increased, including 27 enzymes related to mucin degradation. Among the sixteen M60-like peptidases encoded in *B. caccae* genome, fourteen of these transcripts were detectable *in vivo*. It should be noted that these detected M60-like peptidases are found in PULs.

Overall, these fourteen *B. caccae* glycopeptidases showed differential levels of upregulation across the three diet types tested in this study: fiber-rich, fiber-free, and prebiotic (Desai et al., 2016). The prebiotic diet included host-indigestible glycans, including CS, and a food composition designed to better reflect a typical human meal. Of the fourteen detectable glycopeptidases, thirteen showed diet-dependent upregulation, with some showing as high as a 72-fold increase in a fiber-free diet. In the case of BcM60_F, this glycopeptidase showed a minimum 5-fold increase in expression under all dietary conditions, with the highest expression fold change (15-fold) in response to the prebiotic diet. This suggests that while BcM60_F may have a role in degrading multiple targets, including mucin, it primarily acts on a different glycoprotein that is more prevalent in the probiotic diet. Based on the known activity of BcPUL25, aggrecan is a likely candidate (**Figure 3.8** and **Figure 5.2**) (Alvarez et al., 2025). Curiously, the expression of detectable BcPUL25 CAZymes significantly increased (>10-fold) when mice alternated daily between a fiber-free/glucose or prebiotic/glucose diet, compared mice maintained on a consistent fiber-free or prebiotic diet. This indicates that the BcPUL25-associated HTCS regulatory sensor may respond to the oligosaccharides that come from different complex glycans, which is a regulatory mechanism previously observed in *B. thetaiotaomicron* and *B. ovatus* (Lowe et al., 2012; Martens et al., 2011; Singh, 2019). Collectively, the findings presented by Desai *et. al* and this project reinforce *B. caccae* as a valuable model organism for studying glycoprotein degradation from different nutrient sources (Alvarez et al., 2025; Desai et al., 2016). The characterization of these PULs may provide insight into how the HGM degrades both the glycan and peptide portion of complex substrates, such as mucin and proteoglycans.

5.3. Conclusion

The purpose of this research was to determine the molecular basis of BcPUL25 activity on aggrecan by the *in vitro* characterization of the four CAZymes (BcPL35, BcGH88, BcGH109, and BcGDH) and two M60-like peptidases associated with it. Using functional and structural analyses, we established a unique CS degradation pathway that uses a carbohydrate dehydratase (BcGDH) to prime substrates for *exo*-uronyl hydrolase (BcGH88) removal, allowing for the more efficient depolymerization of chondroitin by the PUL (Alvarez et al., 2025). Additionally, the BcM60-like peptidase results presented here substantially extend the previous work done with BcM60_F and BcM60_G (Deventer, 2021). Functional assessment of both *O*-glycopeptidases using mucin-specific FRET substrates revealed that BcM60_G can accommodate numerous glycan structures, whereas BcM60_F exhibited more stringent peptide specificity than its counterpart and could only bind to linear glycan moieties. Structural analysis of BcM60_G complexed with Core 2 revealed a Phe at the G2 subsite that may be required for recognizing branched glycans. Moreover, the superposition of AMUC_0627 complex with P1 (**PDB:** 7YX8) revealed a Tyr critical in binding to bis-glycosylated substrates, which are found on the mucin and aggrecan peptide backbone.

To date, this study represents the first *in vitro* investigation of any PUL from *B. caccae*. It has expanded our knowledge surrounding their respective substrate specificities, as well as substantiated the foundation on how *B. caccae* interacts with its glycoprotein substrates. Given the findings presented here and the research potential of other *O*-glycopeptidase-containing BcPULs, we propose that future research should explore *B. caccae* capability as a glycoprotein specialist. Its genome encodes sixteen M60-like peptidases, fifteen of which are associated with thirteen PULs (Nakjang et al., 2012; Terrapon et al., 2018). This is a level of redundancy not observed in any other gut microbe. For comparison, the mucin specialist *A. muciniphila* only encodes four M60-like peptidases in its genome. The integration of M60-like peptidases within PULs is also unique to *B. caccae*, further emphasising the microbe's specialization as a glycoprotein-degrading bacterium. Ultimately, studying the molecular basis by which the HGM targets and depolymerize their substrate targets will better guide the research that studies the overall effects of HGM composition and their metabolites in the gut. This understanding, coupled with advancements in modern biotechnology, may inform the development of targeted therapeutics aimed at mitigating HGM-influenced disease states.

References

Abbott, D. W., Gilbert, H. J., & Boraston, A. B. (2010). The active site of oligogalacturonate lyase provides unique insights into cytoplasmic oligogalacturonate β-elimination. *Journal of Biological Chemistry*, *285*(50), 39029–39038. https://doi.org/10.1074/jbc.M110.153981

Abbott, D. W., Martens, E. C., Gilbert, H. J., Cuskin, F., & Lowe, E. C. (2015). Coevolution of yeast mannan digestion: Convergence of the civilized human diet, distal gut microbiome, and host immunity. *Gut Microbes*, *6*(5), 334–339. https://doi.org/10.1080/19490976.2015.1091913

Abdul Rahim, M. B. H., Chilloux, J., Martinez-Gili, L., Neves, A. L., Myridakis, A., Gooderham, N., & Dumas, M. E. (2019). Diet-induced metabolic changes of the human gut microbiome: importance of short-chain fatty acids, methylamines and indoles. *Acta Diabetologica*, *56*(5), 493–500. https://doi.org/10.1007/s00592-019-01312-x

Abka-khajouei, R., Tounsi, L., Shahabi, N., Patel, A. K., Abdelkafi, S., & Michaud, P. (2022). Structures, Properties and Applications of Alginates. *Marine Drugs*, *20*(6). https://doi.org/10.3390/md20060364

Agus, A., Denizot, J., Thévenot, J., Martinez-Medina, M., Massier, S., Sauvanet, P., Bernalier-Donadille, A., Denis, S., Hofman, P., Bonnet, R., Billard, E., & Barnich, N. (2016). Western diet induces a shift in microbiota composition enhancing susceptibility to adherent-invasive E. coli infection and intestinal inflammation. *Scientific Reports*, *6*. https://doi.org/10.1038/srep19032

Al-Biltagi, M., El Amrousy, D., El Ashry, H., Maher, S., Mohammed, M. A., & Hasan, S. (2022). Effects of adherence to the Mediterranean diet in children and adolescents with irritable bowel syndrome. *World Journal of Clinical Pediatrics*, *11*(4), 330–340. https://doi.org/10.5409/wjcp.v11.i4.330

Ali, G., Sharma, M., Salama, E. S., Ling, Z., & Li, X. (2024). Applications of chitin and chitosan as natural biopolymer: potential sources, pretreatments, and degradation pathways. In *Biomass Conversion and Biorefinery* (Vol. 14, Issue 4, pp. 4567–4581). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/s13399-022-02684-x

Alvarez, B., Canil, O. F., Low, K. E., Hettle, A. G., Abbott, D. W., & Boraston, A. B. (2025). Analysis of chondroitin degradation by components of a Bacteroides caccae polysaccharide utilization locus. *Journal of Biological Chemistry*, *110354*. https://doi.org/10.1016/j.jbc.2025.110354

Apweiler, R., Hermjakob, H., & Sharon, N. (1999). On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1473*(1), 4–8. https://doi.org/10.1016/S0304-4165(99)00165-8

Arai, T., Araki, R., Tanaka, A., Karita, S., Kimura, T., Sakka, K., & Ohmiya, K. (2003). Characterization of a cellulase containing a family 30 carbohydrate-binding module (CBM) derived from Clostridium thermocellum CelJ: Importance of the CBM to cellulose hydrolysis. *Journal of Bacteriology*, *185*(2), 504–512. https://doi.org/10.1128/JB.185.2.504-512.2003

Barreto-Bergter, E., & Figueiredo, R. T. (2014). Fungal glycans and the innate immune recognition. In *Frontiers in Cellular and Infection Microbiology* (Vol. 4, Issue OCT). Frontiers Media S.A. https://doi.org/10.3389/fcimb.2014.00145

Bäumgen, M., Dutschei, T., Bartosik, D., Suster, C., Reisky, L., Gerlach, N., Stanetty, C., Mihovilovic, M. D., Schweder, T., Hehemann, J. H., & Bornscheuer, U. T. (2021). A new

carbohydrate-active oligosaccharide dehydratase is involved in the degradation of ulvan. *Journal of Biological Chemistry*, *279*(4). https://doi.org/10.1016/j.jbc.2021.101210

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., Ioannou, A., de Ram, C., Boeren, S., Plugge, C. M., & Belzer, C. (2024). Mucin-driven ecological interactions in an in vitro synthetic community of human gut microbes. *Glycobiology*, *34*(12). https://doi.org/10.1093/glycob/cwae085

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

Berlemont, R., & Martiny, A. C. (2016). Glycoside Hydrolases across Environmental Microbial Communities. *PLoS Computational Biology*, *12*(12). https://doi.org/10.1371/journal.pcbi.1005300

Bhattacharya, T., Ghosh, T. S., & Mande, S. S. (2015). Global profiling of carbohydrate active enzymes in human gut microbiome. *PLoS ONE*, *10*(11), 1–20. https://doi.org/10.1371/journal.pone.0142038

Blum, M., Chang, H. Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G. A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D. H., Letunic, I., Marchler-Bauer, A., … Finn, R. D. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research*, *49*(D1), D344–D354. https://doi.org/10.1093/nar/gkaa977

Bolam, D. N., & Koropatkin, N. M. (2012). Glycan recognition by the Bacteroidetes Sus-like systems. *Current Opinion in Structural Biology*, *22*(5), 563–569. https://doi.org/10.1016/j.sbi.2012.06.006

Boraston, A. B., Bolam, D. N., Gilbert, H. J., & Davies, G. J. (2004). Carbohydrate-binding modules: Fine-tuning polysaccharide recognition. *Biochemical Journal*, *382*(3), 769–781. https://doi.org/10.1042/BJ20040892

Bradley, E., & Haran, J. (2024). The human gut microbiome and aging. *Gut Microbes*, *16*(1). https://doi.org/10.1080/19490976.2024.2359677

Breyer, W. A., & Matthews, B. W. (2001). *A structural basis for processivity*. https://doi.org/10.1101/ps.10301

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

Brockhousen, I., Wandall, H. H., Hagen, K. G. Ten, & Stanley, P. (2022). Chapter 10: O-GalNAc Glycans. In *Cold Spring Harbor* (4th ed.). Cold Spring Harbor Laboratory Press. https://doi.org/10.1101/glycobiology.4e.10

Brown, H. A., & Koropatkin, N. M. (2020). Host glycan utilization within the Bacteroidetes Sus-like paradigm. *Glycobiology*, 1–10. https://doi.org/10.1093/glycob/cwaa054

Carrico, I. S., Carlson, B. L., & Bertozzi, C. R. (2007). Introducing genetically encoded aldehydes into proteins. *Nature Chemical Biology*, *3*(6), 321–322. https://doi.org/10.1038/nchembio878

Cartmell, A., Lowe, E. C., Baslé, A., Firbank, S. J., Ndeh, D. A., Murray, H., Terrapon, N., Lombard, V., Henrissat, B., Turnbull, J. E., Czjzek, M., Gilbert, H. J., & Bolam, D. N.

(2017). How members of the human gut microbiota overcome the sulfation problem posed by glycosaminoglycans. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(27), 7037–7042. https://doi.org/10.1073/pnas.1704367114

Cartmell, A., Muñoz-Muñoz, J., Briggs, J. A., Ndeh, D. A., Lowe, E. C., Baslé, A., Terrapon, N., Stott, K., Heunis, T., Gray, J., Yu, L., Dupree, P., Fernandes, P. Z., Shah, S., Williams, S. J., Labourel, A., Trost, M., Henrissat, B., & Gilbert, H. J. (2018). A surface endogalactanase in Bacteroides thetaiotaomicron confers keystone status for arabinogalactan degradation. *Nature Microbiology*, *3*(11), 1314–1326. https://doi.org/10.1038/s41564-018-0258-8

Cerdà-Costa, N., & Gomis-Rüth, F. X. (2014). Architecture and function of metallopeptidase catalytic domains. *Protein Science*, *23*(2), 123–144. https://doi.org/10.1002/pro.2400

ChemAxon. (n.d.). *Marvin JS: Web-based chemical editor*.

Chen, T., Long, W., Zhang, C., Liu, S., Zhao, L., & Hamaker, B. R. (2017). Fiber-utilizing capacity varies in Prevotella- versus Bacteroides-dominated gut microbiota. *Scientific Reports*, *7*(1), 1–7. https://doi.org/10.1038/s41598-017-02995-4

Cheng, K., Zhou, Y., & Neelamegham, S. (2016). DrawGlycan-SNFG: a robust tool to render glycans and glycopeptides with fragmentation information. *Glycobiology*, *27*(3), 200–205. https://doi.org/10.1093/glycob/cww115

Chia, L. W., Mank, M., Blijenberg, B., Aalvink, S., Bongers, R. S., Stahl, B., Knol, J., & Belzer, C. (2020). Bacteroides thetaiotaomicron fosters the growth of butyrate-producing anaerostipes caccae in the presence of lactose and total human milk carbohydrates. *Microorganisms*, *8*(10), 1–13. https://doi.org/10.3390/microorganisms8101513

Chia, S., Teo, G., Tay, S. J., Loo, L. S. W., Wan, C., Sim, L. C., Yu, H., Walsh, I., & Pang, K. T. (2022). An Integrative Glycomic Approach for Quantitative Meat Species Profiling. *Foods*, *11*(13). https://doi.org/10.3390/foods11131952

Costlow, L., Herforth, A., Sulser, T. B., Cenacchi, N., & Masters, W. A. (2025). Global analysis reveals persistent shortfalls and regional differences in availability of foods needed for health. *Global Food Security*, *44*. https://doi.org/10.1016/j.gfs.2024.100825

Davies, G., & Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure*, *3*(9), 853–859. https://doi.org/10.1016/S0969-2126(01)00220-9

Davies, G. J., & Sinnot, M. L. (2008). Sorting the diverse. *The Biochemical Society*, *August*, 26–32.

Davies, G. J., Wilson, K. S., & Henrissat, B. (1997). Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochemical Journal*, *321*(2), 557–559. https://doi.org/https://doi.org/10.1042/bj3210557

Desai, M. S., Seekatz, A. M., Koropatkin, N. M., Kamada, N., Hickey, C. A., Wolter, M., Pudlo, N. A., Kitamoto, S., Terrapon, N., Muller, A., Young, V. B., Henrissat, B., Wilmes, P., Stappenbeck, T. S., Núñez, G., & Martens, E. C. (2016). A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell*, *167*(5), 1339-1353.e21. https://doi.org/10.1016/j.cell.2016.10.043

Deventer, A. T. (2021). *Structural and functional characterization of M60 peptidases from the gut microbe Bacteroides caccae* (Issue April). University of Victoria.

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

Farach-Carson, M. C., Wu, D., & França, C. M. (2024). Proteoglycans in mechanobiology of tissues and organs: Normal functions and mechanopathology. *Proteoglycan Research*, *2*(2). https://doi.org/10.1002/pgr2.21

Feng, J., Qian, Y., Zhou, Z., Ertmer, S., Vivas, E. I., Lan, F., Hamilton, J. J., Rey, F. E., Anantharaman, K., & Venturelli, O. S. (2022). Polysaccharide utilization loci in Bacteroides determine population fitness and community-level interactions. *Cell Host & Microbe*, 1–16. https://doi.org/10.1016/j.chom.2021.12.006

Flannery, C. R., Little, C. B., & Caterson, B. (1998). Molecular Cloning and Sequence Analysis of the Aggrecan Interglobular Domain from Porcine, Equine, Bovine and Ovine Cartilage: Comparison of Proteinase-Susceptible Regions and Sites of Keratan Sulfate Substitution. *Matrix Biology*, *16*(98), 507–511.

Foley, M. H., Cockburn, D. W., & Koropatkin, N. M. (2016). The Sus operon: a model system for starch uptake by the human gut Bacteroidetes. *Cellular and Molecular Life Sciences*, *73*(14), 2603–2617. https://doi.org/10.1007/s00018-016-2242-x

Garron, M. L., & Cygler, M. (2010). Structural and mechanistic classification of uronic acid-containing polysaccharide lyases. *Glycobiology*, *20*(12), 1547–1573. https://doi.org/10.1093/glycob/cwq122

Garron, M. L., & Cygler, M. (2014). Uronic polysaccharide degrading enzymes. *Current Opinion in Structural Biology*, *28*(1), 87–95. https://doi.org/10.1016/j.sbi.2014.07.012

Garron, M. L., & Henrissat, B. (2019). The continuing expansion of CAZymes and their families. *Current Opinion in Chemical Biology*, *53*, 82–87. https://doi.org/10.1016/j.cbpa.2019.08.004

Ghosh, T. S., Rampelli, S., Jeffery, I. B., Santoro, A., Neto, M., Capri, M., Giampieri, E., Jennings, A., Candela, M., Turroni, S., Zoetendal, E. G., Hermes, G. D. A., Elodie, C., Meunier, N., Brugere, C. M., Pujos-Guillot, E., Berendsen, A. M., De Groot, L. C. P. G. M., Feskins, E. J. M., … O'Toole, P. W. (2020). Mediterranean diet intervention alters the gut microbiome in older people reducing frailty and improving health status: The NU-AGE 1-year dietary intervention across five European countries. *Gut*, *69*(7), 1218–1228. https://doi.org/10.1136/gutjnl-2019-319654

Glover, J. S., Ticer, T. D., & Engevik, M. A. (2022). Characterizing the mucin-degrading capacity of the human gut microbiota. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-11819-z

Golshany, H., Helmy, S. A., Morsy, N. F. S., Kamal, A., Yu, Q., & Fan, L. (2024). The gut microbiome across the lifespan: how diet modulates our microbial ecosystem from infancy to the elderly. *International Journal of Food Sciences and Nutrition*. https://doi.org/10.1080/09637486.2024.2437472

Gong, J., Liu, S., Wang, S., Ruan, H., Mou, Q., Fan, P., Chen, T., Cai, W., Lu, Y., & Lu, Z. (2023). Identification of fecal microbiome signatures associated with familial longevity and candidate metabolites for healthy aging. *Aging Cell*, *22*(6). https://doi.org/10.1111/acel.13848

Götting, C., Kuhn, J., & Kleesiek, K. (2007). Human xylosyltransferases in health and disease. *Cellular and Molecular Life Sciences*, *64*(12), 1498–1517. https://doi.org/10.1007/s00018-007-7069-z

Gray, A. L., Pun, N., Ridley, A. J. L., & Dyer, D. P. (2022). Role of extracellular matrix proteoglycans in immune cell recruitment. In *International Journal of Experimental*

Pathology (Vol. 103, Issue 2, pp. 34–43). John Wiley and Sons Inc. https://doi.org/10.1111/iep.12428

Gray, D. A., White, J. B. R., Oluwole, A. O., Rath, P., Glenwright, A. J., Mazur, A., Zahn, M., Baslé, A., Morland, C., Evans, S. L., Cartmell, A., Robinson, C. V., Hiller, S., Ranson, N. A., Bolam, D. N., & van den Berg, B. (2021). Insights into SusCD-mediated glycan import by a prominent gut symbiont. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-020-20285-y

Grondin, J. M., Tamura, K., Déjean, G., Abbott, D. W., & Brumer, H. (2017a). Polysaccharide utilization loci: Fueling microbial communities. *Journal of Bacteriology*, *199*(15), 1–15. https://doi.org/10.1128/JB.00860-16

Grondin, J. M., Tamura, K., Déjean, G., Abbott, D. W., & Brumer, H. (2017b). Polysaccharide utilization loci: Fueling microbial communities. *Journal of Bacteriology*, *199*(15), 1–15. https://doi.org/10.1128/JB.00860-16

Hameleers, L., Pijning, T., Gray, B. B., Fauré, R., & Jurak, E. (2024). Novel β-galactosidase activity and first crystal structure of Glycoside Hydrolase family 154. *New Biotechnology*, *80*(December 2023), 1–11. https://doi.org/10.1016/j.nbt.2023.12.011

Hansen, L., Husein, D. M., Gericke, B., Hansen, T., Pedersen, O., Tambe, M. A., Freeze, H. H., Naim, H. Y., Henrissat, B., Wandall, H. H., Clausen, H., & Bennett, E. P. (2020). A mutation map for human glycoside hydrolase genes. *Glycobiology*, *30*(8), 500–515. https://doi.org/10.1093/glycob/cwaa010

Hansson, G. C. (2020). Mucins and the Microbiome. *Annual Review of Biochemistry*, *89*, 769–793. https://doi.org/10.1146/annurev-biochem-011520-105053

Harris, H. C., Morrison, D. J., & Edwards, C. A. (2020). Impact of the source of fermentable carbohydrate on SCFA production by human gut microbiota in vitro - a systematic scoping review and secondary analysis. *Critical Reviews in Food Science and Nutrition*, *0*(0), 1–12. https://doi.org/10.1080/10408398.2020.1809991

Hashimoto, W., Kobayashi, E., Nankai, H., Sato, N., Miya, T., Kawai, S., & Murata, K. (1999). Unsaturated glucuronyl hydrolase of Bacillus sp. GL1: Novel enzyme prerequisite for metabolism of unsaturated oligosaccharides produced by polysaccharide lyases. *Archives of Biochemistry and Biophysics*, *368*(2), 367–374. https://doi.org/10.1006/abbi.1999.1305

Helbert, W., Poulet, L., Drouillard, S., Mathieu, S., Loiodice, M., Couturier, M., Lombard, V., Terrapon, N., Turchetto, J., Vincentelli, R., & Henrissat, B. (2019). Discovery of novel carbohydrate-active enzymes through the rational exploration of the protein sequences space. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(13), 6063–6068. https://doi.org/10.1073/pnas.1815791116

Hobbs, J. K., Hettle, A. G., Vickers, C., & Boraston, A. B. (2018). *Biochemical Reconstruction of a Metabolic Pathway from a Marine Bacterium Reveals Its Mechanism of Pectin Depolymerization*. https://journals.asm.org/journal/aem

Hobbs, J. K., Lee, S. M., Robb, M., Hof, F., Barr, C., Abe, K. T., Hehemann, J. H., McLean, R., Abbott, D. W., & Boraston, A. B. (2016). KdgF, the missing link in the microbial metabolism of uronate sugars from pectin and alginate. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(22), 6188–6193. https://doi.org/10.1073/pnas.1524214113

Hou, K., Wu, Z. X., Chen, X. Y., Wang, J. Q., Zhang, D., Xiao, C., Zhu, D., Koya, J. B., Wei, L., Li, J., & Chen, Z. S. (2022). Microbiota in health and diseases. *Signal Transduction and Targeted Therapy*, *7*(1). https://doi.org/10.1038/s41392-022-00974-4

Inokuma, K., Sasaki, D., Kurata, K., Ichikawa, M., Otsuka, Y., & Kondo, A. (2023). Sulfated and non-sulfated chondroitin affect the composition and metabolism of human colonic microbiota simulated in an in vitro fermentation system. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-38849-5

Itoh, T., Hashimoto, W., Mikami, B., & Murata, K. (2006a). Crystal structure of unsaturated glucuronyl hydrolase complexed with substrate: Molecular insights into its catalytic reaction mechanism. *Journal of Biological Chemistry*, *281*(40), 29807–29816. https://doi.org/10.1074/jbc.M604975200

Itoh, T., Hashimoto, W., Mikami, B., & Murata, K. (2006b). Substrate recognition by unsaturated glucuronyl hydrolase from Bacillus sp. GL1. *Biochemical and Biophysical Research Communications*, *344*(1), 253–262. https://doi.org/10.1016/j.bbrc.2006.03.141

Johnson, J. L., Moore, W. E. C., & Moore, L. V. H. (1987). Bacteroides caccae sp. nov., Bacteroides merdae sp. nov., and Bacteroides stercoris sp. nov. Isolated from Human Feces. *International Journal of Systemic Bacteriology*, *36*(4), 499–501. https://doi.org/https://doi-org.ezproxy.library.uvic.ca/10.1099/00207713-36-4-499

Jongkees, S. A. K., Yoo, H., & Withers, S. G. (2014). Mechanistic investigations of unsaturated glucuronyl hydrolase from clostridium perfringens. *Journal of Biological Chemistry*, *289*(16), 11385–11395. https://doi.org/10.1074/jbc.M113.545293

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kaushik, S., He, H., & Dalbey, R. E. (2022). Bacterial Signal Peptides- Navigating the Journey of Proteins. *Frontiers in Physiology*, *13*. https://doi.org/10.3389/fphys.2022.933153

Kiani, C., Chen, L., Wu, Y. J., Yee, A. J., & Yang, B. B. (2002). Structure and function of aggrecan. *Cell Research*, *12*(1), 19–32. https://doi.org/10.1038/sj.cr.7290106

Kim, H. T., Chung, J. H., Wang, D., Lee, J., Woo, H. C., Choi, I. G., & Kim, K. H. (2012). Depolymerization of alginate into a monomeric sugar acid using Alg17C, an exo-oligoalginate lyase cloned from Saccharophagus degradans 2-40. *Applied Microbiology and Biotechnology*, *93*(5), 2233–2239. https://doi.org/10.1007/s00253-012-3882-x

Kim, J., Ryu, C., Ha, J., Lee, J., Kim, D., Ji, M., Park, C. S., Lee, J., Kim, D. K., & Kim, H. H. (2020). Structural and quantitative characterization of mucin-type O-glycans and the identification of Oglycosylation sites in bovine submaxillary mucin. *Biomolecules*, *10*(4). https://doi.org/10.3390/biom10040636

Klassen, L., Xing, X., Tingley, J. P., Low, K. E., King, M. L., Reintjes, G., & Abbott, D. W. (2021). Approaches to Investigate Selective Dietary Polysaccharide Utilization by Human Gut Microbiota at a Functional Level. *Frontiers in Microbiology*, *12*(February), 1–11. https://doi.org/10.3389/fmicb.2021.632684

Klein, J. A., Meng, L., & Zaia, J. (2018). Deep sequencing of complex proteoglycans: A novel strategy for high coverage and sitespecific identification of glycosaminoglycanlinked peptides. *Molecular and Cellular Proteomics*, *17*(8), 1578–1590. https://doi.org/10.1074/mcp.RA118.000766

Koch, C. D., Lee, C. M., & Apte, S. S. (2020). Aggrecan in Cardiovascular Development and Disease. *Journal of Histochemistry and Cytochemistry*, *68*(11), 777–795. https://doi.org/10.1369/0022155420952902

Kostopoulos, I., Aalvink, S., Kovatcheva-Datchary, P., Nijsse, B., Bäckhed, F., Knol, J., de Vos, W. M., & Belzer, C. (2021). A Continuous Battle for Host-Derived Glycans Between a Mucus Specialist and a Glycan Generalist in vitro and in vivo. *Frontiers in Microbiology*, *12*(June), 1–14. https://doi.org/10.3389/fmicb.2021.632454

Köwitsch, A., Zhou, G., & Groth, T. (2018). Medical application of glycosaminoglycans: a review. In *Journal of Tissue Engineering and Regenerative Medicine* (Vol. 12, Issue 1, pp. e23–e41). John Wiley and Sons Ltd. https://doi.org/10.1002/term.2398

Kunisawa, J., Hashimoto, E., Inoue, A., Nagasawa, R., Suzuki, Y., Ishikawa, I., Shikata, S., Arita, M., Aoki, J., & Kiyono, H. (2014). Regulation of Intestinal IgA Responses by Dietary Palmitic Acid and Its Metabolism. *The Journal of Immunology*, *193*(4), 1666–1671. https://doi.org/10.4049/jimmunol.1302944

Kuzmič, P. (1996). Program DYNAFIT for the Analysis of Enzyme Kinetic Data: Application to HIV Proteinase. *Analytical Biochemistry*, *237*, 260–273.

Kuzmič, P. (2009). Application of the Van Slyke-Cullen irreversible mechanism in the analysis of enzymatic progress curves. *Analytical Biochemistry*, *394*(2), 287–289. https://doi.org/10.1016/j.ab.2009.06.040

Kuzmič, P., Sexton, D. J., & Martik, D. (2010). A “hit-and-run” kinetic mechanism for the analysis enzyme progress curves under low substrate concentrations. *Analytical Biochemistry*.

Lebrilla, C. B., Liu, J., Widmalm, G., & Prestegard, J. H. (2022). Chapter 3: Oligosaccharides and Polysaccharides. In *Cold Spring Harbor* (Vol. 4). Cold Spring Harbor Laboratory Press. https://doi.org/10.1101/glycobiology.4e.3

Li, C., Tang, T., Du, Y., Jiang, L., Yao, Z., Ning, L., & Zhu, B. (2023). Ulvan and Ulva oligosaccharides: a systematic review of structure, preparation, biological activities and applications. *Bioresources and Bioprocessing*, *10*(1). https://doi.org/10.1186/s40643-023-00690-z

Li, J., Peng, C., Mao, A., Zhong, M., & Hu, Z. (2024). An overview of microbial enzymatic approaches for pectin degradation. *International Journal of Biological Macromolecules*, *254*. https://doi.org/10.1016/j.ijbiomac.2023.127804

Li, Z., Xiong, W., Liang, Z., Wang, J., Zeng, Z., Kołat, D., Li, X., Zhou, D., Xu, X., & Zhao, L. (2024). Critical role of the gut microbiota in immune responses and cancer immunotherapy. *Journal of Hematology and Oncology*, *17*(1). https://doi.org/10.1186/s13045-024-01541-w

Liu, F., Zhang, N., Li, Z., Wang, X., Shi, H., Xue, C., Li, R. W., & Tang, Q. (2017). Chondroitin sulfate disaccharides modified the structure and function of the murine gut microbiome under healthy and stressed conditions. *Scientific Reports*, *7*(1), 1–14. https://doi.org/10.1038/s41598-017-05860-6

Liu, G., Ma, Y., Yang, Q., & Deng, S. (2020). Modulation of inflammatory response and gut microbiota in ankylosing spondylitis mouse model by bioactive peptide IQW. *Journal of Applied Microbiology*, *128*(6), 1669–1677. https://doi.org/10.1111/jam.14588

Liu, P., Liu, Z., Wang, J., Wang, J., Gao, M., Zhang, Y., Yang, C., Zhang, A., Li, G., Li, X., Liu, S., Liu, L., Sun, N., & Zhang, K. (2024). Immunoregulatory role of the gut microbiota in inflammatory depression. *Nature Communications*, *15*(1). https://doi.org/10.1038/s41467-024-47273-w

Liu, Q. P., Sulzenbacher, G., Yuan, H., Bennett, E. P., Pietz, G., Saunders, K., Spence, J., Nudelman, E., Levery, S. B., White, T., Neveu, J. M., Lane, W. S., Bourne, Y., Olsson, M. L., Henrissat, B., & Clausen, H. (2007). Bacterial glycosidases for the production of

universal red blood cells. *Nature Biotechnology*, *25*(4), 454–464. https://doi.org/10.1038/nbt1298

Loiodice, M., Drula, E., McIver, Z., Antonyuk, S., Baslé, A., Lima, M., Yates, E. A., Byrne, D. P., Coughlan, J., Leech, A., Mesdaghi, S., Rigden, D. J., Drouillard, S., Helbert, W., Henrissat, B., Terrapon, N., Wright, G. S. A., Couturier, M., & Cartmell, A. (2025). Bacterial polysaccharide lyase family 33: Specificity from an evolutionarily conserved binding tunnel. *Proceedings of the National Academy of Sciences of the United States of America*, *122*(7). https://doi.org/10.1073/pnas.2421623122

Lombard, V., Henrissat, B., & Garron, M. L. (2025). CAZac: An activity descriptor for carbohydrate-active enzymes. *Nucleic Acids Research*, *53*(D1), D625–D633. https://doi.org/10.1093/nar/gkae1045

Lowe, E. C., Baslé, A., Czjzek, M., Firbank, S. J., & Bolam, D. N. (2012). A scissor blade-like closing mechanism implicated in transmembrane signaling in a Bacteroides hybrid two-component system. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(19), 7298–7303. https://doi.org/10.1073/pnas.1200479109

Lu, D., Wang, W., Li, X., Wang, L., Guo, Y., Zhu, C., Wang, X., Lian, B., Bai, J., & Zhang, Q. (2024). Identification and characterization of a PL35 GAGs lyase with 4-O-sulfated N-acetylgalactosamine (A-type)-rich structures producing property. *International Journal of Biological Macromolecules*, *266*(P2), 131283. https://doi.org/10.1016/j.ijbiomac.2024.131283

Luis, A. S., Jin, C., Pereira, G. V., Glowacki, R. W. P., Gugel, S. R., Singh, S., Byrne, D. P., Pudlo, N. A., London, J. A., Baslé, A., Reihill, M., Oscarson, S., Eyers, P. A., Czjzek, M., Michel, G., Barbeyron, T., Yates, E. A., Hansson, G. C., Karlsson, N. G., … Martens, E. C. (2021). A single sulfatase is required to access colonic mucin by a gut bacterium. *Nature*, *598*(7880), 332–337. https://doi.org/10.1038/s41586-021-03967-5

Luis, A. S., Yates, E. A., & Cartmell, A. (2023). Functions and specificity of bacterial carbohydrate sulfatases targeting host glycans. *Essays in Biochemistry*, *67*(3), 429–442. https://doi.org/10.1042/EBC20220120

Madeira, F., Madhusoodanan, N., Lee, J., Eusebi, A., Niewielska, A., Tivey, A. R. N., Lopez, R., & Butcher, S. (2024). The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Research*, *52*(W1), W521–W525. https://doi.org/10.1093/nar/gkae241

Malaker, S. A., Pedram, K., Ferracane, M. J., Bensing, B. A., Krishnan, V., Pett, C., Yu, J., Woods, E. C., Kramer, J. R., Westerlind, U., Dorigo, O., & Bertozzi, C. R. (2019). The mucin-selective protease StcE enables molecular and functional analysis of human cancer-associated mucins. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(15), 7278–7287. https://doi.org/10.1073/pnas.1813020116

Malaker, S. A., Pedram, K., Ferracane, M. J., Woods, E. C., Kramer, J. R., Dorigo, O., & Bertozzi, C. R. (2016). *A mucin-specific protease enables molecular and functional analysis of human cancer- associated mucins*. 1–12.

Mann, E. R., Lam, Y. K., & Uhlig, H. H. (2024). Short-chain fatty acids: linking diet, the microbiome and immunity. In *Nature Reviews Immunology* (Vol. 24, Issue 8, pp. 577–595). Nature Research. https://doi.org/10.1038/s41577-024-01014-8

Marchix, J., Goddard, G., & Helmrath, M. A. (2018). Host-Gut Microbiota Crosstalk in Intestinal Adaptation. *Cmgh*, *6*(2), 149–162. https://doi.org/10.1016/j.jcmgh.2018.01.024

Marcobal, 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

Martens, E. C., Chiang, H. C., & Gordon, J. I. (2008). Mucosal Glycan Foraging Enhances Fitness and Transmission of a Saccharolytic Human Gut Bacterial Symbiont. *Cell Host and Microbe*, *4*(5), 447–457. https://doi.org/10.1016/j.chom.2008.09.007

Martens, E. C., Lowe, E. C., Chiang, H., Pudlo, N. A., Wu, M., McNulty, N. P., Abbott, D. W., Henrissat, B., Gilbert, H. J., Bolam, D. N., & Gordon, J. I. (2011). Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biology*, *9*(12). https://doi.org/10.1371/journal.pbio.1001221

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

Martín, R., Rios-Covian, D., Huillet, E., Auger, S., Khazaal, S., Bermúdez-Humarán, L. G., Sokol, H., Chatel, J. M., & Langella, P. (2023). Faecalibacterium: a bacterial genus with promising human health applications. In *FEMS Microbiology Reviews* (Vol. 47, Issue 4). Oxford University Press. https://doi.org/10.1093/femsre/fuad039

Maruyama, Y., Nakamichi, Y., Itoh, T., Mikami, B., Hashimoto, W., & Murata, K. (2009). Substrate specificity of streptococcal unsaturated glucuronyl hydrolases for sulfated glycosaminoglycan. *Journal of Biological Chemistry*, *284*(27), 18059–18069. https://doi.org/10.1074/jbc.M109.005660

McNulty, N. P., Wu, M., Erickson, A. R., Pan, C., Erickson, B. K., Martens, E. C., Pudlo, N. A., Muegge, B. D., Henrissat, B., Hettich, R. L., & Gordon, J. I. (2013). Effects of Diet on Resource Utilization by a Model Human Gut Microbiota Containing Bacteroides cellulosilyticus WH2, a Symbiont with an Extensive Glycobiome. *PLoS Biology*, *11*(8). https://doi.org/10.1371/journal.pbio.1001637

Medley, B. J., Leclaire, L., Thompson, N., Mahoney, K. E., Pluvinage, B., Parson, M. A. H., Burke, J. E., Malaker, S., Wakarchuk, W., & Boraston, A. B. (2022). A previously uncharacterized O-glycopeptidase from Akkermansia muciniphila requires the Tn-antigen for cleavage of the peptide bond. *Journal of Biological Chemistry*, *298*(10). https://doi.org/10.1016/j.jbc.2022.102439

Megazyme. (2019). *D-Glucuronic Acid & D-Galacturonic Acid Assay Procedure*.

Merry, C. L. R., Lindahl, U., Couchman, J., & Esko, J. D. (2022). Chapter 17: Proteoglycans and Sulfated Glycosaminoglycans. In *Essentials of Glycobiology* (4th ed.). Cold Spring Harbor Laboratory Press. https://doi.org/10.1101/glycobiology.4e.17

Meslier, V., Laiola, M., Roager, H. M., De Filippis, F., Roume, H., Quinquis, B., Giacco, R., Mennella, I., Ferracane, R., Pons, N., Pasolli, E., Rivellese, A., Dragsted, L. O., Vitaglione, P., Ehrlich, S. D., & Ercolini, D. (2020). Mediterranean diet intervention in overweight and obese subjects lowers plasma cholesterol and causes changes in the gut microbiome and metabolome independently of energy intake. *Gut*, *69*(7), 1258–1268. https://doi.org/10.1136/gutjnl-2019-320438

Mikami, T., & Kitagawa, H. (2013). Biosynthesis and function of chondroitin sulfate. *Biochimica et Biophysica Acta - General Subjects*, *1830*(10), 4719–4733. https://doi.org/10.1016/j.bbagen.2013.06.006

Minic, Z. (2008). Physiological roles of plant glycoside hydrolases. In *Planta* (Vol. 227, Issue 4, pp. 723–740). https://doi.org/10.1007/s00425-007-0668-y

Muthana, S. M., Campbell, C. T., & Gildersleeve, J. C. (2012). Modifications of glycans: Biological significance and therapeutic opportunities. *ACS Chemical Biology*, *7*(1), 31–43. https://doi.org/10.1021/cb2004466

Myron, P., Siddiquee, S., & Al Azad, S. (2014). Fucosylated chondroitin sulfate diversity in sea cucumbers: A review. *Carbohydrate Polymers*, *112*, 173–178. https://doi.org/10.1016/j.carbpol.2014.05.091

Nagasawa, K., Inoue, Y., & Tokuyasu, T. (1979). An improved method for the preparation of chondroitin by solvolytic desulfation of chondroitin sulfates. *Journal of Biochemistry*, *86*(5), 1323–1329. https://doi.org/10.1093/oxfordjournals.jbchem.a132648

Nakamichi, Y., Maruyama, Y., Mikami, B., Hashimoto, W., & Murata, K. (2011). Structural determinants in streptococcal unsaturated glucuronyl hydrolase for recognition of glycosaminoglycan sulfate groups. *Journal of Biological Chemistry*, *286*(8), 6262–6271. https://doi.org/10.1074/jbc.M110.182618

Nakamichi, Y., Mikami, B., Murata, K., & Hashimoto, W. (2014). Crystal structure of a bacterial unsaturated glucuronyl hydrolase with specificity for heparin. *Journal of Biological Chemistry*, *289*(8), 4787–4797. https://doi.org/10.1074/jbc.M113.522573

Nakjang, S., Ndeh, D. A., Wipat, A., Bolam, D. N., & Hirt, R. P. (2012). A novel extracellular metallopeptidase domain shared by animal host-associated mutualistic and pathogenic microbes. *PLoS ONE*, *7*(1). https://doi.org/10.1371/journal.pone.0030287

Natalini, J. G., Singh, S., & Segal, L. N. (2023). The dynamic lung microbiome in health and disease. *Nature Reviews Microbiology*, *21*(4), 222–235. https://doi.org/10.1038/s41579-022-00821-x

Ndeh, D. A., Nakjang, S., Kwiatkowski, K. J., Koropatkin, N. M., Hirt, R. P., & Bolam, D. N. (2024). A genetic locus in the gut microbe Bacteroides thetaiotaomicron encodes activities consistent with mucin-O-glycoprotein processing and plays a critical role in N-acetylgalactosamine metabolism. *BioRxiv*. https://doi.org/10.1101/2024.02.01.578401

Ndeh, D., Baslé, A., Strahl, H., Yates, E. A., McClurgg, U. L., Henrissat, B., Terrapon, N., & Cartmell, A. (2020). Metabolism of multiple glycosaminoglycans by Bacteroides thetaiotaomicron is orchestrated by a versatile core genetic locus. *Nature Communications*, *11*(1), 1–12. https://doi.org/10.1038/s41467-020-14509-4

Ndeh, D., Munoz, J. M., Cartmell, A., Bulmer, D., Wills, C., Henrissat, B., & Gray, J. (2018). The human gut microbe Bacteroides thetaiotaomicron encodes the founding member of a novel glycosaminoglycan-degrading polysaccharide lyase family PL29. *Journal of Biological Chemistry*, *293*(46), 17906–17916. https://doi.org/10.1074/jbc.RA118.004510

Nie, K., Ma, K., Luo, W., Shen, Z., Yang, Z., Xiao, M., Tong, T., Yang, Y., & Wang, X. (2021). Roseburia intestinalis: A Beneficial Gut Organism From the Discoveries in Genus and Species. In *Frontiers in Cellular and Infection Microbiology* (Vol. 11). Frontiers Media S.A. https://doi.org/10.3389/fcimb.2021.757718

Nilsson, B., De Luca, S., Lohmander, S., & Hascall, V. C. (1982). Structures of N-linked and O-linked oligosaccharides on proteoglycan monomer isolated from the Swarm rat chondrosarcoma. *Journal of Biological Chemistry*, *257*(18), 10920–10927. https://doi.org/10.1016/s0021-9258(18)33911-5

Noach, I., Ficko-Blean, E., Pluvinage, B., Stuart, C., Jenkins, M. L., Brochu, D., Buenbrazo, N., Wakarchuk, W., Burke, J. E., Gilbert, M., & Boraston, A. B. (2017). Recognition of protein-linked glycans as a determinant of peptidase activity. *Proceedings of the National Academy of Sciences*, *114*(5), E679–E688. https://doi.org/10.1073/pnas.1615141114

O'Neill, M. A., Darvill, A. G., Etzler, M. E., Mohnen, D., Perez, S., Mortimer, J. C., & Pauly, M. (2022). Chapter 24: Viridiplantae and Algae. In *Essentials of Glycobiology* (4th ed.). Cold Spring Harbor Laboratory Press. https://doi.org/10.1101/glycobiology.4e.24

Onishi, S., Shionoya, K., Sato, K., Mubuchi, A., Maruyama, S., Nakajima, T., Komeno, M., Miyata, S., Yoshizawa, K., Wada, T., Linhardt, R. J., Toida, T., & Higashi, K. (2023). Fucosylated heparan sulfate from the midgut gland of Patinopecten yessoensis. *Carbohydrate Polymers*, *313*. https://doi.org/10.1016/j.carbpol.2023.120847

Overbeeke, A., Hausmann, B., Nikolov, G., Pereira, F. C., Herbold, C. W., & Berry, D. (2022). Nutrient niche specificity for glycosaminoglycans is reflected in polysaccharide utilization locus architecture of gut Bacteroides species. *Frontiers in Microbiology*, *13*. https://doi.org/10.3389/fmicb.2022.1033355

Pan, L., Ai, X., Fu, T., Ren, L., Shang, Q., Li, G., & Yu, G. (2021). In vitro fermentation of hyaluronan by human gut microbiota: Changes in microbiota community and potential degradation mechanism. *Carbohydrate Polymers*, *269*(June), 118313. https://doi.org/10.1016/j.carbpol.2021.118313

Paone, P., & Cani, P. D. (2020). Mucus barrier, mucins and gut microbiota: The expected slimy partners? *Gut*, *69*(12), 2232–2243. https://doi.org/10.1136/gutjnl-2020-322260

Park, D., Jagtap, S., & Nair, S. K. (2014). Structure of a PL17 family alginate lyase demonstrates functional similarities among exotype depolymerases. *Journal of Biological Chemistry*, *289*(12), 8645–8655. https://doi.org/10.1074/jbc.M113.531111

Pérez-Cruz, C., Moraleda-Montoya, A., Liébana, R., Terrones, O., Arrizabalaga, U., García-Alija, M., Lorizate, M., Martínez Gascueña, A., García-Álvarez, I., Nieto-Garai, J. A., Olazar-Intxausti, J., Rodríguez-Colinas, B., Mann, E., Chiara, J. L., Contreras, F. X., Guerin, M. E., Trastoy, B., & Alonso-Sáez, L. (2024). Mechanisms of recalcitrant fucoidan breakdown in marine Planctomycetota. *Nature Communications* , *15*(1). https://doi.org/10.1038/s41467-024-55268-w

Plaas, A. H. K., Moran, M. M., Sandy, J. D., & Hascall, V. C. (2023). Aggrecan and Hyaluronan: The Infamous Cartilage Polyelectrolytes – Then and Now. In *Advances in Experimental Medicine and Biology* (Vol. 1402, pp. 3–29). Springer. https://doi.org/10.1007/978-3-031-25588-5_1

Pluvinage, B., Ficko-Blean, E., Noach, I., Stuart, C., Thompson, N., McClure, H., Buenbrazo, N., Wakarchuk, W., & Boraston, A. B. (2021). Architecturally complex O-glycopeptidases are customized for mucin recognition and hydrolysis. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(10). https://doi.org/10.1073/pnas.2019220118

Pudlo, N. A., Pereira, G. V., Parnami, J., Cid, M., Markert, S., Tingley, J. P., Unfried, F., Ali, A., Varghese, N. J., Kim, K. S., Campbell, A., Urs, K., Xiao, Y., Adams, R., Martin, D., Bolam, D. N., Becher, D., Eloe-Fadrosh, E. A., Schmidt, T. M., … Martens, E. C. (2022). Diverse events have transferred genes for edible seaweed digestion from marine to human gut bacteria. *Cell Host and Microbe*, *30*(3), 314-328.e11. https://doi.org/10.1016/j.chom.2022.02.001

Qin, H. M., Miyakawa, T., Inoue, A., Nishiyama, R., Nakamura, A., Asano, A., Sawano, Y., Ojima, T., & Tanokura, M. (2017). Structure and polymannuronate specificity of a eukaryotic member of polysaccharide lyase family 14. *Journal of Biological Chemistry*, *292*(6), 2182–2190. https://doi.org/10.1074/jbc.M116.749929

Rahfeld, P., & Withers, S. G. (2020). Toward universal donor blood: Enzymatic conversion of A and B to O type. *Journal of Biological Chemistry*, *295*(2), 325–334. https://doi.org/10.1074/jbc.REV119.008164

Ravcheev, D. A., & Thiele, I. (2017). Comparative genomic analysis of the human gut microbiome reveals a broad distribution of metabolic pathways for the degradation of host-synthetized mucin glycans and utilization of mucin-derived monosaccharides. *Frontiers in Genetics*, *8*(AUG), 1–22. https://doi.org/10.3389/fgene.2017.00111

Rawat, P. S., Seyed Hameed, A. S., Meng, X., & Liu, W. (2022). Utilization of glycosaminoglycans by the human gut microbiota: participating bacteria and their enzymatic machineries. *Gut Microbes*, *14*(1). https://doi.org/10.1080/19490976.2022.2068367

Rawlings, N. D., Barrett, A. J., & Finn, R. (2016). Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*, *44*(D1), D343–D350. https://doi.org/10.1093/nar/gkv1118

Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A., & Finn, R. D. (2018). The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Research*, *46*(D1), D624–D632. https://doi.org/10.1093/nar/gkx1134

Rawlings, N. D., & Bateman, A. (2021). How to use the MEROPS database and website to help understand peptidase specificity. *Protein Science*, *30*(1), 83–92. https://doi.org/10.1002/pro.3948

Rios-Covian, D., Salazar, N., Gueimonde, M., & de los Reyes-Gavilan, C. G. (2017). Shaping the metabolism of intestinal Bacteroides population through diet to improve human health. In *Frontiers in Microbiology* (Vol. 8, Issue MAR). Frontiers Research Foundation. https://doi.org/10.3389/fmicb.2017.00376

Rodriguez, E., Roland, S. K., Plaas, A., & Roughley, P. J. (2006). The glycosaminoglycan attachment regions of human aggrecan. *Journal of Biological Chemistry*, *281*(27), 18444–18450. https://doi.org/10.1074/jbc.M512531200

Ross, F. C., Patangia, D., Grimaud, G., Lavelle, A., Dempsey, E. M., Ross, R. P., & Stanton, C. (2024). The interplay between diet and the gut microbiome: implications for health and disease. *Nature Reviews Microbiology*. https://doi.org/10.1038/s41579-024-01068-4

Roughley, P. J. (2006). The structure and function of cartilage proteoglycans. *European Cells and Materials*, *12*, 92–101. https://doi.org/10.22203/eCM.v012a11

Roughley, P. J., & Mort, J. S. (2014). The role of aggrecan in normal and osteoarthritic cartilage. *Journal of Experimental Orthopaedics*, *1*(1), 1–11. https://doi.org/10.1186/s40634-014-0008-7

Routier, F. H., Doering, T. L., Cummings, R. D., & Aebi, M. (2022). Chapter 23: Fungi. In *Essentials of Glycobiology* (4th ed.). Cold Spring Harbor Laboratory Press. https://doi.org/10.1101/glycobiology.4e.23

Saeed, N. K., Al-Beltagi, M., Bediwy, A. S., El-Sawaf, Y., & Toema, O. (2022). Gut microbiota in various childhood disorders: Implication and indications. *World Journal of Gastroenterology*, *28*(18), 1875–1901. https://doi.org/10.3748/wjg.v28.i18.1875

Salimi, M., Channab, B. eddine, El Idrissi, A., Zahouily, M., & Motamedi, E. (2023). A comprehensive review on starch: Structure, modification, and applications in slow/controlled-release fertilizers in agriculture. In *Carbohydrate Polymers* (Vol. 322). Elsevier Ltd. https://doi.org/10.1016/j.carbpol.2023.121326

Sayers, E. W., Beck, J., Bolton, E. E., Brister, J. R., Chan, J., Connor, R., Feldgarden, M., Fine, A. M., Funk, K., Hoffman, J., Kannan, S., Kelly, C., Klimke, W., Kim, S., Lathrop, S., Marchler-Bauer, A., Murphy, T. D., O'Sullivan, C., Schmieder, E., … Pruitt, K. D. (2025). Database resources of the National Center for Biotechnology Information in 2025. *Nucleic Acids Research*, *53*(D1), D20–D29. https://doi.org/10.1093/nar/gkae979

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

Seeberger, P. H. (2022). Chapter 2: Monosaccharide Diversity. In *Cold Spring Harbor* (4th ed.). Cold Spring Harbor Laboratory Press. https://doi.org/10.1101/glycobiology.4e.2

Shin, J. H., Tillotson, G., MacKenzie, T. N., Warren, C. A., Wexler, H. M., & Goldstein, E. J. C. (2024). Bacteroides and related species: The keystone taxa of the human gut microbiota. *Anaerobe*, *85*. https://doi.org/10.1016/j.anaerobe.2024.102819

Shionoya, K., Suzuki, T., Takada, M., Sato, K., Onishi, S., Dohmae, N., Nishino, K., Wada, T., Linhardt, R. J., Toida, T., & Higashi, K. (2022). Comprehensive analysis of chondroitin sulfate and aggrecan in the head cartilage of bony fishes: Identification of proteoglycans in the head cartilage of sturgeon. *International Journal of Biological Macromolecules*, *208*, 333–342. https://doi.org/10.1016/j.ijbiomac.2022.03.125

Shon, D. J., Fernandez, D., Riley, N. M., Ferracane, M. J., & Bertozzi, C. R. (2022). Structure-guided mutagenesis of a mucin-selective metalloprotease from Akkermansia muciniphila alters substrate preferences. *Journal of Biological Chemistry*, *298*(5), 101917. https://doi.org/10.1016/j.jbc.2022.101917

Shon, D. J., Malaker, S. A., Pedram, K., Yang, E., Krishnan, V., Dorigo, O., & Bertozzi, C. R. (2020). An enzymatic toolkit for selective proteolysis, detection, and visualization of mucin-domain glycoproteins. *Proceedings of the National Academy of Sciences*, *117*(35), 202012196. https://doi.org/10.1073/pnas.2012196117

Silva, Y. P., Bernardi, A., & Frozza, R. L. (2020). The Role of Short-Chain Fatty Acids From Gut Microbiota in Gut-Brain Communication. *Frontiers in Endocrinology*, *11*. https://doi.org/10.3389/fendo.2020.00025

Singh, R. P. (2019). Glycan utilisation system in Bacteroides and Bifidobacteria and their roles in gut stability and health. *Applied Microbiology and Biotechnology*, *103*(18), 7287–7315. https://doi.org/10.1007/s00253-019-10012-z

Sirotek, K., Slováková, L., Kopečný, J., & Marounek, M. (2004). Fermentation of pectin and glucose, and activity of pectin-degrading enzymes in the rabbit caecal bacterium Bacteroides caccae. *Letters in Applied Microbiology*, *38*(4), 327–332. https://doi.org/10.1111/j.1472-765X.2004.01492.x

Soldán, M., Argalášová, Ľ., Hadvinová, L., Galileo, B., & Babjaková, J. (2024). The Effect of Dietary Types on Gut Microbiota Composition and Development of Non-Communicable Diseases: A Narrative Review. *Nutrients*, *16*(18). https://doi.org/10.3390/nu16183134

Stam, M., Lelièvre, P., Hoebeke, M., Corre, E., Barbeyron, T., & Michel, G. (2023). SulfAtlas, the sulfatase database: state of the art and new developments. *Nucleic Acids Research*, *51*(1), D647–D653. https://doi.org/10.1093/nar/gkac977

Suhre, K., & Sanejouand, Y. H. (2004). ElNémo: A normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Research*, *32*(WEB SERVER ISS.). https://doi.org/10.1093/nar/gkh368

Sulzenbacher, G., Liu, Q. P., Bennett, E. P., Levery, S. B., Bourne, Y., Ponchel, G., Clausen, H., & Henrissat, B. (2010). A novel α-N-acetylgalactosaminidase family with an NAD +-dependent catalytic mechanism suitable for enzymatic removal of blood group A antigens. *Biocatalysis and Biotransformation*, *28*(1), 22–32. https://doi.org/10.3109/10242420903424259

Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A. L., Fang, T., Doncheva, N. T., Pyysalo, S., Bork, P., Jensen, L. J., & Von Mering, C. (2023). The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, *51*(1 D), D638–D646. https://doi.org/10.1093/nar/gkac1000

Tailford, L. E., Crost, E. H., Kavanaugh, D., & Juge, N. (2015). Mucin glycan foraging in the human gut microbiome. *Frontiers in Genetics*, *5*(FEB). https://doi.org/10.3389/fgene.2015.00081

Taleb, V., Liao, Q., Narimatsu, Y., García-García, A., Compañón, I., Borges, R. J., González-Ramírez, A. M., Corzana, F., Clausen, H., Rovira, C., & Hurtado-Guerrero, R. (2022). Structural and mechanistic insights into the cleavage of clustered O-glycan patches-containing glycoproteins by mucinases of the human gut. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-32021-9

Terrapon, N., Lombard, V., Drula, É., Lapébie, P., Al-Masaudi, S., Gilbert, H. J., & Henrissat, B. (2018). PULDB: The expanded database of Polysaccharide Utilization Loci. *Nucleic Acids Research*, *46*(D1), D677–D683. https://doi.org/10.1093/nar/gkx1022

Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., Winther, O., Brunak, S., von Heijne, G., & Nielsen, H. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature Biotechnology*, *40*(7), 1023–1025. https://doi.org/10.1038/s41587-021-01156-3

Teze, D., Shuoker, B., Chaberski, E. K., Kunstmann, S., Fredslund, F., Nielsen, T. S., Stender, E. G. P., Peters, G. H. J., Karlsson, E. N., Welner, D. H., & Hachem, M. A. (2020). The Catalytic Acid-Base in GH109 Resides in a Conserved GGHGG Loop and Allows for Comparable α-Retaining and β-Inverting Activity in an N-Acetylgalactosaminidase from Akkermansia muciniphila. *ACS Catalysis*, *10*(6), 3809–3819. https://doi.org/10.1021/acscatal.9b04474

The 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

Tuncil, Y. E., Xiao, Y., Porter, N. T., Reuhs, B. L., Martens, E. C., & Hamaker, B. R. (2017). Reciprocal prioritization to dietary glycans by gut bacteria in a competitive environment promotes stable coexistence. *MBio*, *8*(5), 1–17. https://doi.org/10.1128/mBio.01068-17

Ulmer, J. E., Vilén, E. M., Namburi, R. B., Benjdia, A., Beneteau, J., Malleron, A., Bonnaffé, D., Driguez, P. A., Descroix, K., Lassalle, G., Le Narvor, C., Sandström, C., Spillmann, D., & Berteau, O. (2014). Characterization of glycosaminoglycan (GAG) sulfatases from the human gut symbiont Bacteroides thetaiotaomicron reveals the first GAG-specific bacterial endosulfatase. *Journal of Biological Chemistry*, *289*(35), 24289–24303. https://doi.org/10.1074/jbc.M114.573303

Wang, J., Chen, M. S., Wang, R. S., Hu, J. Q., Liu, S., Wang, Y. Y. F., Xing, X. L., Zhang, B. W., Liu, J. M., & Wang, S. (2022). Current Advances in Structure-Function Relationships

and Dose-Dependent Effects of Human Milk Oligosaccharides. *Journal of Agricultural and Food Chemistry*, *70*(21), 6328–6353. https://doi.org/10.1021/acs.jafc.2c01365

Wang, X., Sun, G., Feng, T., Zhang, J., Huang, X., Wang, T., Xie, Z., Chu, X., Yang, J., Wang, H., Chang, S., Gong, Y., Ruan, L., Zhang, G., Yan, S., Lian, W., Du, C., Yang, D., Zhang, Q., … Geng, M. (2019). Sodium oligomannate therapeutically remodels gut microbiota and suppresses gut bacterial amino acids-shaped neuroinflammation to inhibit Alzheimer's disease progression. *Cell Research*, *29*(10), 787–803. https://doi.org/10.1038/s41422-019-0216-x

Wardman, J. F., Bains, R. K., Rahfeld, P., & Withers, S. G. (2022). Carbohydrate-active enzymes (CAZymes) in the gut microbiome. In *Nature Reviews Microbiology* (Vol. 20, Issue 9, pp. 542–556). Nature Research. https://doi.org/10.1038/s41579-022-00712-1

Wardman, J. F., Sim, L., Liu, J., Howard, T. A., Geissner, A., Danby, P. M., Boraston, A. B., Wakarchuk, W. W., & Withers, S. G. (2023). A high-throughput screening platform for enzymes active on mucin-type O-glycoproteins. *Nature Chemical Biology*, *19*(10), 1246–1255. https://doi.org/10.1038/s41589-023-01405-3

Wargaki, A. J., Leonard, E., Nyan Win, M., Regitsky, D. D., Santos, C. N. S., Kim, P. B., Cooper, S. R., Raisner, R. M., Herman, A., Sivitz, A. B., Lakshmanaswamy, A., Kashiyama, Y., Baker, D., & Yoshikuni, Y. (2012). An Engineered Microbial Platform for Direct Biofuel Production from Brown Macroalgae. *Science*, *335*(6066), 303–308. https://doi.org/10.1126/science.1214707

Wei, B., Dalwadi, H., Gordon, L. K., Landers, C., Bruckner, D., Targan, S. R., & Braun, J. (2001). Molecular Cloning of a Bacteroides caccae TonB-Linked Outer Membrane Protein Identified by an Inflammatory Bowel Disease Marker Antibody. *Infection and Immunity*, *69*(10), 6044–6054. https://doi.org/10.1128/IAI.69.10.6044-6054.2001

Wei, L., Cao, H.-Y., Du, M., Zhang, Q., Lu, D., Xu, X., Xu, Y., Wang, W., Chen, X.-L., Zhang, Y.-Z., & Li, F. (2024). Crystal structure and catalytic mechanism of PL35 family glycosaminoglycan lyases with an ultrabroad substrate spectrum. *ELife*. https://doi.org/10.1016/j.jbc.2024.107466

Wei, L., Zou, R., Du, M., Zhang, Q., Lu, D., Xu, Y., Xu, X., Wang, W., Zhang, Y. Z., & Li, F. (2024). Discovery of a class of glycosaminoglycan lyases with ultrabroad substrate spectrum and their substrate structure preferences. *Journal of Biological Chemistry*, *300*(7). https://doi.org/10.1016/j.jbc.2024.107466

Wu, S., Bekhit, A. E. D. A., Wu, Q., Chen, M., Liao, X., Wang, J., & Ding, Y. (2021). Bioactive peptides and gut microbiota: Candidates for a novel strategy for reduction and control of neurodegenerative diseases. *Trends in Food Science and Technology*, *108*, 164–176. https://doi.org/10.1016/j.tifs.2020.12.019

Yang, Q., Chang, S., Zhang, X., Luo, F., Li, W., & Ren, J. (2024). The fate of dietary polysaccharides in the digestive tract. *Trends in Food Science and Technology*, *150*(January), 104606. https://doi.org/10.1016/j.tifs.2024.104606

Ye, M., Yu, J., Shi, X., Zhu, J., Gao, X., & Liu, W. (2020). Polysaccharides catabolism by the human gut bacterium -Bacteroides thetaiotaomicron: advances and perspectives. *Critical Reviews in Food Science and Nutrition*, *0*(0), 1–20. https://doi.org/10.1080/10408398.2020.1803198

Yoshida, C. A., Kawane, T., Moriishi, T., Purushothaman, A., Miyazaki, T., Komori, H., Mori, M., Qin, X., Hashimoto, A., Sugahara, K., Yamana, K., Takada, K., & Komori, T. (2014). Overexpression of Galnt3 in chondrocytes resulted in dwarfism due to the increase of

mucin-type O-glycans and reduction of glycosaminoglycans. *Journal of Biological Chemistry*, *289*(38), 26584–26596. https://doi.org/10.1074/jbc.M114.555987

Yüksel, E., Kort, R., & Voragen, A. G. J. (2024). Structure and degradation dynamics of dietary pectin. *Critical Reviews in Food Science and Nutrition*. https://doi.org/10.1080/10408398.2024.2437573

Zhang, C., Yu, L., Zhai, Q., Zhao, R., Zhao, J., Zhang, H., Chen, W., & Tian, F. (2023). In vitro fermentation of heparin by the human gut microbiota: Changes in the microbiota community and metabolic functions. *Food Chemistry*, *406*. https://doi.org/10.1016/j.foodchem.2022.135010

Zhang, Z., Park, Y., Kemp, M. M., Zhao, W., Im, A. R., Shaya, D., Cygler, M., Kim, Y. S., & Linhardt, R. J. (2009). Liquid chromatography-mass spectrometry to study chondroitin lyase action pattern. *Analytical Biochemistry*, *385*(1), 57–64. https://doi.org/10.1016/j.ab.2008.10.014