FUNCTIONAL DISSECTION OF BETA-GLUCAN UTILIZATION BY PROMINENT HUMAN GUT SYMBIONTS

by

Kazune Tamura

B.Sc., The University of British Columbia, 2015

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Biochemistry and Molecular Biology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2021

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Functional dissection of beta-glucan utilization by prominent human gut symbionts

submitted by Kazune Tamura in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

in Biochemistry and Molecular Biology

Examining Committee:

Prof. Harry Brumer, Chemistry, UBC
Supervisor

Prof. Filip Van Petegem, Biochemistry and Molecular Biology, UBC
Supervisory Committee Member

Prof. Nobuhiko Tokuriki, Biochemistry and Molecular Biology, UBC
Supervisory Committee Member

Prof. Natalie Strynadka, Biochemistry and Molecular Biology, UBC
University Examiner

Prof. William Mohn, Microbiology and Immunology, UBC
University Examiner

Abstract

The human gut microbiota (HGM) is a remarkably dense and dynamic community of microbes with an incredible collective metabolic capacity. Complex glycans ("dietary fiber") evade digestion by the human host and feed the HGM, driving its composition, and in turn influencing diverse facets of host health. In order to access these otherwise recalcitrant glycans, Bacteroidetes, a dominant bacterial phylum in the HGM, co-localize genes encoding a membrane-associated machinery that work in concert to bind, deconstruct, and sequester target glycan into co-regulated Polysaccharide Utilization Loci (PULs). One important class of complex glycans with numerous documented health benefits is the β-glucans. In this thesis, I undertake holistic functional characterization, combining biochemistry, structural biology, microbiology, and (meta)genomics, of PULs targeting two distinct subclasses of β-glucan: mixed-linkage β-glucan (MLG), and β(1,3)-glucan.

The MLG utilization locus (MLGUL) from *B. ovatus*, necessary to enable growth on MLG, encodes an endo-β-glucanase anchored to the cell surface to fragment MLG, and a periplasmic exo-β-glucosidase to completely saccharify the oligosaccharides imported by the TonB-dependent outer-membrane transporter. The process is aided by two cell surface glycan-binding proteins (SGBPs) which employ binding platforms shaped to complement that of target MLG to recruit and retain the glycan at the cell surface. Growth analysis combined with comparative genomics reveal MLGULs serve as genetic markers for ability to grow on MLG. Metagenomic analysis further suggest MLGUL presence in, and consequent ability to utilize MLG by, the HGM of a majority of humans worldwide.

A distinct set of syntenic β(1,3)-glucan utilization loci (1,3GULs) from three prominent Bacteroides species (*B. uniformis*, *B. thetaiotaomicron*, and *B. fluxus*) were subject to similar holistic functional characterization. Differential ability to grow on β-glucan congeners is driven by synergy between enzymes and SGBPs: a particular 1,3GUL can mediate utilization of a β-glucan congener if it encodes both an enzyme that can hydrolyze the target, as well as an SGBP that can bind the target. Detailed structure-function analysis of glycoside hydrolases (GH), including a family-first structure of a GH158, and a suite of SGBPs reveal the molecular basis of catalytic and binding specificities that together give rise to species-differential specificity of 1,3GULs.

Lay Summary

The trillions of microbial organisms that reside in our large intestine, collectively called the human gut microbiota (HGM), have an incredible collective metabolic capacity to degrade "dietary fiber" (complex glycans) that we humans cannot intrinsically digest. Consequently, these indigestible glycans we consume shape HGM composition, and in turn influence diverse facets of our health. Beta-glucans are one such class of indigestible glycans, and a common part of the human diet worldwide (cereals, mushrooms, seaweeds, yeasts), whose consumption has been linked to numerous health benefits. In this thesis, I investigate the enzymes and associated proteins that work together to endow Bacteroidetes, one of the dominant members of the HGM, the ability to utilize and grow on beta-glucans. A holistic, multidisciplinary approach combining biochemistry, structural biology, microbiology, and (meta)genomics sheds unprecedented light on this process and will inform the development of novel strategies to manipulate the HGM for health-promoting and therapeutic purposes.

Preface

Chapter 1 incorporates content published in a review article: [Tamura, K.], Brumer, H. (2021) Glycan utilization systems in the human gut microbiota: A gold mine for structural discoveries., *Curr. Opin. Struct. Biol.*, **68**, 26-40.

I conducted the literature review, composed the manuscript, created the figures, and contributed to revision of the article. Prof. Harry Brumer refined and revised the article.

A version of Chapter 2 has been published: [Tamura, K.], Hemsworth, G.R., Déjean, G., Rogers, T.E., Pudlo, N.A., Urs, K., Jain, N., Davies, G.J., Martens, E.C., Brumer, H. (2017) Molecular mechanism by which prominent human gut Bacteroidetes utilize mixed-linkage beta-glucans, major health-promoting cereal polysaccharides, *Cell Rep.*, **21**, 417-430.

I cloned, expressed, and purified recombinant enzymes; conducted and analyzed kinetics for hydrolysis of polysaccharides, oligosaccharides, and chromogenic substrates; determined hydrolysis products; conducted phylogenetic and structural analyses; and carried out metagenomics survey. I created all figures (besides Fig. 2-5 and Fig. A-1), drafted the entire manuscript (besides the crystallography and microbiology methods) and contributed to revision of the article. Dr. Glyn R. Hemsworth determined the crystal structures and drafted the crystallography methods. Dr. Guillaume Dejean conducted enzyme localization analyses (microscopy and western blot) and drafted the enzyme localization methods. Dr. Theresa E. Rogers conducted whole-MLGUL knockout in *B. ovatus* and growth analysis. Dr. Nicholas A. Pudlo and Karthik Urs conducted meta-growth analysis of Bacteroidetes and Dr. Nicholas A. Pudlo drafted the microbiology methods. Namrata Jain synthesized the G3G-CNP substrate. Prof. Gideon J. Davies supervised structural biology and provided feedback on the manuscript. Prof. Eric C. Martens supervised microbiology, created Fig. 5 and Fig. S1, and provided feedback on the manuscript. Prof. Harry Brumer coordinated overall research and revised the article.

A version of Chapter 3 has been published: [Tamura, K.], Foley, M.H., Gardill, B.R., Dejean, G., Schnizlein, M., Bahr, C.M.E., Creagh, A.L., Van Petegem, F., Koropatkin, N.M., Brumer, H. (2019) Surface glycan binding proteins are essential for cereal beta-glucan utilization by the human gut symbiont *Bacteroides ovatus*. *Cell. Mol. Life. Sci.*, **76**, 4319–4340.

I cloned, expressed, and purified recombinant SGBPs, GFP-fusion proteins, and site-directed mutants, conducted affinity gel electrophoresis and pull-down depletion isotherm experiments, produced and purified MLG partial digest oligosaccharides, and solved crystal structures of $BoSGBP_{MLG}$-A, $BoSGBP_{MLG}$-A_MLG7 and $BoSGBP_{MLG}$-B_MLG7 with assistance from Dr. Bernd. R. Gardill. I created all figures (besides Fig. 3-5 and 3-6, and Fig. B-11, B-12, and B-14), and co-drafted the manuscript. Dr. Matthew H. Foley conducted SGBP gene knockouts and complementations in *B. ovatus* and growth analyses, created Fig. 3-5 and 3-6, and Fig. B-11, B-12, and B-14, and co-drafted the manuscript. Dr. Bernd R. Gardill taught me all aspects of X-ray crystallography and assisted significantly in the determination of $BoSGBP_{MLG}$-A, $BoSGBP_{MLG}$-A_MLG7 and $BoSGBP_{MLG}$-B_MLG7 crystal structures. Dr. Guillaume Dejean conducted ITC experiments. Matthew Schnizlein expressed, purified and crystallized the selenomethionine variant of $BoSGBP_{MLG}$-B. Dr. Constance M. E. Bahr crystalized $BoSGBP_{MLG}$-A_cellohexaose and native $BoSGBP_{MLG}$-B_cellohexaose. Dr. Louise A. Creagh assisted with ITC data collection and analysis. Prof. Filip Van Petegem supervised structural biology and provided feedback on the manuscript. Prof. Nicole M. Koropatkin supervised microbiology, determined crystal structures of $BoSGBP_{MLG}$-A_cellohexaose and $BoSGBP_{MLG}$-B_cellohexaose, and provided feedback on the manuscript. Prof. Harry Brumer coordinated overall research and revised the article.

A version of Chapter 4 has been published: Déjean, G., [Tamura, K.], Cabrera, A., Jain, N., Pudlo, N.A., Pereira, G., Viborg, A.H., Van Petegem, F., Martens, E.C., Brumer, H. (2020) Synergy between cell surface glycosidases and glycan-binding proteins dictates the utilization of specific beta(1,3)-glucans by human gut *Bacteroides*. *mBio*, **11**, e00095-20.

This work was spearheaded by Dr. Guillaume Dejean who performed the bacterial growth analyses, qPCR transcript analyses, comparative bioinformatics, gene cloning and recombinant

protein production, and enzymatic and biophysical analyses. I performed recombinant protein production and purification, all crystallography and tertiary structural analyses, phylogenetic analysis of GH16, PFAM13004 domain production and analysis, assisted with characterization of *B. uniformis* enzymes (all pH- and temperature-rate analyses, and BuGH3 Glc-*p*NP kinetics), and conducted GH158 product anomer analysis with Dr. Namrata Jain. I contributed significantly to drafting the manuscript (enzyme biochemistry and structural biology; section 4.3.2), heavily edited remaining sections drafted by Dr. Guillaume Dejean, and contributed to revisions pre- and post-submission (which Dr. Guillaume Dejean was not involved in). Adriana Cabrera assisted Dr. Guillaume Dejean with GH cloning, production, and characterization. Namrata Jain synthesized laminaribiose-CNP and performed NMR analysis of GH158 product with me. Nicholas A. Pudlo assisted Dr. Guillaume Dejean with bacterial growth analyses. Dr. Gabriel Pereira performed human metagenome analysis. Alexander H. Viborg performed phylogenetic analysis of GH158. Prof. Filip Van Petegem supervised structural biology and provided feedback on the manuscript. Prof. Eric C. Martens supervised microbiology, contributed to project conceptualization, and provided feedback on the manuscript. Prof. Harry Brumer coordinated overall research and revised the article.

A version of Chapter 5 has been published: [Tamura K.], Déjean, G., Van Petegem, F. Brumer, H. (2021) Distinct protein architectures mediate species-specific beta-glucan binding and metabolism in the human gut microbiota. *J. Biol. Chem.*, **296**, 100415.

I conceptualized the project, designed experiments, generated gene expression constructs, produced and purified protein, performed all biochemistry and crystallography, and analyzed data. I also created all figures and composed the article. Dr. Guillaume Dejean cloned full-length genes and contributed to project conceptualization. Prof. Filip Van Petegem supervised structural biology and provided feedback on the manuscript. Prof. Harry Brumer coordinated overall research and revised the article.

Chapter 6 incorporates content published in a review article: Grondin, J.M., [Tamura, K.], Déjean, G., Abbott, D.W., Brumer, H. (2017) Polysaccharide Utilization Loci: Fuelling microbial communities. *J. Bacteriol.*, **99**, e00860-16.

I created all figures, contributed to literature review, drafting of manuscript, and revision of the article.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AA: auxiliary activity

AGE: affinity gel electrophoresis

APS: Advanced Photon Source

BBG: barley beta-glucan

BCA: bicinchoninic acid

BSA: bovine serum albumin

CAZyme: carbohydrate-active enzyme

CBM: carbohydrate-binding module

CE: carbohydrate esterase

CLS: Canadian Lightsource

CMC: carboxymethylcellulose

CNP: 2-chloro-4-nitrophenyl

DMSO: dimethyl sulfoxide

DNA: deoxyribonucleic acid

DP: degree of polymerization

EDTA: ethylenediaminetetraacetic acid

F_{calc}: calculated structure factor

F_{obs}: experimental structure factor

G6PDH: glucose-6-phosphate dehydrogenase

GH: glycoside hydrolase

GT: glycosyl transferase

HEC: hydroxyethylcellulose

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HK: hexokinase

HPAEC-PAD: high-performance anion exchange chromatography – pulsed amperometric detection

HPLC: high-performance liquid chromatography

HTCS: hybrid two-component system

IG: immunoglobulin

IMG: Integrated Microbial Genomics

IPTG: isopropyl β-D-1-thiogalactopyranoside

ITC: isothermal titration calorimetry

JGI: Joint Genome Institute

LacNAc: N-acetyl lactosamine

LB: lysogeny broth

LC-MS: liquid chromatography – mass spectrometry

LN2: liquid nitrogen

m/z: mass-to-charge ratio

MAD: multiple anomalous dispersion

MLG: mixed-linkage glucan

MLGO: mixed-linkage glucan oligosaccharide

MR: molecular replacement

NADPH: nicotinamide adenine dinucleotide phosphate reduced

NCBI: National Center for Biotechnology Information

Ni-NTA: nickel nitrilotriacetic acid

NMR: nuclear magnetic resonance

PCR: polymerase chain reaction

PDB: Protein Data Bank

PEG: polyethylene glycol

PKD: polycystic kidney disease domain

PL: polysaccharide lyase

*p*NP: para-nitrophenyl

PUL: polysaccharide utilization locus

R_{free}: residual factor of test set of reflections

R_{meas}: residual factor independent of multiplicity

RMSD: root-mean-square deviation

RNA: ribonucleic acid

R_{work}: residual factor of all reflections

SAD: single anomalous dispersion

SDS-PAGE: sodium dodecyl sulfate – polyacrylamide gel electrophoresis

SEC: size-exclusion chromatography

SeMet: selenomethionine

sfGFP: super folder green fluorescent protein

SGBP: surface glycan-binding protein

SSRL: Stanford Synchrotron Radiation Lightsource

Sus: starch utilization system

TB: terrific broth

TBDT: TonB-dependent transporter

TPR: tetratricopeptide repeat

UV: ultraviolet

XyG: xyloglucan

XyGO: xyloglucan oligosaccharide

Acknowledgements

I would like to extend my sincere gratitude to Dr. Harry Brumer for his guidance, mentorship, and support during my graduate studies. I appreciate that I was afforded substantial freedom to explore my interests, chase my curiosities, and decide on the directions to proceed. The privilege to exercise discretion and independence from early on undoubtedly contributed to my growth as a scientist and to the productivity in my learning as well as scientific output. I am thankful to my committee members Dr. Filip Van Petegem and Dr. Nobuhiko Tokuriki for all the valuable advice and discussions regarding not just my PhD project but also my career and future in science. I want to thank UBC for all the financial support, especially through the Four-Year Doctoral Fellowship, which enabled me to work and study while also enjoying my many (too expensive) outdoor hobbies.

I must also express my enduring gratitude to Dr. Filip Van Petegem for – in addition to serving on my committee – opening the door to the world of structural biology. I spent just as much time in the Van Petegem Lab as I did in the Brumer Lab in the later years as I pursued various X-ray crystallography endeavours. I am grateful to the members of his group for being welcoming and helpful, and in particular to Dr. Bernd Gardill for showing me the ropes during the early innings.

I am fortunate to have had the opportunity to engage with numerous amazing collaborators from around the world. I am particularly grateful to Dr. Nicole Koropatkin, Dr. Matthew Foley, Dr. Glyn Hemsworth, Dr. Eric Martens, and Dr. Gideon Davies for sharing their expertise and for the fruitful collaborations. Meeting in person at various conferences and on collaboration visits after having worked remotely together in professional capacities really taught me what it means to be a great person in addition to being a great scientist.

I would like to thank all members of the Brumer Lab that I had the privilege of working alongside and sharing my time with. Particularly, my fellow "classmates" Sean McDonald, Maria Cleveland, Roderick Xia, Hila Behar, and James Li; we all started on our graduate journeys at the same time and went through many ups and downs together. I must also thank Dr. Guillaume Dejean, an invaluable teammate on the gut microbiota PUL team during the bulk of my time in the lab. I also want to thank Dr. Shaheen Shojania for always being there for me in the absence of having family in town, especially on my (too frequent) visits to the hospital for one reason or another.

Chapter 1: Introduction

1.1 Human gut microbiota

Trillions of microorganisms are estimated to live in and on us, the vast majority of which reside in our gastrointestinal (GI) tract. This community of microbes, collectively known as the human gut microbiota (HGM), is one of the most dense and dynamic on the planet (1). The HGM provides many symbiotic benefits to the host including protection from pathogens, stimulation of the immune system, and the production of short-chain fatty acids which are absorbed and utilized as the preferred energy source by the epithelial cells of the intestinal lining (2). Also, robust correlation between breakdown in the proper balance of the HGM (aka dysbiosis) and numerous diseases have been established and continue to be uncovered. Prominent examples include metabolic diseases (diabetes, hypertension, obesity, etc.), autoimmune disorders (asthma, atopy, allergies, etc.), inflammatory bowel disease, cancers, and even neurological disorders (3-7). In many cases, direct causal links have also been demonstrated (8,9).

The composition and population of the HGM is easily altered on a short timescale (9,10), making their manipulation an attractive therapeutic route to improve health and treat diseases (11,12). However, the prospect of harnessing microbial intervention strategies as a means to correct dysbiosis is currently hampered by limited mechanistic insights into interactions between nutrients, microbes, and host.

One of the main factors that shape HGM composition is metabolism of complex glycans (*vide infra*) – i.e. indigestible "dietary fiber" from various plant, whole-grain, and microbial sources – that we consume through our diet (13,14). Complex glycans are defined by their extraordinary diversity of monosaccharide components and glycosidic linkages, rendering them recalcitrant to the limited set of human-encoded carbohydrate-active enzymes (CAZymes; *vide infra*). In stark contrast, the HGM collectively harbors an expanded repertoire of CAZymes (Fig. 1-1) (15), enabling microbes to thrive on complex glycans that evade digestion by the human host. In fact, the gut microbiome (collective genomic complement of the HGM) encodes on the order of one hundred times as many genes as the human genome (16). Due to the resulting enormous metabolic throughput and the growing appreciation of its influence on human physiology, the HGM has often been called "the forgotten organ" (17).

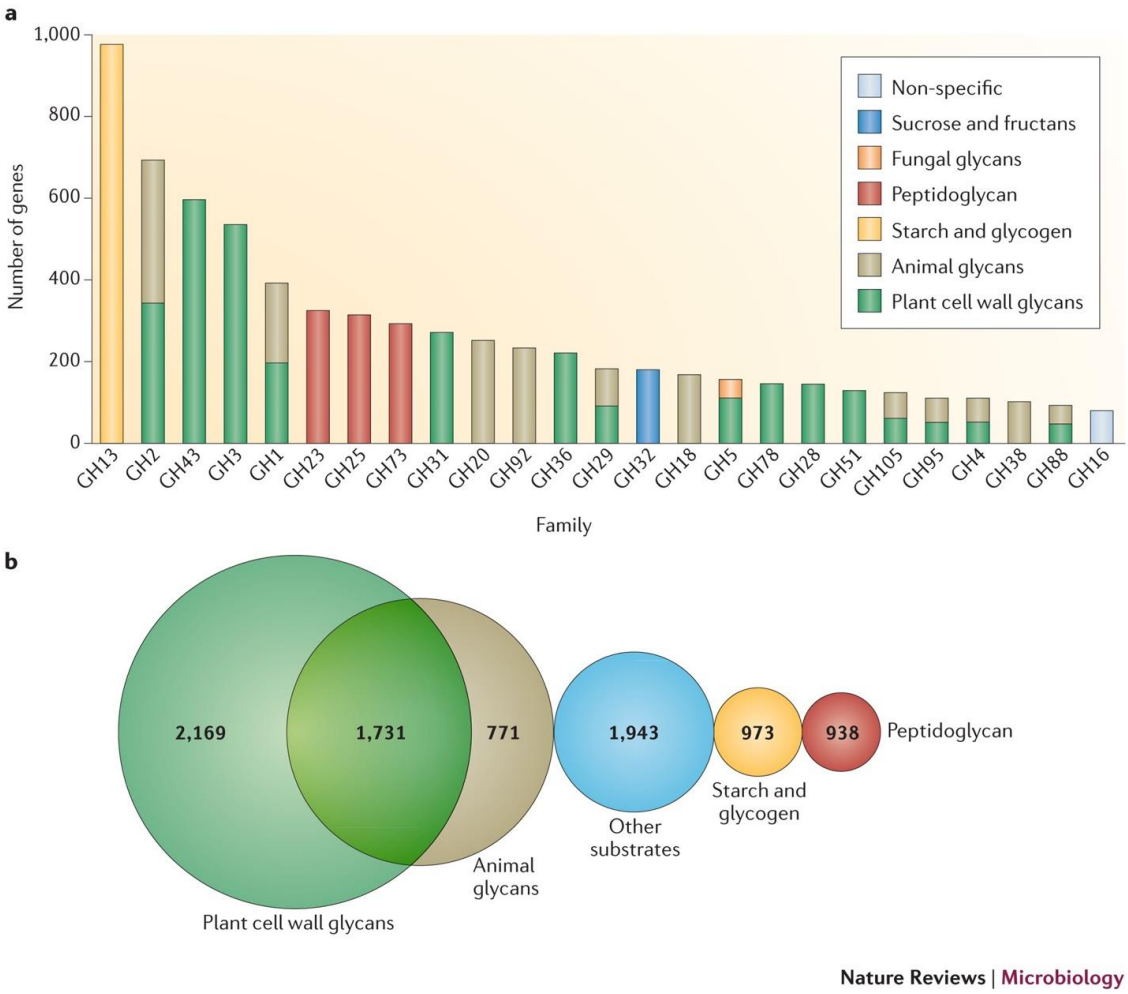

Figure 1-1 Expansion of GHs in the human gut microbiota.

(A) Abundance of GHs and variety of GH families in a representative human gut microbiota. Target substrate categories in different colors as specified in the legend. (B) Venn diagram of substrate specificity categories of GHs and PLs revealing particular enrichment of those targeting plant glycans. Reproduced from El Kaoutari et al. (15)

1.2 Glycans in the human gut

Carbohydrates are one of three major classes of macronutrients that comprise the human diet alongside lipids and proteins. Among the three, carbohydrates account for the majority of daily caloric supply worldwide (18). The simplest building blocks of dietary carbohydrates are monosaccharides, which upon cyclization can have its anomeric hydroxyl group in the alpha or beta stereochemical configuration (19). These monosaccharides, which can have a number of different chemical modifications on any of the free hydroxyl groups, can be linked by glycosidic bonds to form oligo- and polysaccharides. These glycosidic bonds are formed between the anomeric carbon of one monosaccharide and any of the hydroxyl groups on another

monosaccharide (including between two anomeric centers such as in the case of sucrose; *vide infra*). Multiple glycosidic bonds to a single monosaccharide can also occur, leading to branching, and glycosidic bonds to non-carbohydrate moieties are also common. The variety of linkage options combined with the diverse palette of monosaccharides that could be strung together combine to give rise to the enormous diversity of glycans found in nature (19).

Despite this diversity of available glycans in nature, and indeed in our diet, humans are only intrinsically capable of digesting a handful of them (15). These "digestible" glycans can be broken down to constituent monosaccharides by CAZymes expressed by the human genome and absorbed in the small intestine. All other "indigestible" glycans transit to the large intestine where they have the potential to feed the HGM (13). Commonly referred to collectively as "complex glycans", these prebiotic dietary fibers are ubiquitously derived from plant cell wall and storage polysaccharides, as well as from a diversity of other sources including marine plants and edible microbes/fungi.

1.2.1 Human-digestible glycans

Prominent simple sugars in the human diet include sucrose, a disaccharide formed by a $_D$-glucose monomer joined by an $\alpha(1,2)$ glycosidic bond to a $_D$-fructose monomer, and lactose, another disaccharide made up of a $_D$-galactose $\beta(1,4)$-linked to a $_D$-glucose (19). Another major source of calories in the human diet is starch, a polysaccharide comprised of repeating $\alpha(1,4)$-linked $_D$-glucose subunits (with $\alpha(1,6)$-linked branching in the case of amylopectin; unbranched starch is called amylose) (20). Though most forms of starch in our diet is degraded in the upper GI tract, some dietary starch cannot be processed by host starch-hydrolyzing CAZymes and arrives intact in the large intestine as resistant starch, which also has prebiotic properties (21). (See Table 1 in ref. (21) for different types of resistant starch). Sucrose and starch are storage polysaccharides derived from diverse plant sources whereas lactose is a common sugar found in milk from various mammals.

1.2.2 Glycans available to the HGM

1.2.2.1 Plant cell wall polysaccharides

Terrestrial plant cells are protected by a robust coat of polysaccharide-rich cell wall, outside the plasma membrane, which provides structural integrity, protection against pathogens, as well as a platform for intracellular signalling. Plant cell walls are comprised of multiple distinct

layers comprised of different ratios and organization of constituent cellulose, hemicellulose, pectin, lignin, and proteins (22).

Cellulose is the predominant polysaccharide in the plant cell wall, as well as the most abundant store of carbon on the planet (23). It is a linear polysaccharide comprised of repeating β(1,4)-linked $_{D}$-glucose units with a regular interchain hydrogen bonding network that forms a crystalline fibril. Cellulose is largely recalcitrant to degradation by human-encoded CAZymes as well as those encoded by the HGM, as is lignin, a phenolic polymer that together with cellulose significantly contribute to the structural rigidity of the plant cell wall (22). Hemicellulose and pectin are heterogenous polysaccharides that cross link cellulose microfibrils, which serves to contribute elastic properties to complement the robust nature of the cell wall (24). These two classes of glycans are very diverse and complex, and are potentially prebiotic dietary fibers that can be deconstructed by members of the HGM.

1.2.2.1.1 Hemicelluloses

Structurally, hemicelluloses are very complex and diverse but can be broadly categorized into four groups based on linkage and monosaccharide composition: xylan, xyloglucan, β-mannan, and β-glucan. Their abundance and distribution are highly variable depending on plant species, tissue type, and the exact layer within the cell wall.

Xylans have a backbone structure comprised of β(1,4)-linked xylose, which can be decorated with a huge diversity of branching options depending on source (Fig 1-2) (24,25). For example, relatively simple glucuronoxylan from hardwoods like birch have single 4-O-methyl-α-glucuronic acid decorations at the O2 position of xylose. On the opposite end of the complexity spectrum, glucuronoarabinoxylans from for example maize have additional single or double α-arabinofuranosyl substitutions at the O2 and/or O3 positions, which can be further extended with β-xylosyl and α-$_{L}$-galactosyl residues. Backbone xylose residues can also be decorated with acetyl groups.

Xyloglucans consist of a core structure of β(1,4)-linked glucose backbone with two or three consecutive α(1,6)-xylose branches per every four consecutive glucose, depending on source (Fig. 1-2). The former core structure is more prevalent in solanaceous plants such as tomatoes with the latter more prevalent in dicots such as cabbage (27). The xylosyl branches can be further substituted with β-galactosyl, α-arabinofuranosyl, and α-fucosyl residues depending on plant

tissue and species of origin (28). Xyloglucans have high affinity for cellulose and play an important role in tethering cellulose microfibrils together within the plant cell wall (24).

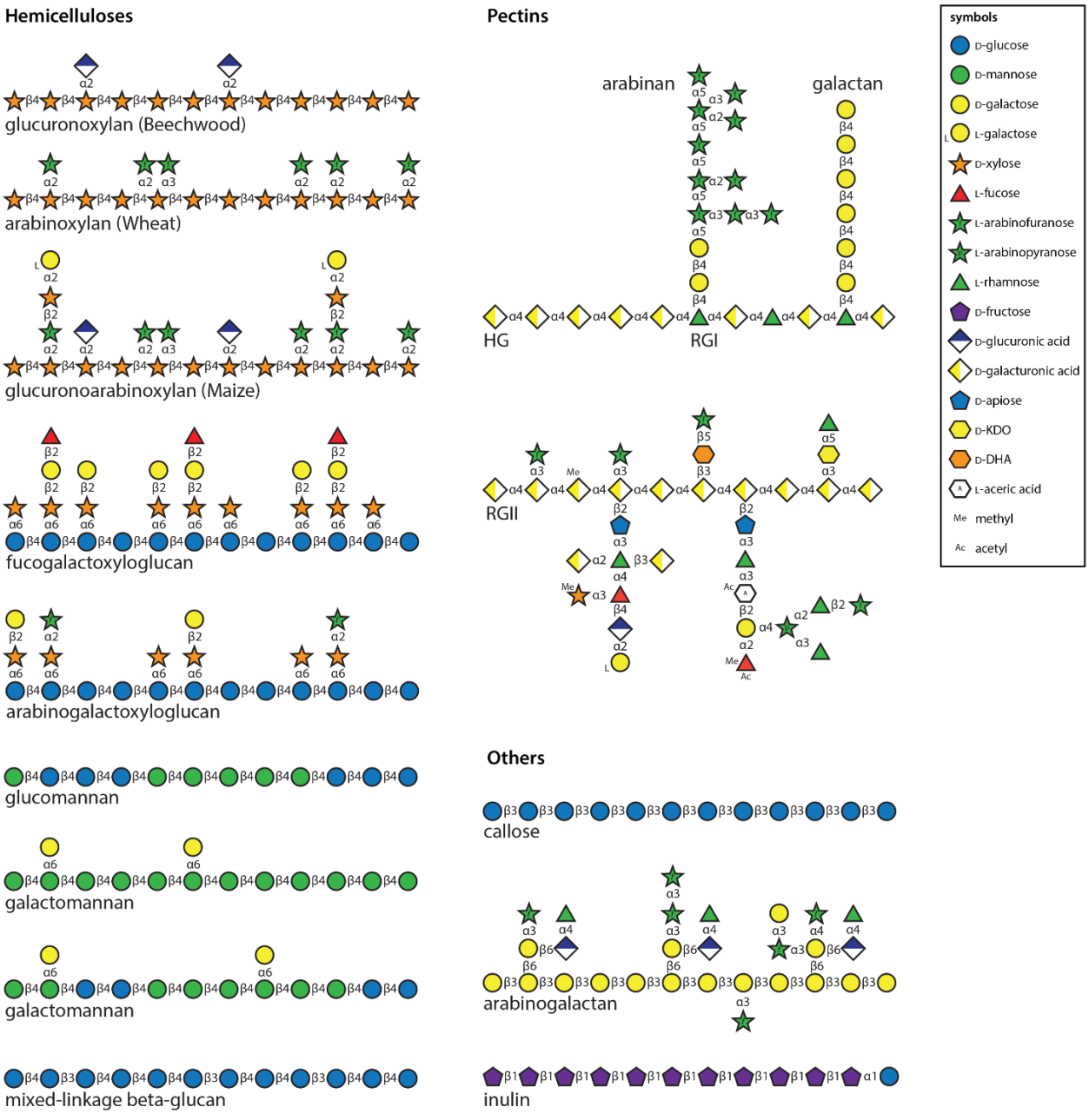

Figure 1-2 Diversity of glycans available to the human gut microbiota.

Monosaccharides are represented by Consortium for Functional Glycomics symbols (26).

β-mannans are comprised of either a homopolymeric β(1,4)-linked glucose backbone (pure mannan and galactomannan) or a heteropolymeric backbone that additionally contains short stretches of β(1,4)-linked glucan at non-regular intervals (glucomannan and galactoglucomannan) (Fig. 1-2) (29). Galactomannan and galactoglucomannan additionally possess single α(1,6)-galactose branches on their respective linear backbones (24). All types of β-mannan can also be decorated with acetyl groups on their backbone mannose residues (29).

Mixed-linkage β-glucans (MLGs) are unbranched homopolymeric glucans with single β(1,3)-linkages separating tracts of either two or three consecutive β(1,4)-linkages (Fig. 1-2, 1-3) (30). This effectively means the structure is comprised of non-regular repeats of cellotriosyl and cellotetraosyl units, the distribution of which varies by source (31). MLGs are commonly found in the cell walls of grasses and cereals (32).

1.2.2.1.2 Pectins

Structurally, pectins are similarly heterogenous and complex, but in contrast to hemicelluloses, are acidic due to the high content of galacturonic acid (33). The most abundant and structurally simple pectin is homogalacturonan (HG), comprised of a linear chain of α(1,4)-linked galacturonic acid residues, which are often methyl esterified at the C6 carboxylate and/or acetylated at free hydroxyls (Fig. 1-2) (34). HG is often, but not always, covalently associated with rhamnogalacturonan I (RGI), which are disaccharide repeats comprised of alternating galacturonic acid and rhamnose linked by α(1,4)- and α(1,2)-linkages (33). RGI contains poly β(1,4)-linked galactan sidechains as well as α(1,5)-linked arabinan sidechains, which are further branched with additional single or poly α(1,2)- or α(1,3)-linked arabinofuranose residues. Rhamnogalacturonan II (RGII) is a highly complex pectic glycan with a backbone resembling HG but with multiple sidechains comprised of an enormous diversity of constituent monosaccharides and glycosidic linkages (Fig. 1-2) (35). Despite its complexity, the structure of RGII is highly conserved. Pectins are abundantly present in the cell walls of diverse fruits including apples, pears, and various citrus fruits, as well as in smaller quantities in soft fruits such as strawberries and grapes (34).

1.2.2.1.3 Other plant polysaccharides

Many other complex glycans can be found in plants that do not neatly fit either of the prior major categories. Callose for instance is an unbranched homopolymeric β-glucan like MLG but

comprised entirely of β(1,3)-linkages (Fig. 1-2). Callose is insoluble and plays an important role as temporary plugs at sites of cell wall damage (36).

A similar function is carried out by arabinogalactan proteins (AGP), members of the hydroxyproline-rich cell wall glycoprotein superfamily (37). The glycan portion of AGP is comprised of a β(1,3)-linked galactan backbone with varying degrees of branching containing α(1,3)-linked arabinofuranose, β(1,6)-linked galactose, β(1,6)-linked glucuronic acid, and α(1,4)-linked rhamnose in a diversity of permutations (Fig. 1-2).

The storage polysaccharide inulin is a type of fructan found in the stems of various dicots, notable sources of which include chicory roots, garlic, and artichoke (38). Structurally, inulin is a β(2,1)-linked fructose polymer with an α(2,1)-linked reducing end glucose cap (Fig. 1-2).

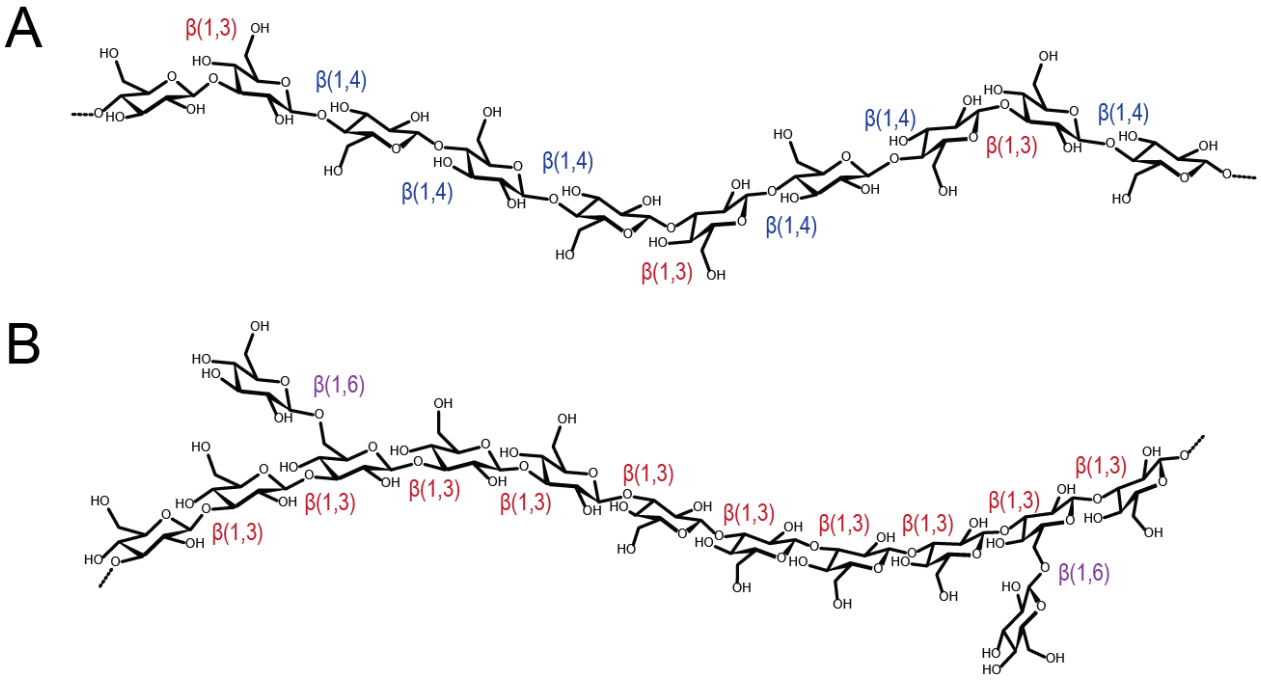

Figure 1-3 Representative chemical structures of β-glucans.

(A) Mixed-linkage β-glucan (MLG). MLGs from diverse sources have the same general structure but differ in the ratio of cellotriosyl to cellotetraosyl units separated by β(1,3)-linkages. (B) β(1,3)-glucan. β(1,3)-glucans from diverse sources vary in the length and frequency of β(1,6)-linked glucose branching. Shown as representative is laminarin from *Laminaria digitata*, which contain single β(1,6)-glucose branching at a frequency of around once per every seven β(1,3)-linked glucose.

1.2.2.2 Other common polysaccharides available to the HGM

Another major source of complex glycans is marine plants (i.e. seaweeds/algae) and marine animals. Algal polysaccharides are important components of the cell walls as well as forms of energy storage, as in terrestrial plants, and similarly display extraordinary diversity. For example: the polysaccharide ulvan is found in green seaweeds; agar, carrageenan, and porphyran are found in red seaweeds; and alginate, fucoidan, and laminarin are found in brown seaweeds (39). Of relevance to this thesis, laminarin is a storage glucan with a β(1,3)-linked glucose backbone, similar to callose, but in contrast contains single β(1,6)-glucose branches (Fig. 1-3) (40).

Bacterial species also represent a rich source of an extreme diversity of complex glycans, serving structural roles within the cell wall in addition to molecular signalling/recognition and adhesion. Bacteria produce an extraordinary diversity of cell surface-bound extracellular polysaccharides (capsular polysaccharides (CPS) and lipopolysaccharides (LPS)); their unique structural composition reflected in the specific antigenicity of strains (41). Bacteria also release loosely bound exopolysaccharide (EPS) that forms an adherent matrix called biofilm (of note, CPS can also form biofilm and their distinction with EPS can be blurry in this context). Prominent examples include dextran, xanthan, levan (a fructan), as well as the β(1,3)-glucan curdlan, an unbranched and insoluble glucan comprised entirely of β(1,3)-linkages, identical to callose (Fig. 1-2).

Diverse fungi (unicellular and multicellular) harbor unique complex glycans as well and represent a major part of the cell wall dry weight. In yeasts, linear β(1,3)-glucans and highly branched α-mannan glycoproteins represent the major cell wall constituents (42). Edible fungi (mushrooms) also contain various glycans such as galactans, xylans, and by far the most prominent, β(1,3)-glucans such as lentinan, scleroglucan, and schizophyllan (43). The yeast- and mushroom-derived β(1,3)-glucans have structures similar to laminarin with different frequencies of single β(1,6)-glucose branches and in some cases longer branches extended by β(1,3)-linked glucose units.

Host-derived glycans are also an abundantly available and diverse class of complex glycans available to the HGM. Mucins are viscous gel-forming glycoproteins constantly released by the epithelial cells of the large intestine forming a continually renewing protective layer (44). Glycosaminoglycans, such as heparan, chondroitin sulfate, and keratan sulfate are proteoglycans that occur in connective tissue and commonly derived from consumption of animal products (45-

48). Human milk oligosaccharides derived from mothers' milk represent a particularly important class of prebiotic glycans that shape the unweaned infant gut (49).

Glycans are extraordinarily diverse macromolecules and the examples above by no means represent an exhaustive list. Indeed, new glycans are constantly being discovered and their structural characteristics and biological/chemical properties actively researched.

1.3 Carbohydrate-active enzymes

In order to contend with the sheer diversity of complex glycans, a correspondingly vast array of specific CAZymes is required to effect complete deconstruction. CAZymes are abundantly present in all domains of life and broadly refer to a large class of enzymes involved in synthesis, modification, and deconstruction of carbohydrate substrates. Known CAZymes are classified into families based on amino acid sequence similarity in the continuously updated and curated Carbohydrate-Active EnZYmes (CAZy) database (50,51). Tertiary fold, key active site residues, and catalytic mechanism within a family are conserved (with very few exceptions) (52). The CAZy classification thus provides valuable ability to predict some of the functional aspects of uncharacterized enzymes based solely on primary sequence. Substrate specificity however can be diverse within a single family, and the delineation of substrate-specific subfamilies can be helpful, especially in some of the larger families (53-57). Such predictive power has been transformative in bioinformatic approaches such as large-scale genome annotation efforts. However, functional prediction from sequence alone will always have limitation and is certainly not exception-proof. Thus, for the enzymologist, the most valuable aspect of the CAZy classification is in providing a starting point upon which to base testable hypotheses and thereby directing functional characterization efforts, rather than completely eliminating the need for functional characterization. In addition, due to the paucity of biochemically and structurally characterized members relative to the vast bulk of (meta)genomic sequence data now available, there is considerable potential for the discovery of as-yet unclassified families (50,51).

The CAZy classification system currently defines five enzyme classes: glycoside hydrolases, glycosyltransferases, polysaccharide lyases, carbohydrate esterases, and auxiliary activities. In addition, one class of non-catalytic modules associated with CAZyme domain(s) – carbohydrate-binding modules – is also classified into families within which overall fold and key binding site residues are similarly conserved.

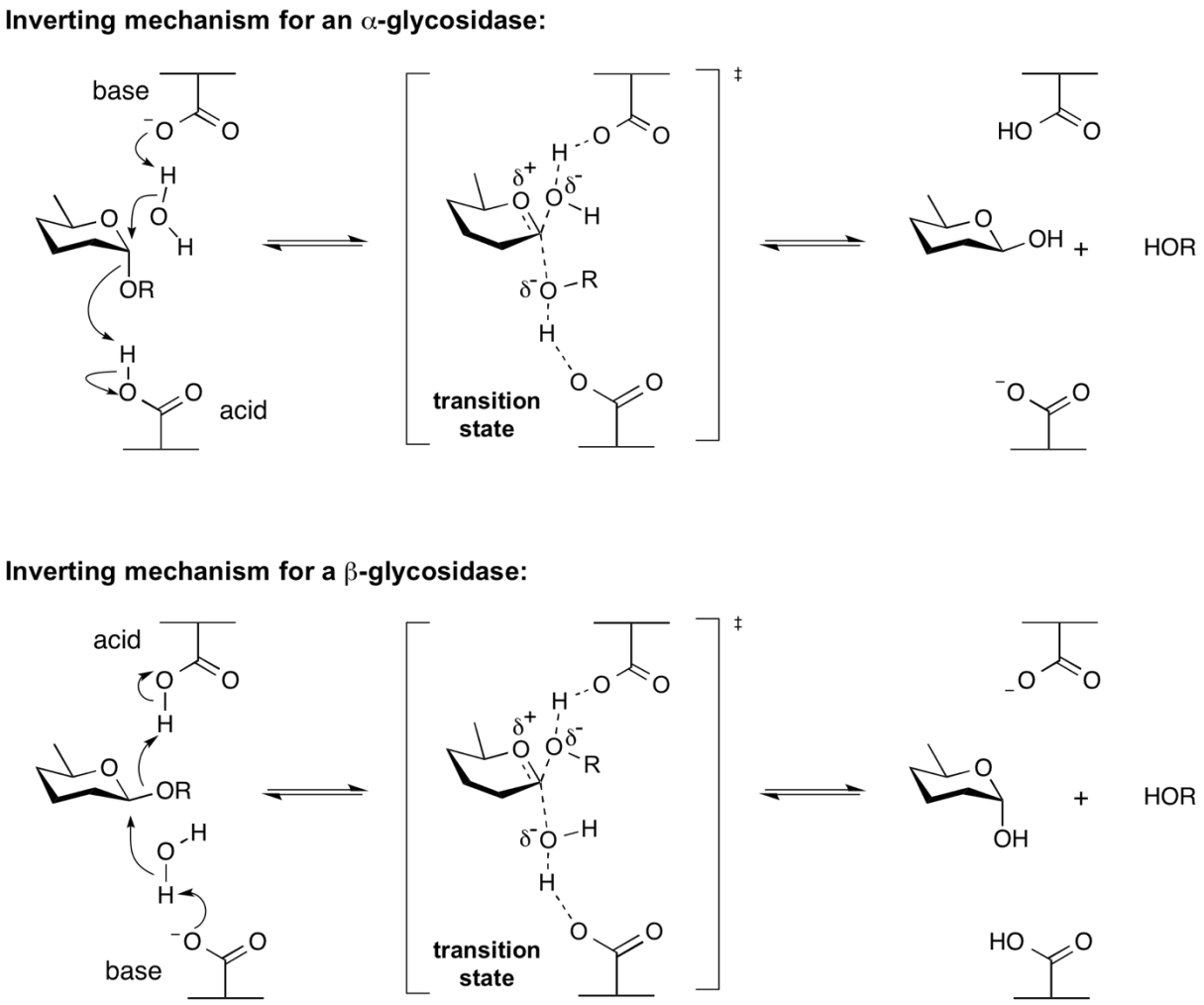

Figure 1-4 Inverting mechanism of glycoside hydrolases.

Reproduced from "Glycoside Hydrolases" in *Cazypedia* (58).

1.3.1 Glycoside hydrolases

Glycoside hydrolases (GHs) catalyze the hydrolysis and/or rearrangement of glycosidic bonds between monosaccharides as well as those that occur in glycoconjugates. GHs are the main CAZyme class of interest in the context of complex glycan deconstruction by the HGM.

1.3.1.1 Catalytic mechanism

Glycosidic bonds are hydrolyzed with either net inversion or retention of anomeric stereochemistry of the resulting product reducing end monosaccharide (59). The most common mechanisms found in each category utilize a pair of carboxylate-containing side chains (glutamate or aspartate) to carry out the reaction.

In the inverting mechanism, the two carboxylates are positioned 6 – 11 Å apart, in between which is accommodated the scissile glycosidic linkage and a water molecule. The enzyme has an active site structure in which one of these residues is deprotonated to assume the role of catalytic base and the other is protonated to play the role of catalytic acid (Fig. 1-4). The one-step reaction involves the simultaneous deprotonation of the water molecule by the catalytic base and the protonation of the leaving group hydroxyl by the acid (60,61). Catalysis proceeding through an oxocarbenium ion-like transition state where for hexoses, the sugar puckering is distorted from the relaxed $^4C_1$ chair conformation into a higher energy half-boat or skew-boat in which the endocyclic oxygen lies in a plane with C1, C2, and C5 (62,63). Enzyme active sites are generally structured to stabilize these high-energy intermediates.

In the retaining mechanism, the two carboxylates are located ~5.5 Å apart from each other, closer than in inverting GHs because the acceptor water molecule is not simultaneously accommodated as the carbohydrate substrate (60,61). Instead of activating a water molecule as in the inverting mechanism, one of the carboxylates is positioned adjacent to the anomeric carbon for direct nucleophilic attack (Fig. 1-5). Leaving group departure is facilitated by donation of a proton by the catalytic acid/base carboxylate, playing the role of acid in this step. This reaction results in a stable, covalent glycosyl-enzyme intermediate, which is subsequently deglycosylated with the acid/base this time deprotonating a water molecule as it attacks, restoring original anomeric configuration (60,61). Each of the two steps in this double displacement mechanism proceeds through an oxocarbenium ion-like transition state. In the second step, it is possible for the acceptor molecule to be another saccharide instead of water, which results in transglycosylation rather than hydrolysis (64). Many retaining GHs will display varying degrees of transglycosylation activities at high enough substrate concentrations, and some have optimized active sites to exclude water to preferentially catalyze transglycosylation (65). These GHs are important in the context of plant cell wall remodelling and is largely not applicable to complex glycan degradation by the HGM.

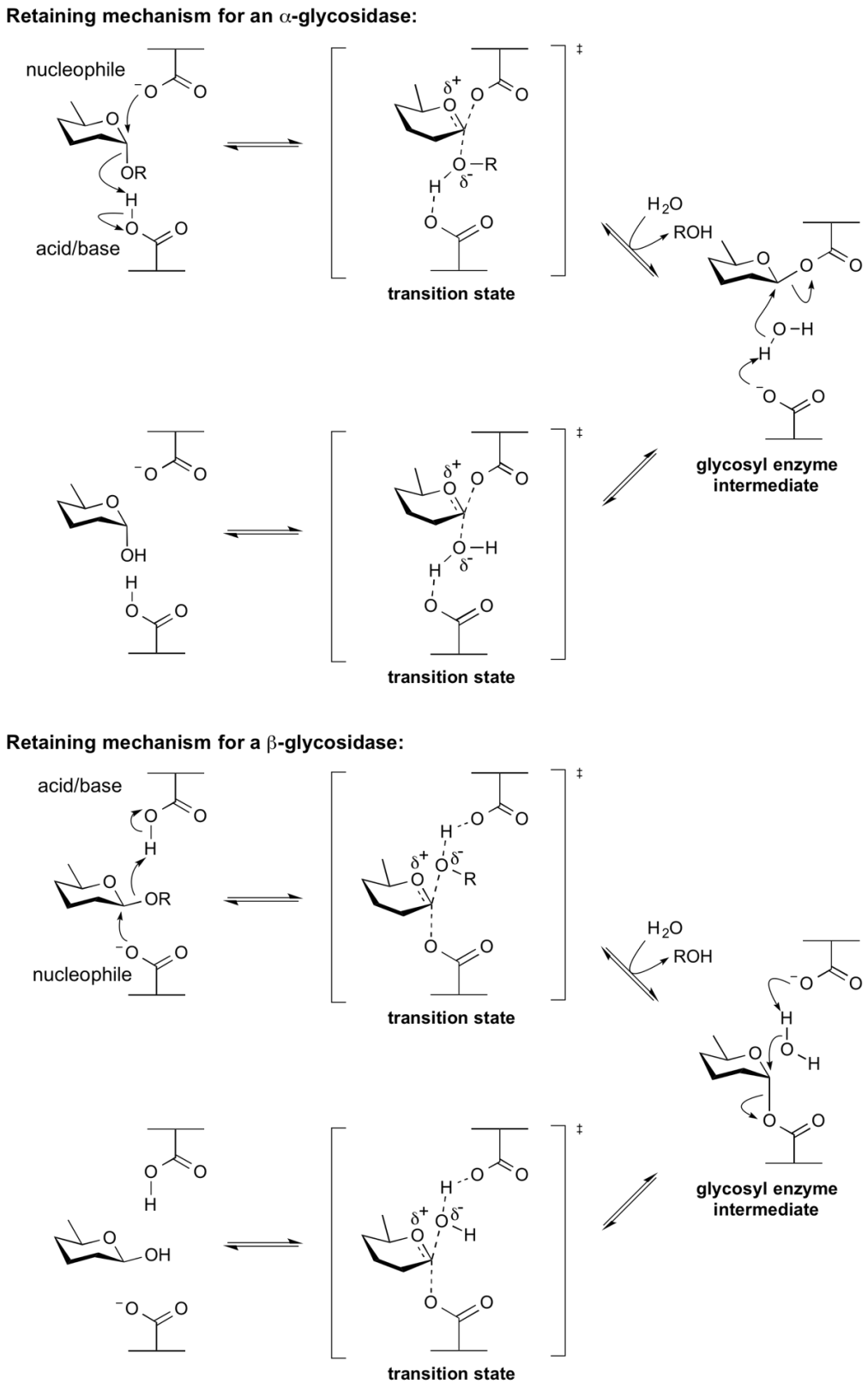

Figure 1-5 Retaining mechanism of glycoside hydrolases.

Reproduced from "Glycoside Hydrolases" in *Cazypedia* (58).

1.3.1.2 Mode of action

Regardless of their mechanism of hydrolysis, GHs are also distinguished based their specificity for glycosidic bonds at the substrate chain ends (exo-acting) or for those found internally (endo-acting) (60). Endo-GHs can further be classified into endo-dissociative enzymes, which stochastically hydrolyze polysaccharide releasing both new products each reaction cycle, and endo-processive, which move along the polysaccharide after each hydrolysis event to sequentially release short oligosaccharides before disengaging. The pattern of product accumulation over successive reaction cycles is thus more similar between the exo-GHs and endo-processive GHs, which consistently release products of same size, versus endo-processive GHs, which release a range of product sizes until limit digest is reached (66).

Exo-GHs typically have pocket-shaped active sites optimized to recognize saccharide termini (often the non-reducing end; rarely the reducing end) to release monosaccharides or short oligosaccharides (60,63). Endo-GHs in contrast have active sites resembling a cleft open to either side, optimized to accommodate a non-terminal segment of polysaccharide.

1.3.2 Glycosyltransferases

Glycosyltranferases (GTs) are responsible for catalyzing the formation of glycosidic bonds (64). The energy required for synthesis is stored in sugar-1-phosphate derivatives (including nucleotide phosphates and lipid phosphates) formed in an upstream energy-utilizing process. GTs catalyze the transfer of these "activated" sugar phosphates as glycosyl donors to a nucleophilic acceptor (67). This reaction can occur with retention or inversion of anomeric stereochemistry of the activated donor and thus GTs are classified as retaining or inverting enzymes, as with GHs.

1.3.3 Polysaccharide lyases

Polysaccharide lyases (PLs) cleave glycosidic bonds similar to GHs, but do so using a non-hydrolytic, β-elimination mechanism. PLs act specifically on hexuronic acids (hexose sugars oxidized at the C-5 position) with a glycosidic bond at the C4 position (68,69). Taking advantage of the C-5 carboxylic acid, the resulting relatively acidic proton at the C-5 position is abstracted by a catalytic base residue, leading to an enolate intermediate. Subsequent electron transfer from the carboxylate to form a double bond between C-4 and C-5 results in elimination of the glycosidic bond (68,69). In contrast to GHs, the newly created nonreducing end is unsaturated with a double bond between C-4 and C-5, while the newly created reducing end remains saturated. PLs feature

prominently in the deconstruction of various pectins and GAGs, as well as alginate and xanthan due to their high hexuronic acid content.

1.3.4 Carbohydrate esterases

Carbohydrate esterases (CEs) catalyze hydrolysis of acyl ester modifications specifically on saccharides. Ester linkages can occur between a hydroxyl group on a monosaccharide and an acyl group, in which case the carbohydrate plays the role of alcohol (70). Many common dietary complex glycans including xylans, xyloglucans, pectins, and β-mannans are often acetylated at various positions in a non-regular fashion. Ester linkages can also occur with the carbohydrate acting as the acid, specifically in uronic acids. Pectin methyl esters and lignin-xylan ester linkages are common targets of CEs in this category. CEs also include carbohydrate amidases, common substrates of which include N-acetylated glycosamines such as components in chitin (71). A number of CE reaction mechanisms have been described with the most common the involving a classical serine-histidine-aspartate catalytic triad, analogous to the serine proteases (72). Briefly, histidine abstracts the proton from the serine nucleophile causing attack on the electrophilic carbonyl carbon of the acyl group, resulting in its transfer to serine via an unstable tetrahedral intermediate. Hydrolysis of the acyl-enzyme intermediate to regenerate free serine is also facilitated by the histidine, whose acid/base properties and orientation are modulated by the aspartate. In less common cases, CEs use a catalytic dyad in which the aspartate "helper" is absent or use a fundamentally different metal ion-dependent mechanism (73).

1.3.5 Auxiliary activities

Members of auxiliary activity (AA) classes are redox enzymes whose activities facilitate other CAZymes to gain access to carbohydrates that comprise the complex plant cell wall (74). They can be broadly divided into the ligninolytic enzymes and the lytic polysaccharide monooxygenases (LPMOs). The substrates of lignolytic redox enzymes are diverse, directly targeting lignols and phenolic components of lignin, as well as free hydroxyl groups on saccharides and alcohols within the plant cell wall. These reactions release reactive molecules like hydrogen peroxide and contribute to non-specific oxidative lignolysis (74). LPMOs - which oxidatively cleave glycosidic linkages - in particular have been the subject of significant recent interest for its ability to boost efficiency of cellulase cocktails (75).

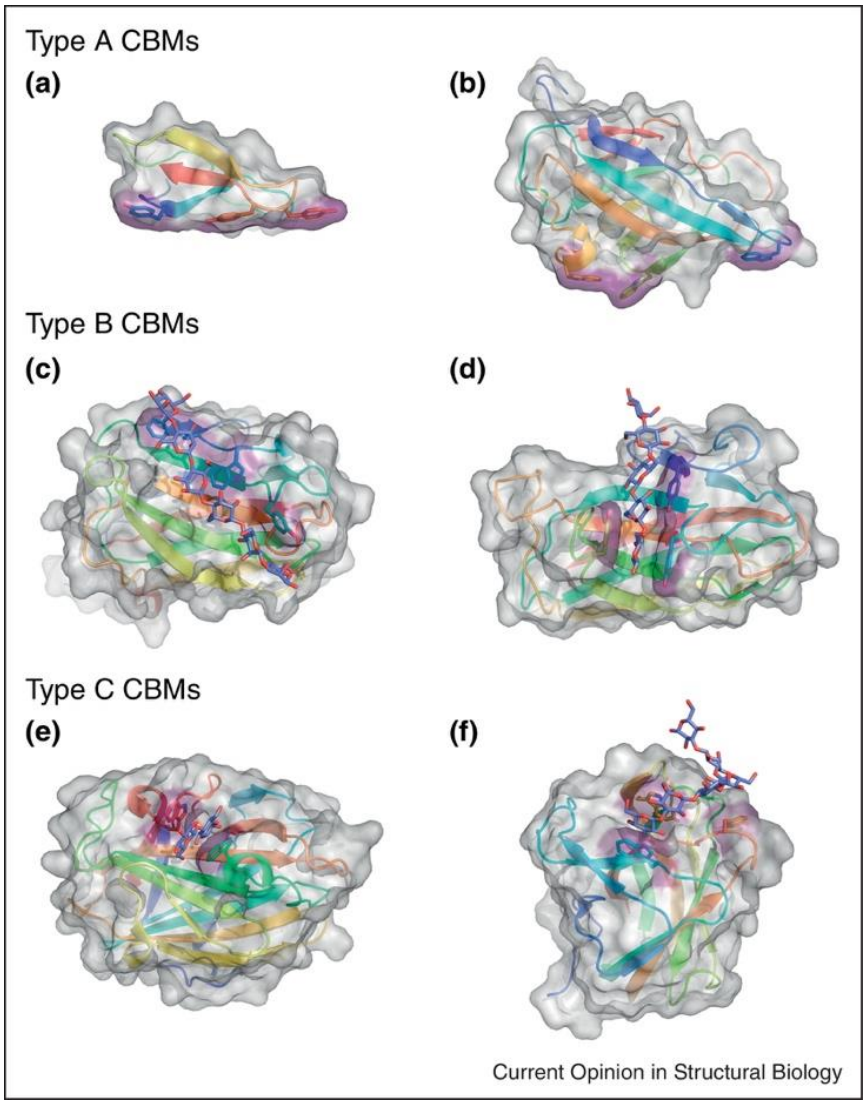

Figure 1-6 Ligand binding mode by Types A, B and C carbohydrate binding modules.

(A-B) Examples of Type A CBM from *Hypocrea jecorina* (CBM family 1; PDB ID: 1CBH) and *Cellulomonas fimi* (CBM family 2; PDB ID: 9EXG), respectively. (C-D) Examples of Type B CBM from *Piromyces equii* (CBM family 29; PDB ID: 1GWL) and *Cellulomonas fimi* (CBM family 4; PDB ID: 1GU3), respectively. Examples of Type C CBM from *Thermotoga maritima* (CBM family 9; PDB ID: 1I82) and *Bacillus halodurans* (CBM family 6; PDB ID: 1W9W), respectively. Taken from Gilbert et al. (76).

1.3.6 Carbohydrate-binding modules.

Carbohydrate-binding modules (CBMs) represent non-catalytic accessory modules with discrete folds and capacity to bind carbohydrates, which are found as part of larger sequences that contain catalytic CAZyme modules. CBMs play an important role in enhancing catalytic activity of modular CAZymes, and only in very rare cases are CBMs found as independent proteins. The

modularity is a defining feature of CBMs and distinguishes them from other carbohydrate-binding proteins such as lectins, antibodies, and surface glycan-binding proteins (*vide infra*). Indeed, multiple CBM domains are often associated within parent CAZyme sequences. CBMs target and maintain proximity of the CAZyme to the substrate, thereby increasing local concentration and potentiating polysaccharide deconstruction.

CBMs are classified into numerical families based on sequence similarity, similar to the enzymes. Of the 87 current families of CBMs, the vast majority share a conserved compact β-sandwich fold, the abundance of which is vested in the plasticity of this structural scaffold. Binding sites can be located on one of the β-sheets (CFS: concave face site) or formed from loops joining β-strands at one end of the β-sandwich (VLS: variable loop site) (77). Furthermore, three distinct modes of glycan binding are grouped into types. Type A CBMs recognize surfaces of crystalline polysaccharides such as cellulose (flat topology), Type B CBMs bind internal regions of single glycan chains (cleft topology; "endo-type"), and Type C CBMs bind glycan chain termini (pocket topology; "exo-type") (Fig. 1-6) (76).

1.4 Glycan utilization systems

A taxonomically diverse consortium of microbes has evolved various strategies of employing the abundant CAZymes to compete for nutrients in the highly dynamic environment of the human gut.

1.4.1 Polysaccharide utilization loci

The genomes of Bacteroidetes are particularly enriched in CAZymes (15) with most clustered into multi-gene operons that are co-expressed to form a membrane-associated machinery that work in concert to bind, break down, and sequester target glycan (Fig. 1-7A). These polysaccharide utilization loci (PULs) are a unique feature of this phylum of Gram-negative bacteria which are among the most dominant in the HGM. PULs are comprised of specific N-terminally lipidated surface glycan-binding proteins (SGBPs), outer membrane-spanning TonB-dependent transporters (TBDTs), and carbohydrate-sensor/transcriptional regulators, in addition to the cohort of CAZymes (most frequently GHs, but also PLs and CEs where substrate-appropriate). The complexity of PULs often scales with that of their cognate substrates (78,79), and may include ancillary enzymes such as proteases, sulfatases, and phosphatases. These elegant systems constitute the major nutrient acquisition strategy deployed by Bacteroidetes and thus are intrinsically linked to their successful colonization of the HGM.

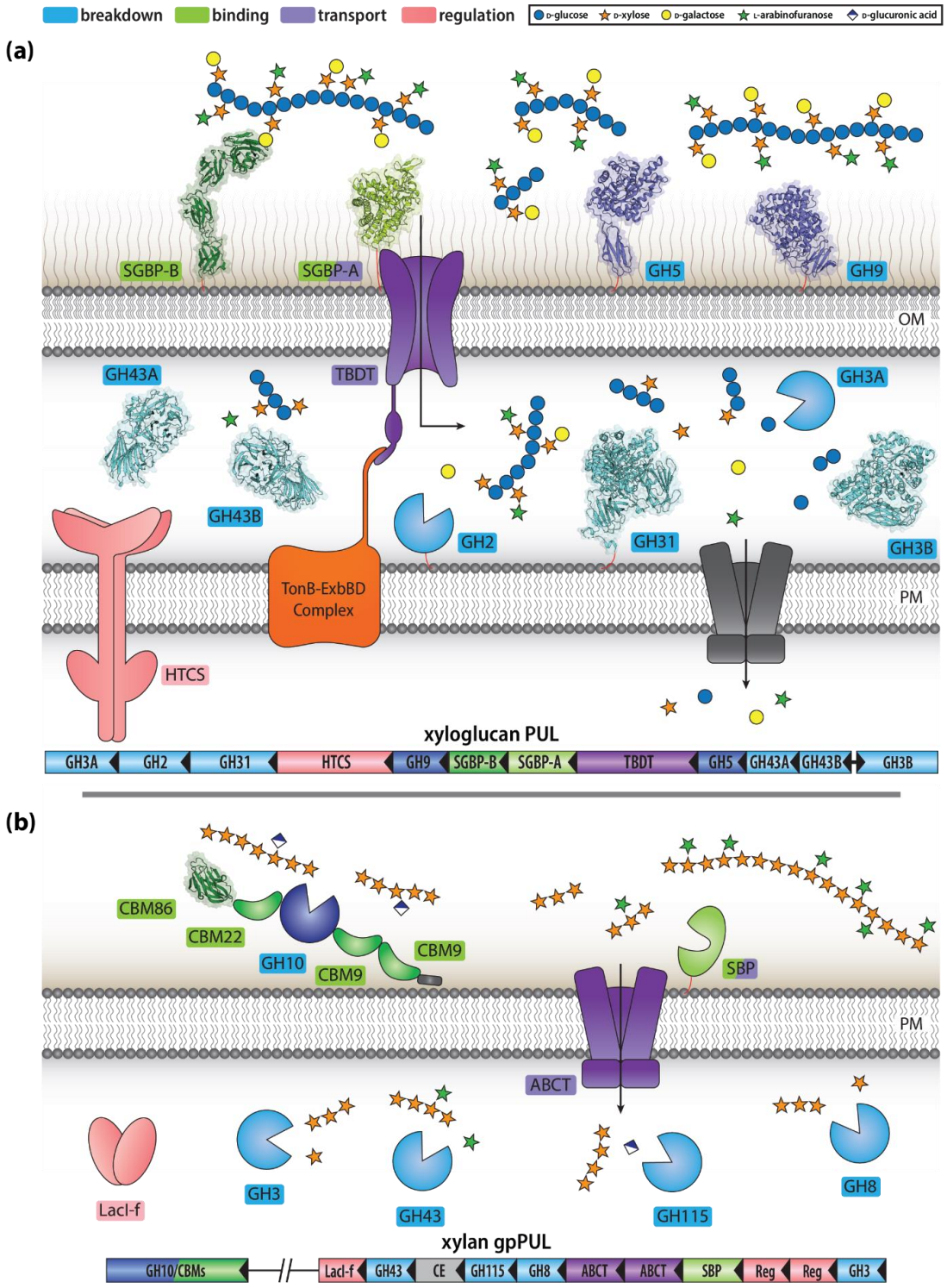

Figure 1-7 Models of representative glycan processing machineries of prominent human gut microbes.

(A) Xyloglucan utilization locus from *Bacteroides ovatus*. (B) Xylan utilization locus from *Roseburia intestinalis*. Loci gene contents are shown below each model. Endo-acting GHs are colored blue and exo-acting GHs are colored cyan. GH: glycoside hydrolase, SGBP: surface glycan-binding protein, TBDT: TonB-dependent transporter, HTCS: hybrid two-component sensor, CBM: carbohydrate-binding module, SBP: solute-binding protein, ABCT: ATP-

binding cassette transporter, LacI-f: LacI family regulator, CE: carbohydrate esterase, OM: outer membrane, PM: plasma membrane. PDB accession codes – SGBP-A: 5E76, SGBP-B: 5E7G, GH5: 3ZMR, GH9: 6DHT, GH31: 5JOV, GH43A: 5JOX, GH43B: 5JOZ, GH3B: 5JP0, CBM86: 6SGF.

The starch utilization system (Sus) is the archetypal PUL first identified in *Bacteroides thetaiotaomicron* through pioneering work by Abigail Salyers and co-workers starting in the late 1980s (80-90). Subsequent studies, which are ongoing to this day by various groups, have illuminated a precisely coordinated system that serves as a general cellular model for the study of other PULs.

1.4.1.1 TonB-dependent transporters and surface glycan-biding proteins A

A defining feature of canonical PULs is the presence of at least one tandem pair of *susC* and *susD* homologs (91,92), which encode a TBDT and a partner SGBP (referred to as SGBP-A to distinguish from other SGBPs; *vide infra*), respectively (Fig. 1-8). PUL-associated TBDTs have a large, structurally conserved 22-stranded β-barrel architecture with an N-terminal plug domain as well as an N-terminal extension (NTE) (93). The functionally essential nature of the NTE was recently demonstrated in fructan-transporting TBDTs (94). In between the plug domain and the NTE domain is the TonB box, which plays an essential role in interacting with the C-terminal TonB, which is a part of the TonB-ExbB-ExbD complex. The proton gradient across the inner membrane is harnessed by this complex and mechanical force is transduced to the outer membrane via this TonB protein to energize the active transport process (93). SGBPs-A also have a highly conserved structure comprised of a single rigid-body architecture with four tetratricopeptide repeat (TPR) motifs forming a structural scaffold. A more variable mix of mostly loops and α-helices, and some β-strands are cradled in the concave groove of the TPR superhelix, and houses the highly tunable ligand binding site. A structure comprised almost entirely of α-helices is highly unusual relative to other carbohydrate-binding proteins in diverse contexts (76,95).

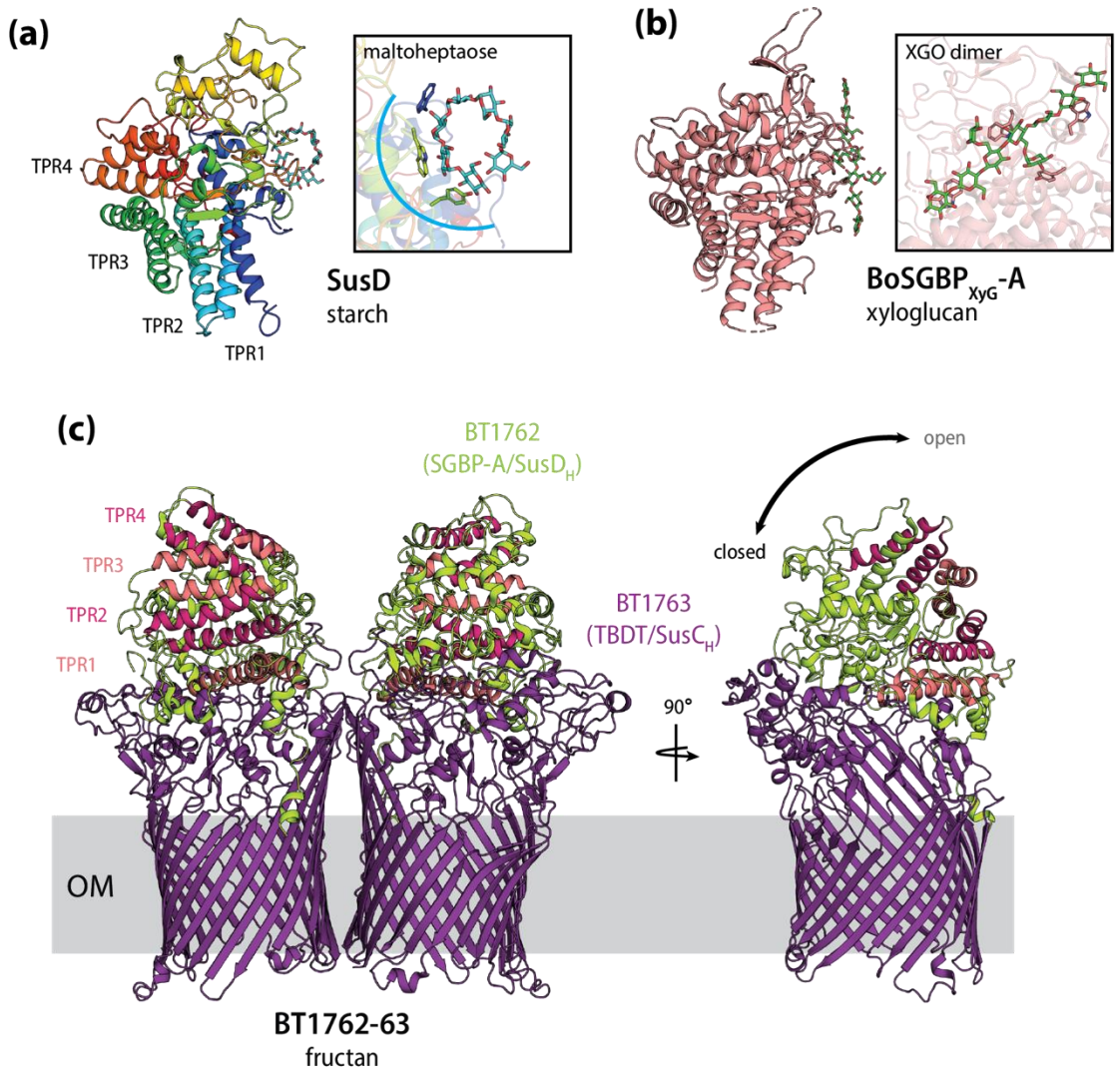

Figure 1-8 Liganded and TBDT-complexed SGBPs-A.

(A-B) Ligand-complexed SGBP-A structures solved to date with cognate oligosaccharides bound. SusD structures have also been solved in complex with cyclodextrins but only the maltoheptaose complex is shown since their overall shapes are similar. XGO dimer: glucooctaose backbone-based xyloglucooligosaccaharide, MLG: mixed-linkage β-glucan, $SusC_H$: SusC-homolog $SusD_H$: SusD-homolog. PDB accession codes – BT1762-63: 5T4Y, SusD: 3CK9, $BoSGBP_{XyG}$-A: 5E76. (C) Structure of the glycan transporting SGBP-A/TBDT complex from a fructan PUL. Only one of the dimeric complexes is shown in the orthogonal view on the right.

Evidence of tight association of these two PUL components was available from the original investigations into the Sus as well as subsequent biophysical studies (80,82,83,86). Because TPR motifs are often associated with protein-protein interactions, it was hypothesized that this part of SGBPs-A was involved in binding to the TBDT (95,96). The question was laid to rest with

determination of a seminal TBDT structure from a fructan PUL in *B. thetaiotaomicron* (BT1763), which revealed the SGBP-A (BT1762) forms a lid that sits atop the extracellular opening of the β-barrel, crucially with its TPR motifs actually facing outward (Fig. 1-8C) (97). TPR1 makes contacts with adjacent extracellular loops of the TDBT (TPRs 2-4 are fully exposed) and molecular dynamics simulations showed that the SGBP-A lid can swing open about this "hinge" contact point (97). Dynamic motion of the lid was recently experimentally confirmed by single particle cryo-EM reconstructions, which captured lid-open and -closed conformations (94,98). The ligand binding site opposite the TPR faces the interior of the TDBT β-barrel and may play a role in specificity determination and facilitating transport. Genetic studies have shown that SGBPs-A are essential to proper PUL function; indeed, in-frame gene deletions often result in defective growth of the mutant on target glycan in minimal medium (99,100) (with exceptions (101)). Interestingly, in select cases, growth can be rescued in these deletion mutants by complementation with SGBP-A variants in which substrate binding has been eliminated by site-directed mutagenesis (99,102), pointing to their physical presence being of primary importance with glycan binding being secondary. Whether the highly conserved TPR motif in SGBPs-A (96) simply serves as a structural scaffold or plays a functionally significant role in interacting with other catalytic or non-catalytic cell-surface proteins is still unclear. Evidence of interaction of PUL surface-anchored proteins and SGBPs-A exist (89,103) but structural insights remain elusive as these interactions may indeed be too transient and/or require certain environmental conditions to elicit (104).

1.4.1.2 Auxiliary surface glycan binding proteins

PULs usually encode one or more auxiliary SGBPs adjacent to the *TBDT/SGBP-A* genes (*susC/susD* homolog), often immediately downstream (105). However, unlike the SusD-like SGBPs-A, these additional SGBPs have such low sequence similarity and tertiary structural homology that they cannot be confidently identified by bioinformatic approaches. Despite this lack of sequence conservation, crystal structures obtained to-date reveal that they share extended, multimodular architectures with proline residues demarcating discrete β-sandwich domains (Fig. 1-9) (106). These beads-on-a-string structures present ligand binding sites on one or more distal C-terminal domains. SusE and SusF from the archetypal Sus possess two and three ligand binding domains, respectively (107). In contrast, BoSGBP$_{XyG}$-B from *B. ovatus* XyGUL and BT4661 from a *B. thetaiotaomicron* PUL targeting heparin/heparan sulfate present a single binding site on their C-terminal domains. BoSGBP$_{XyG}$-B is preceded by three β-sandwich domains, while BT4661

possesses a binding site spanning two of its C-terminal domains preceded by four β-sandwich domains (99,103). These auxiliary SGBPs assist in the glycan capture process and in increasing local substrate concentration. Thought another way, auxiliary SGBPs aid the bacterium in seeking out and staying attached to the target glycan. Though generally not as indispensable to growth as the SGBPs-A in monocultures, widespread evolutionary maintenance of auxiliary SGBPs in PULs suggests they do impart fitness advantage (99,100). Indeed, a mutant lacking SusE was found to be at a competitive disadvantage when grown in co-cultures alongside wildtype *B. thetaiotaomicron* (108).

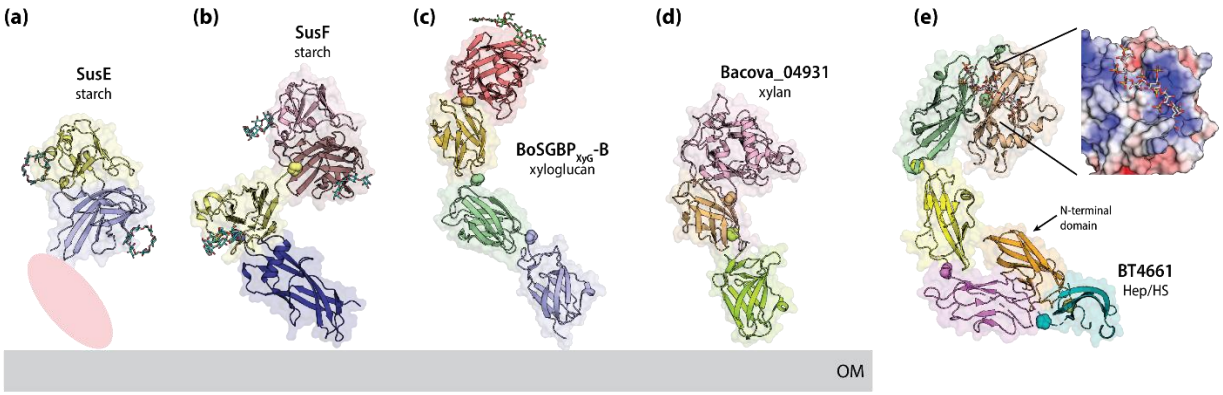

Figure 1-9 Auxiliary SGBP structures.

(A-D) All structures solved to date with individual domains shown in different colors and interdomain proline residues shown as spheres. The N-terminal domain of SusE was not resolved in the crystal structure. Hep/HS: heparin/heparan sulfate. PDB accession codes – SusE: 4FEM, SusF: 4FE9, BoSGBP$_{XyG}$-B: 5E7G, BT4661: 4AK1, 4AK2, Bacova_04931: 3ORJ.

1.4.1.3 Surface and periplasmic CAZymes

PULs compartmentalize the CAZymes specific to the dismantling of its cognate polysaccharide in a very deliberate manner. One or more extracellular endo-glycanases, anchored to the cell-surface via N-terminal lipidation, initiate glycan breakdown. The resulting fragments are actively shuttled into the periplasmic space by the cognate TBDT where, additional linkage-specific, exo-acting GHs complete the saccharification process (79,109). Component monosaccharides can then be moved to the cytosol to fuel fermentation pathways. Exceptions exist where exo-GHs are found anchored to the cell surface and endo-GHs are localized to the periplasm,

but the general strategy results in sequestration of breakdown intermediates to the periplasm, away from competition in the HGM environment (110,111).

1.4.1.4 Sensor/regulator

The enormous number of PULs encoded by Bacteroidetes (109) to match the diversity of glycans encountered by the HGM means that rapid and specific regulation of appropriate PULs is paramount. Hybrid two-component sensors (HTCSs) and extracytoplasmic function sigma (ECF-σ) factor/anti-σ factors are the predominant regulatory systems found in PULs. HTCS regulators are commonly associated with PULs targeting a variety of plant cell wall carbohydrates and ECF-σ/anti-σ factor pairs with those targeting host-derived glycans.

HTCSs contain all the components found in classical two-component systems (TCS) in a single polypeptide (Fig. 1-10) (112). TCSs, as the name implies, are comprised of two parts: a membrane spanning sensor histidine kinase (HK) and a cytoplasmic response regulator (RR) (113). HKs are homodimeric integral membrane proteins with a N-terminal sensor domain which bind a specific inducing ligand (114). This binding action causes conformational changes that lead to the C-terminal catalytic and ATP-binding (CA) domain to autophosphorylate a conserved histidine residue on the dimerization and histidine phosphorylation (DHp) domain. This phosphoryl group is then transferred to a conserved aspartate residue on the receiver domain of the RR. As with the HK, phosphorylation results in conformational changes, which are transmitted to the effector domain (also commonly referred to as the output domain) to regulate its activities. A diversity of outputs exists in TCSs, ranging from ligand binding to enzymatic activity, but the great majority exhibit DNA binding, which is the case for HTCSs (115). Thus, HTCSs are inner-membrane spanning transcriptional regulators that recruit target DNA sequences proximal to the membrane.

In transcriptional regulation by ECF-σ/anti-σ factors, the primary σ subunit of RNA polymerase is exchanged with an alternative ECF-σ factor redirecting the RNA polymerase to transcription of the cognate PUL (116). Before this can happen, the ECF-σ factor must be released from the single-pass transmembrane anti-σ factor, which occurs upon sensing of the appropriate signaling molecule. In the case of PULs, regulation by ECF-σ/anti-σ factors is believed to occur through “trans-envelope signaling”, whereby a specialized TBDT relays signal recognition via protein-protein interaction to the periplasmic portion of the anti-σ factor (117).

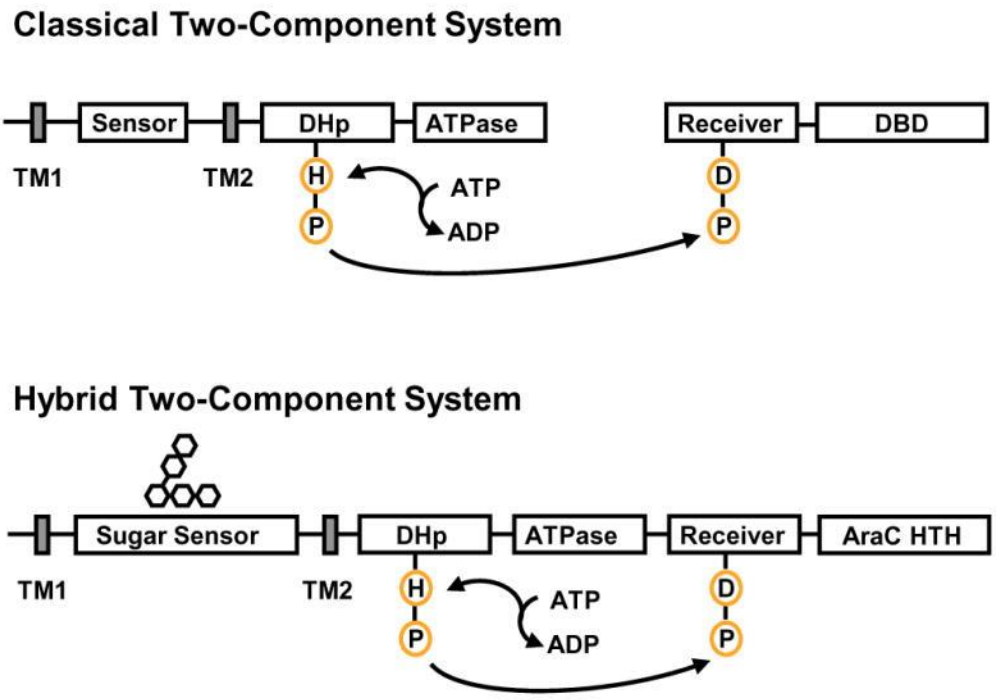

Figure 1-10 Schematic of classical and hybrid two-component systems.

Reproduced from Raghavan et al. (115).

1.4.2 Other glycan utilization systems

In contrast to Bacteroidetes, prominent Gram-positive members of the HGM such as Firmicutes and Actinobacteria are in general more specialist utilizers of certain glycans. They do however harbor similar systems called Gram-positive PULs (gpPULs) (118), though the distinct cell membrane architecture necessarily requires a slight variation on the PUL paradigm (Fig 1-7B). In place of standalone non-catalytic SGBPs, the glycan-binding function is vested in carbohydrate-binding modules (CBMs) expressed in-train with cell-surface CAZymes on a single polypeptide, which are often not colocalized with the remainder of the gpPUL genes. Specific transporters, most often an ATP-binding cassette (ABC) transporter, and cognate solute-binding proteins (SBPs) are analogous to the TBDT and SGBP-A of PULs. Other common transporters associated with gpPULs include major facilitator superfamily transporters, cation symporter family transporters, and phosphoenolpyruvate-phosphotransferase system transporters (119). Finally, the regulatory proteins in gpPULs are generally soluble, instead of membrane-spanning as in PULs, with mechanisms ranging from LacI-type to AraC-like.

A notable member of the Firmicutes, *Ruminococcus bromii*, are particularly adept at degrading resistant starch due to their secreted amylosome strategy (120), analogous to the cellulosome paradigm (121,122). Briefly, the GH13 amylases that are abundant in their genome are expressed in-train with multiple starch-binding CBMs, and crucially with dockerins. Scaffold proteins, which may or may not be cell surface anchored, contain cohesin domains with which dockerins are involved in high affinity interaction. Multiple dockerin-containing amylases and scaffold proteins assemble into large multi-subunit RS-degrading complexes (120). Referred to as primary degraders, *R. bromii* enable secondary degraders such as *B. thetaiotaomicron* (via the Sus) to scavenge partially degraded starch granules and solubilized fragments (21).

1.5 Aim of investigation

The advent of the whole genome sequencing era combined with advances in bioinformatics have provided us with a wealth of genetic information. As introduced above, *susC/susD* pairs are hallmarks of PULs, and have been used to enumerate PUL complements among the genomes of key human gut symbionts, including *B. thetaiotaomicron* (88 PULs), *B. ovatus* (126 PULs), and *B. cellulosilyticus* WH2 (113 PULs) (109,123). The abundance and diversity of PULs has been well-documented as a result of these initiatives, which have enabled the comparative analysis of PULs from various gut Bacteroidetes and provided an essential foundation to understand nutrient niche colonization and community dynamics. For example, despite the enormous number of PULs present in the genomes of *B. thetaiotaomicron* and *B. ovatus*, strikingly few homologous PULs are shared between them, suggesting that these two symbionts have distinct glycan niches (109).

Large-scale (meta)genomics approaches have clearly been instrumental for PUL discovery, as well as predicting the metabolic potential of diverse Bacteroidetes. However, refined functional characterization at the molecular and cellular levels remains critical for a full understanding of the roles of PULs in microbial communities.

To contribute to this end, the overarching aim of my thesis is to provide functional insights into the utilization of β-glucans by prominent members of Bacteroidetes. Two broad subclasses of β-glucans – mixed-linkage β-glucans and β(1,3)-glucans – are the focus of this investigation. I hypothesized that employing a multidisciplinary approach synergistically combining enzymology, biophysics, structural biology, and microbial genetics, in harness with meta(genomics) will shed unprecedented light on details of PUL-mediated assimilation of these common dietary glycans in the HGM.

For each β-glucan subclass, the focus of PUL functional characterization is divided into two parts: 1) enzymatic analysis of the GHs, and 2) biophysical analysis of the SGBPs. Thorough characterization of the GHs includes Michaelis-Menten kinetics on a panel of polysaccharides, defined oligosaccharides and chromogenic substrates, analysis of reaction products by high performance liquid chromatography, and localization studies by fluorescence microscopy and western blotting. SGBPs are subject to affinity gel electrophoresis on a library of polysaccharides and isothermal titration calorimetry on defined oligosaccharides. Subsequent structural analysis of GHs and SGBPs by X-ray crystallography provide molecular insights into kinetic and binding properties. These biochemical and molecular biological initiatives are then in turn combined with reverse genetics and microbial growth studies to provide functional insights. Furthermore, comparative genetics and metagenomic analyses are employed to broaden the scope of the detailed biochemistry and microbiology, and overall provide holistic insight into the utilization of these ubiquitous glycan and its implications.

Chapter 2: Molecular mechanism by which prominent human-gut Bacteroidetes utilize mixed-linkage beta-glucans, major health-promoting cereal polysaccharides

2.1 Introduction

The composition and homeostasis of the human gut microbiota has a profound and intimate connection to various aspects of our physiology, health and wellbeing (124). Indeed, a multitude of diseases such as type 2 diabetes, inflammatory bowel diseases (IBDs), and cancer have been linked to alterations in the population and proportion of microbes in this highly complex and dynamic ecosystem that exists in our large intestine (7,125-127). The molecular mechanisms by which the microbiota exerts influence on human health are largely unresolved and undoubtedly complex, yet may hold the key to personalized medicine through therapeutics that target the gut microbial ecosystem (128-131).

A major factor in shaping the composition and physiology of the gut microbiota is the influx of complex glycans – popularly known as "dietary fibre" – that evade degradation by the limited set of human genome-encoded glycoside hydrolases (13,15,78). Indeed, regular ingestion of plant polysaccharides is integral to maintaining a healthy balance of microbes in our lower gastrointestinal tract (132,133). Members of the Bacteroidetes, a dominant phylum in the human gut, possess an arsenal of Polysaccharide Utilization Loci (PUL) to target a wide range of complex glycans (15). Analogous to the archetypal *Bacteroides thetaiotaomicron* starch utilization system (Sus), a hallmark of all Bacteroidetes PULs is the organization of genes clustered around tandem *susC*/*susD* homologs (encoding a TonB dependent transporter, TBDT, and a cell-surface glycan-binding protein, SGBP, respectively) (134). Additional co-localized and co-regulated SGBP(s), a cohort of enzymes, and a transcriptional regulator typically make up a machinery that acts in concert to sense, break down, and import complex glycans (105,135). Many such PULs, each targeting specific glycan structures, have been identified by genomics and transcriptomics (see, *e.g.*, the seminal study by (109), but detailed functional characterization lags severely behind (reviewed in (79,105)). Developing a precise understanding of the molecular details of complex glycan utilization by individual members of the microbiota is essential to designing targeted therapies based on prebiotics, probiotics, and symbiotic (136,137), as well as novel therapeutic interventions.

Recently, comprehensive functional analysis has revealed the detailed molecular mechanisms by which individual PULs enable human gut Bacteroidetes to utilize predominant plant polysaccharides, including the matrix glycans, xyloglucan (99,135,138), xylan (110), β-mannan (139), and rhamnogalacturonan II (140). Mixed-linkage β(1,3)/β(1,4)-glucans (MLGs, Fig. 1A) from cereal grains constitute an additional key group of dietary glycans, whose utilization by gut microbes was previously unresolved at the molecular level. MLGs are abundant in the aleurone layer of common cereals, including oats (3-8 % dry weight) and barley (2-20 % dry weight) (141). Beyond their obvious potential to contribute to energy intake (142,143), MLGs have been linked to a range of health benefits, *e.g.* promoting healthy cholesterol and blood glucose levels, ameliorating insulin resistance, and mitigating metabolic syndrome (141). In particular, the cholesterol lowering effect of oat MLG has long been recognized by the United States Food and Drug Administration (FDA) as well as the United Kingdom Joint Health Claims Initiative (JHCI), and been confirmed by subsequent studies (144).

The mechanisms behind these systemic benefits of MLG are, however, incompletely understood, in part due to a lack of understanding of MLG metabolism by individual members of the human gut microbiota. Thus, we report here the molecular characterization of a mixed-linkage glucan utilization locus (MLGUL) in the common symbiont *B. ovatus*. Identifying syntenic MLGUL in other Bacteroidetes revealed that as the archetype, this MLGUL serves as a molecular marker for MLG utilization across the Bacteroidetes phylum, thereby enabling future functional prediction across species.

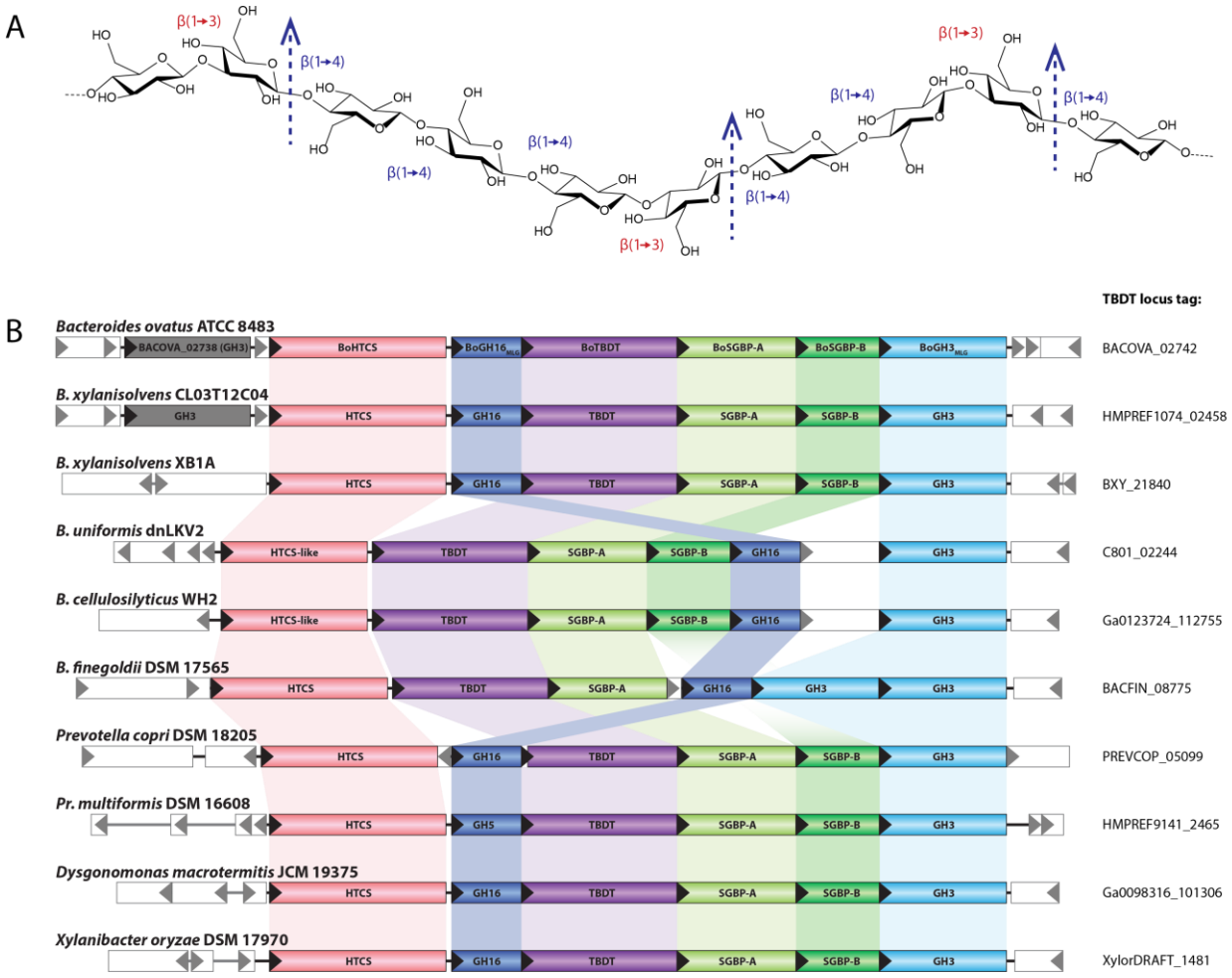

Figure 2-1. Cereal mixed-linkage β(1,3)/β(1,4)-glucan (MLG) and MLG Utilization Locus (MLGUL) structures.

(A) Chemical structure of MLG, consisting of a linear glucan chain of β(1,4)-linked cellotriosyl and cellotetraosyl units linked by β(1,3) bonds. MLGs from various sources (barley, oat, lichenin, etc.) vary in the ratio of cellotriose to cellotetraose units (145). Arrows indicate the specific site of hydrolysis by the vanguard endo-glucanase of the MLGUL, BoGH16$_{MLG}$. (B) Genetic organization of the *B. ovatus* MLGUL and syntenic loci in select Bacteroidetes species. Triangles represent gene directionality. Homologous genes are connected by colored bars and the locus tag of the TBDT of each syntenic MLGUL is given on the right.

2.2 Materials and Methods

2.2.1 Microbiology

2.2.1.1 Bacteroidetes reverse genetics and growth analysis

Strains of Bacteroidetes used in this study include a number of type strains from ATCC and DSMZ and a large number that were a generous gift from Dr. Abigail Salyers (see https://www.ericmartenslab.org/salyers for a full list).

The *B. ovatus Δtdk* strain has the gene encoding a thymidine kinase deleted for the purpose of engineered strain selection (146). For the purposes of this work, the *Δtdk* strain is treated as the 'wild-type' control. *B. ovatus* gene deletions were constructed in the *Δtdk* background using allelic exchange as previously described (146).

Anaerobic growth profiles were measured as previously described (109,145). Flat bottom 96-well plates (Costar) were loaded with 100 µL of sterilized carbohydrate stock at 2× concentration. A 24-hour culture was centrifuged to pellet bacteria, resuspended in 2× MM-no carbohydrate (MM-NC) and used to inoculate MM-NC at a ratio of 1:50. Each carbohydrate array was loaded with 100 µL of the inoculated 2× medium to produce 200 µL cultures at a final bacteria ratio of 1:100. Assay plates were sealed with an optically clear gas-permeable polyurethane membrane (Diversified Biotech, Boston, MA) in an anaerobic chamber (Coy manufacturing, Grass Lake, MI). Plates were loaded into a Biostack automated plate handling device coupled to a Powerwave HT absorbance reader (both devices from Biotek Instruments, Winooski, VT). Absorbance at 600 nm (A_{600}) was measured for each well at 10–15 minute intervals.

2.2.1.2 Enzyme localization

2.2.1.2.1 Immunofluorescence microscopy

Fluorescence microscopy was performed on fixed *Bacteroides ovatus Δtdk* and *ΔMLGUL* cells. The cells were grown to mid-exponential phase (A_{600} = 0.5-0.6) in Minimal Media (MM) with bMLG or glucose (0.5% w/v) as the sole carbon source. The cultures were then pelleted, and washed with phosphate-buffered saline (PBS). The cells were then fixed by incubation in formalin (4.5% formaldehyde in PBS) for 1.5 h at room temperature, washed with PBS, and blocked for 16 hours at 4 °C in blocking solution (2% goat serum (Sigma-Aldrich), 0.02% NaN_3, PBS). The cells were then incubated with individual polyclonal antibodies raised against recBoGH16$_{MLG}$, recBoGH3$_{MLG}$, and recBACOVA_02738(GH3) (Cedarlane Laboratories, Burlington, ON) for 2 hours at room temperature (1:500 dilution of the antibody in blocking solution). For secondary labelling, cells were pelleted, washed three times in 1 mL of PBS and resuspended in 0.4 mL goat anti-rabbit IgG Alexa-Fluor 488 (Thermo Fisher Scientific), diluted 1:500 in blocking solution, and incubated 1 hour at room temperature in the dark. The cells were then washed three more times and resuspended in 0.05 mL of PBS containing ProLong Gold Antifade (Thermo Fisher Scientific). Cells were mounted on agarose pads on glass slides and capped with coverslips. Fluorescence was

imaged on an Olympus IX70 inverted microscope (Olympus, Tokyo, Japan) at 100 X magnification.

2.2.1.2.2 Immunoblotting analysis

Bacteroides ovatus ⊿tdk cells were grown as described above in MM on bMLG (0.5 % w/v) or glucose (0.5 % w/v) as a sole carbon source. The cells were then centrifuged at 10,000 g for 45 minutes, resuspended in Tris-buffered saline (TBS), and lysed. After cell disruption, the membranes and cell debris were harvested by centrifugation for 1 hour at 42,000 rpm (TLA 100.3 Beckman) at 4 °C. To prepare the total membrane fraction, the pellet was resuspended in 60 mM of n-octyl β-D-glucopyranoside, agitated for 1 hour at room temperature and centrifuged at 35,000 rpm for 45 minutes at 4 °C. The supernatant was then harvested for further analysis.

The appropriate dilution of the culture supernatant, the lysate supernatant, and the total membrane fraction were added to 4X Laemmli buffer, boiled for 10 minutes, and run on an SDS polyacrylamide gel (Mini-PROTEAN® TGX^{TM} gels, Bio-Rad). Transfer to a western blot Polyvinylidene difluoride (PVDF) membrane (Immobilon®-P) was performed for 45 minutes at 20 volts using a semi-dry transfer cells (Trans-Blot SD, Bio-Rad). The membranes were then blocked for 1 hour at room temperature with blocking buffer (5% milk in TBST buffer (Tris-Buffered Saline (TBS) with 0.1% Tween20)). The membranes were then washed three times with TBST buffer and the proteins of interest were revealed by incubation with the primary antibodies generated for BoGH16$_{MLG}$, BoGH3$_{MLG}$, and BACOVA_02738(GH3), diluted in blocking buffer (1:15000, 1:20000, and 1:15000 dilution respectively). After three more washes, the membranes were incubated for 1 hour at room temperature with the secondary antibody goat anti-rabbit IgG H&L (Alkaline Phosphatase; Abcam), diluted 1:25000 in blocking buffer solution. The membranes were then washed another three times and the immunodetection of the alkaline phosphatase enzyme on the membrane was revealed with Novex® AP Chromogenic Substrate (ThermoFisher Scientific).

2.2.2 Cloning, expression, and purification of recombinant enzymes

Gene sequences were obtained from the *B. ovatus* ATCC 8483 draft genome available on the Integrated Microbial Genomes database from the Joint Genome Institute. PCR primers were synthesized by Integrated DNA technologies.

2.2.2.1 Cloning

Open reading frames encoding BACOVA_02738, BACOVA_02742, and BACOVA_02745 were amplified by PCR using Q5 high fidelity polymerase (NEB) with appropriate primers (Table A-1) and genomic *B. ovatus* DNA as template. All primers were designed to amplify constructs truncated to exclude predicted signal peptides (prediction by SignalP 4.1) and N-terminal lipidation cysteine residues (prediction by LipoP 1.0). NdeI and XhoI restriction sites were included in the forward and reverse primers of BACOVA_02742 for subsequent digestion (NdeI and XhoI from NEB) and ligation (T4 ligase from Thermo Scientific) into the pET28 vector. pMCSG complementary sequences were included in the forward and reverse primers of BACOVA_02738 and BACOVA_02745 for subsequent ligation independent cloning into pMCSG53 plasmids as per Eschenfeldt *et al.* (147). All three constructs were designed to harbor an N-terminal His_6-tag fusion in the translated recombinant peptide. Successful cloning was confirmed by colony PCR (GoTaq polymerase from Promega) and sequencing (Genewiz).

2.2.2.2 Expression

Plasmids harboring the gene of interest were transformed into chemically competent *E. coli* BL21 (DE3) and cultured in lysogeny broth (LB) containing 50 μg/mL kanamycin for BACOVA_02742 or 100 μg/mL ampicillin for BACOVA_02738 and BACOVA_02745. Cells were grown on a large scale at 37 °C until mid-logarithmic growth phase was reached (A_{600} = 0.4-0.6) at which point protein expression was induced by addition of isopropyl β-D-thioglactopyranoside to a final concentration of 0.5 mM and temperature was lowered to 16 °C. Induction of recombinant protein production continued overnight after which the cells were collected by centrifugation at 4000 g for 20 minutes.

2.2.2.3 Purification

The harvested cell pellet was resuspended in binding buffer (20 mM sodium phosphate pH 7.4, 500 mM sodium chloride, 20 mM imidazole) and lysed using a Sonic Dismembrator F550 Ultrasonic Homogenizer (Fisher Scientific). Cell debris was pelleted by centrifugation at 15000 rpm for 45 minutes and the supernatant was loaded onto a 2 × 1 mL HisTrap IMAC FF nickel-nitrilotriacetic acid column (GE Healthcare), using a BioLogic FPLC system (BioRad). After washing with 10 column volumes of binding buffer, His_6-tagged protein was eluted using a linear gradient of 0 - 100% elution buffer (20 mM sodium phosphate pH 7.4, 500 mM sodium chloride, 500 mM imidazole) over 10 column volumes. Fractions were monitored by A_{280} and eluted protein

fractions were pooled and buffer exchanged into 50 mM sodium phosphate pH 7.0 using Vivaspin centrifugal filters (GE Healthcare). After concentrating, aliquots were flash frozen in liquid nitrogen and stored at -80°C. Protein purity was determined by SDS-PAGE analysis and mass was confirmed by intact protein mass spectrometry on a Waters Xevo Q-TOF with nanoACQUITY UPLC system, as described previously (148). Protein concentrations were determined by spectrophotometry on an Epoch Microplate Spectrophotometer (BioTek) using the following molar extinction coefficients, which were calculated using ProtParam tool from the ExPASy Bioinformatics Resource Portal (149): 108555 $M^{-1}cm^{-1}$ for BACOVA_02738, 54890 $M^{-1}cm^{-1}$ for BACOVA_02742, and 108180 $M^{-1}cm^{-1}$ for BACOVA_02745.

Typical yields after purification were 80 mg of recBoGH16$_{MLG}$ from 1 L of lysogeny broth (LB) culture, 70 mg of recBoGH3$_{MLG}$ from 1 L of LB culture, and 8 mg of recBACOVA_02738(GH3) from 1 L terrific broth culture (Fig. S3).

2.2.3 Enzyme kinetics and product analysis

2.2.3.1 Substrates and polysaccharides

2.2.3.1.1 Polysaccharides

Beta-glucan (barley) high viscosity (bMLG; >94 % purity), yeast beta-glucan (>80 % purity), curdlan (>99 % purity), tamarind xyloglucan (~95 % purity), konjac glucomannan (>95 % purity), carob galactomannan (>95 % purity), wheat arabinoxylan (~95 % purity), beechwood xylan (>95 % purity) were purchased from Megazyme International (Bray, Ireland). Laminarin (from *Laminaria digitata*) was purchased from Sigma Aldrich (St. Louis, MO, USA). Carboxymethyl cellulose (>99.5 % purity) was purchased from Acros Organics (Morris Plains, NJ, USA). Hydroxyethyl cellulose was purchased from Amresco (Solon, OH, USA). Xanthan gum (99% purity) was purchased from Spectrum (New Brunswick, NJ, USA). Ulvan (from Ulva sp.) was purchased from Elicityl (Crolles, France). Laminarin was reduced to laminaritol as described previously (150).

2.2.3.1.2 Oligosaccharides

Cellobiose (G4G) was purchased from Acros Organics. Cellotriose (G4G4G), cellotetraose (G4G4G4G), cellopentaose (G4G4G4G4G), cellohexaose (G4G4G4G4G4G), laminaribiose (G3G), laminaritriose (G3G3G), laminaritetraose (G3G3G3G), laminaripentaose (G3G3G3G3G), mixed-linkage glucotriose A (G3G4G), mixed-linkage glucotriose B (G4G3G), mixed-linkage glucotetraose A (G3G4G4G), mixed-linkage glucotetraose B (G4G4G3G), mixed-linkage

glucotetraose C (G4G3G4G) were purchased from Megazyme. Gentiobiose (G6G) was purchased from Carbosynth (Compton, UK).

2.2.3.1.3 Chromogenic substrates

para-nitrophenyl (*p*NP) glycosides of β-glucoside (G-β-*p*NP), α-glucoside, β-galactoside, β-mannoside, and β-xyloside were purchased from Sigma Aldrich. *ortho*-chloro-*para*-nitrophenyl (CNP) glycosides of G4G3G (G4G3G-CNP) and G4G4G3G (G4G4G3G-CNP), and *p*NP β-laminaribioside (G3G-*p*NP) were purchased from Megazyme. CNP glycosides of cellobioside (G4G-CNP) and that of cellotriose (G4G4G-CNP) were purchased from Carbosynth. G3G-CNP was synthesized by glycosylation of the known α-laminaribiosyl bromide (151) and the corresponding phenol under phase-transfer conditions (151,152), the details of which will be published elsewhere.

2.2.3.2 Enzyme kinetics

2.2.3.2.1 BCA endpoint assay

Polysaccharide hydrolysis was quantified using the bicinchoninic acid (BCA) reducing sugar assay. All reactions were carried out in 100 μL volumes in the optimum pH buffer (50 mM sodium citrate pH 6.5 for $BoGH16_{MLG}$) at 37 °C unless otherwise specified. Reactions were initiated by adding 10 μL of enzyme solution to 90 μL of the remaining assay mixture, which had been pre-incubated at 37 °C. Reactions were terminated by addition of equal volume (100 μL) of BCA reagent and developing the color by boiling at 80 °C for 20 minutes. Absorbance at 563 nm (A_{563}) was measured in 96-well plates on an Epoch Microplate Spectrophotometer (BioTek). Blank absorbance readings were determined for each polysaccharide at each concentration by using inactivated enzyme (denatured by boiling at 100 °C for 10 minutes). Reducing ends released were quantified with a glucose standard curve (25 – 150 μM). All kinetic assays were conducted in technical triplicates.

Activity on a library of polysaccharides was initially screened by incubating 10 μM $BoGH16_{MLG}$ with 1.0 mg/mL substrate for 24 hours. The polysaccharide was determined to be a substrate for $BoGH16_{MLG}$ if the A_{563} increased significantly compared to the blank.

The pH optimum of $BoGH16_{MLG}$ was determined by incubating 7.5 nM enzyme with 1.0 mg/mL bMLG for 10 minutes in different buffers at 50 mM: sodium citrate (pH 3.0 – 6.5), sodium phosphate (pH 6.5 – 8.5), glycylglycine (pH 7.5 – 9.0), glycine (pH 9.0 – 10.5). Released reducing ends were measured as described above.

The temperature optimum of $BoGH16_{MLG}$ was determined by incubating 7.5 nM enzyme with 1.0 mg/mL bMLG for 10 minutes at various temperatures ranging from 30 to 70 °C. Released reducing ends were measured as described above.

For initial-rate saturation kinetics, the following concentrations of enzyme were used: 4.9 nM for bMLG, 48.6 nM for laminarin, 485.6 nM for yeast beta-glucan, and 4.9 μM for curdlan. These are the concentration that were optimized for the reaction to be in the initial, linear stage of the reaction (less than 10% conversion) after 12 minutes of hydrolysis. To determine Michaelis-Menten parameters, eight different concentrations of each substrate were hydrolyzed by appropriate concentration of enzyme for 10 minutes after which the reaction was quenched and reducing ends released were quantified as described above.

2.2.3.2.2 Chromogenic substrate assay

Reaction with *p*NP and CNP glycoside substrates was used to quantify the hydrolysis of chromophore from the aglycone. Enzyme concentrations used to maintain initial-rate conditions were 6.9 nM for $BoGH3_{MLG}$ against G-*p*NP, 941 nM for BACOVA_02738(GH3) against G-*p*NP, 9.4 nM for $BoGH16_{MLG}$ against G4G3G-CNP and G4G4G3G-CNP, 446 nM for $BoGH16_{MLG}$ against G3G-CNP, and 942 nM for $BoGH16_{MLG}$ against G3G-*p*NP.

Endpoint assays were used for pH and temperature optima of $BoGH3_{MLG}$ and BACOVA_02738(GH3). Enzyme, 1 mM G-*p*NP, and 50 mM of the same range of different pH buffers described above were mixed to a final volume of 100 μL. The reactions were also carried out in optimal pH buffer (50 mM sodium phosphate pH 7.5 for $BoGH3_{MLG}$ and 50 mM sodium citrate pH 6.5 for BACOVA_02738(GH3)) at various temperatures ranging from 30 to 70 °C. Reactions were terminated after 10 minutes by addition of 100 μL of 1 M Na_2CO_3 to raise the pH and absorbance at A_{405} was measured in 96-well plates on an Epoch Microplate Spectrophotometer (BioTek). An extinction coefficient of 18,100 M^{-1} cm^{-1} was used for these assays.

Continuous assays were used for initial-rate saturation kinetics. Reactions, carried out in 250 μL volumes in the optimum pH buffer at 37 °C, were initiated by adding 25 μL of enzyme solution to 225 μL of the remaining assay mixture, pre-incubated at 37 °C. Release of *p*NP or CNP was monitored by following absorbance at 405 nm in quartz cuvettes using a Cary 60 UV-Vis spectrophotometer (Agilent Technologies). Eight different substrate concentrations were assayed and rate was calculated using an extinction coefficient of 15298 M^{-1} cm^{-1} for CNP in sodium citrate

pH 6.5, 3311 M^{-1} cm^{-1} for *p*NP in sodium citrate pH 6.5, and 12511 M^{-1} cm^{-1} for *p*NP in sodium phosphate pH 7.5

2.2.3.2.3 HK/G6PDH coupled assay

Release of glucose monosaccharides was quantified using the D-Glucose HK Assay Kit from Megazyme, modified for use as a continuous assay. All reactions were carried out in 250 µL volumes at 37 °C in the triethylamine pH 7.6 buffer provided in the kit. $BoGH3_{MLG}$ concentrations used to maintain initial-rate conditions were 9.2 nM for laminari-oligosaccharides and mixed-linkage oligosaccharides, 50.1 nM for cello-oligosaccharides, and 1.28 µM for gentiobiose. Reactions were initiated by adding 25 µL of enzyme solution to 225 µL of the remaining assay mixture containing hexokinase, glucose-6-phosphate dehydrogenase, ATP, and $NADP^+$, pre-incubated at 37°C. The release of glucose monosaccharides corresponds stoichiometrically with the reduction of a molecule of $NADP^+$ to NADPH, which was monitored by following absorbance at 340 nm on a Cary 60 UV-Vis spectrophotometer. An extinction coefficient of 6,220 M^{-1} cm^{-1} was used to convert to rate of hydrolysis.

2.2.3.3 Enzyme limit digest assay

To determine limit-digestion products of $BoGH16_{MLG}$, 10 µM enzyme was incubated with 1.0 mg/mL polysaccharide in 1 mL of 50 mM sodium citrate pH 6.5 for 24 hours at 37 °C. To determine limit-digestion products of $BoGH3_{MLG}$ and BACOVA_02738(GH3), 10 µM enzyme was incubated with the limit digest product of $BoGH16_{MLG}$ hydrolysis of 1.0 mg/mL polysaccharide in 1 mL of the appropriate buffer for 24 hours at 37 °C. 10 µL of the reaction was diluted into 1 mL of ultrapure water and analyzed on HPAEC-PAD and HILIC-MS as described below.

The same experiment was conducted to observe reaction progress, except 10 nM of $BoGH16_{MLG}$ and 12 nM of $BoGH3_{MLG}$ were used and reactions were stopped at various time points by taking 100 µL of the reaction mixture and adding to 100 µL of NH_4OH. 20 µL of the reaction was diluted into 1 mL of ultrapure water and analyzed on HPAEC-PAD as described below.

2.2.3.4 Carbohydrate analytical methods

2.2.3.4.1 HPAEC-PAD product analysis

HPAEC-PAD was performed on a Dionex ICS-5000 HPLC system operated by Chromelion software version 7. Samples were separated on a 3 × 250 mm Dionex Carbopac PA200

column (Thermo Scientific). Solvent A was ultrapure water, solvent B was 1 M sodium hydroxide, and solvent C was 1 M sodium acetate. Conditions used were 0 – 5 min, 10 % B (initial conditions); 5 – 12 min, 10 % B, linear gradient from 0 – 30 % C; 12.0 –12.1 min, linear gradient from 10 – 50 % B, linear gradient from 30 – 50 % C; 12.1 – 13.0 min, exponential gradient of B and C back to initial conditions; 13 – 17 min, initial conditions.

2.2.3.4.2 HILIC-MS product analysis

Samples were separated by hydrophilic interaction liquid chromatography on a TSKgel Amide-80 column (Tosoh Bioscience). Solvent A was ultrapure water and solvent B was 1 M acetonitrile. The mobile phase used was a linear gradient of 35 % A and 65 % B to 50 % A and 50 % B over 30 minutes. The eluent was split between an evaporative light scattering detector (ELSD) (Agilent Technologies) and the Bruker Esquire 3000 Plus ion trap mass spectrometer (Bruker Daltonics). The eluent was ionized in positive mode by electrospray ionization before detection by ion trap. The ELSD and total ion count chromatograms were identical to the HPAEC-PAD trace. Esquire HyStar software was used to process the mass spectrometry data (Bruker Daltonics).

2.2.4 X-ray crystallography

Purified $BoGH16_{MLG}$ (>90 % as estimated from SDS-PAGE) at 23 mg/ml in 50 mM sodium phosphate pH 7, was used to set up initial sitting drop crystal screens using a Mosquito robot (TTP Labtech). An initial hit condition was identified in the PACT screen (Qiagen) condition B9: 100 mM MES pH 6, 200 mM LiCl, 20 % w/v PEG-6000. Crystals were readily reproduced by hand in larger sitting drops by screening around this condition, varying only the PEG-6000 concentration from 15 to 25%. The crystals obtained from these optimizations were used in all subsequent work.

Crystals of the apo protein were cryo-cooled for data collection by first soaking in a solution of mother liquor supplemented with 18% ethylene glycol for 30 seconds before plunging in liquid nitrogen. Diffraction data were collected from these crystals at Diamond Light Source, beamline I03 at a wavelength of 0.976 Å. Data were indexed and integrated using XDS (153) with all subsequent data processing performed using the CCP4 software suite (154). A search model for molecular replacement was prepared using a single subunit from the *Zobellia galactanivorans* laminarinase $ZgLamC_{GH16\text{-}E142S}$ (PDB code 4CRQ; 38% amino acid identity with BoGH16) (155) and using CHAINSAW (156) to trim any sidechains in the model to the nearest common atom

based on a sequence alignment. The structure was then determined using this model by molecular replacement in PHASER (157). Following density modification in PARROT (158), BUCCANEER (159) was used to construct an initial model before further model building and refinement were performed in COOT (160) and REFMAC5 (161) respectively.

To obtain the G4G4G3G complex structure, crystals were soaked for 30 minutes in cryo-protectant solution (100 mM MES pH 6, 200 mM LiCl, 25 % w/v PEG-6000, 18 % w/v ethylene glycol) in which the ligand had been dissolved at 50 mM. Crystals were then plunged in liquid nitrogen ready for data collection. X-ray data were collected from these crystals at Diamond Light Source, beamline I02 at a wavelength of 0.979 Å. Data were processed as above using XDS (153) for indexing and integration followed by subsequent processing in the CCP4 software suite (154). Since the crystals were isomorphous to the apo-structure, the apo model with waters and flexible loops removed was refined against these new data. The model was rebuilt and refined using COOT (162) and REFMAC5 (161).

For both structures, the quality of the model was monitored throughout using MOLPROBITY (163) - the final models having no outliers, 98.5 % and 98.7 % of residues in the favored region of the Ramachandran plot for the apo- and G4G4G3G-complex respectively. Additionally, the sugar conformations in the G4G4G3G-BoGH16$_{MLG}$ complex were all confirmed as $^{4}C_{1}$ chairs using PRIVATEER (164) and the generated restraints applied during structure refinement. Data processing and refinement statistics for both structures can be found in Table S4. The apo- and G4G4G3G-complex structures have been deposited in the Protein Data Bank with accession codes 5NBO and 5NBP respectively.

2.2.5 Bioinformatics

2.2.5.1 Phylogenetic analysis

Glycoside Hydrolase Family 16 sequences with EC number 3.2.1.6 and 3.2.1.39 were extracted from the CAZy Database (URL http://www.cazy.org) using the Extract Sequences tool (URL http://research.ahv.dk/cazy/extract). The sequences were initially aligned by MUSCLE (165) in AliView (166) and manually trimmed to remove amino acids outside of the GH16 catalytic domain. The resulting sequences were structurally aligned using T-Coffee Expresso (167), then further manually refined in AliView, guided by available three-dimensional structures. A maximum-likelihood phylogenetic tree was constructed using MEGA6 v6.06 (168) and

reliability of the nodes was tested by bootstrap analysis using 100 resampling. Five cellulases from Glycoside Hydrolase Family 7 were used as an outgroup to root the tree.

All Glycoside Hydrolase Family 3 sequences listed as "characterized" as well as those that are structurally characterized with EC number 3.2.1.21 (β-glucosidases) were similarly extracted from the CAZy database. The roughly 300 characterized GH3 sequences were aligned by MUSCLE and a maximum-likelihood phylogenetic tree constructed using MEGA v6.06. The structurally characterized β-glucosidase sequences were combined with $BoGH3_{MLG}$ and BACOVA_02738 and aligned by T-Coffee Expresso.

2.2.5.2 Survey of metagenomic datasets

Human metagenomic sequence data sets (169-172) were searched by BLAST for the presence of MLGUL nucleotide sequences from *B. ovatus* (13.4kb), *B. uniformis* (14.4kb), *B. cellulosilyticus* (14.1kb), *B. finegoldii* (16.2kb), and *Pr. copri* (13.9kb). Each BLAST probe was first searched against the NCBI Refseq genomes database to determine the background thresholds for BLAST hits and subsequently trimmed to remove any sequences that may return off-target hits. *B. ovatus* and *B. xylanisolvens* MLGULs could not be distinguished due to their very high nucleotide identity (97%). Otherwise, this analysis failed to reveal any off-target hits with length >75 bp, nucleotide identity >90%, and E value $<1^{-20}$. Thus, we considered a metagenome to be positive for a particular MLGUL probe if it returned two or more hits >100 bp in length with >90% identity and E value $<1^{-20}$, or one hit >1000 bp in length with the same identity and E value cut-offs.

2.3 Results

2.3.1 Identification of a multi-gene locus responsible for MLG utilization by *B. ovatus*

A putative MLGUL was previously identified in *B. ovatus* (Fig. 2-1B) based on the presence of a tandem *susC/susD* homolog signature (134) and high-level expression of select genes in the presence of bMLG (109). Individual genes in the locus, BACOVA_02741-02745, were all substantially upgregulated (125 to 298-fold) during growth on bMLG vs. glucose as sole carbon sources (Table A-2). BACOVA_02742 and BACOVA_02743 encode the signature TBDT/SGBP pair with 28% and 19% protein sequence identity to SusC and SusD, respectively. The putative MLGUL was additionally predicted to encode a second, non-homologous SGBP (BACOVA_02744), a hybrid two-component sensor/transcriptional regulator (HTCS, BACOVA_02740), and up to three glycoside hydrolases (GHs).

BoGH16$_{MLG}$ (BACOVA_02741) is a member of Glycoside Hydrolase Family 16 (GH16), a family of *endo*-β-glucanases in the Carbohydrate Active Enzymes (CAZy) classification (50). GH16 notably includes canonical bacterial MLG *endo*-glucanases (*endo*-MLGase) (173), along with a diversity of *endo*-glucanases and *endo*-galactanases. BoGH3$_{MLG}$ (BACOVA_02745) is classified into Glycoside Hydrolase Family 3 (GH3), whose members include *exo*-β-glucosidases. Notably, we have determined that BACOVA_02738, which is predicted to encode a second GH3 *exo*-β-glucosidase, is unlikely to be part of the MLGUL for three reasons: (1) BACOVA_02738 was not significantly upregulated on MLG (1.6-fold vs. glucose control, Table A-2), (2) a corresponding gene is not found among syntenic loci (Fig. 2-1B), and (3) the encoded enzyme was catalytically feeble compared to BoGH3$_{MLG}$ on β-glucosides relevant to MLG saccharification (*vide infra*).

To determine the correlation between the presence of the predicted MLGUL and growth of *B. ovatus* on MLG, we constructed an isogenic mutant of *B. ovatus Δtdk* (138) in which a contiguous region of DNA encoding genes BACOVA_02738-02745 was deleted (*B. ovatus ΔMLGUL*). Vis-à-vis the parent strain, the *B. ovatus ΔMLGUL* was able to grow equally well on glucose as the sole carbon source, however the ability to grow on bMLG was completely abolished (Fig. A-1). Thus, the putative MLGUL is necessary to confer *B. ovatus* the ability to utilize MLG.

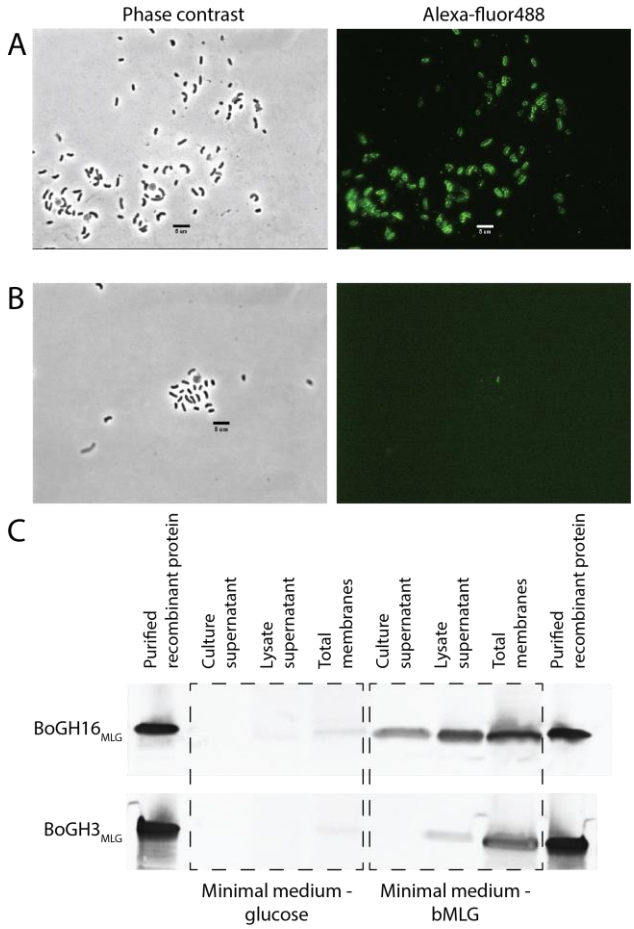

Figure 2-2. Enzyme localization analysis.

Phase contrast microscopy and corresponding fluorescence microscopy images of *B. ovatus Δtdk* cells grown in minimal medium with bMLG as the sole carbon source probed with custom polyclonal antibodies against (A) recBoGH16$_{MLG}$ and (B) recBoGH3$_{MLG}$. (C) Western blots of protein collected from the culture supernatant, cell lysate supernatant, and cell lysate membrane fraction of *B. ovatus Δtdk* cells grown in minimal medium with glucose or bMLG as a sole carbon source.

2.3.2 Enzymology and structural biology of BoGH16$_{MLG}$, the vanguard MLGase

2.3.2.1 Cellular localization

The GH family membership of BoGH16$_{MLG}$ suggested a potential role as the vanguard enzyme catalyzing polysaccharide backbone cleavage at the cell surface as the essential first step in MLG utilization. Indeed, the presence of a predicted Type II signal sequence (determined with LipoP 1.0 (174)) suggested that the protein is membrane-anchored via lipidation of the N-terminal cysteine residue (175). To validate this prediction, *B. ovatus Δtdk* was grown on minimal medium with either glucose or bMLG as a sole carbon source prior to immunolocalization of BoGH16$_{MLG}$.

As shown in Fig. 2-2A, BoGH16$_{MLG}$ was clearly visualized on the outer surface of cells in which the presence of the polysaccharide induced MLGUL expression, but was absent from cells grown on glucose (Fig. A-2C, A-2F). Further analysis of cellular fractions by Western blotting revealed the presence of BoGH16$_{MLG}$ in the membrane fraction, corroborating its attachment to the outer membrane (Fig. 2-2C). Interestingly, BoGH16$_{MLG}$ was also detected in the lysate supernatant (soluble periplasmic or cytoplasmic proteins) and in the culture supernatant (secreted protein) (Fig. 2-2C). While the former may represent anchored protein released into the soluble fraction during cell lysis, detection in the culture supernatant could result from packaging and release in outer membrane vesicles, which has previously been observed for other Bacteroidetes glycoside hydrolases (176).

2.3.2.2 Substrate and product specificity

To verify the leading catalytic role of BoGH16$_{MLG}$ in MLG utilization and determine the specificity of the enzyme for individual β-glucans, recombinant BoGH16$_{MLG}$ produced in *E. coli* (recBoGH16$_{MLG}$, Fig. A-3A, A-3B) was screened for hydrolytic activity against a library of polysaccharides. No activity was observed on tamarind xyloglucan, beechwood xylan, wheat arabinoxylan, carob galactomannan, konjac glucomannan, synthetic carboxymethylcellulose, synthetic hydroxyethylcellulose, *Xanthomonas campestris* xanthan gum, or *Ulva sp.* ulvan. In this initial screen, BoGH16$_{MLG}$ was minimally active on all-β(1,3)-glucans, including *Laminaria digitata* laminarin, yeast β-glucan, and *Alcaligenes faecalis* curdlan, but exhibited high specific activity on bMLG. The optimum pH of 6.5 (consistent with function in the distal human colon) and maximum temperature range of 45 to 55 °C was determined using bMLG as substrate (Fig. A-4A, A-4B).

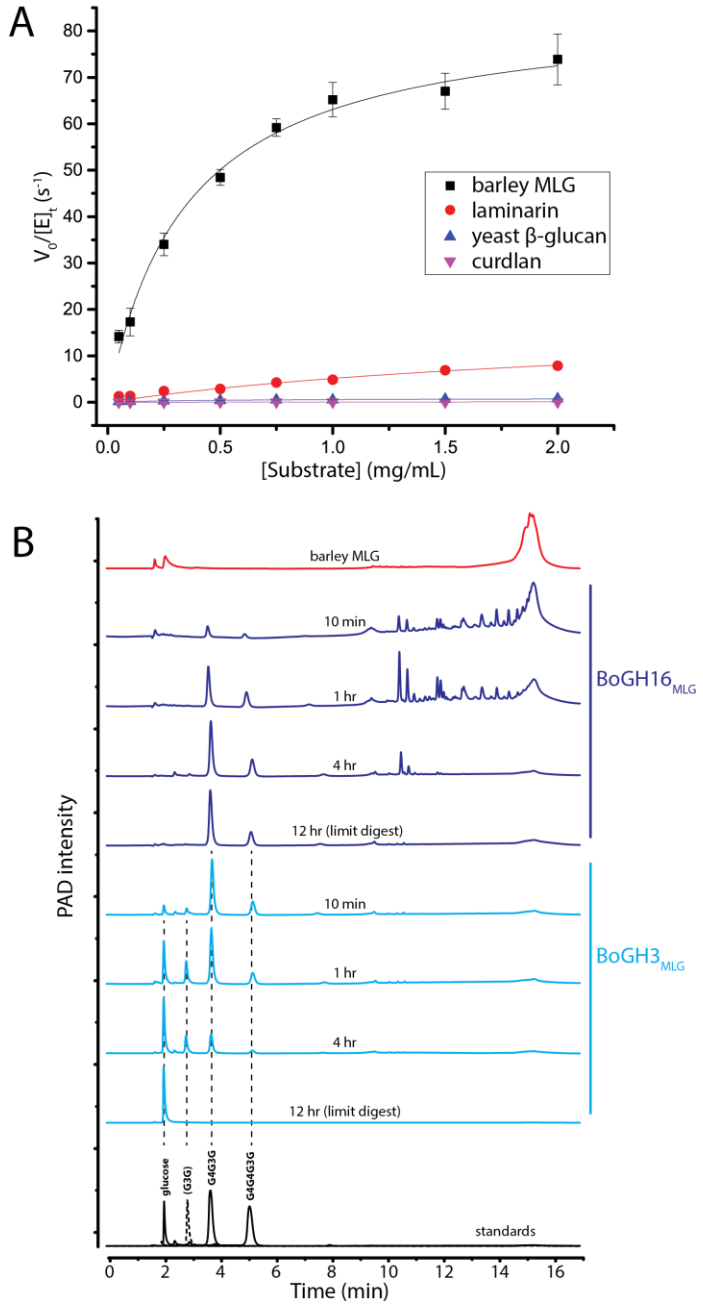

Figure 2-3. BoGH16$_{MLG}$ kinetics and MLGUL GHs product analysis.

(A) BoGH16$_{MLG}$ initial-rate kinetics curves fitted to the Michaelis-Menten equation for β-glucan polysaccharide substrates on which it is active. Laminarin was reduced to laminaritol by sodium borohydride reduction to reduce background in the BCA assay. Curve fitting was done on OriginPro 2015 and error bars represent standard deviations from the mean. (B) Chromatograms of bMLG and its hydrolysis products by BoGH16$_{MLG}$ and BoGH3$_{MLG}$ separated by HPAEC-PAD. Red: full length bMLG polysaccharide. Dark blue: reaction progress time course and limit digest of bMLG hydrolysis by 10 nM BoGH16$_{MLG}$. Cyan: reaction progress time course and limit digest of BoGH16$_{MLG}$ products hydrolysis by 25 nM BoGH3$_{MLG}$. Standards are shown below in black: solid lines are those corresponding to limit digest products and dotted line to intermediate products. See also Figures S3, S4 and S5; Tables S2 and S3.

Subsequent Michaelis-Menten kinetic analysis at the pH optimum and 37 °C demonstrated that BoGH16$_{MLG}$ is a highly predominant mixed-linkage β(1,3)/β(1,4)-glucanase (MLGase), with a 33-fold higher specificity constant, k_{cat}/K_m, for bMLG (Fig. 2-1A) over laminarin, an all-β(1,3)-glucan with single β(1,6)-linked glucosyl branches (Fig. 2-3A, Table A-3) (177). BoGH16$_{MLG}$ was even less efficient on the other two all-β(1,3)-glucans for which activity was initially observed: The k_{cat}/K_m, was 147-fold higher for bMLG than yeast β-glucan (similar in structure to laminarin but with longer β(1,6)-linked glucose branches (178)) and nearly four orders of magnitude higher than that of high curdlan, a 22 kDa, non-branched β(1,3)-glucan (179) (Fig. 2-3A, Table A-3).

Detailed product analysis was employed to determine the mode of hydrolysis, *endo* vs. *exo*, and linkage specificity of recBoGH16$_{MLG}$ to gain information on the nature of the MLG cleavage products at the *B. ovatus* cell surface. HPLC analysis at selected time points in the hydrolysis showed the initial production of very large oligosaccharide fragments, which were progressively converted into shorter species and ultimately to two distinct oligosaccharides in the limit-digest (Fig. 2-3B). This product evolution indicates that BoGH16$_{MLG}$ operates through an *endo*-dissociative mode of action in which the MLG polysaccharide is stochastically cleaved along the backbone.

Comparison with oligosaccharide standards (Fig. 2-3B) and additional LC-MS analysis (data not shown) revealed that the limit-digest products were the mixed-linkage trisaccharide, G4G3G [Glcβ(1-4) Glcβ(1-3) Glc], and the mixed-linkage tetrasaccharide, G4G4G3G [Glcβ(1-4) Glcβ(1-4) Glcβ(1-3) Glc]. Thus, BoGH16$_{MLG}$ specifically hydrolyzes β(1,4)-linkages of glycosyl residues that are immediately preceded by a β(1,3)-linked glucosyl residue (toward the non-reducing chain end). This specificity is typical of bacterial *endo*-MLGases within GH16 (173,180,181).

To provide more refined insight into BoGH16$_{MLG}$ substrate specificity, Michaelis-Menten kinetics were determined for a series of chromogenic glycosides (Fig. A-5A, A-5B; Table A-4). recBoGH16$_{MLG}$ had no activity on the *ortho*-chloro-*para*-nitrophenyl (CNP) β-glycosides of glucose (G-CNP), cellobiose (G4G-CNP), cellotriose (G4G4G-CNP), nor on *para*-nitrophenyl (*p*NP) β-glucoside (G-*p*NP). Weak activity was observed on the *p*NP and CNP β-glycosides of laminaribiose (G3G), consistent with a requirement for a β(1,3) linkage spanning the -2 to -1 active-site subsites (GH subsite nomenclature according to (182)), as was indicated by the bMLG limit-digest analysis (*vide supra*). Likewise, G4G3G-CNP and G4G4G3G-CNP were specifically

and efficiently hydrolyzed to release the aglycone, with no cleavage of the internal glycosidic bonds. The specificity constants (k_{cat}/K_m values) for CNP release from these mixed-linkage tri- and tetrasaccharides were 800- and 1500-fold greater than that of G3G-CNP, respectively, which indicate that potential -3 and -4 subsites contribute 17 kJ/mol and 1.6 kJ/mol to transition state stabilization ($\Delta\Delta G^{\ddagger}$). Indeed, a very significant contribution from the -3 subsite is a common feature of GH16 *endo*-MLGases (173,180,183).

2.3.2.3 BoGH16$_{MLG}$ tertiary structure

Three-dimensional structures of recBoGH16$_{MLG}$ were solved by X-ray crystallography to reveal the molecular basis for MLG recognition and hydrolysis. The apo structure of recBoGH16$_{MLG}$ was determined to a resolution of 1.8 Å by molecular replacement using the structure of *Zobellia galactanivorans* laminarinase ZgLamC$_{GH16\text{-}E142S}$ (PDB code 4CRQ) (155) as a search model (See Table A-5 for processing and refinement statistics). The crystal contained two polypeptide chains in the asymmetric unit corresponding to residues I35-L271 of wild-type BoGH16$_{MLG}$ for both chains (residue numbering is from transcriptional start site according to the genomic sequence). No electron density was observed for the N-terminal His$_6$-tag and subsequent 15 amino acids in either chain of the recombinant protein, which suggests that residues C19-D34 of the wild-type enzyme constitute a flexible linker sequence to distance the catalytic module from the outer membrane surface (residues M1-S18 comprise the predicted signal peptide); the sidechain of C19 would constitute the site of N-terminal lipidation (175). The overall fold of BoGH16$_{MLG}$ consists of a β-jelly roll architecture typical of other GH16 members (184): Two antiparallel β-sheets stack on top of each other with the concave face forming the polysaccharide substrate binding cleft. The end-on arrangement of the two chains in the asymmetric unit hinted at the possibility of the formation of a dimer (Fig 2-4A). Size-exclusion chromatography, however, indicated that BoGH16$_{MLG}$ exists as a monomer in solution (data not shown) which, together with steric considerations of polysaccharide binding through the active-site cleft, indicates that end-on contacts observed between Chains A and B are artifacts of crystal packing.

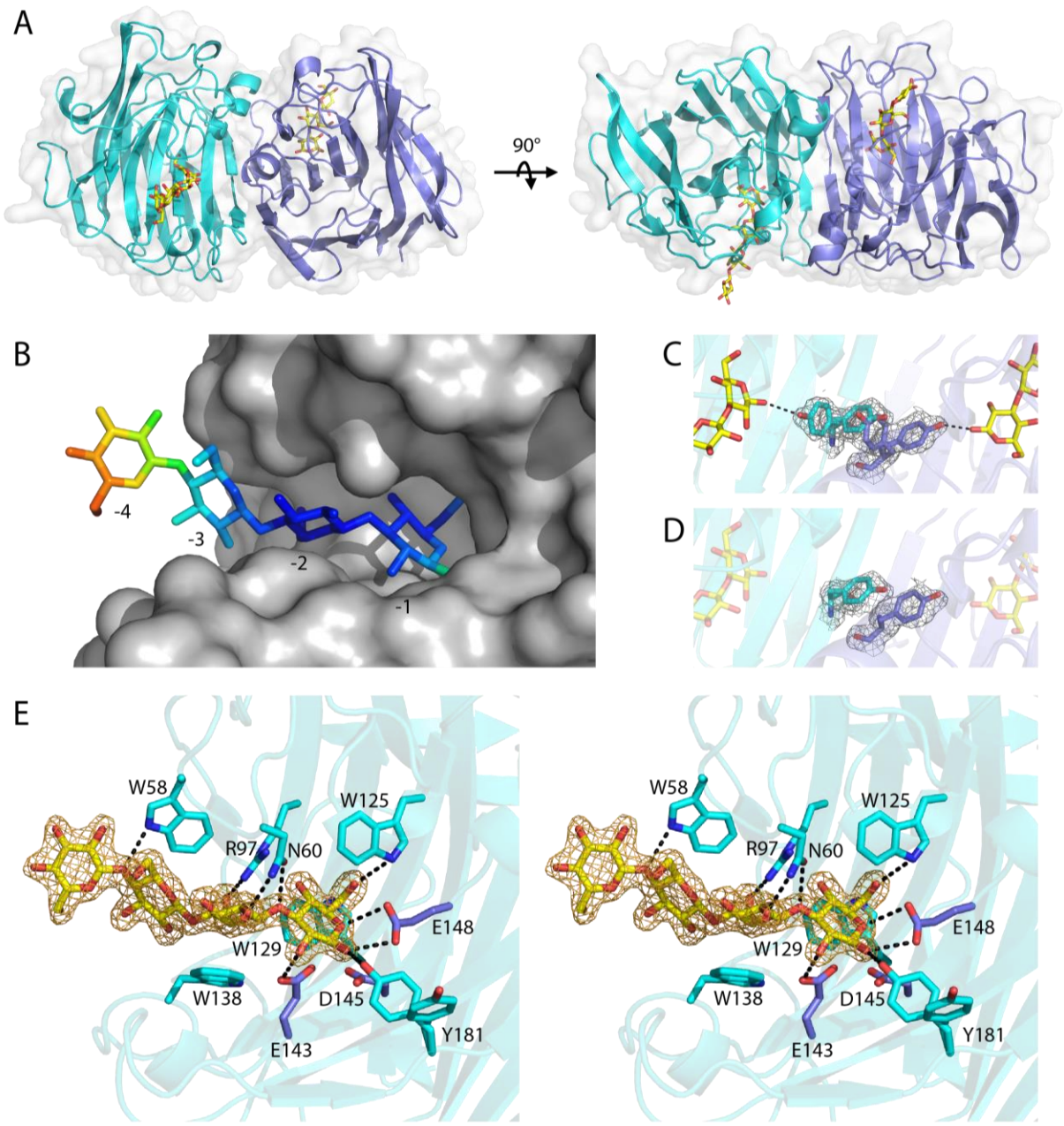

Figure 2-4. BoGH16$_{MLG}$ structural biology.

(A) the overall structure of the BoGH16$_{MLG}$:G4G4G3G asymmetric unit containing two polypeptide chains shown from orthogonal views with the bound oligosaccharides in yellow and the transparent surface representation in white. Chain A cartoon is shown in cyan, and chain B cartoon is shown in slate blue throughout the figure. (B) Mixed-linkage tetrasaccharide ligand modelled into chain A of BoGH16$_{MLG}$ with the opaque surface representation in gray and the oligosaccharide colored according to B-factors. The glucose in subsite -4 is outside of the active site cleft and has significantly higher B factor than the glucose units in subsites -1 to -3. (C) Tyr-181 rotamers observed in the complex structure with the 2Fo-Fc map of the tyrosines shown contoured at 0.5σ in grey. (D) Tyr-181 residues observed in the apo structure with the 2Fo-Fc map of the tyrosines shown contoured at 0.5σ in grey. (E) Wall-eyed stero view of the

active site of chain A of the $BoGH16_{MLG}$:G4G4G3G complex. Hydrogen bonding interactions are shown as dashed black lines, sugars are shown in yellow with its 2Fo-Fc map contoured at 1σ in orange, and the conserved GH16 active site residues shown in purple. Hydrophobic stacking interactions in addition to hydrogen bonds position the mixed-linkage oligosaccharide in the negative subsite of $BoGH16_{MLG}$.

The sidechains of the conserved GH16 catalytic residues (173), comprising Glu-143 (nucleophile), Asp-145 (electrostatic "helper") and Glu-148 (acid/base) are presented on the same face of one β-strand (β8), pointing into the active-site cleft (see Fig. 1-5 for hydrolysis mechanism). Notably, these residues are contained in a EXDXXE consensus sequence that is typical of bacterial GH16 laminarinases (β(1,3)-specific endo-glucanases). The insertion of an extra amino acid (underlined), typically methionine, results in a so-called "β-bulge" secondary structural motif that is not found among canonical bacterial MLGases, which instead possess a regular β-strand (185,186).

Commensurate with this observation, the closest eight structural homologs identified using the Dali server (187) feature a β-bulge active-site motif (Table A-6). Specifically, the top match (Z-score = 29.3) was the structure of laminarinase "$ZgLamC_{GH16\text{-}E142S}$" from *Zobellia galactanivorans* (PDB code 4CTE) (155), which has 38% amino acid identity and superimposed with $BoGH16_{MLG}$ with a root mean square deviation (RMSD) value of 2.0 Å over 211 out of 231 Cα pairs. In comparison, the closest GH16 homolog with a regular active-site β-strand was the lichenase (MLGase) from *Paenibacillus macerans* (PDB code 1MAC) (188), which has a comparable Z-score of 25.1 and an RMSD value of also 2.0 Å over 200 out of 212 Cα pairs, despite having only 22% amino acid identity with $BoGH16_{MLG}$.

Soaking crystals of the wild-type enzyme with G4G4G3G yielded a product complex with 1.8 Å resolution (Table A-5). The complete tetrasaccharide was modelled in electron density spanning subsites -1 to -4 in the active-site cleft of Chain A, while the electron density for the fourth glucosyl residue in subsite -4 was not resolved in chain B. This is most likely due to disorder of this residue since the corresponding -4 Glc in Chain A is fully solvent exposed, makes no contact with the enzyme, and has significantly higher B-factors (Fig. 2-4B). These structural observations are consistent with kinetic data for chromogenic MLG oligosaccharides (Table 2-1), which likewise suggest the existence of three primary negative subsites, -1 to -3, and a weakly interacting -4 subsite.

In both Chain A and B, the three glucosyl residues spanning subsites -1 to -3 are well defined and virtually identical. The reducing-end glucosyl residue in the -1 subsite is in the β-conformation, with the C1 hydroxyl hydrogen bonded to Tyr-181, which is observed in a dual conformation in both chains of the G4G4G3G-complex (Fig. 2-4C). Interestingly, this dual conformation is not observed in the apo-form of the enzyme; Tyr-181 is "swung in" to the active site in chain B, while it is "swung out" in chain A, the sidechain from chain A stacking on top of the chain B sidechain (Fig. 2-4D). The conformation of this sidechain will be key to determining the nature of the positive substrate binding subsites, indeed, comparison with other GH16 *endo*-glucanases clearly suggests the presence of two positive subsites (173,180). Whether the dynamics observed for Tyr-181 are an artefact of crystallisation, or perhaps play a role in substrate binding and product release is unclear in the absence of an enzyme-substrate complex spanning the positive subsites.

With regard to specific interactions in the negative subsites, subsite -1 is further characterized by hydrogen bonds between Glu-143 and the C2 hydroxyl, Trp-125 and the C6 hydroxyl, as well as Glu-148 and the ring oxygen and the C1 hydroxyl. This glucose is also positioned by a stacking interaction with Trp-125 and Trp-129 (Fig. 2-4E), both of which are conserved across all GH16 laminarinases. At subsite -2, highly conserved Arg-97 forms a hydrogen bond with the C6 hydroxyl, and Asn-60 hydrogen bonds to the C2 hydroxyl as well as to the glucosidic bond oxygen between the -1 and -2 sugars. Another conserved residue, Trp-138, serves as a platform that stacks with the subsite -2 glucose. In subsite -3, the main interaction is stacking against Trp-58, which also forms a hydrogen bond to the glucosidic bond oxygen between the -3 and -4 sugars (Fig. 2-4E). Together, these interactions in subsite -3 are responsible for 17 kJ/mol of transition-state stabilization (*vide supra*).

2.3.3 Downstream saccharification of mixed-linkage oligosaccharides produced by BoGH16$_{MLG}$

To elucidate the mechanism for the downstream conversion of the oligosaccharide products of BoGH16$_{MLG}$ to glucose for primary metabolism, we examined the biochemistry of the two predicted exo-β-glucosidases, BoGH3$_{MLG}$ and BACOVA_02738(GH3) associated with the MLGUL.

2.3.3.1 Cellular localization of BoGH3$_{MLG}$ and the BACOVA_02738(GH3) gene product

BoGH3$_{MLG}$ and BACOVA_02738(GH3) were unambiguously predicted by SignalP 4.0 (189) to contain a secretion signal peptide, while LipoP 1.0 (174) additionally indicated a Type II lipoprotein signal sequence (175) in BoGH3$_{MLG}$ only. The same *B. ovatus Δtdk* cultures used for BoGH16$_{MLG}$ localization, containing glucose or bMLG as the sole carbon source, were probed using polyclonal antibodies independently raised against recombinant BoGH3$_{MLG}$ and the BACOVA_02738(GH3) gene product. Neither protein was detected on the cell surface by fluorescence microscopy, especially in the presence of bMLG which induces BoGH16$_{MLG}$ production (Fig. 2-2B, Fig. A-2A). BoGH3$_{MLG}$ induction by bMLG was confirmed by a Western blot of cellular fractions, which also confirmed that this enzyme is membrane anchored (Fig. 2C).

In contrast, the BACOVA_02738(GH3) gene product was detected to a higher degree in *B. ovatus* cells grown in minimal medium with glucose as a sole carbon source compared to cells induced with bMLG (Fig. A-2B). The lack of upregulation by bMLG is consistent with transcriptional analysis which showed a limited change in expression in bMLG vs. glucose (1.6-fold), which was two orders of magnitude lower than definitive MLGUL genes (Table A-2). The higher detection in uninduced cells is explained by the high basal expression of BACOVA_02738(GH3) (more than an order of magnitude greater than all MLGUL members; Table S1). The lack of detection in minimal medium containing bMLG is due to high amounts of induced MLGUL proteins diminishing the presence of the BACOVA_02738(GH3) gene product when normalized to total protein (Fig. A-2B).

2.3.3.2 Substrate product specificity of BoGH3$_{MLG}$ and BACOVA_02738(GH3)

Initial activity screening on chromogenic *p*NP glycosides (see Experimental Procedures) revealed that both recBoGH3$_{MLG}$ and recBACOVA_02738(GH3) are specific *exo*-β-glucosidases (activity on other *p*NP glycosides was undetectable at micromolar enzyme concentrations). However, recBACOVA_02738(GH3) is strikingly feeble compared to recBoGH3$_{MLG}$ on G-β-*p*NP (k_{cat}/K_m values of 0.084 mM^{-1} s^{-1} versus 20 mM^{-1} s^{-1}; Fig. A-5C, A-5D, Table 2-1). Further, measuring Michealis-Menten kinetic parameters on cello- and laminari-oligosaccharides was not feasible due to overall poor activity and low production yields (data not shown). These kinetic results corroborate the above comparative genetic and transcriptional analyses, collectively suggesting BACOVA_02738(GH3) is not integral to the MLGUL. Hence, this enzyme was not characterized further.

To investigate oligosaccharide substrate preference of the BoGH3$_{MLG}$, we conducted initial-rate kinetics analyses on a series of gluco-oligosaccharides of distinct linkage composition and degrees of polymerization (d.p.). The non-reducing-end glucose was hydrolyzed from all-β(1,4)-linked cello-oligosaccharides (d.p. 2-6), all-β(1,3)-linked laminari-oligosaccharides (d.p. 2-5), and mixed-linkage β(1,3)/β(1,4)-gluco-oligosaccharides (d.p. 3-4, 5 examples) with comparable efficiencies, according to classic Michaelis-Menten saturation kinetics (Fig. A-5E, A-5F; Table 2-1). In this series, only cellobiose (G4G) was poorly hydrolyzed by BoGH3$_{MLG}$ *vis-à-vis* the β(1,3)-linked congener laminaribiose (G3G) and all other gluco-oligosaccharides (*e.g.*, G4G has a k_{cat}/K_m value 20-fold lower than G3G, Table 1). The β(1,6)-linked diglucoside gentiobiose (G6G) was also a very poor substrate, with a k_{cat}/K_m value 260-fold lower than that of G3G. Gluco-oligosaccharides with a β (1,3)-linked glucosyl unit at the non-reducing end all have slightly higher k_{cat} values compared to those with a β(1,4)-linkage in this position, which typically contributes to higher k_{cat}/K_m values for the former, when substrates of equal d.p. are compared. However, the magnitude of these differences, which are often less than 2-fold, indicate that BoGH3$_{MLG}$ is essentially agnostic to β(1,3) versus β(1,4) linkages. These results also suggest that in addition to a single negative subsite (-1) commensurate with its *exo*-activity, BoGH3$_{MLG}$ has only two positive subsites that contribute to catalysis: in each gluco-oligosaccharide series, tetrasaccharides and larger are hydrolyzed with identical k_{cat}/K_m values to the corresponding trisaccharides.

Product analysis following extended incubation of BoGH3$_{MLG}$ with G4G4G3G and G4G3G demonstrated that BoGH3$_{MLG}$ completely degrades the BoGH16$_{MLG}$ limit-digest products to glucose. HPLC also revealed that laminaribiose (G3G) is the only new intermediate formed during the course of hydrolysis (Fig. 2-3B). This demonstrates that BoGH3$_{MLG}$ sequentially hydrolyzes one glucose unit at a time from the non-reducing end of MLG oligosaccharides, *viz.*: G4G4G3G $\rightarrow$ G + G4G3G (also present in the starting mixture) $\rightarrow$ G+ G3G $\rightarrow$ G + G. Notably, the individual k_{cat} and K_m values for each step are nearly identical (Table 2-1).

Table 2-1. Summary of kinetic analyses of BoGH3MLG and BACOVA_02738(GH3) exo-glucosidases.

Enzyme	Substrate	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (s^{-1} mM^{-1})	Assay
$BoGH3_{MLG}$	β-Glc-*p*NP	59.5 ± 1.46	2.95 ± 0.14	20.2	*p*NP
	gentiobiose (G6G)	ND	ND	0.0571	HK/G6PDH
	cellobiose	5.52 ± 0.19	7.47 ± 0.48	0.739	HK/G6PDH
	cellotriose	22.1 ± 0.3	0.859 ± 0.033	25.7	HK/G6PDH
	cellotetraose	17.3 ± 0.5	0.687 ± 0.044	25.2	HK/G6PDH
	cellopentaose	19.4 ± 0.8	0.777 ± 0.060	25.0	HK/G6PDH
	cellohexaose	17.4 ± 0.4	0.747 ± 0.041	23.3	HK/G6PDH
	laminaribiose	28.0 ± 1.1	1.90 ± 0.12	14.7	HK/G6PDH
	laminaritriose	34.2 ± 1.0	0.911 ± 0.052	37.5	HK/G6PDH
	laminaritetraose	31.3 ± 2.3	0.898 ± 0.135	34.9	HK/G6PDH
	laminaripentaose	39.5 ± 3.4	1.27 ± 0.20	31.1	HK/G6PDH
	$MLGO_3$ A (G3G4G)	61.6 ± 1.6	0.997 ± 0.040	61.8	HK/G6PDH
	$MLGO_3$ B (G4G3G)	24.7 ± 1.3	0.521 ± 0.064	47.4	HK/G6PDH
	$MLGO_4$ A (G3G4G4G)	55.7 ± 2.7	1.33 ± 0.12	41.9	HK/G6PDH
	$MLGO_4$ B (G4G4G3G)	30.8 ± 2.0	0.736 ± 0.106	41.8	HK/G6PDH
	$MLGO_4$ C (G4G3G4G)	15.7 ± 0.3	0.601 ± 0.031	26.1	HK/G6PDH
BACOVA_02738 (GH3)	β-Glc-*p*NP	0.212 ± 0.004	2.53 ± 0.13	0.0838	*p*NP

ND: not determined (in cases where Michealis-Menten curve fitting was not feasible, individual k_{cat} and K_m values are not reported and k_{cat}/K_m value was determined from linear curve fit to initial-rate data in the $[S] << K_{m(apparent)}$ range). Data are represented as mean ± standard deviation. Highlighted in bold are the biologically relevant substrates that $BoGH3_{MLG}$ encounters in the periplasmic space.

2.3.3.3 $BoGH3_{MLG}$ and BACOVA_02738(GH3) primary structures

Despite extensive efforts, we were unable to crystallize the key β-glucosidase $BoGH3A_{MLG}$ for experimental tertiary structure determination. However, $BoGH3_{MLG}$ has 63% sequence identity to a *B. ovatus* β-glucosidase (BoGH3B, PDB code 5JP0 (135)), from the xyloglucan utilization locus (Fig. A-6A) and, as such, was amenable to tertiary structure homology modelling. Phyre2 (190) utilized PDB code 5JP0 as the sole template and 728 out of 742 residues (98% of the sequence, excluding the signal peptide) were modelled with 100 % confidence. The model suggests that $BoGH3_{MLG}$ possesses a homologous three-domain architecture with the active site being formed at the interface of the $(\alpha/\beta)_8$ TIM barrel and α/β sandwich domains (Fig. A-6B). The predicted catalytic nucleophile and acid/base residues are Asp-309 and Glu-453, respectively,

based on primary and tertiary alignment (Fig. A-6A, A-6C). Two tryptophan residues were modelled on opposite sides of the entrance to the active site pocket (Fig. A-6D), forming a narrow "coin slot", which may effect a preference towards β(1,3)- and β(1,4)-linked glucans and poor activity against β(1,6)-linked gentiobiose (Table 2-1). In constrast, enzymes that lack a homologous Trp-453 have a more open entrance to the active site and show broad activity against β(1,2)- and β(1,6)-linked glucans in addition to β(1,3)- and β(1,4)-linked glucans (191,192).

In comparison, BACOVA_02738(GH3) possess catalytic residues homologous to $BoGH3_{MLG}$ and similar GH3 members, despite having only 31% sequence identity to $BoGH3_{MLG}$ (Fig. A-6A). The most similar characterized GH3 member to BACOVA_02738(GH3) among ca. 300 members identified in the CAZy is a *Chrysosporium lucknowense* β-glucosidase with 39% sequence identity (193).

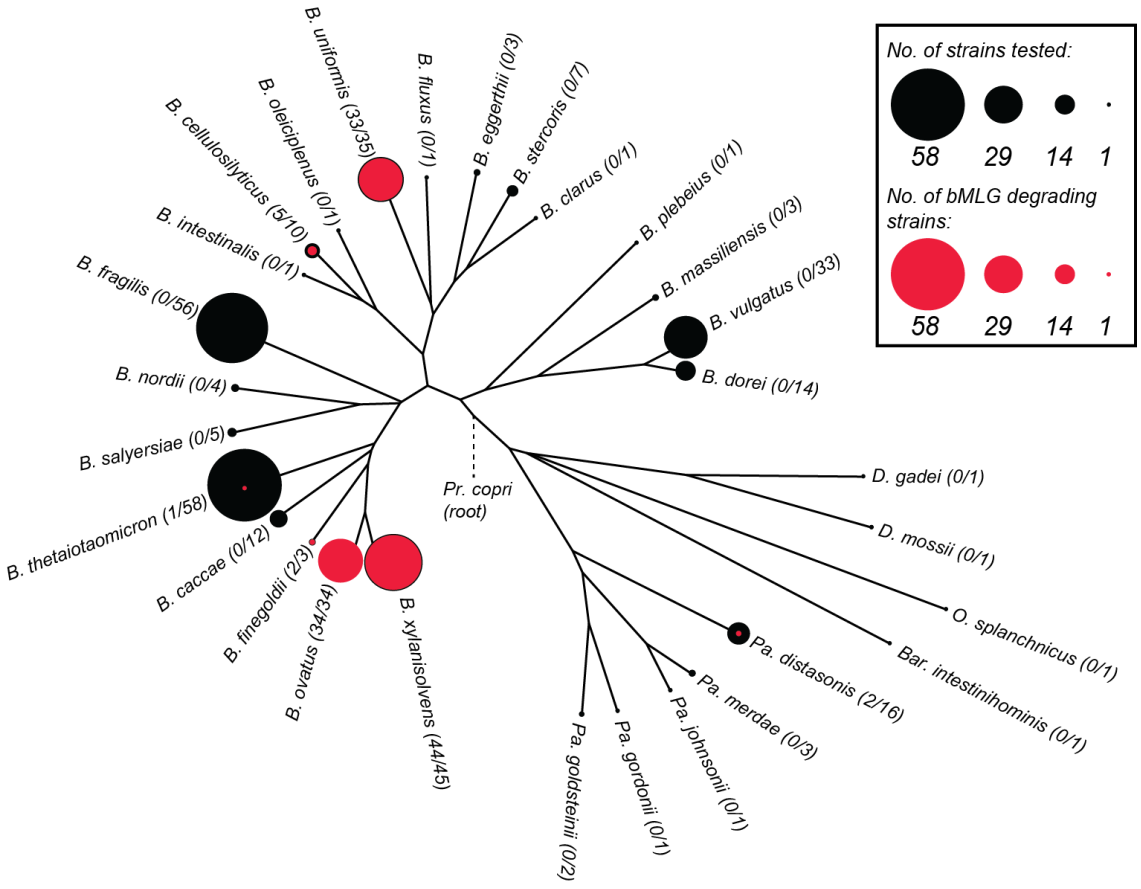

Figure 2-5. Penetrance map of MLG utilization ability across diverse human gut Bacteroidetes.

The phylogenetic tree was constructed from fully sequenced strains of the species shown. The number of strains of each species tested for growth is depicted to scale as a black circle at each leaf. The number of those strains that grew on bMLG as a sole carbon source is shown to scale in red within the black circle.

2.3.4 Syntenic MLGUL are molecular markers of MLG utilization across the Bacteroidetes

Refined functional characterization of the catalytic specificity of GH components significantly increases confidence in the use of individual PULs as genetic markers of complex carbohydrate metabolism among Bacteroidetes (101,110,138,194). The MLGUL characterized here represents the only PUL in *B. ovatus* that confers growth on MLG from cereals. To understand the wider distribution of MLG metabolic capacity among symbiotic Bacteroidetes in the human gut, we correlated the presence of a syntenic MLGUL with growth on bMLG for 354 individual Bacteroidetes strains representing 29 different species.

A total of 121 strains across just 7 of the species were able to grow on bMLG (Fig. 2-5). In particular, 33 of 33 *B. ovatus* strains (including the type strain ATCC 8483) grew well on bMLG, as well as 44 of 45 strains of the closely related *B. xylanisolvens*. The majority of *B. uniformis* strains tested (33 out of 35) were also competent bMLG utilizers. The limited penetrance of the MLGUL across the genus clearly demonstrates nutrient-niche specialization among individual *Bacteroides* species.

Comparative analysis of available genomic sequences revealed that strains able to grow on bMLG as the sole carbon source harbor a syntenic MLGUL (Fig. 2-1B). Previous transcriptional analysis demonstrated that the syntenic MLGUL found in *B. cellulosilyticus* is also activated during growth on bMLG (123). Concordance between the presence of a syntenic MLGUL and the ability to utilize MLG is further highlighted by the lack of a MLGUL in the *B. uniformis* ATCC 8492, one of only two strains of *B. uniformis* that could not grow on bMLG. Based on this analysis, we can also predict MLG utilization ability in two sequenced species of *Prevotella*, *Pr. copri* DSM 18205 and *Pr. multiformis* DSM16608, important members of the Bacteroidetes from the human gut and oral cavity, respectively (Fig. 2-1B).

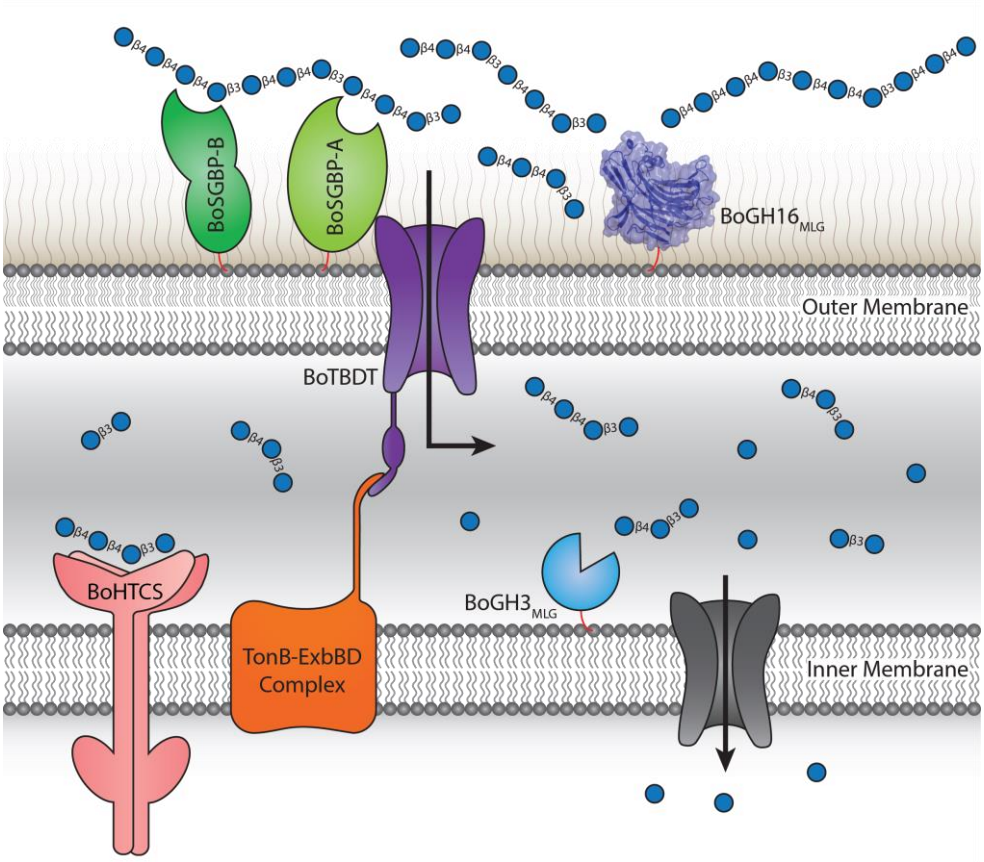

Figure 2-6. Model of mixed-linkage β-glucan saccharification by the concerted action of the MLGUL machinery.

Gene products are colored analogously to the gene locus in Fig. 2-1. The cell surface localized endo-MLGase BoGH16$_{MLG}$ cleave large mixed-linkage β-glucan polysaccharides into shorter fragments which are imported into the periplasm via the TonB dependent transporter, BoTBDT. This glycan capture and transport process at the cell surface is aided by the two surface glycan binding proteins BoSGBP-A and BoSGBP-B. The smaller mixed-linkage β-glucan fragments in the periplasm bind the sensor domain of the hybrid two-component sensor BoHTCS to induce upregulation of the system. Periplasmic exo-β-glucosidases BoGH3$_{MLG}$ and BACOVA_02738(GH3) act from the non-reducing ends to liberate individual glucose monomers which are imported into the cell and metabolized.

2.4 Discussion

2.4.1 A molecular model for MLG utilization by *B. ovatus*

Our current suite of data suggests a model by which the MLGUL gene products work in concert to enable the utilization of MLG (Fig. 2-6), analogous to that of other PUL-encoded systems (105). Thus, BoGH16$_{MLG}$ is anchored to the outer membrane where it plays a leading role in fragmenting large MLG polysaccharide chains (typical d.p. 700 – 5000, depending upon the plant species of origin (145,195)) into oligosaccharides that can be imported into the periplasm via

the TBDT. Notably, the specific limit-digest products of $BoGH16_{MLG}$ *endo*-hydrolysis identified here, *viz.* the trisaccharide G4G3G and the tetrasaccharide G4G4G3G (Fig. 2-3B), have been shown previously to bind the periplasmic sensor domain of the HTCS encoded by BACOVA_02740 (K_D 300 μM and 400 μM, respectively), while the intact MLG polysaccharide does not (109). Monomeric glucose, all-β(1,4)-linked cello-oligosaccharides, and all-β(1,3)-linked laminari-oligosaccharides are also not bound by the HTCS (109), indicating that the unique linkages present in MLG are central to inducing the MLGUL system. Thus, there is an essential, yet distant, interplay between the outer-membrane localized MLGase and the HTCS in specific nutrient sensing.

It is therefore likely that the $BoGH16_{MLG}$ limit-digest products, or minimal repeats of these structures $[(G4G4G3G)_m(G4G3G)_n]$, comprise the main products transported through the TBDT *in vivo*. Recent studies on inulin (β(2,1)-fructan) utilization suggest that some TBDTs are able to import longer polysaccharide chains (e.g. d.p. >20) (196). Regardless of length, the efficient *exo*-hydrolytic activity of $BoGH3_{MLG}$ in the periplasm is sufficient to completely saccharify all imported oligosaccharides to glucose (Fig. 2-3B), to feed primary metabolism in the cytosol. In this process, the trisaccharide substrate of the HTCS, G4G3G, will always be generated regardless of the imported saccharide chain length, ensuring continual production of the MLGUL up-regulation signal until substrate is exhausted. Interestingly, $BoGH3_{MLG}$ will never encounter cellobiose (G4G), towards which it has relatively weak activity (Fig. A-4F; Table 2-1), in this process; the final step of saccharification of MLGOs is the hydrolysis of the competent substrate laminaribiose (G3G).

2.4.2 Structural enzymology reveals complex trajectories for the evolution of MLG activity in GH16

Previous phylogenetic analyses of GH16 have suggested that the evolution of the active-site β-bulge motif EXDXXE, which is widespread among Clan GH-B (comprising GH16 and GH7), to a regular β-strand motif EXDXE is a defining feature that delineates *endo*-β(1,3)-glucanases (laminarinases, EC 3.2.1.39 and EC 3.2.1.6) from mixed-linkage *endo*-β(1,3)/β(1,4)-glucanases (licheninases, EC 3.2.1.73), respectively (185,186). In this context, the observation that $BoGH16_{MLG}$ is highly specific for MLG, despite having a β-bulge motif, was surprising.

Using the CAZy Database as a starting point (http://www.cazy.org/GH16_characterized.html) together with mining of the primary literature,

we generated a contemporary maximum-likelihood phylogeny of all biochemically characterized GH16 members (Fig. A-7). This analysis indicates that although canonical, normal β-strand MLGases do form a monophyletic group as previously observed, MLGase activity is in fact widespread among the historical "laminarinase" group, in which $BoGH16_{MLG}$ is itself positioned. Despite currently limited and disparate kinetic data for individual enzymes, it also appears that it is not possible to define further substrate-specific clades within this group based on molecular phylogeny alone, in light of weak bootstrap support. This precludes defining any single evolutionary event giving rise to unique trajectories for the further diversification of extant laminarinases and MLGases in this clade. Instead, it appears that diverse, subtle mutations have allowed the independent evolution of predominant laminarinase or MLGase activity numerous times across a flat evolutionary landscape. As such, we suggest that this GH16 subgroup should be more generally referred to as the "laminarinase/MLGase group" going forward.

Detailed tertiary structural comparison of 10 β-bulge-containing members of this laminarinase/MLGase group revealed, however, that predominant laminarinases harbor a significantly more protruding loop (which is often, but not always, longer) connecting strands β2 and β3 than predominant MLGases (Fig. A-8A, A-8B). Structural superposition with the $BoGH16_{MLG}$:G4G4G3G complex indicates that this loop in predominant laminarinases would clash with MLG in the negative subsites, instead favoring binding of an all-β(1,3)-glucan that curves away from this loop. Such curvature is exemplified by the $ZgLamC_{GH16\text{-}E142S}$:thio-β-1,3-trisaccharide structure (Fig. A-8A, PDB code 4CTE) (155). Indeed, Ilari *et al.* observed that shortening the homologous loop in LamA from the archeon *Pyrococcus furiosus* (Fig. A-8A, PDB code 2VY0) by 4 amino acids increased the activity towards MLG by 10-fold (197). Likewise, BglF from *Nocardiopsis* sp. F96 (Fig. A-8B, PDB code 2HYK) and LamR from *Rhodothermus marinus* (Fig. A-8B, PDB code 3ILN), which have a 3.3- and 8.5-fold greater specificity constant and specific activity, respectively, toward MLG than laminarin, also have a smaller loop, similar to $BoGH16_{MLG}$, in this position. The canonical, regular-β-strand MLGase from *Paenibacillus macerans* (Fig. A-8C, PDB code 1MAC) and *Bacillus licheniformis* (Fig. A-8C, PDB code 1GBG), similarly have a small loop at this position.

Taken together, these analyses reveal a complex evolutionary landscape that computational phylogenetic analysis fails to resolve. Despite using a manually curated, structure-based sequence alignment as input, the maximum-likelihood numerical approach did not delineate the members of

the laminarinase/MLGase group on the basis of the distinct active-site loop differences observed in tertiary structures (Fig. A-8). Instead, the phylogeny was likely obfuscated by diverse, random variations in amino acid composition across the entire β-sandwich domain, which clearly limits large-scale, unsupervised phylogenetic analysis of these GH16 members. Moreover, analysis of both MLG and laminarin specificity (as a minimum) for individual members of this group, in light of their tertiary structures, is essential to avoid potential mis-annotation of these enzymes.

2.4.3 Mining metagenomic data reveals the ubiquity of MLG utilization in the human gut and beyond

Using syntenic MLGULs as genetic markers, we surveyed the publicly available metagenome data of 426 adults to understand the capacity of human populations to derive nutrition from cereal MLGs. We were able to distinguish the species of origin based on nucleotide sequence except for MLGULs from *B. ovatus* and *B. xylanisolvens*, which were strikingly similar at 97% nucleotide identity. The *B. ovatus*/*B. xylanisolvens* and *B. uniformis* MLGULs are the most prevalent; both are observed in about 70% of the total human cohort (Fig. 2-7). The *Pr. copri* MLGUL is more often the sole MLGUL of an individual than that of *B. cellulosilyticus* when only one is present (Fig. 2-7, cyan lines), despite the latter being more frequent in total. Overall, 92.5% of the subjects harbor at least one of the five different MLGULs identified in this study, irrespective of nationality or whether they have been diagnosed with IBD. MLGULs are ubiquitously detectable despite variability in sampling depth across different metagenomics sequencing projects (Fig. 2-7). The prevalence of MLGULs across different nationalities is consistent with MLG from cereal grains being a ubiquitous component of the human diet. Indeed, the importance of MLG utilization is underscored by the upregulation of the MLGUL in the ceca of mice fed a complex plant cell wall diet (109). Similar widespread global distribution in human populations has been observed for xyloglucan utilization loci (138). These observations are sharply contrasted by the PUL that mediates utilization of the red algal polysaccharide porphyran, which is essentially confined to subjects from Japan, where seaweed is a common part of the diet (138,198). Interestingly, we were unable to detect MLGULs in four unweaned infants sampled in the Japanese metagenome project (data not included in our analysis of adult metagenomes). This is consistent with a dearth of Bacteroidetes in infants who receive the bulk of their nutrition from milk and are not yet consuming plant polysaccharides (169)

Moving beyond the human microbiota, we can likewise predict MLG utilization ability in *Dysgonomonas gadei* and *Pr. oryzae* (formerly *Xylanibacter oryzae*) based on the presence of a syntenic MLGUL. These species are commonly found in the termite hindgut and decomposing rice straw, respectively. This provides direct evidence that an analogous MLG utilization system is employed by Bacteroidetes operating in environments beyond the human gut.

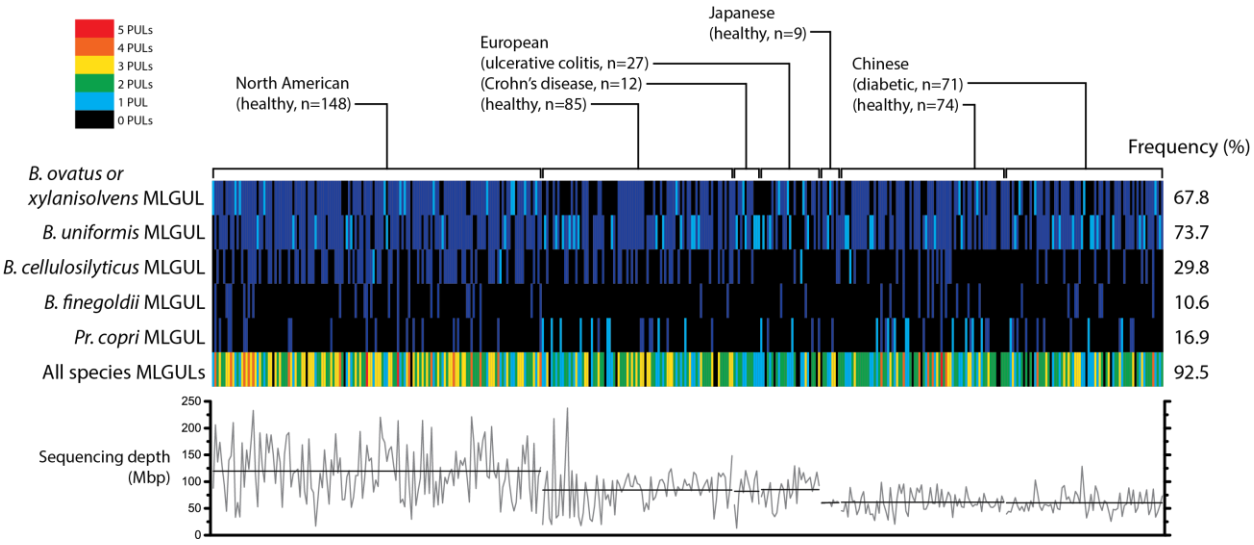

Figure 2-7. Bacteroidetes MLGULs from a survey of 426 adult human gut metagenomes.

Vertical lines represent the presence (cyan when unique, blue when one of multiple) or absence (black) of a corresponding species-related MLGUL in a single individual. The total number of MLGULs observed in an individual is shown in the bottom row, colored according to the legend in the top left corner. The frequency of MLGUL occurrence across all 426 individuals is shown on the right. Variation in sequencing depth in megabase pair is illustrated in the graph below: grey lines show the depth for individual subjects and black lines show the average depth of each metagenomics project.

2.5 Conclusion

Complex carbohydrates that promote the growth of beneficial microbes in our distal large intestine are a cornerstone of a healthy diet. MLGs in particular have long been known to impart healthful effects (144), yet its mechanism of utilization for fermentation by gut microbes was unknown. Our work here sheds light on the fine-tuned mechanism that *B. ovatus* and other Bacteroidetes has evolved to efficiently utilize MLGs in the highly competitive environment of the human gut microbiota. The finding that a majority of humans possess microbes that can utilize this ubiquitous cereal polysaccharide highlights the relevance of potential therapeutic interventions

based on MLG utilization to the general population. The present study also sets the stage for future work to understand the quantitative contributions of individual members of the microbiota and their cognate enzymes to MLG utilization in the human gut (199,200).

Chapter 3: Surface glycan binding proteins are essential for cereal beta-glucan utilization by the human gut symbiont *Bacteroides ovatus*

3.1 Introduction

Trillions of microbial (mainly bacterial) cells make the human gut microbiota one of the most complex and dynamic ecosystems on the planet (1). This microbial community has far-reaching influences on diverse aspects of human physiology and health (2) including, but not limited to, links to obesity (8,201), asthma (202), allergies (6), and cancer (7). The abundance of individual members of the human gut microbiota is driven by our diet, especially the influx of complex polysaccharides into the large intestine (10,13,125,203,204). Indeed, regular ingestion of plant polysaccharides is integral to maintaining a healthy balance of microbes in our lower gastrointestinal tract (9,133,205).

Strikingly, the human genome is remarkably bereft of genes encoding carbohydrate-active enzymes (CAZymes) targeting dietary glycans. As such, we are critically dependent on members of the human gut microbiota to metabolize otherwise indigestible "dietary fibers" (15,205). Of the dominant bacterial phyla that comprise the microbiota, the Bacteroidetes in particular possess an enormous arsenal of CAZyme genes (206), which are co-localized along with genes encoding cognate surface glycan binding proteins (SGBPs), TonB-dependent transporters (TBDTs), and transcriptional regulators into Polysaccharide Utilization Loci (PULs) (105). The gene products of a single PUL work in concert to sense, bind, cleave, and import a particular complex polysaccharide. Reflecting the large natural diversity of complex glycans encountered by the human gut microbiota, Bacteroidetes possess a plethora of PULs (e.g., 88 and 112 PULs in *Bacteroides thetaiotaomicron* and *B. ovatus*, respectively (109)), each of which is specifically upregulated by a target polysaccharide (109). The abundance and broad distribution of PULs underscores their importance to human gut microbiota metabolism and, consequently, human nutrition and health.

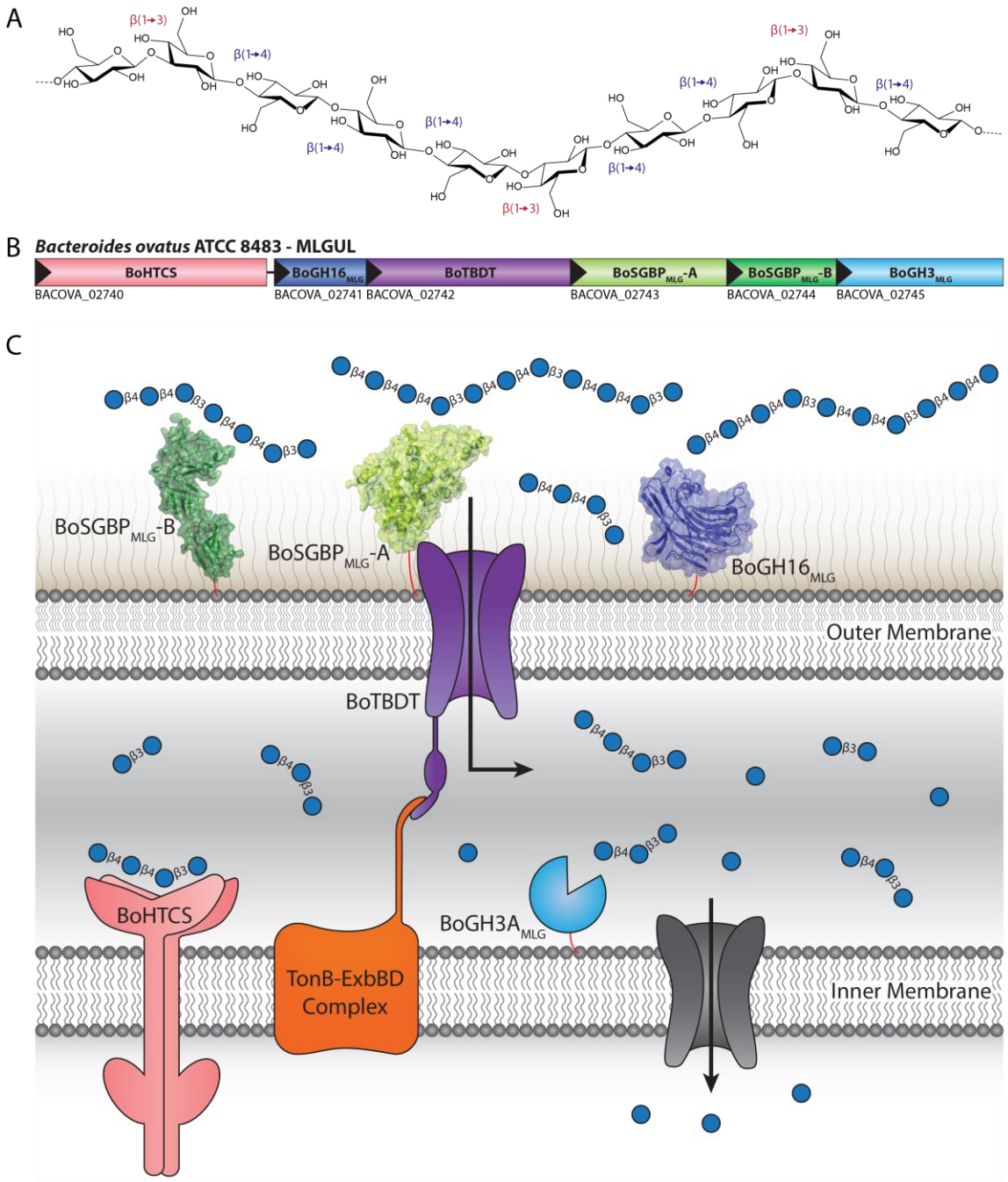

Figure 3-1. Mixed-linkage glucan utilization locus (MLGUL).

(A) chemical structure of MLG targeted by this PUL, consisting of a linear glucan chain of β(1,4)-linked cellotriosyl and cellotetraosyl units separated by single β(1,3) bonds. (B) Genetic organization of the *B. ovatus* MLGUL with the locus tag shown below each gene. (C) Updated MLGUL model at the cell envelope, based on Fig. 2-6, including SGBP structures. Protein products are colored analogously to the gene locus. HTCS: hybrid two-component system sensor/regulator; GH: glycoside hydrolase, with family number indicated; TBDT: TonB-depdendent transporter (SusC homolog); SGBP: cell surface glycan-binding protein (SGBP-A is a SusD homolg, SGBP-B is highly sequence-divergent); BACOVA_*n*: *Bacteroides ovatus* ATCC 8483 gene locus tag.

Mixed-linkage β-glucans (β(1,3)/β(1,4)-glucans, MLGs) are abundant in cereal grains such as oats and barley (Fig. 3-1A), and comprise a key group of human dietary glycans with recognized healthful effects. For example, the benefits of cereal-derived MLGs in ameliorating hypertension, diabetes mellitus, cardiovascular diseases, and cholesterol levels have been reported (141,205,207,208). Although the mechanisms behind these health benefits are not fully understood, the prebiotic activity of MLGs (209), a direct result of fermentability by the human gut microbiota (210), is likely to be a major factor.

We recently identified a MLG utilization locus (MLGUL) in *B. ovatus* (Fig. 3-1B), syntenic homologs of which enable select *Bacteroides* species in the human gut microbiota to metabolize this cereal glycan, and biochemically characterized its cognate glycoside hydrolases (GHs). Structural enzymology detailed the high specificity of the outer-membrane GH16 *endo*-β-glucanase for MLG and supported a concerted model in which polysaccharide cleavage at the cell surface, oligosaccharide transport via a SusC-like TBDT, and periplasmic hydrolysis by GH3 *exo*-β-glucosidase enable complete MLG saccharification to glucose (Fig. 3-1C). As part of this study, we also demonstrated that MLGUL are essentially ubiquitous in human gut metagenomes (211).

This model also predicts the involvement of two cell-surface glycan-binding proteins (SGBPs) encoded by the MLGUL (Fig. 3-1B) in the initial capture of the polysaccharide at the cell surface as a prelude to backbone hydrolysis, and/or facilitating product transport through the TBDT (Fig. 3-1C). Indeed, previous studies have outlined these roles for SGBPs in the archetypal starch utilization system (Sus) and other PULs (95,99,103,107,110,146), and SusC/SusD (TBDT/SGBP-A) homologs are known to form an intimate structural association in the membrane (97). However, the molecular structures, substrate specificities, and individual contributions to MLG utilization are currently unknown for $BoSGBP_{MLG}$-A (a SusD homolog) and $BoSGBP_{MLG}$-B (a sequence divergent "SusE-positioned" gene product) from the MLGUL. Here we present the comprehensive biochemical, crystallographic, and reverse-genetic characterization of these two SGBPs to reveal the distinct roles the two SGBPs play in MLG metabolism by *B. ovatus* in the context of the human gut microbiota.

3.2 Materials and Methods

3.2.1 Substrates and polysaccharides

Polysaccharides. Barley beta-glucan (high viscosity), yeast beta-glucan, curdlan, tamarind xyloglucan, konjac glucomannan, carob galactomannan, wheat arabinoxylan, beechwood xylan

were purchased from Megazyme International (Bray, Ireland). Laminarin (from *Laminaria digitata*) was purchased from Sigma Aldrich (St. Louis, MO, USA). Carboxymethyl cellulose was purchased from Acros Organics (Morris Plains, NJ, USA). Hydroxyethyl cellulose was purchased from Amresco (Solon, OH, USA). Xanthan gum was purchased from Spectrum (New Brunswick, NJ, USA). Ulvan (from Ulva sp.) was purchased from Elicityl (Crolles, France).

Oligosaccharides. Cellobiose (G4G) was purchased from Acros Organics. Cellotriose (G4G4G), cellotetraose (G4G4G4G), cellopentaose (G4G4G4G4G), cellohexaose (G4G4G4G4G4G), laminaribiose (G3G), laminaritriose (G3G3G), laminaritetraose (G3G3G3G), laminaripentaose (G3G3G3G3G), mixed-linkage glucotriose A (G3G4G), mixed-linkage glucotriose B (G4G3G), mixed-linkage glucotetraose A (G3G4G4G), mixed-linkage glucotetraose B (G4G4G3G), mixed-linkage glucotetraose C (G4G3G4G) were purchased from Megazyme. Gentiobiose (G6G) was purchased from Carbosynth (Compton, UK). MLG partial digest mixture, mixed-linkage hexasaccharide (MLG6) and mixed-linkage heptasaccharide (MLG7) were produced in-house as described by McGregor, et al. (212) using $BoGH16_{MLG}$ (211) in 50 mM sodium phosphate pH 7.0.

3.2.2 Cloning, expression, and purification of recombinant proteins

Gene sequences were obtained from *B. ovatus* ATCC 8483 genome available on the Integrated Microbial Genomes database from the Joint Genome Institute. PCR primers were synthesized by Integrated DNA technologies.

Cloning. Open reading frames encoding BACOVA_02743, and BACOVA_02744 were amplified by PCR using Q5 high fidelity polymerase (NEB) with appropriate primers (Table B-1) and genomic *B. ovatus* DNA as template. All primers were designed to amplify constructs truncated to exclude predicted signal peptides (prediction by SignalP 4.1 (189,213)) and N-terminal lipidation cysteine residues (prediction by LipoP 1.0 (174)). NdeI and XhoI restriction sites were included in the forward and reverse primers for subsequent digestion (all restriction enzymes from NEB) and ligation (T4 ligase from Thermo Scientific) into the pET28 vector. Both constructs were designed to harbor an N-terminal his_6-tag fusion in the translated recombinant peptide. The gene encoding sfGFP was fused to their N-termini by restriction enzyme-based cloning using a BamHI site between the sfGFP and BACOVA_02743 or BACOVA_02744, and NheI and XhoI to the corresponding sites on the pET28 vector. Site-directed mutant constructs were generated using the QuikChange II site-directed mutagenesis kit (Agilent) according to the

manufacturer's instructions. Oligonucleotides used in this study are listed in Table S1. Success of clones and mutants were verified by Sanger sequencing (Genewiz).

Expression. Plasmids harboring the gene of interest were transformed into chemically competent *E. coli* BL21 (DE3) and cultured in lysogeny broth (LB) containing 50 μg/mL kanamycin. Cells were grown on a large scale at 37 °C until mid-logarithmic growth phase was reached ($O.D._{600}$ = 0.4-0.6) at which point protein expression was induced by addition of isopropyl β-D-thioglactopyranoside (IPTG) to a final concentration of 0.5 mM and temperature was lowered to 16 °C. Induction of recombinant protein production continued overnight after which the cells were collected by centrifugation at 4000 g for 20 minutes.

Purification. The harvested cell pellet was resuspended in binding buffer (20 mM sodium phosphate pH 7.4, 500 mM sodium chloride, 20mM imidazole) and lysed using a Sonic Dismembrator F550 Ultrasonic Homogenizer (Fisher Scientific). Cell debris was pelleted by centrifugation at 15,000 rpm for 45 minutes and the supernatant was loaded onto a 2 mL HisTrap IMAC FF nickel-nitrilotriacetic acid column (GE Healthcare), a nickel based matrix, using a BioLogic FPLC system (BioRad). After washing with 10 column volumes of binding buffer, his_6-tagged protein was eluted using a linear gradient of 0 - 100% elution buffer (20 mM sodium phosphate pH 7.4, 500 mM sodium chloride, 500 mM imidazole) over 10 column volumes. Fractions were monitored by A_{280} and eluted protein fractions were pooled and buffer exchanged into 50 mM sodium phosphate pH 7.0 using Vivaspin centrifugal filters (GE Healthcare). After concentrating, aliquots were flash frozen in liquid nitrogen and stored at -80°C. Protein purity was determined by SDS-PAGE analysis and mass was confirmed by intact protein mass spectrometry on a Waters Xevo Q-TOF with nanoACQUITY UPLC system (data not shown), as described previously (148). Protein concentrations were determined by spectrophotometry on an Epoch Microplate Spectrophotometer (BioTek) using the following molar extinction coefficients: 106690 $M^{-1}cm^{-1}$ for BACOVA_02743, and 51340 $M^{-1}cm^{-1}$ for BACOVA_02744. Typical yields were around 50 mg for $BoSGBP_{MLG}$-A and 40 mg for $BoSGBP_{MLG}$-B from 1 L of lysogeny broth (LB) culture.

Selenomethionine Protein Expression and Purification. For selenomethionine-substituted $BoSGBP_{MLG}$-B the pET28-$BoSGBP_{MLG}$-B plasmid was transformed into Rosetta (DE3) pLysS *E. coli*, and plated onto LB supplemented with kanamycin (50 μg/mL) and chloramphenicol (20 μg/mL). After 16 hours of growth at 37 °C, colonies were harvested from the plates and used to

inoculate 100 mL of M9 minimal media supplemented with kanamycin (30 μg/mL) and chloramphenicol (20 μg/mL) and grown at 37 °C for 16 hours. This overnight culture was used to inoculate a 2 L baffled flask containing 1 L of Molecular Dimensions SelenoMet Premade Medium supplemented with 50 mL of the recommended sterile nutrient mix, chloramphenicol, and kanamycin. Cultures were grown at 37 °C to an $O.D._{600} \approx 0.45$ before adjusting the temperature to 20 °C, and supplementing each flask with 100 mg each of L-lysine, L-threonine, L-phenylalanine, and 50 mg each of L-leucine, L-isoleucine, L-valine, and L-selenomethionine (214). After 20 additional minutes of growth, the cells were induced with 0.5 mM IPTG, and cultures were grown for an additional 48 hours. For the purification of selenomethionine-substituted protein, cells were thawed and lysed via sonication in His-Buffer (25 mM NaH_2PO_4, 500 mM NaCl, 20 mM imidazole pH 7.5) and purified via immobilized Nickel affinity chromatography (His-Trap, GE Healthcare) using a gradient of 20-300 mM imidazole, according to the manufacturer's instructions. The fractions were inspected for purity via SDS-PAGE, then pooled and dialyzed against 20 mM HEPES, 100 mM NaCl (pH 7.0) and concentrated using Vivaspin 15 (10,000 MWCO) centrifugal concentrators (Vivaproducts, Inc.).

3.2.3 Affinity gel electrophoresis

Qualitative assessment of binding was carried out on the following soluble substrates: barley β-glucan, laminarin, yeast β-glucan, curdlan, xyloglucan, glucomannan, galactomannan, xylan, arabinoxylan, xanthan gum, dextran, carboxymethylcellulose, and hydroxyethylcellulose. Native polyacrylamide gels consisting of 10% (w/v) acrylamide in 40 mM Tris-HCl at pH 8.8 were prepared. Final concentration of 0.05 % or 0.1 %(w/v) substrate (or water for control) was added to the gel solution prior to polymerization. 5 μg of $BoSGBP_{MLG}$-A and $BoSGBP_{MLG}$-B, along with bovine serum albumin (BSA) as a non-interacting negative control, were loaded on the gels and subjected to electrophoresis under non-denaturing conditions at 100 V for 3 hours at room temperature. Proteins were visualized by staining with Coomassie Brilliant Blue.

3.2.4 Isothermal titration calorimetry

All isothermal titration calorimetry (ITC) experiments were performed using a MicroCal VP-ITC titration calorimeter calibrated to 25 °C. All titrations were performed in 50 mM sodium phosphate pH 7.0 with the exception of $BoSGBP_{MLG}$-B with bMLG which was performed in 10 mM HEPES pH 7.0. Proteins (20 – 100 μM) were placed in the sample cell and a first injection of 2 μL was performed followed by 24 subsequent injections of 10 μL of 2.5 to 4.0 mg/mL

polysaccharide or 1 to 2 mM oligosaccharides (see Fig. S3-5 for exact compositions of each protein-ligand pair). The solution was stirred at 280 rpm and the resulting heat of reaction was recorded. Data were analyzed using the OriginPro graphing software; data are averages and standard deviations of at least triplicate titrations. K_a values were calculated on a molar basis from MLG polysaccharide concentrations in g/L by assuming a hexasaccharide binding motif, based on crystal complex structures with MLGOs. K_a values were calculated on a molar basis from XyG polysaccharide concentrations in g/L by assuming a Glc_8-backbone (XyGO dimer) oligosaccharide binding motif.

3.2.5 Insoluble polysaccharide binding assay

Qualitative assessment of binding to insoluble polysaccharides, cellulose and mannan, were carried out by a pull-down assay. 10 mg of substrate and 100 μg of protein were mixed in 200 μL of 50 mM sodium phosphate pH 7.0. After incubation at 4 °C for 4 hours with end-over-end rotation, the samples were centrifuged at 16,000 G for 5 minutes. The supernatant containing unbound protein was collected, and the pellet was washed three times with 200 μL of 50 mM sodium phosphate pH 7.0. Bound protein was released from the substrate by resuspending the pellet in 200 μL of 1X SDS running buffer and heating to 80 °C for 10 minutes. Eluted bound protein was collected by centrifugation and both fractions were subjected to SDS-PAGE. BSA was used as a non-binding negative control.

Quantitative assessment of binding to cellulose was conducted by depletion isotherm. 10 mg/mL avicel, 5.9 to 140.4 μg of GFP_BoSBGP-A or 4.1 to 196.8 μg of GFP_BoSGBP$_{MLG}$- B, and 0.1 mg/mL BSA were mixed in 1 mL of 50 mM sodium phosphate pH 7.0. After incubation at 4 °C for 4 hours with end-over-end rotation, the samples were centrifuged at 16,000 G for 5 minutes and the supernatant was collected. Concentration of unbound GFP-fused protein was determined by fluorescence using an Infinite M1000 Pro multifunction plate reader (Tecan Ltd.) with an excitation filter of 485 nm and an emission filter of 510 nm. Fluorescence measurement of the same concentration range of GFP-fused protein in the absence of avicel was determined to construct a standard curve of total protein and unbound protein was subtracted to determine concentration of bound protein. OriginPro graphing software was used to fit the isotherms to the equation $[PC] = [FP][PC]_{max}/(K_d + [FP])$ where PC represents concentration of protein bound to Avicel and FP represents concentration of free protein in the supernatant.

3.2.6 X-ray crystallography

Selenomethionine-substituted BoSGBP$_{MLG}$-B protein crystals with cellohexaose were obtained directly from the SaltRx crystallization screen (Hampton Research) via hanging drop vapor diffusion at room temperature. The protein (21.1 mg/ml with 10mM cellohexaose) was mixed 1:1 with a crystallization solution comprised of 1.8 M ammonium phosphate monobasic and 0.1 M sodium acetate trihydrate pH 4.6. The selenomethionine-substituted BoSGBP$_{MLG}$-B crystals were flash frozen in a cryo-protectant comprised of mother liquor supplemented with 20% ethylene glycol. Anomalous diffraciton data were collected using X-rays with a wavelength of 0.979 Å at the Life Sciences Collaborative Access Team (LSCAT) beamline 21-ID-F of the Advanced Photon Source (APS) at Argonne National Laboratory. X-ray data were processed in HKL2000 (215) and scaled with Scalepack (215). Phasing via the anomalous selenium signal was performed in AutoSol (216) from the Phenix package (217), followed by partial refinement in phenix.refine (218) for use with the native protein data collected for BoSGBP$_{MLG}$-B.

Crystals of the BoSGBP$_{MLG}$-A with cellohexaose were obtained directly from Crystal Strategy Screen I (Molecular dimensions) via hanging drop vapor diffusion at room temperature. The protein (14.9 mg/mL with 10 mM cellohexaose) was mixed 1:1 with the crystallization solution comprised of 0.2 M $MgCl_2$, 0.1 M Tris acetate pH 8.5, 25% PEG 2000 monomethylether. Native crystals of BoSGBP$_{MLG}$-B were obtained directly from JCSG Plus screen (Molecular dimensions) via hanging drop vapor diffusion at room temperature. BoSGBP$_{MLG}$-B (20.1 mg/mL with 10 mM cellohexaose) was mixed 1:1 with the crystallization solution comprised of 0.8 M NaH_2PO_4, 0.8 M KH_2PO_4, 0.1M sodium HEPES pH 7.5. All crystals were flash frozen in a cryo-protectant comprised of mother liquor supplemented with 20% ethylene glycol and x-ray data were collected at the LSCAT beamline 21-ID-G of the APS at Argonne National Laboratory. X-ray data for BoSGBP$_{MLG}$-A were processed in HKL2000 and scaled with Scalepack, while the x-ray data for BoSGBP$_{MLG}$-B were processed and scaled in DIALS/Xia2 (219,220). Molecular replacement was performed in Phaser (157) from the Phenix package using the homologous SGBP-A structure 6DK2 as the search model for BoSGBP$_{MLG}$-A with cellohexaose and the partially refined selenomethionine-substituted model of BoSGBP$_{MLG}$-B was used with the native data collected for BoSGBP$_{MLG}$-B with cellohexaose. The native BoSGBP$_{MLG}$-A and BoSGBP$_{MLG}$-B structures with cellohexaose were refined in Refmac5 (161), with alternate rounds of manual model building in

Coot (162). Validation of the carbohydrates was performed with Privateer (164), and validation of the model fit with Phenix.validate (217).

Initial sitting drop crystal screens for unliganded $BoSGBP_{MLG}$-A at 21.7 mg/mL, $BoSGBP_{MLG}$-A at 20.3 mg/mL with 6.6 mM MLG7, and $BoSGBP_{MLG}$-B at 20.4 mg/mL with 6.3 mM MLG7 were set up using a Phoenix robot (Art Robbins). Crystals of unliganded $BoSGBP_{MLG}$-A were obtained at room temperature in the JSCS+ screen (Qiagen) condition H1: 0.2 M $MgCl_2$, 0.1 M Bis-Tris pH 5.5, 25% (w/v) PEG 3350. Crystals of $BoSGBP_{MLG}$-A co-crystallized with MLG7 were obtained at room temperature in the Classics II screen (Qiagen) condition G12: 0.2 M $MgCl_2$, 0.1 M HEPES pH7.5, 25% (w/v) PEG 3350. Crystals of $BoSGBP_{MLG}$-B co-crystallized with MLG7 were obtained at room temperature in the JCSG+ screen (Qiagen) condition A6: 0.2 M Li_2SO_4, 0.1 M phosphate-citrate pH4.2, 20% (w/v) PEG 1000. For all three initial hits, crystals were readily reproduced by hand in larger hanging drops by screening around the condition varying the buffer pH in one dimension and PEG concentration in the other. The crystals obtained from these optimizations were used for data collection by flash freezing in cryo-protectant comprised of mother liquor supplemented with 25% ethylene glycol. X-ray data for unliganded $BoSGBP_{MLG}$-A and MLG7-bound $BoSGBP_{MLG}$-A were collected at the Stanford Synchrotron Radiation Lightsource (SSRL) beamline 9-2. X-ray data for MLG7-bound $BoSGBP_{MLG}$-B at the Canadian Macromolecular Crystallography Facility (CMCF) beamline 08B1-1 at the Canadian Light Source (CLS). All three data sets were indexed and integrated using XDS (153). $BoSGBP_{MLG}$-A (unliganded and MLG7-bound) were determined by molecular replacement in Phaser from the CCP4i2 package (154,221) using the cellohexaose-bound $BoSGBP_{MLG}$-A structure (with the ligand and waters removed) as the search model, and refined in Refmac5 with alternate rounds of manual model building in Coot. MLG7-bound $BoSGBP_{MLG}$-B structure was determined by molecular replacement in Molrep (222) using the cellohexaose-bound $BoSGBP_{MLG}$-B (with the ligand and waters removed) structure as the search model, and refined in Buster (223) with alternate rounds of manual model building in Coot. Data processing and refinement statistics for all structures can be found in Table 3-2.

3.2.7 *B. ovatus* genetics and anaerobic growth study

Bacterial strains and culture conditions. For these experiments and to generate all of the mutant MLGUL strains used in these experiments, the *B. ovatus* ATCC-8483 Δ*tdk* (*ΔBACOVA_03071*) strain was employed to facilitate allelic exchange, as previously described

(107,146). For clarity we refer to the Δ*tdk* strain as wild-type, as this parent strain retains a wild-type MLGUL. Mutations were generated using the counter-selectable allelic exchange vector pExchange-*tdk* as previously described (146). Oligonucleotides used in this study are listed in Table B-1.

B. ovatus was cultured in a 37 °C Coy anaerobic chamber (5 % H_2/10 % CO_2/85 % N_2) from freezer stocks into tryptone-yeast extract-glucose (TYG) medium and grown for 24 hours, to an $O.D._{600}$ ~1.0. The following day cells were back-diluted 1:100 into Bacteroides minimal media (MM) including 5 mg ml^{-1} glucose (Sigma) as noted and grown overnight (16 hours). For kinetic growth experiments in a plate reader, MM-glucose grown cells were back-diluted 1:200 into MM with the experimental carbohydrate, and in parallel to MM with glucose. Thus both glucose controls and experimental MLG and oligosaccharide grown cultures were started at the same initial $O.D._{600}$ in the plate reader. Kinetic growth experiments were performed at 37 °C in 96 well plates and $O.D._{600}$ were recorded every 10-30 min. All plate reader growth experiments were performed in 3 replicates and the averages are reported in each figure. However, all biological experiments were repeated at least twice to verify consistent growth phenotypes from day to day.

Quantitative RT-PCR (qRT-PCR). B. ovatus strains were cultured in 5 mL of MM containing 5 mg/mL glucose or MLG. Duplicate bacterial cultures were arrested at mid-log phase ($O.D._{600}$ ~0.8) with RNAprotect (Qiagen), then stored at −80 °C overnight, before purification with RNeasy kit (Qiagen). RNA purity was assessed spectrophotometrically, and 1 μg of RNA was used immediately for reverse transcription (QuantiTect Reverse Transcription kit, Qiagen). RT-qPCR was performed in a 96-well plate on a LightCycler 480 System (Roche) with FastStart Essential DNA Green Master (Roche) using the standard primer. Reactions were carried out in 10 μL, consisting of 5 μL of SYBR Green mix, 20 ng of cDNA and 1 μM (MLGUL TBDT gene Bacova_02742) or 0.125 μM (16S ribosomal RNA) primer mix. Reaction conditions were 95 °C for 600 s, followed by 45 cycles of 95 °C for 10 s, 55 °C for 10 s, 72 °C for 10 s. *C*q values (cycle at which an amplification signal is first detected) were calculated using a LightCycler 480 SW 1.5. Data were normalized to 16S rRNA transcript levels, and a change in expression level was calculated as a fold change compared with MM-glucose cultures.

3.3 Results

3.3.1 MLGUL SGBPs are highly specific for mixed-linkage β(1,3)/β(1,4)-glucans

Recombinant BoSGBP$_{MLG}$-A and BoSGBP$_{MLG}$-B were produced in *E. coli* from constructs designed to exclude the predicted signal peptide and N-terminal lipidation site (Cys-1 of the mature protein, Fig B-1). Carbohydrate binding was first screened qualitatively against a library of soluble polysaccharides (Fig. S2) by affinity gel electrophoresis (AGE). As suggested by their context in the MLGUL, the migration of both BoSGBP$_{MLG}$-A and BoSGBP$_{MLG}$-B was strongly retarded in a gel containing barley mixed-linkage β-glucan (bMLG, Fig. 3-2A). Qualitatively weaker interactions with tamarind xyloglucan (Xyg), konjac glucomannan, and hydroxyethylcellulose were also observed (Fig. 3-2A), all of which contain stretches of β(1,4)-linked backbone glucosyl residues (Fig. B-2). No binding was observed to a range of other polysaccharides, including diverse β(1,3)-glucans, mannans, xylans, dextran, ulvan, and anionic carboxymethylcellulose (Fig. B-3). Isothermal titration calorimetry (ITC) confirmed the specific binding of both BoSGBP$_{MLG}$-A and BoSGBP$_{MLG}$-B to bMLG (K_a (4.41 ± 0.65) x 10^5 M^{-1} and (1.04 ± 0.1) x 10^4 M^{-1}, respectively (Fig. B-4, Table 3-1). In comparison, the affinity of BoSGBP$_{MLG}$-A for xyloglucan was two orders-of-magnitude lower, while BoSGBP$_{MLG}$-B binding to xyloglucan was too weak to be quantified (Fig. B-4, Table 3-1).

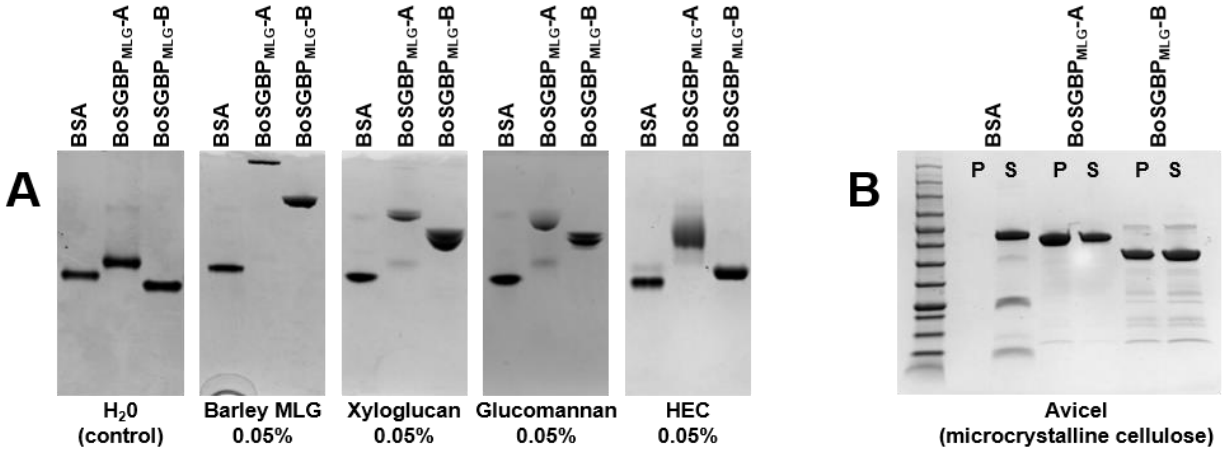

Figure 3-2. BoSGBP$_{MLG}$-A and BoSGBP$_{MLG}$-B binding analyses.

(A) affinity electrophoresis gels against soluble polysaccharides that were bound by BoSGBP$_{MLG}$-A and BoSGBP$_{MLG}$-B with BSA as a non-binding control in each gel. (B) SDS-PAGE gel of pull-down assay against insoluble polysaccharide bound by BoSGBP$_{MLG}$-A and BoSGBP$_{MLG}$-B with BSA as a non-binding control.

Following the observed weak binding towards the soluble cellulose derivative hydroxyethylcellulose (Fig. 3-2A), pull down assays were conducted to assess binding to insoluble cellulose (β(1,4)-glucan) and β(1,4)-mannan. Both $BoSGBP_{MLG}$-A and $BoSGBP_{MLG}$-B indeed bind crystalline cellulose (Avicel, Fig. 3-2B), whereas neither bound insoluble β(1,4)-mannan (Fig. B-2B). To quantify binding affinities to crystalline cellulose, green fluorescent protein (GFP) fusion proteins were produced to conduct depletion binding-isotherm experiments. GFP-$BoSGBP_{MLG}$-A and GFP-$BoSGBP_{MLG}$-B bound Avicel with lower affinity (K_a (2.04 ± 0.54) x 10^4 M^{-1} and (8.52 ± 1.50) x 10^3 M^{-1}, respectively) compared to bMLG, but with higher affinity than xyloglucan (Fig. B-7, Table 3-1).

The GH16 endo-β-glucanase of the MLGUL ($BoGH16_{MLG}$, Fig. 3-1) specifically cleaves β(1,4) bonds of glucosyl residues to which a β(1,3)-linked glucosyl residue is attached, producing G4G3G and G4G4G3G as the limit-hydrolysis products from MLG (211). As such, these would be the smallest mixed-linkage gluco-oligosaccharides (MLGOs) that could be encountered by the SGBPs. ITC analysis however showed that neither $BoSGBP_{MLG}$-A nor $BoSGBP_{MLG}$-B bind these mixed-linkage trisaccharide and tetrasaccharide products (Fig. B-5, Table 3-1). Hence, longer MLGOs were obtained by controlled hydrolysis of oat MLG with $BoGH16_{MLG}$ followed by fractionation with size exclusion chromatography. The mixed-linkage hexasaccharide G4G3G4G4G3G (MLG6) and heptasaccharide G4G4G3G4G4G3G (MLG7) were bound by $BoSGBP_{MLG}$-A and $BoSGBP_{MLG}$-B with respective affinities ca. one order-of-magnitude less than those for the full-length polysaccharide, with marginally stronger binding observed for the heptasaccharide for both SGBPs. (Fig. B-5, Table 3-1).

Analogous to results for short MLGOs, neither SGBP was able to bind the all-β(1,4)-linked cellotetraose and cellopentaose (Fig. B-5, Table 3-1). Reflecting the observed binding to insoluble cellulose and the apparent requirement for longer oligosaccharides, both SGBPs quantifiably bound cellohexaose (Fig. B-6, Table 3-1). Notably, the affinity of $BoSGBP_{MLG}$-A to cellohexaose was one order-of-magnitude weaker (K_a 10^3) than to the MLG hexa- and heptasaccharides (K_a 10^4), while the binding of $BoSGBP_{MLG}$-B to cellohexaose and these MLGOs was comparably weak (K_a 10^3, Table 3-1).

Table 3-1. Summary of thermodynamic parameters for wild-type $BoSGBP_{MLG}$-A and $BoSGBP_{MLG}$-B obtained by isothermal titration calorimetry at 25 °C[a]

	K_a (M^{-1})		ΔG (kcal.mol^{-1})		ΔH (kcal.mol^{-1})		TΔS (kcal.mol^{-1})		*n*	*n*
Carbohydrate	$BoSGBP_{MLG}$-A	$BoSGBP_{MLG}$-B	$BoSGBP_{MLG}$-A	$BoSGBP_{MLG}$-B	$BoSGBP_{MLG}$-A	$BoSGBP_{MLG}$-B	$BoSGBP_{MLG}$-A	$BoSGBP_{MLG}$-B	$BoSGBP_{MLG}$-A	$BoSGBP_{MLG}$-B
bMLG[b]	(4.41 ± 0.65) x 10^5	(1.04 ± 0.1) x 10^4	-7.7	-5.5	-45.3 ± 1.6	-14.5 ± 1.6	-37.6	-9	1.11 ± 0.03	0.79 ± 0.07
XyG[c]	(7.19 ± 1.9) x 10^3	Weak[d]	-5.2	Weak	-5.9 ± 1.8	Weak	-0.7	Weak	1.69 ± 0.53	Weak
Cellotetraose	NB[d]	NB	NB	NB	NB	NB	NB	NB	NB	NB
Cellopentaose	NB	NB	NB	NB	NB	NB	NB	NB	NB	NB
Cellohexaose	(3.39 ± 0.03) x 10^3	(3.1 ± 0.06) x 10^3	-4.8	-4.8	-11.4 ± 0.05	-6.3 ± 0.09	-6.6	-1.5	1	1
G3G4G4G	NB	NB	NB	NB	NB	NB	NB	NB	NB	NB
G4G4G3G	NB	NB	NB	NB	NB	NB	NB	NB	NB	NB
G4G3G4G	NB	NB	NB	NB	NB	NB	NB	NB	NB	NB
G4G3G4G4G3G	(1.44 ± 0.04) x 10^4	(3.16 ± 0.19) x 10^3	-5.7	-2.4	-11.8 ± 0.3	-7.0 ± 0.4	-6.1	-2.2	1.49 ± 0.04	1
G4G4G3G4G4G3G	(4.51 ± 0.34) x 10^4	(4.69 ± 0.13) x 10^3	-6.3	-5	- 11.9 ± 0.4	-13.9 ± 0.2	-5.6	-8.9	1.19 ± 0.05	1
G4G4G3G4G4G3G		(4.07 ± 0.08) x 10^3		-4.9		-29 ± 0.4		-24.1		0.5

[a] Corresponding thermograms are shown in Figures S4-6. All parameters were allowed to vary independently during data fitting, with the exception of $BoSGBP_{MLG}$-A and $BoSGBP_{MLG}$-B binding cellohexaose, and $BoSGBP_{MLG}$-B binding the MLGOs G4G3G4G4G3G and G4G4G3G4G4G3G, for which *n* was fixed at 1 or 0.5 as indicated.

[b] K_a values were calculated on a molar basis from MLG polysaccharide concentrations in g/L by assuming a hexasaccharide binding motif, based on crystal complex structures with MLGOs.

[c] K_a values were calculated on a molar basis from XyG polysaccharide concentrations in g/L by assuming a Glc_8-backbone (XyGO dimer) oligosaccharide binding motif.

[d] Weak: K_a <500 M^{-1}

[e] NB: no binding observed

Table 3-2. Crystallographic data and refinement statistics table

	$BoSGBP_{MLG}$-A (6E60)	$BoSGBP_{MLG}$-A Cellohexaose (6DMF)	$BoSGBP_{MLG}$-A MLG7 (6E61)	$BoSGBP_{MLG}$-B Cellohexaose (6E57)	$BoSGBP_{MLG}$-B MLG7 (6E9B)
Data collection					
Beamline	SSRL 9-2	APS 21-ID-G	SSRL 9-2	APS 21-ID-F	CLS 08B1-1
Wavelength (Å)	0.979	0.979	0.979	0.979	0.979
Resolution range (Å)	37.25 – 1.50 (1.59 – 1.50)	49.60 - 2.40 (2.49 - 2.40)	39.65 – 2.51 (2.66 – 2.51)	78.34 - 2.71 (2.81 - 2.71)	34.33 – 3.15 (3.34 – 3.15)
Space group	P 2_1 2_1 2_1	P 6_1	P 2_1 2_1 2_1	P 2_1 2_1 2_1	P 2_1 2_1 2_1
Unit cell					
a, b, c (Å)	47.35 89.48 121.10	228.85 228.85 246.52	86.98 93.14 155.59	156.44 243.65 76.06	155.62 241.10 74.88
α, β, γ (°)	90 90 90	90 90 120	90 90 90	90 90 90	90 90 90
Total reflections	363678 (57305)	2422857 (206443)	279444 (34379)	447089 (44223)	371772 (58717)
Unique reflections	81232 (12744)	283613 (28197)	43053 (6405)	78048 (7714)	50094 (7894)
Redundancy	4.5 (4.5)	8.5 (7.3)	6.5 (5.4)	5.7 (5.7)	7.4 (7.4)
Completeness (%)	97.6 (95.6)	99.9 (99.5)	98.6 (92.3)	97.7 (98.3)	99.8 (99.7)
I/σI	9.8 (1.5)	8.4 (1.1)	14.6 (3.5)	10.4 (2.4)	15.6 (1.7)
Wilson B-factor ($Å^2$)	14.5	34.9	28.7	50.1	90.3
R-meas	0.119 (0.975)	0.239 (2.007)	0.12 (0.462)	0.135 (0.914)	0.136 (1.277)
CC(1/2)	0.997 (0.605)	0.996 (0.580)	0.997 (0.887)	0.992 (0.700)	0.999 (0.651)
Molecules in AU	1	10	2	4	4
Refinement					
R-work	0.158 (0.309)	0.185 (0.285)	0.173 (0.255)	0.194 (0.297)	0.221 (0.272)
R-free	0.184 (0.310)	0.239 (0.339)	0.223 (0.310)	0.239 (0.357)	0.269 (0.324)
Number of non-hydrogen atoms					
All	4987	42441	8765	12333	11352
Macromolecules	4374	40685	8181	11933	11161
Ligands	16	812	164	260	123

Water	597	944	420	140	68
RMS deviations					
Bonds lengths (Å)	0.012	0.010	0.005	0.009	0.010
Bond angles (°)	1.581	1.240	0.958	1.270	1.310
Ramachandran statistics					
Ramachandran favored (%)	98	96	97	95	87
Ramachandran allowed (%)	2		3		11
Ramachandran outliers (%)	0	0.16	0	0.58	2
Average B-factor ($Å^2$)					
All	15.1	40.4	28.6	58.7	100.4
Macromolecules	14.9	40.2	29.9	58.6	100.5
Carbohydrate ligands		50.1	30.4	74.2	74.2
Solvent	25.7	37.1	25.4	43.8	60.5

3.3.2 SGBP crystallography illuminates the molecular basis of MLG specificity

To reveal the molecular basis of the substrate specificity of the SGBPs in the context of their contribution to MLG utilization by *B. ovatus*, we solved the three-dimensional structure of these proteins in several unliganded and oligosaccharide-complexed forms by X-ray crystallography (Table 3-2).

3.3.2.1 BoSGBP$_{MLG}$-A is a canonical SusD homolog

The crystal structure of unliganded BoSGBP$_{MLG}$-A (1.50 Å, R_{work} = 15.8%, R_{free} = 18.4%; Table 3-2) revealed a single globular domain with a canonical "SusD-like" fold (146), which is dominated by α-helices (Fig. 3-3A). As in the amylose-binding SusD, a series of these α helices are organized into four tetratricopeptide repeat (TPR) units that form a super-helical fold along the convex surface of the protein. The concave side of the TPR motif cradles the remainder of the polypeptide chain comprising many loops and short α helices, where the ligand binding site is found. This region is variable among SusD homologs, thus providing a tunable platform optimized to discriminate cognate substrates (95), e.g. starch (146), xyloglucan (99), mucin O-glycan (224), sialic acid (225), laminarin/pustulan (226), and chitin (227).

Co-crystallization of BoSGBP$_{MLG}$-A with cellohexaose (2.40 Å, R_{work} = 18.5%, R_{free} = 24.0%; Table 3-2) and the MLG7 (2.51 Å, R_{work} = 17.3%, R_{free} = 22.3%; Table 3-2) clearly revealed this platform as the substrate-binding site. Superposition of the two complexes with the unliganded structure reveals that there are no major changes in global conformation or sidechain positioning upon substrate binding (Fig. B-8). This is similar to the SGBP-A homolog of the xyloglucan utilization locus (XyGUL) (23) but contrary to SusD, in which two loops undergo a large conformational change to enable a tyrosine sidechain to stack against the α-glucan ligand (21). Four surface-exposed aromatic residues (Y266, W77, W350, and W353) of BoSGBP$_{MLG}$-A are arranged in a linear fashion to constitute a long (ca. 36 Å as measured from Y266 to W353) binding platform (Fig. 3-3B, 3-3C, 3-3D, 3-3E).

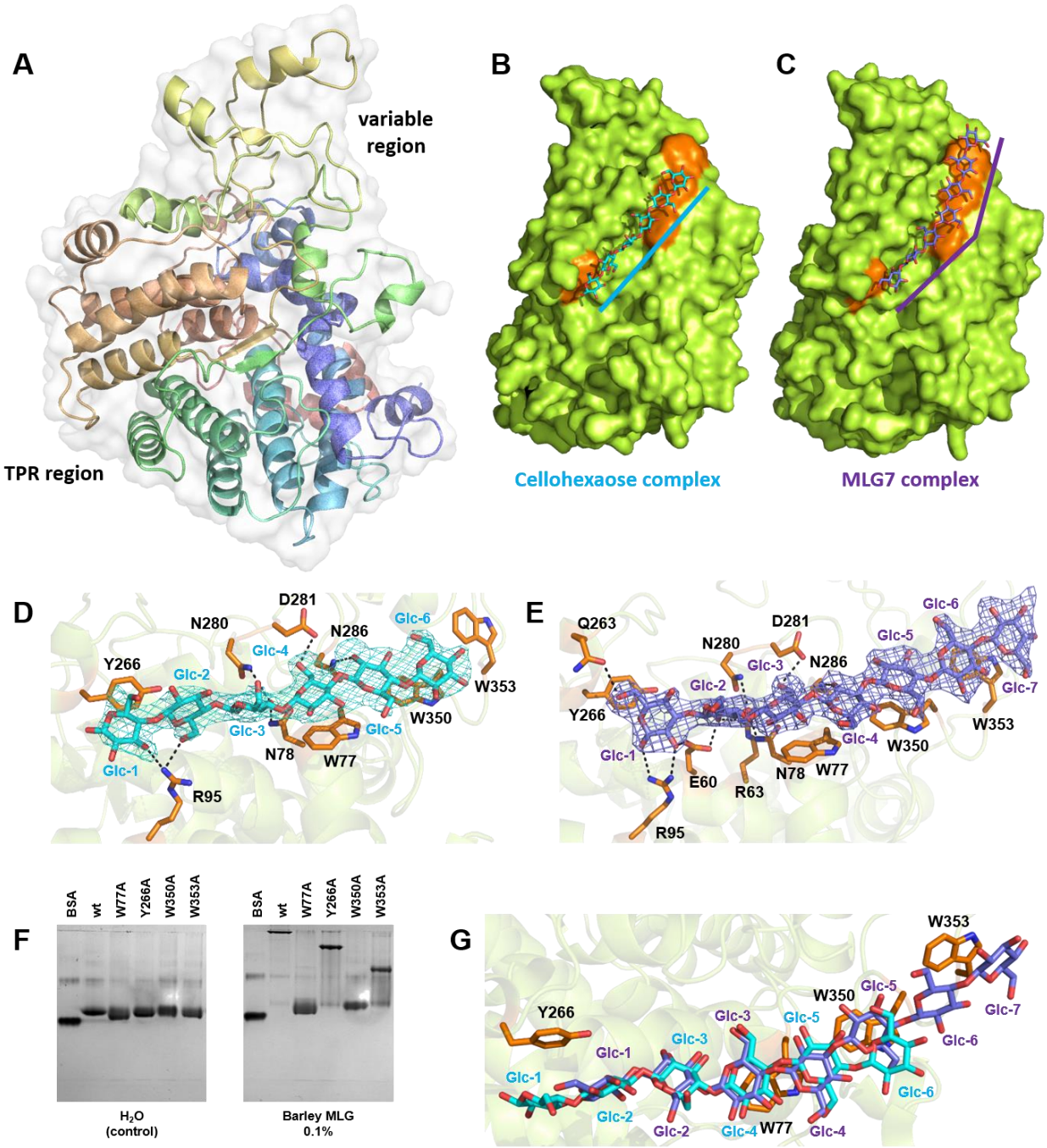

Figure 3-3. BoSGBP$_{MLG}$-A crystal structure.

(A) Overall structure of BoSGBP$_{MLG}$-A in cartoon representation with the polypeptide color ramped from blue to red (from N- to C-termini) and the transparent surface shown in white. (B) Surface representation of the cellohexaose complex with ligand shown in cyan and the aromatic residues that form the binding platform shown in orange. Ten molecules are found in the asymmetric unit, each displaying no significant structural differences from the others (<0.2 Å RMSD across equivalent Cα pairs); chain E is shown as a representative. (C) Surface representation of the MLG7 complex with ligand shown in slate and the aromatic residues that form the binding platform shown in orange. Two

molecules are found in the asymmetric unit, each displaying no significant structural difference from the other; chain A is shown as a representative. (D) Close-up of the binding site of the cellohexaose complex with interacting residues shown as opaque orange sticks. Cellohexaose is colored cyan and potential hydrogen bonding interactions are shown as black dashed lines (within 3.5 Å of the ligand). Omit map for the ligand (generated by Privateer (164)) is shown contoured to 3σ. (E) Close-up of the binding site of the MLG7 complex with interacting residues shown as opaque orange sticks. MLG7 is colored slate and potential hydrogen bonding interactions are shown as black dashed lines (within 3.5 Å of the ligand). Omit map for the ligand is shown contoured to 3σ. (F) Affinity gel electrophoresis of binding platform site-directed mutants. (G) Overlay of cellohexaose and MLG7 bound to the binding platform; cellohexaose is shown in cyan and MLG7 in slate.

In the $BoSGBP_{MLG}$-A:cellohexaose complex, electron density was observed for all six glucose residues in 9 of 10 protein molecules in the asymmetric unit, with five glucose residues observed in the remaining molecule. All glucose residues were in the lowest energy $^4C_1$ (chair) conformation (Fig. 3-3B, 3-3D; Table B-2). The reducing end glucose, Glc-1, displays an aromatic stacking interaction (228,229) with Y266 of the binding platform, Glc-4 is positioned over W77, and the non-reducing end glucose, Glc-6, stacks against W350 (Fig. 3-3B, 3-3D). In the $BoSGBP_{MLG}$-A:MLG7 complex, all seven glucosyl residues could be modelled into the electron density in the $^4C_1$ conformation in both molecules in the asymmetric unit (Fig. 3-3C, 3-3E; Table B-2), with the oligosaccharide chain oriented in the same direction. In contrast to cellohexaose, MLG7 spans the entire length of the binding platform with the sugar rings of Glc-1, Glc-3, Glc-5, Glc-7 stacking against the aromatic side chains of Y266, W77, W350 and W353, respectively (Fig. 3-3C, 3-3E). A limited number of potential hydrogen-bonding interactions were observed with both oligosaccharide ligands. These hydrogen bonds are confined to one half of the binding platform flanked by Y266 and W77; no additional hydrogen-bonding interactions are observed to either ligand in the vicinity of W350 and W353 (Fig. 3-3D, 3-3E). Additionally, most heteroatom interatomic distances are greater than 3.0 Å, suggesting that these hydrogen bonds are moderate/weak and mostly electrostatic (230).

The importance of each aromatic residue comprising the binding platform is underscored by AGE analysis of single site-directed mutants (W77A, Y266A, W350A, and W353A). The two central tryptophan residues are critical as both W77A and W350A variants independently fail to bind bMLG and other β(1,4)-glucosyl-containing polysaccharides (Fig. 3-3F, B-9). The two flanking residues, Y266 and W350, though not as critical as the central tryptophan residues, also contribute to binding MLG as evidenced from the diminished binding relative to wild-type when

either is replaced with an alanine. Although AGE analysis as performed here is only semi-quantitative, the data suggest that W350 may have a greater contribution to binding than Y266 (Fig. 3-3F).

3.3.2.2 BoSGBP$_{MLG}$-B is a novel extended, multimodular MLG-binding protein

Distinct from BoSGBP$_{MLG}$-A, crystal structures of BoSGBP$_{MLG}$-B in complex with cellohexaose (2.71Å, R_{work} = 19.4%, R_{free} = 23.8%; Table 3-2) and MLG7 (3.15 Å, R_{work} = 22.1%, R_{free} = 26.9%, residues 12 to 399; Table 3-2) reveal a multimodular architecture comprised of four discrete immunoglobulin (Ig)-like domains in an extended arrangement: domain A (residues 22 to 133), domain B (residues 134 to 219), domain C (residues 220 to 309), and domain D (residues 310 to 420) (Fig. 3-4A). Such multi-domain architecture is typical of SGBPs, which are usually encoded directly downstream of the corresponding SGBP-A (SusD homolog) in PULs (99,103,105,107,110), but is difficult to predict due to very low sequence similarity among these proteins. Likewise, there is little tertiary structural homology among these proteins, such that a Dali search (187,231) of the Protein Data Bank using the full-length structure or individual domains failed to return any matches with other SGBPs (99,103,107); results were limited to unrelated proteins with Z-scores less than 11.

The number of domains in these SGBPs is variable (99,103,107), ranging from three in SusE (107) to six in the SGBP-B from the heparin/heparan-sulfate PUL (103). The four domains of BoSGBP$_{MLG}$-B are arranged in a right-handed helical configuration in which the domains are rotated 120° around the central axis (three domains per helical turn such that domains A and D overlap when looking down the axis, Fig. 3-4A). Whereas the presence of a single proline residue in each inter-domain linker is a feature of the previously solved SGBP-B structures (99,103,107), in BoSGBP$_{MLG}$-B a proline residue was found only in the linker after domain A (Fig. 3-4A), which may suggest reduced conformational rigidity vis-à-vis related SGBPs (99). However, the identical domain arrangement of all four molecules in the asymmetric unit (for both cellohexaose and MLG7 complexes, Table 3-2) could be evidence of a lack of conformational flexibility.

In the structure of BoSGBP$_{MLG}$-B co-crystallized with cellohexaose, electron density was observed for five of the expected six glucosyl residues, all in the favored $^4C_1$ conformation (Fig. 3-4B, Table B-2). Oligosaccharide binding was mediated by the following aromatic stacking interactions: Glc-1 (reducing end) and Glc-2 with Y371, Glc-3 and Glc-4 with W373 and Glc-5 (non-reducing end) with W322. In the structure of BoSGBP$_{MLG}$-B in complex with MLG7,

electron density was observed for all seven glucosyl residues in the $^4C_1$ conformation (Fig. 3-4C, Table B-2). Reducing end Glc-1 only makes contact with a residue from a molecule in a neighboring asymmetric unit (*vide infra*), Glc-2 and Glc-3 exhibit stacking interaction with Y371, Glc-4 and Glc-5 with W372, Glc-6 with W322, and Glc-7 with Y362. As for $BoSGBP_{MLG}$-A, very few hydrogen bonding interactions were observed with either oligosaccharide (Fig. 3-4B, 3-4C).

Notably, in the $BoSGBP_{MLG}$-B cellohexaose and MLG7 complexes (Table 3-2) two out of the four molecules in the asymmetric unit bound ligand that is shared with molecules from neighboring asymmetric units, such that the oligosaccharides are sandwiched between two binding platforms presented on the C-terminal domain D (Fig. 3-4D, B-10). In the other two molecules in the asymmetric unit that are not involved in the ligand-sharing crystal contact, density for the ligand was very poor and no sugars were modelled. The molecules that share a common ligand are symmetry-related through a two-fold rotation operation about an axis orthogonal to the length of the oligosaccharide ligand, straight through the central Glc-4 of MLG7 (Fig. 3-4D). As such, the binding platforms are oriented in opposite directions to either side of the oligosaccharide chain, which suggests a degree of plasticity in MLG recognition.

As suggested by the complexed structures, domain D provides the only substrate-binding site in $BoSGBP_{MLG}$-B. AGE analysis of the four domains produced independently demonstrated that indeed only domain D binds bMLG (Fig. 3-4E, B-9). This feature is similar to the XyGUL SGBP-B and the heparin/heparan-sulfate PUL SGBP, although in the case of the latter, the binding platform spans two of the distal C-terminal domains, D5 and D6 (99,103). In contrast, the archetypal SusE and SusF SGBPs contain two and three starch-binding sites, respectively, on individual domains and thus represent the only SGBPs known to possess multiple binding sites (107).

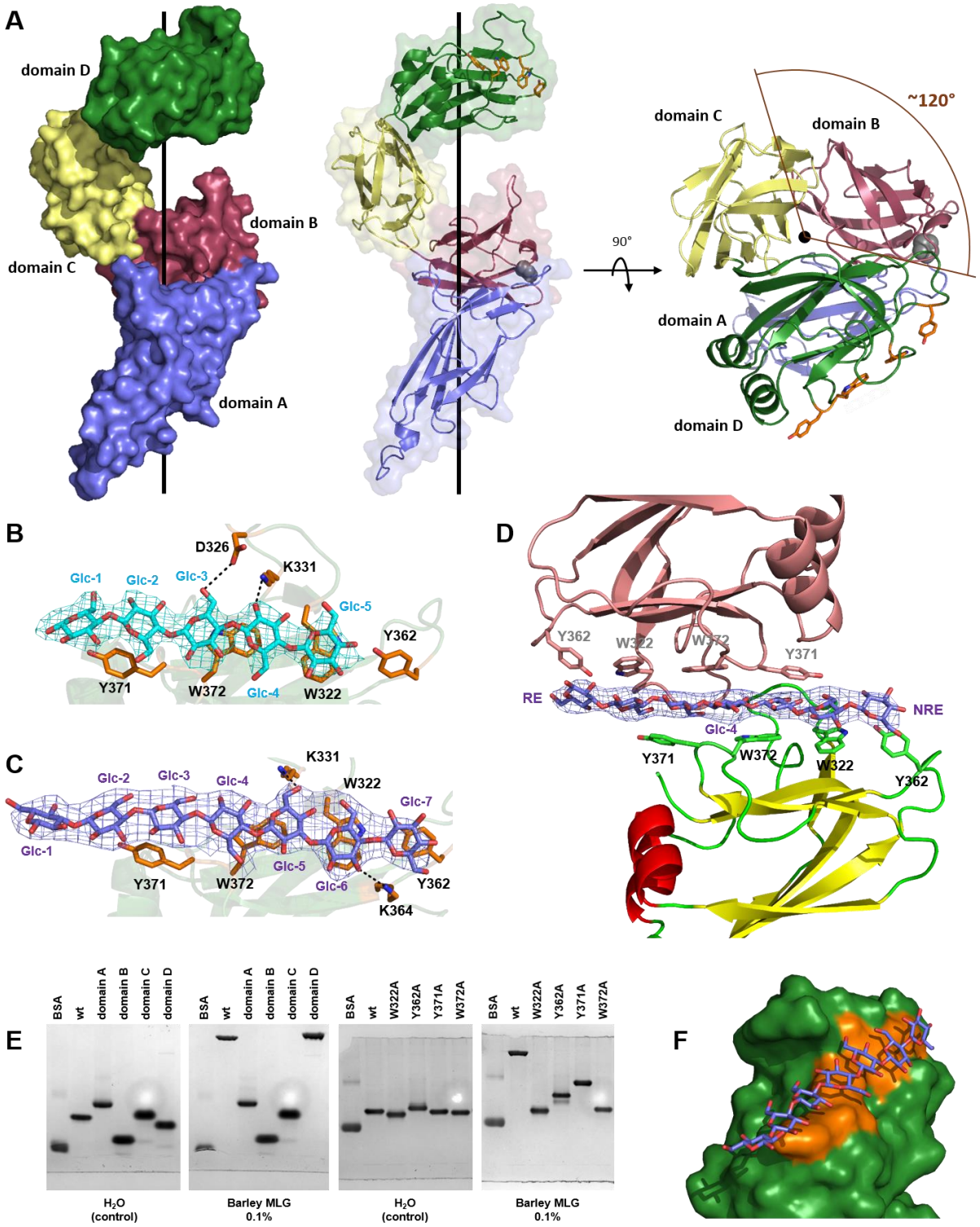

Figure 3-4. BoSGBP$_{MLG}$-B crystal structure.

(A) Overall structure of BoSGBP$_{MLG}$-B in cartoon representation and transparent surface with each domain colored differently: domain A – blue, domain B – raspberry, domain C – pale yellow, domain D – forest. A side view and a top view are shown with the black line representing the imaginary axis around which the domains wrap. The single

interdomain proline is shown as gray spheres and the aromatic sidechains comprising the binding platform are shown as orange sticks. (B) Close-up of the binding site of the cellohexaose complex with interacting residues shown as opaque orange sticks. Cellohexaose is colored cyan and potential hydrogen bonding interactions are shown as black dashed lines (within 3.5 Å of the ligand). Omit map for the ligand (generated by Privateer (164)) is shown contoured to 3σ. Of the four molecules in the asymmetric unit, data from chain A is shown a representative. (C) Close-up of the binding site of the MLG7 complex with interacting residues shown as opaque orange sticks. MLG7 is colored slate and potential hydrogen bonding interactions are shown as black dashed lines (within 3.5 Å of the ligand). Omit map for the ligand is shown contoured to 3σ. Of the four molecules in the asymmetric unit, data from chain D is shown a representative. (D) A single MLG7 ligand being shared between two $BoSGBP_{MLG}$-B molecules belonging to neighbouring asymmetric units. The bottom molecule is colored according to secondary structure (yellow β-strands, red α-helices, and green loops), the top molecule from a different asymmetric unit is colored salmon, and the ligand is colored slate. Omit map for the ligand is shown contoured to 3σ. RE = reducing end, NRE = non-reducing end. An analogous orientation was observed for the cellohexaose complex (not shown, PDB ID 6E57). (E) Affinity gel electrophoresis of individual $BoSGBP_{MLG}$-B domains and binding platform site-directed mutants. (F) Surface representation of the binding platform of domain D in complex with MLG7. Aromatic sidechains comprising the binding platform are colored orange and MLG7 is colored slate.

The β-sandwich fold of domain D is comprised of two sheets of four and three antiparallel β-strands, with two of the connecting loops incorporating α-helices. The substrate-binding site is located above the smaller, top β-sheet although residues important for binding are all borne on loops connecting the β-strands (Fig. 3-4D). The aromatic sidechains of Y371, W372, W322, and Y362 constitute a ca. 28 Å (as measured from Y362 to Y371), flat binding platform (Fig. 3-4F). These aromatic residues are critical for substrate binding, as demonstrated by AGE analysis: mutation of either of the central tryptophans (W322A and W372A) completely abrogates binding whereas mutation of the tyrosines at the end of the platform (Y362A and Y371A) severely diminishes binding to bMLG, xyloglucan, and hydroxyethylcellulose (Fig. 3-4E, B-9).

3.3.3 SGBP-A and SGBP-B have distinct functions in vivo

With a firm understanding of substrate specificity and tertiary structures of $BoSGBP_{MLG}$-A and $BoSGBP_{MLG}$-B, we then sought to determine how polysaccharide binding at the cell surface contributes to the growth of *B. ovatus* on MLG. Strains with in-frame deletions and binding-site deficient alleles of both genes were generated by performing allelic exchange in a Δ*tdk* strain of *B. ovatus* (146), which is subsequently referred to as "wild-type." To monitor anaerobic growth, cells were first cultured overnight in minimal media (MM) containing glucose, then back-diluted 1:200 into parallel cultures containing glucose or glucan substrate. Wild-type *B. ovatus* (Δ*tdk*)

grows on high viscosity MLG, while cells in which the complete MLGUL (Fig. 1) was deleted (ΔMLGUL) do not, which confirmed that the MLGUL is essential and solely responsible for growth on this β-glucan (Fig. 3-5A, 3-5B) (211).

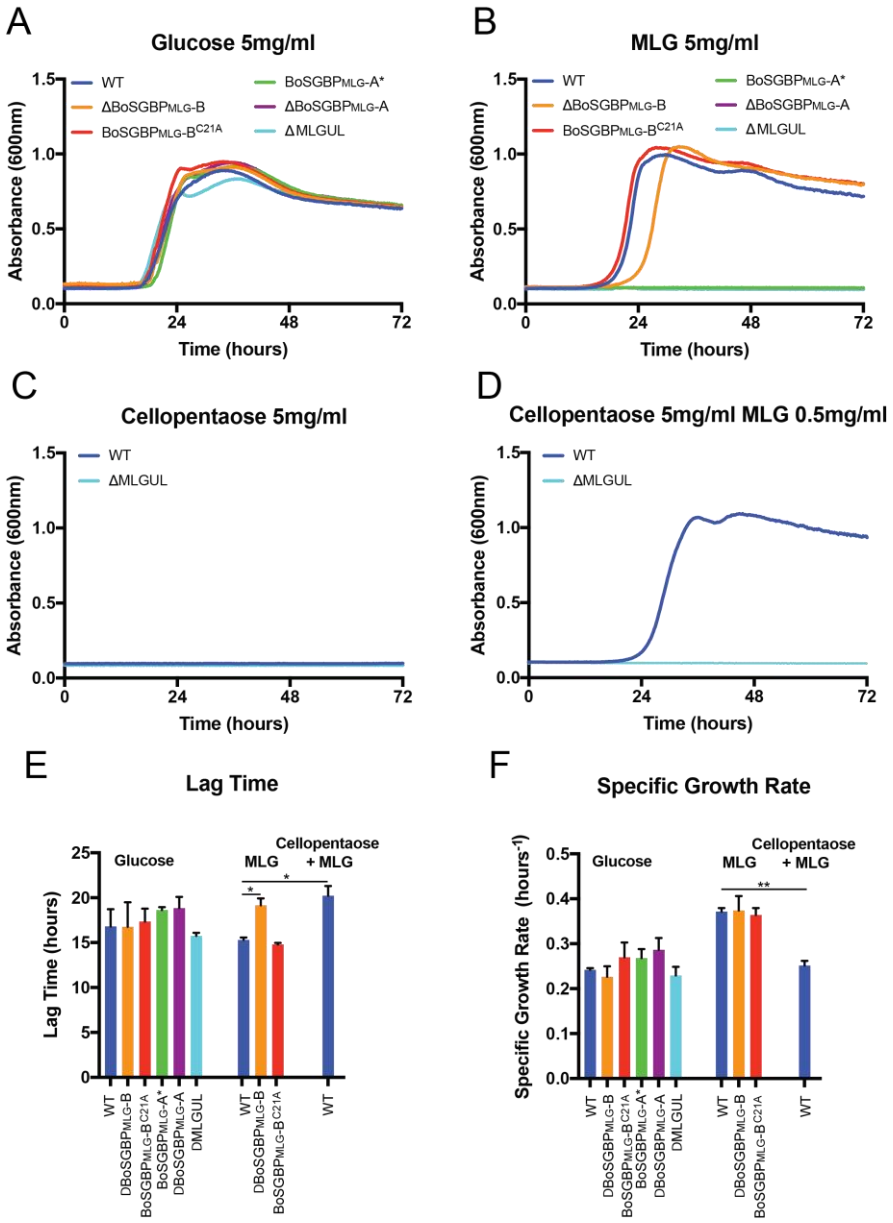

Figure 3-5. MLGUL surface glycan binding proteins facilitate growth on MLG and cellopentaose.
Average growth curve of *B. ovatus* MLGUL strains on (A) 5 mg/mL glucose, (B) 5 mg/mL high viscosity MLG, (C) 5 mg/mL cellopentaose, (D) 5 mg/mL cellopentaose with 0.5 mg/mL MLG. (E) Lag time from the growth curves in A, B, and D. (F) Specific growth rates were calculated at $O.D._{600}$ = 0.5 for growth curves in A, B, and D. Bars denoted with a '*' have a $p < 0.05$ and that with a '**' has a $p < 0.005$. Statistically significant differences were determined using the two-tailed unpaired Student's *t* test.

The *BoSGBP$_{MLG}$-A* single-gene knock-out (ΔSGBP$_{MLG}$-A) likewise cannot grow on MLG. This finding is consistent with the essential role of the SGBP-A (SusD) homologs for the uptake of starch via the Sus of *B. thetaiotaomicron* and the uptake of xyloglucan via the XyGUL of *B. ovatus* (99,102). In both the Sus and the XyGUL, an allele encoding a glycan-binding-deficient version of the native protein (SusD* and SGBP$_{XyG}$-A*, respectively) restored growth on the cognate substrate (99,102). However, when the binding-deficient *BoSGBP$_{MLG}$-A** allele (a W77A/W350A double mutant) was exchanged into the ΔBoSGBP$_{MLG}$-A background to restore the MLGUL genetic structure, growth was not observed on MLG (Fig. 3-5A, 3-5B). These data indicate that glycan-binding activity is essential to BoSGBP$_{MLG}$-A function for growth on MLG, and furthermore suggests an importance to glycan import that is distinct from homologs in the Sus and XyGUL.

While BoSGBP$_{MLG}$-A must be both present and functional for growth on MLG, BoSGBP$_{MLG}$-B is not required. Deletion of *BoSGBP$_{MLG}$-B* results in a longer lag before entering exponential phase on MLG compared to wild-type, but does not affect the specific growth rate (Fig. 3-5A, 3-5B, 3-5E, 3-5F). In light of this result and based on previous experience, we were concerned that the *BoSGBP$_{MLG}$-B* gene deletion might affect transcription of the MLGUL. Our previous work on the Sus of *B. thetaiotaomicron* indicated that deletion of the gene encoding the SGBP SusF decreases the transcription of the downstream *susG*, which encodes the essential, cell-surface amylase. We therefore created a Cys21-Ala mutant to mis-traffick BoSGBP$_{MLG}$-B to the periplasm by removing the N-terminal lipidation site of the mature protein (107,138), while retaining the overall MLGUL genetic structure. Indeed, cells expressing BoSGBP$_{MLG}$-B^{C21A} grow similarly to wild-type cells on MLG, suggesting that the lag seen for ΔBoSGBP$_{MLG}$-B is likely due to a negative impact on transcription of other MLGUL genes. Correspondingly, qPCR on *B. ovatus* ΔBoSGBP$_{MLG}$-B and BoSGBP$_{MLG}$-B^{C21A} strains grown to mid-exponential phase on MLG revealed a 1000-fold decrease in the TBDT transcript in the knock-out versus the wild-type and BoSGBP$_{MLG}$-B^{C21A} strains (Fig. S11). The magnitude of this defect was surprising because all cells were harvested at the same O.D.$_{600}$ (after the lag growth defect of the knock-out had passed), but clearly indicates the *BoSGBP$_{MLG}$-B* deletion influences overall transcript stability (Fig. 3-1B, B-11). Moreover, this observation also suggests that sustained upregulation of this PUL is not required to support wild-type level growth rates on MLG, as long as sufficient amounts of MLGUL components eventually populate the cell surface and periplasm (Fig. 3-5F).

A key question regarding the MLGUL and related Sus-like systems is the size of the transported oligosaccharide. Our biochemical data demonstrates that the BoSGBP$_{MLG}$-A and BoSGBP$_{MLG}$-B proteins preferentially bind oligosaccharides with degree of polymerization ≥ 6 and, in particular, do not bind the limit-digest products of BoGH16$_{MLG}$ (Fig. B-4, B-5, Table 3-1). To mimic the likely MLG fragments that are transported by a putative TBDT/BoSGBP$_{MLG}$-A(/BoSGBP$_{MLG}$-B) complex, MLG was partially digested by recombinant BoGH16$_{MLG}$ *in vitro* to generate a profile of oligosaccharides that span a broad range of lengths (Fig. B-13). Control experiments revealed that although a *B. ovatus* ΔGH16 mutant cannot grow on the native MLG polysaccharide, it does grow similarly to wild-type on this mixture, thereby demonstrating competent uptake through the TBDT/BoSGBP$_{MLG}$-A complex (Fig. 3-6, B-14). Underscoring the essential role of the BoSGBP$_{MLG}$-A in capturing longer oligosaccharides, the ΔBoSGBP$_{MLG}$-A and BoSGBP$_{MLG}$-A* mutants cannot grow on the MLG digest (Fig. 3-6A, 3-6B). This phenotype is identical to that observed on native MLG polysaccharide (Fig. 3-5). Interestingly, the mistrafficked BoSGBP$_{MLG}$-B^{C21A} mutant displays a longer lag time compared to the wild-type, regardless of the presence of BoGH16$_{MLG}$, although specific growth rates were similar (Fig. 3-6D, 3-6E, 3-6F).

B. ovatus is unable to grow on insoluble cellulose as a sole carbon source (109). However, because the MLGUL-encoded SGBPs have weak affinity for cello-oligosaccharides, we wanted to explore the contribution of these proteins to growth. Wild-type cells fail to grow on cellopentaose as the sole carbon source (Fig. 3-5C), yet the addition of 0.5 mg/mL MLG to upregulate MLGUL (109) enables growth on cellopentaose to a cell density greater than that achievable on 0.5 mg/ml MLG alone (Fig. 3-5D, A-12). These results are concordant with the observation that the HTCS of the MLGUL (Fig. 3-1) is not activated by cello-ologosaccharides (109). The addition of 0.5 mg/mL MLG does not support growth on cellopentaose in the *ΔMLGUL* strain, demonstrating that growth on cello-oligosaccharides is indeed dependent on MLGUL expression (Fig. 3-5D). Approximately 10 % of the MLG structure is composed of longer regions (d.p. 5-9) of β(1,4)-glucosyl residues lacking β(1,3) kinks (232,233) and the ability to utilize cellopentaose via the MLGUL undoubtedly is a consequence of the known catalytic promiscuity of the GH3 *exo*-β-glucosidase for both β(1,3)/β(1,4)-mixed-linkage gluco-oligosaccharides and all-β(1,4)-linked gluco-oligosaccharides (211).

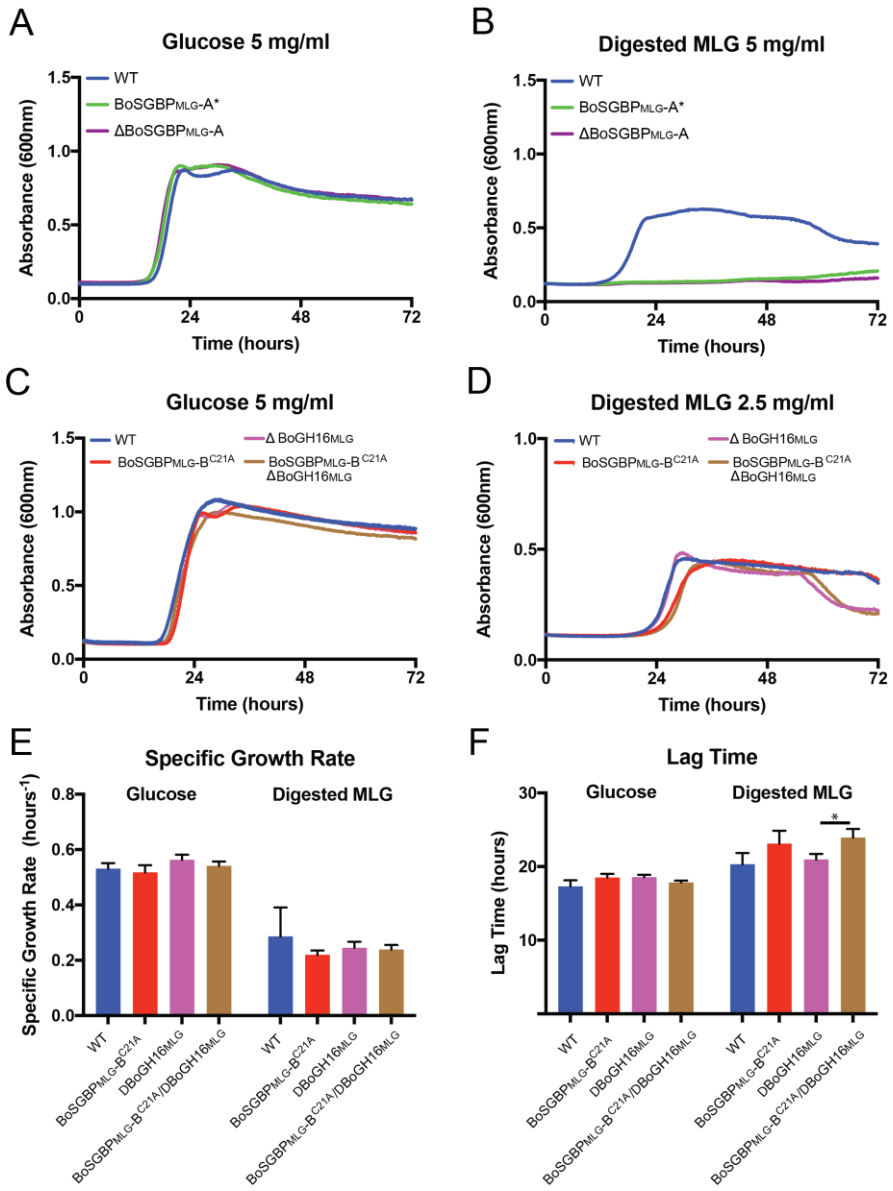

Figure 3-6. MLGUL surface glycan binding proteins capture BoGH16$_{MLG}$-digested MLG.

Average growth curve of *B. ovatus* MLGUL strains on (A,C) 5 mg/mL glucose, (B) 5 mg/mL digested MLG, and (D) 2.5 mg/mL digested MLG. (E) Lag time and from the growth curves in C and D. (F) Specific growth rates were calculated at O.D.$_{600}$ = 0.25 for growth curves in C and D. Bars denoted with a '*' have a $p < 0.05$. Statistically significant differences were determined using the two-tailed unpaired Student's *t* test.

3.4 Discussion

SGBPs are an important class of PUL components that effect target glycan capture at the outer membrane surface for backbone cleavage and import (Fig. 3-1). All PULs encode at least one SGBP: a SusD homolog, SGBP-A, which forms a functional complex with a partner TBDT (97,99,102). Indeed, the tandem *TBDT/SGBP-A* (*susC/susD*-homolog) pair is a signature feature that can be used to identify PULs in sequenced genomes (92). As exemplified by the $BoSGBP_{MLG}$-A structures reported here (Fig. 3-3), SGBP-A homologs are single-domain globular proteins built upon prominent tetratricopeptide repeats (TPRs), as first revealed by the structure of SusD from the starch utilization system (99,146,224,226,227). Recent crystallography of two SGBP-A/TBDT homologs demonstrated that SGBPs-A associate closely with the extracellular side of the TBDTs, forming a "pedal bin" lid (97).

Although this structural arrangement might suggest a role in recognizing and directing cognate substrates into the transporter (97), previous studies on the Sus and XyGUL using non-binding SGBP-A* mutants *in vivo* have failed to demonstrate that substrate binding is a prerequisite for growth (99,102). In these systems, the expression of additional SGBPs (i.e. SusE, $BoSGBP_{XyG}$-B) have been shown to supplement the loss of substrate binding by SusD and $BoSGBP_{XyG}$-A deletion, respectively (99,102,108). Deletion of the corresponding *SGBP-A* in these systems does cripple growth on the cognate substrates, consistent with the critical structural role in complex formation with the TBDT. Here, we showed that either complete removal of the $BoSGBP_{MLG}$-A from the cell surface (by gene knock-out, ΔSGBP-A) or abrogation of its ability to bind substrate (by removing crucial amino acid sidechains, SGBP-A*), prevents MLG utilization by *B. ovatus* (Fig. 3-5). Thus, the present study is, to our knowledge, the first to demonstrate directly an essential role of substrate binding by an SGBP-A (SusD) homolog, which extends beyond the obvious requirement for its physical presence at the cell surface (97,99,102).

Additional SGBP(s) can also be encoded proximal to SGBP-A homologs in PULs (usually immediately downstream), although these proteins generally have such low sequence similarity and tertiary structural homology that they cannot be confidently identified as SGBPs by bioinformatic approaches. Of the functionally characterized representatives, the archetypal Sus system contains two such SGBPs (SusE and SusF) (107), whereas the XyGUL (99), heparin/heparan-sulfate PUL (103), xylan PULs (PUL-XylL and PUL-XylS) (110), and the MLGUL studied here, contain only one additional SGBP (denoted "SGBP-B" or "SusE-

positioned" in the literature). As exemplified by the structures of BoSGBP$_{MLG}$-B (Fig. 3-4), these proteins are generally comprised of a multi-domain "beads-on-a-string" arrangement that presents a critical carbohydrate-binding site on the C-terminal domain most distal to the N-terminal membrane-anchoring lipid.

Vis-à-vis BoSGBP$_{MLG}$-A, the role of the BoSGBP$_{MLG}$-B is less easily deduced in the context of the MLG utilization system encoded by the MLGUL. BoSGBP$_{MLG}$-B is dispensable for growth on MLG and is not able to compensate for the BoSGBP$_{MLG}$-A* mutant (Fig. 3-5). Although it is clearly highly specific for MLG, the affinity of BoSGBP$_{MLG}$-B for the cognate polysaccharide of the MLGUL is ca. 10-fold lower than that of BoSGBP$_{MLG}$-A. This comparatively weaker binding is consistent with XyGUL SGBPs but in the case of the Sus SGBPs, SusD (SGBP-A) and SusE (SGBP-B) have comparable affinities towards α-cyclodextrin while SusF (SGBP-C) has lower affinity (99,107,146). *In vivo*, SGBP-B-like proteins have been shown to play multiple roles during glycan capture. For example, the SGBP SusE influences the lengths of malto-oligosaccharide that can be taken up by the TBDT of the starch utilization system, SusC (108), while both SusE and SusF appear to offset the diffusion barrier at the cell surface established by the capsular polysaccharide, thereby aiding in starch capture (102).

The slight but reproducible increase in growth lag on pre-digested MLG (Fig. 3-6) suggests that BoSGBP$_{MLG}$-B functions in the capture of medium-length MLGOs, possibly including those in the extracellular environment that are not generated proximal to the cell surface by the BoGH16$_{MLG}$. This could allow *B. ovatus* to access MLGOs liberated from neighboring species. Indeed, work studying cooperative growth between *Bacteroides* species and other members of the microbiota on xylans, inulin, and dietary pectins has demonstrated that members of this genus benefit from the uptake of oligosaccharides released by neighbors (111,234). These studies have primarily focused on the roles of glycosidases in the communal breakdown of polysaccharides, however the potential contributions of glycan-binding proteins in these processes remain uncharacterized.

Our structural studies of BoSGBP$_{MLG}$-B revealed an additional aspect of MLG recognition that suggests a more complex interplay of molecular interactions at the cell surface than previously observed for this class of proteins. Specifically, the observation of bi-directional ligand recognition (Fig. 3-4D) *in crystallo* implies the possibility of avidity effect through dual binding of one MLG chain from opposite faces of the polysaccharide. This binding mode may be biologically relevant:

when the symmetry operation bisecting the ligand is applied to the entire $BoSGBP_{MLG}$-B molecule, both N-termini (through which the protein is anchored to the outer cell membrane via a canonical Cys lipidation site (175)) are orientated in the same direction (Fig. B-10). $BoSGBP_{MLG}$-B does not dimerize on its own in solution, as determined by size exclusion chromatography (data not shown), and there is minimal direct contact between the two symmetry related protein molecules *in crystallo*. Unfortunately, ITC analysis was unable to resolve the binding stoichiometry within the limits of the current data (Table 3-1). However, we also note that no ligand density is observed in protein molecules that do not participate in this sandwich interaction.

Together, biochemical, structural, and microbiological data for $BoSGBP_{MLG}$-A and $BoSGBP_{MLG}$-B suggest that the external recognition machinery for MLG acquisition is tailored for the capture of longer oligo/polysaccharides, rather than the limit-digest products (tri- and tetrasaccharides) generated by BoGH16 of the MLGUL (211). Other PUL-encoded SGBP systems similarly target longer partial-digest saccharides: XyGUL SGBPs display stronger affinity towards xyloglucan oligosaccharides with a backbone of eight Glc residues or more, versus limit-digest oligosaccharides with Glc_4 backbones (99); xylan PUL SGBPs bind xylo-oligosaccharides with affinities that increase with chain length (110); and the substrate binding affinity of the heparin/heparan sulfate PUL SGBP-B likewise increases with oligosaccharide degree of polymerization (103). Preferential targeting and transport of longer oligosaccharides pre- and post-hydrolysis by the vanguard endo-glycanase encoded by a PUL is anticipated to be an effective strategy for *Bacteroides* species to rapidly acquire multiple glucose-equivalents with minimal loss in a competitive environment.

The present study also provides specific molecular insight into the selectivity of the SGBPs of the MLGUL for mixed-linkage β(1,3)/β(1,4)-glucans from cereal grains over all-β(1,4)-glucans, i.e. cellulose and cello-oligosaccharides. Notably, *B. ovatus* does not grow on crystalline cellulose (109) and exhibits a significant growth lag on soluble cello-oligosaccharides ((109) and Fig. 3-5E). Comparison of the MLGO and cello-oligosaccharide complex structures of both $BoSGBP_{MLG}$-A and $BoSGBP_{MLG}$-B (Fig. 3-3, Fig. 3-4) reveals that shape complementarity of the glycan with the binding platforms rationalizes the primary specificity for MLG (Fig. 3-2, Table 3-1).

On one hand, the faces of the aromatic platform residues are angled with respect to each other in both SGBPs (Fig. 3-3D, 3-3E, 3-3G; Fig. 3-4B, 3-4C, 3-4F), which complements the

twisted conformation that β-glucans natively adopt in solution (235). In comparison, canonical cellulose-binding modules ("type A" CBMs) generally present aromatic sidechains in a flat, co-linear arrangement that matches the planar cellulose crystal (76). On the other hand, the non-linear topology of the binding surface of both SGBPs is complementary to the intrinsically bent shape of the MLG chain. In the case of $BoSGBP_{MLG}$-A, this enables optimal interaction with all four aromatic sidechains of the binding platform, which is then unavailable to the strictly linear β(1,4)-linked cello-oligosaccharides (Fig. 3-3B, 3-3C). In the case of $BoSGBP_{MLG}$-B, an inherent curvature of the binding surface likewise complements the twist of the MLG chain. (Fig. 3-4F). Similar conformation-dependent specificity was observed for the archetypal SGBP-A homolog, SusD, which bound cyclodextrins with higher affinity than linear malto-oligosaccharides, due to an arced binding platform evolved to fit the natural helical conformation of amylose (146). Thus, although both $BoSGBP_{MLG}$-A and $BoSGBP_{MLG}$-B broadly recognizes polysaccharides containing contiguous β(1,4)-linked glucosyl residues, they do so with significantly lower affinity than for MLG (Fig. 3-2, B-6, Table 3-1). We anticipate that future structural studies may reveal contrary specificity determinants in SGBPs that predominantly bind cellulose (236-238).

Although $BoSGBP_{MLG}$-A and $BoSGBP_{MLG}$-B exhibit some degree of off-target affinity vis-à-vis the cognate substrate of the MLGUL, this is unlikely to be evolutionarily disadvantageous in the context of the human gut where a diverse array of dietary glycans are present. Indeed, as one component of the plant cell wall, MLG is generally found associated with cellulose and other hemicelluloses (239), such that non-specific binding may improve bacterial adhesion to insoluble particles. Additionally, in the context of the gut ecosystem, the ability to scavenge the products of possible para-crystalline cellulose degradation may allow *B. ovatus* to compete for an additional, privileged nutrient niche, as many gut species lack the ability to utilize cellooligosaccharides larger than cellobiose (109,123,240).

3.5 Conclusion

Effective manipulation of the gut microbiota for therapeutic purposes – a topic of growing recent interest – will be significantly informed by a holistic understanding of the metabolism of the microbiota, including the mechanisms of complex dietary polysaccharide utilization which fuel this ecosystem (204,205,209,241). The present study reveals the essential roles that two SGBPs play in cereal β-glucan utilization by working in concert with the glycoside hydrolases and TBDT of the MLGUL from the human symbiont *B. ovatus*. On one hand, this study contributes directly

to the currently limited but growing pool of structure-function relationships among SGBPs. On the other, the identification of syntenic MLGUL in other members of the human gut microbiota indicates that this data may be extrapolated more broadly in metagenomic analyses (138,211,242). Expanding our knowledge of the interplay of SGBP and other PUL components is a critical step towards developing novel strategies to manipulate microbial communities (243-245), in the human gut and beyond.

Chapter 4: Synergy between cell-surface glycosidases and glycan-binding proteins dictates the utilization of specific beta(1,3)-glucans by human gut *Bacteroides*

4.1 Introduction

The human gut microbiota (HGM) is a complex community that underpins our nutrition and overall well-being (246,247), yet is also associated with some diseases (7,248-250), depending on its particular composition and physiology. A key challenge in manipulating HGM dynamics toward healthful outcomes is a limited understanding of the ecological forces that shape this community within individuals (251,252). The catabolism of complex dietary carbohydrates is a key driver of HGM structure and metabolic function (10,13). Thus, resolving a detailed roadmap of the glycan utilization mechanisms deployed by individual members of the HGM is central to the development of dietary and microbial interventions to promote human health.

An explosion of (meta)genome sequence data continues to reveal substantial taxa-level variation in the metabolic potential of the HGM, yet a lack of functional data restricts our ability to fully explain or predict these differences and eventually to use this knowledge to engineer changes to the HGM. *Bacteroides* species, in particular, are predominant autochthonous members of the HGM that metabolize a wide variety of complex glycans into short-chain fatty acids (SCFAs) (253), which is enabled by a plethora of PULs in their genomes (254). As exemplified by the β(1,3)-glucan utilization loci (1,3GULs) elucidated here, PULs encode concerted molecular systems of surface-glycan binding proteins (SGBPs), carbohydrate-active enzymes (CAZymes), TonB-dependent transporters (TBDTs), and sensor/regulators to recognize, capture, import, and saccharify individual substrates (105) (Fig. 4-1). Recently, a number of seminal integrated PUL studies combining genetics, biochemistry, and structural biology have highlighted how strain-level genomic variation dictates nutrient specificities (100-103,110,139,140,194,211,255-259).

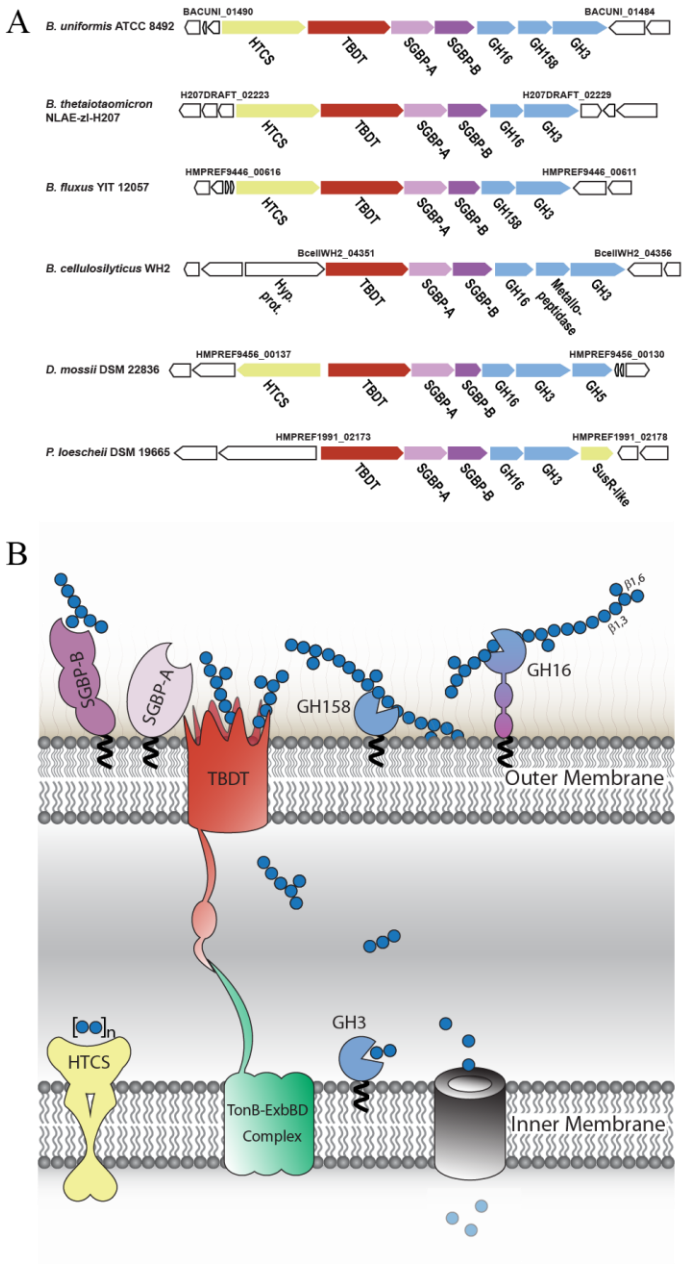

Figure 4-1. β(1,3)-glucan utilization systems in the Order *Bacteroidales*.

(A) Strain-specific, syntenic β(1,3)-glucan utilization loci (1,3GUL) from *Bacteroides uniformis*, *B. thetaiotaomicron*, *B. fluxus*, *B. cellulosilyticus*, *Dysgonomonas mossii*, and *Prevotella loescheii*. Genome locus tags constituting 1,3GUL boundaries are indicated. Predicted or confirmed (this work) functional annotations are denoted below each gene: HTCS, hybrid two-component system sensor/regulator; Sus-R like sensor-regulator; TBDT, SusC-like TonB-dependent transporter; SGBP-A, SusD-like cell-surface glycan-binding protein, SGBP-B, sequence-divergent cell-surface glycan-binding protein; GH*n*, member of Glycoside Hydrolase family *n*. (B) Model of the *B. uniformis* laminarin utilization based on the present study and by analogy with the archetypal starch utilization system (Sus)(13,105). Gene products are colored analogously to panel A, and predicted N-terminal lipidation following signal peptidase II cleavage is indicated with a black squiggle.

Amorphous β-glucans are ubiquitous polysaccharides in the human diet, which can be delineated broadly by backbone linkage (Fig. 4-2A): Mixed-linkage β(1,3)/β(1,4)-glucans (MLGs) are commonly found in cereal crops such as oats and barley, while the β(1,3)-glucan callose is found as a component of plant cell walls (260). (Edible) fungi, including yeasts, contain β(1,6)-glucans (256) and β(1,3)-glucans (260). β(1,3)-glucans also occur in the cell walls of seaweeds (261). Many β(1,3)-glucans also contain β(1,6)-linked branches, the length and frequency of which varies according to source (*e.g* yeast β-glucan, laminarin) (260). It is known generally that a range of taxa in the HGM, including from the phyla *Firmicutes*, *Actinobacteria*, and *Bacteroidetes*, metabolize β-glucans to produce short-chain fatty acids (118,211,256,262,263). In addition to this nutritional benefit (250), β-glucans have been associated with health-promoting effects against cancer, diabetes/metabolic syndrome, and inflammation (141,264-266). However, the molecular mechanisms underpinning these effects have not been fully elucidated.

Previous studies on MLG and β(1,6)-glucan utilization by the symbiotic *Bacteroides*, revealed the molecular details by which the different, dedicated PUL-encoded machineries target these two distinct classes of β-glucans (100,211,256). Here we functionally dissected an exemplar 1,3GUL from *Bacteroides uniformis* ATCC 8492 to provide molecular insight into β(1,3)-glucan utilization, thereby resolving a key outstanding deficit in our understanding of β-glucan metabolism by the HGM. Notably, this included solving the first tertiary structure and resolving the catalytic mechanism of a member of the new glycoside hydrolase family 158 (GH158). Building upon these foundational results, we subsequently demonstrated that the individual abilities of three *Bacteroides* species to metabolize distinct β(1,3)-glucans and/or MLG is dictated by the cumulative specificities and contributions of their respective SGBPs, cell-surface GHs, and other sensor/transport functions encoded by partially homologous 1,3GULs. Finally, we found through metagenomic analysis that prevalence of 1,3GULs in human gut microbiomes is species-dependent but broadly distributed worldwide.

4.2 Materials and Methods

4.2.1 Substrates

Tamarind seed xyloglucan, barley beta-glucan (267), konjac glucomannan, carob galactomannan, *Alcaligenes faecalis* curdlan, yeast beta-glucan (268) were purchased from Megazyme International (Bray, Ireland). Carboxymethyl cellulose was purchased from Acros Organics (Morris Plains, NJ, and USA). Hydroxyethyl cellulose was purchased from Amresco

(Solon, OH, USA). *Xanthomonas campestris* xanthan gum was purchased from Spectrum (New Brunswick, NJ, USA). Laminarins from *Laminaria digitate* (269) and *Eisenia bicyclis* (270) were purchased from Sigma Aldrich (St. Louis, MO, USA) and Carbosynth (Compton, UK), respectively. Laminarin from *Laminaria digitata* was reduced to laminaritol as described previously (271), to reduce background response in the BCA reducing-sugar enzyme kinetics assay.

Laminaribiose (G3G), laminaritriose (G3G3G), laminaritetraose (G3G3G3G), laminaripentaose (G3G3G3G3G), mixed-linkage glucotriose A (G3G4G), mixed-linkage glucotriose B (G4G3G), cellotriose (G4G4G) were purchased from Megazyme. Gentiobiose (G6G) was purchased from Carbosynth (Compton, UK). Cellobiose (G4G) was purchased from Acros Organics. G3G-CNP was synthesized by glycosylation of α-laminaribiosyl bromide(272) with 2-chloro-4-nitrophenol under phase-transfer conditions (152,272), the details of which will be published elsewhere.

4.2.2 Bacterial growth experiments

Bacteroides uniformis ATCC8492 (NCBI:txid411479, JGI:641380447), *Bacteroides thetaiotaomicron* NLAE-zl-H207 (NCBI:txid818, JGI:2515154063), and *Bacteroides fluxus* YIT12057 (NCBI:txid763034, JGI:651324011) were grown in tryptone yeast extract glucose (TYG) medium at 37 ºC under anaerobic conditions (Coy anaerobic chamber; 85 % N_2, 10 % H_2, 5 % CO_2). These cultures were used to inoculate Minimal Medium containing glucose as the sole carbon source (MM-Glc), followed by incubation at 37 ºC for 20 h. At this time, 1 mL samples were centrifuged and the bacteria pellets was gently resuspended in MM containing no carbohydrate (MM-NC). These suspensions were diluted 1:50 in MM-NC before being used to inoculate MM containing 0.5 % of carbohydrate. Growth experiments were performed in replicates of 6 (laminarin, yeast β-glucan, barley β-glucan, and curdlan) or 3 (glucose and H_2O) in 96-well plates. The growth was monitored by measuring A_{600} and data were processed using Gen5 software (BioTek). Growth was quantified in each assay by first identifying a minimum time point (A_{min}) at which A_{600} had increased by 10% over a baseline reading taken during the first 500 min of incubation. Next, the time point was identified at which A_{600} reached its maximum (A_{max}) immediately after the exponential growth. The growth rate for each well was defined by (A_{max} - A_{min}) / (T_{max} - T_{min}), where T_{max} and T_{min} are the corresponding time values for each absorbance.

Cultures where the density did not increase by at least 0.1 (A_{600}) were considered to have no growth.

4.2.3 Quantitative RT-PCR

B. uniformis was capable of growth on all carbon sources of interest in this study, and was therefore cultured directly in 3 ml of MM containing 0.5% (w/v) carbohydrate, as described above. *B. thetaiotaomicron* and *B. fluxus* were precultured on MM-Glc, pelleted, washed, and resuspended twice in MM-NC, and inoculated to A_{600} ~ 0.3 in 4 ml of MM containing 0.5% (w/v) carbohydrate. Bacterial cultures were harvested in triplicate (at mid-log phase (A_{600} ~ 0.6) for *Bu* and after 5 hours of incubation for *Bt* and *Bf*), placed in RNA protect (Qiagen) for immediate stabilization of RNA, and then stored at -20 ºC. RNA was extracted and purified with the RNeasy mini kit (Qiagen), and RNA purity was assessed by spectrophotometry. 1 μg of RNA was used for reverse transcription and synthesis of the cDNA (SuperScript VILO™ Master Mix, Invitrogen). Quantitative PCR reactions (20 μl final volume) using specific primers were performed with SensiFast SYBR Lo-ROX kit (Bioline) on a 7500 Fast Real-Time PCR System (Applied Biosystems) (Table C-9). Data were normalized to 16S rRNA transcript levels, and changes in expression level were calculated as fold change compared with cultures of MM containing glucose.

4.2.4 Bioinformatics, gene cloning and site-directed mutagenesis

Potential 1,3GULs were identified by homology searches of sequences available in the Joint Genome Institute Integrated Microbial Genomes and Metagenomes (JGI-IMG/M) Database (273) and other sources (123). Signal peptides and subcellular localization were predicted by protein sequence analysis (174,213,274). The gene fragments corresponding to BACUNI_01484 to BACUNI_01488 (encoding *Bu*GH3, *Bu*GH158, *Bu*GH16, *Bu*SGBP-B, and *Bu*SGBP-A), H207DRAFT_02225 to H207DRAFT_02227 (encoding *Bt*SGBP-A, *Bt*SGBP-B, and *Bt*GH16), HMPREF9446_00612 to HMPREF9446_00614 (encoding *Bf*GH158, *Bf*SGBP-B, and *Bf*SGBP-A), HMPREF1991_02176 (encoding *Pl*GH16), were PCR amplified from genomic DNA using the Q5 high fidelity polymerase (NEB) with primers designed to exclude signal peptides and lipidation cysteines (174,175) (Table C-10). The PCR introduced SalI/XhoI (NEB) restriction sites to the flanks of BACUNI_01484 and _01486 target genes and the amplified DNA products were ligated into the expression vector pET28a such that the encoded recombinant proteins contain an N-terminal His_6-tag. The rest of the PCR products contained appropriate pMCSG complementary sequences for subsequent ligation independent cloning into pMCSG53 or pMCSG-GST plasmids

providing an N-terminal His_6-tag or an N-terminal His_6-GST-tag (275). Successful cloning was confirmed by colony PCR (GoTaq polymerase from Promega) and sequencing (Genewiz). The site-directed mutants (276) *Bu*GH158 N136A, E137A, and E220A were generated using pMCSG53::*Bu*GH158 as a template DNA (Table C-11).

4.2.5 Recombinant protein production and purification

Recombinant proteins were produced in *E. coli* BL21 (DE3) cells cultured in TB broth containing ampicillin (50 $\mu g.ml^{-1}$) or kanamycin (50 $\mu g.ml^{-1}$) at 37 °C (200 rpm). Cells were grown to mid-exponential phase (OD_{600} ~ 0.4 to 0.6). Overexpression was induced by adding isopropyl β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM and the cultures were further grown at 16 °C (200 *rpm*) for 18 h. The cells were harvested by centrifugation, sonicated and His_6-tagged recombinant proteins were purified via immobilized nickel affinity chromatography (His-Trap; GE Healthcare) utilizing a gradient elution up to 100% elution buffer containing 20 mM sodium phosphate, pH 7.4, 500 mM NaCl, and 500 mM imidazole in an BioLogic FPLC system (BioRad). The purity of the recombinant proteins was determined by SDS/PAGE (Fig. C-15, C-16) and their concentrations were determined from calculated molar extinction coefficients at 280 nm using an Epoch Microplate Spectrophotometer (BioTek).

Selenomethionine-labelled protein was produced by inhibition of methionine biosynthesis in *E. coli* BL21 (DE3) (214). Briefly, cells were grown in 1 L of M9 minimal media supplemented with 100 $\mu g\ ml^{-1}$ ampicillin at 37 °C with shaking until OD_{600} reached 0.6. At this point, 100 mg each of L-lysine, L-threonine, L-phenylalanine, and 50 mg each of L-leucine, L-isoleucine, L-valine, and L-selenomethionine were added to the media and shaken for a further 15 minutes before inducing expression with 0.5 mM IPTG. The culture was transferred to 16 °C and incubated for an additional 24 hours. Nickel affinity chromatography was conducted as described above using HEPES buffer instead of sodium phosphate. Native and selenomethionine-labelled proteins for crystallography were further purified by size exclusion chromatography through a Superdex 75 resin (GE Healthcare Life Sciences) in a XK 16/100 column (GE Healthcare Life Sciences) run in 10 mM HEPES pH 7.0 at 0.8 mL min^{-1}.

4.2.6 Affinity gel electrophoresis

Affinity PAGE was performed for 180 min at 80 V and room temperature on non-denaturing 10% (w/v) polyacrylamide gels containing a polysaccharide concentration of 0.1% (wt/vol) (or water for control), essentially as previously described (99,100,138). 5 μg of the tested

SGBP proteins, along with bovine serum albumin (BSA) used as non-interacting negative control protein were loaded on the gels.

4.2.7 Isothermal titration calorimetry

Isothermal titration calorimetry was performed using the MicroCal VP-ITC titration calorimeter equilibrated to 25 °C, essentially as previously described (99,100,138). The proteins (20 to 100 μM) were placed in the sample cell, and the syringe was loaded with 2.5 mg/ml polysaccharide or 0.5 to 2 mM oligosaccharide. Following an initial injection of 2 μl, 25 subsequent injections of 10 μl were performed with stirring at 280 rpm, and the resulting heat of reaction was recorded. Integrated heats were fit to a single-site model using Microcal Origin v7.0 to derive n, K_a, and ΔH values.

4.2.8 Carbohydrate analytical method

High-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) was performed on Carbopac PA200 guard and analytical columns connected in series on a Dionex ICS-5000 HPLC system operated by Chromeleon software version 7 (Thermo Scientific), essentially as previously described (181). Solvent A was ultrapure water, solvent B was 1 M sodium hydroxide, and solvent C was 1 M sodium acetate (anhydrous Bio Ultra-grade, Sigma-Aldrich). The injection volume was 10 μL and the gradient comprised: 0 – 5 min, 10 % B and 3.5% C; 5 – 12 min, 10% B, linear gradient from 3.5 – 30 % C; 12.0 –12.1 min, 50 % B, 50 % C; 12.1 – 13.0 min, exponential gradient (curve setting 9) of B and C back to initial conditions; 13 – 17 min, initial conditions.

4.2.9 Enzymatic assays

Polysaccharide hydrolysis was quantified using a bicinchoninic acid (BCA) reducing-sugar assay (277). Assays were conducted in a final volume of 100 μL at the optimum pH and 37 °C for 10 min. Reactions were terminated by addition of an equal volume (100 μL) of BCA reagent. Color was developed by heating to 80 °C for 20 min before reading A_{563}. Glucose (25–150 μM) was used to generate a standard curve for quantitation. The pH and temperature optimum of each enzyme was initially determined using the same enzyme reaction assay to quantify reducing ends over 10 min of incubation with 1.0 mg/mL laminarin in different buffers at 50 mM: sodium citrate (pH 3.0 – 6.5), sodium phosphate (pH 6.5 – 8.0), glycine (pH 9.0 – 10.5). To determine Michaelis-Menten parameters, different concentrations of polysaccharide solutions were used over the range 0.025–

3 mg/mL with the appropriate concentration of enzyme for 10 min and reducing ends released were quantified as described above.

The release of glucose from oligosaccharides was quantified using the D-Glucose HK Assay Kit from Megazyme (Bray, Ireland), modified for use as a continuous assay exactly as previously described (211).

To measure enzyme activity on chromogenic glycosides, the release of *para*-nitrophenyl was monitored by following A_{405} in a 1 cm path length quartz cuvette with a Cary 60 UV-Vis spectrophotometer (Agilent Technologies). Reactions in 250 μL at the optimum pH and 37 °C were assayed with nine different substrate concentrations and rate was calculated using an extinction coefficient determined according to the buffer used. Endpoint assays were used for pH and temperature optima (same range as described above) of *Bu*GH3. Reactions were terminated after 10 minutes by addition of 100 μL of 1 M Na_2CO_3 to raise the pH and absorbance at A_{405} was measured in 96-well plates on an Epoch Microplate Spectrophotometer (BioTek). An extinction coefficient of 18,100 M^{-1} cm^{-1} was used for these assays. Continuous assays were used for initial-rate saturation kinetics and reactions, initiated by adding 25 μL of enzyme solution to 225 μL of the remaining assay mixture, in the optimum pH buffer at 37 °C.

4.2.10 Crystallization and structure determination

Initial sitting drop crystal screens were set up in 96-well plates using a Phoenix robot (Art Robbin Instruments) and were stored at room temperature. A hit was obtained in pHClear (Qiagen) condition E12 (0.1 M bicine pH 9.0, 1.6 M ammonium sulfate) with purified *Bu*GH158 at 22.7 mg mL^{-1}, and optimized in larger hanging drops in a grid screen by varying pH and ammonium sulfate concentrations in 24-well plates. Crystals were cryoprotected in crystallization solution supplemented with 2 M lithium acetate before flash freezing with liquid nitrogen, and diffraction data were collected at the Advanced Photon Source (APS) beamline 23ID-D. Selenomethionine-labelled crystals were obtained in the same grid screen as the native crystals but required macroseeding with fine needles in order to obtain crystals of sufficient thickness. Crystals were cryoprotected in crystallization solution supplemented with 4 M lithium chloride before flash freezing with liquid nitrogen. Diffraction data were collected at the Stanford Synchrotron Radiation Lightsource (SSRL) beamline 9-2 at the selenium absorption peak, inflection, and high energy remote wavelengths determined by a fluorescence scan. Datasets were indexed and integrated with XDS (153), space groups were determined with Pointless (278), and data reduction

was performed with Aimless (279). Phasing by multiple anomalous dispersion, density modification, and initial model building were performed in AutoSol (216), within the Phenix suite (217). This selenomethionine-labelled structure was used as the search model for molecular replacement with the native crystal data using Phaser (157). After initial refinement in Phenix.refine (218), iterative rounds of manual model building and refinement were conducted with Coot (280) and Refmac5 (161), respectively, within the CCP4i2 suite (221). The quality of the model was monitored throughout using Molprobity (163). A structure similarity search was conducted using the Dali server (231).

4.2.11 ^{1}H-NMR determination of catalytic mechanism

*Bu*GH158 in 50 mM sodium phosphate pH 7.0 and laminaribiose-β-CNP were independently lyophilized and resuspended in 99.9% D_2O. After recording an initial ^{1}H-NMR spectrum of the substrate (Bruker Avance 400 MHz spectrometer), the enzyme was added to obtain final concentrations 20 μM *Bu*GH158 and 10 mM laminaribiose-β-CNP. Spectra were recorded at appropriate time intervals thereafter to observe the first-formed product anomer and subsequent mutarotation.

4.2.12 Survey of human metagenomic datasets

Available cohorts of human gut metagenomic sequence data (National Center for Biotechnology Information projects: PRJNA422434 (172), PRJEB10878 (281), PRJEB12123 (282), PRJEB12124 (283), PRJEB15371 (284), PRJEB6997 (285), PRJDB3601 (286), PRJNA48479 (287), PRJEB4336 (288), PRJEB2054 (170), PRJNA392180 (289), and PRJNA527208 (290)) were searched for the presence of 1,3GUL nucleotide sequences from *B. fluxus* (12.5kb), *B. thetaiotaomicron* (12.6kb), and *B. uniformis* (14.5kb) using the following workflow: Each 1,3GUL nucleotide sequence was used separately as a template and then magic-blast v1.5.0 (291) was used to recruit raw Illumina reads from the available metagenomic datasets with an identity cutoff of 97%. Next, the alignment files were used to generate a coverage map using bedtools v2.29.0 (292) to calculate the percentage coverage of each sample against each individual reference. We considered a metagenomic data sample to be positive for a particular 1,3GUL if it had at least 70% of the corresponding 1,3GUL nucleotide sequence covered (since the three 1,3GUL sequences were very similar in size, no normalization was made for PUL template variation).

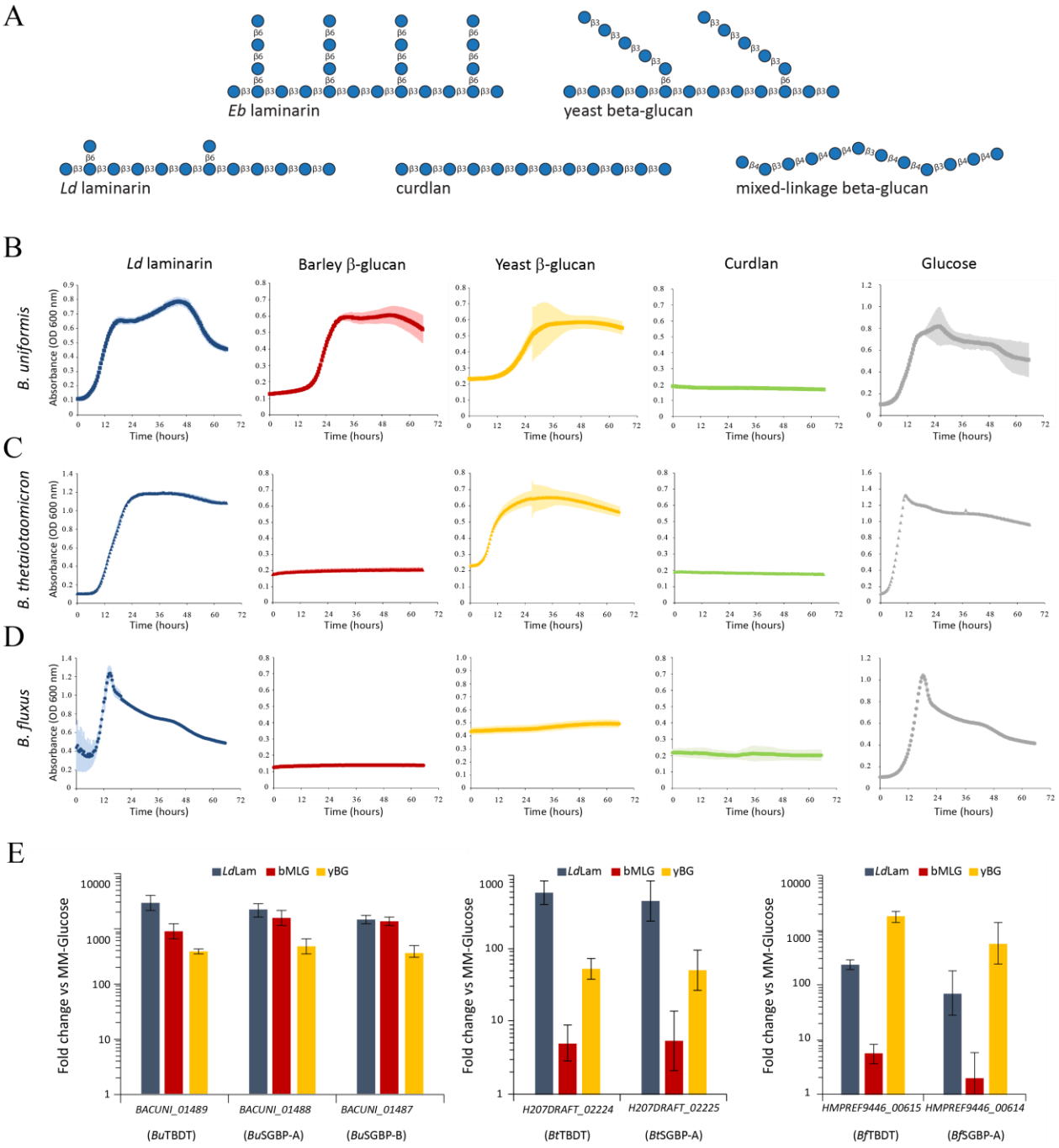

Figure 4-2. β(1,3)-glucan utilization by *Bacteroides*.

(A) Structures of β-glucans used in this study. Bacterial curdlan is a representative unbranched β(1,3)-glucan, analogous to plant callose. Algal laminarins are β(1,3)-glucans with β(1,6)-linked branches. *Laminaria digitata* laminarin has infrequent, single β(1,6)-glucosyl branches (β(1,3) to β(1,6) molar ratio 7:1), while *Eisenia bicyclis* laminarin has a high frequency of branches with degrees of polymerization of up to three β(1,6)-glucosyl residues (β(1,3) to β(1,6) molar ratio 3:2). Yeast β(1,3)-glucan contains longer β(1,3)-glucan branches extended from β(1,6)-linked branch points. Cereal mixed-linkage β(1,3)/β(1,4)-glucans (MLGs) are linear chains of β(1,4)-linked

cellotriosyl and cellotetraosyl units linked by β(1,3) bonds. (B), (C), and (D) Growth curves of *B. uniformis* ATCC 8492, *B thetaiotaomicron* NLAE-zl-H207, and *B. fluxus* YIT 12057, respectively, in minimal medium containing the indicated carbon source at 0.5% w/v. Points represent averages of $n = 6$ technical replicates (microplate wells) for polysaccharides, $n = 3$ for glucose; error bars represent the standard error of the mean. (E) RNA abundance for core PUL genes of *Bacteroides* spp. quantified by qRT-PCR. Bacteria were grown to mid-log phase in minimal medium containing glucose as the sole carbon source and subsequently exposed to different β-glucans R ($n = 3$, expression measurements from individual cultures, relative to glucose control; error bars represent the standard errors of the mean).

4.3 Results

4.3.1 *B. uniformis* possesses a distinct PUL upregulated during growth on β(1,3)-glucans

B. uniformis ATCC 8492 (hereafter, *Bu*) exhibited robust growth in minimal medium containing either glucose or the branched β(1,3)-glucans from *Laminaria digitata* (*Ld* laminarin, *Ld*Lam, which contains single β(1,6)-linked glucosyl branches) and *Saccharomyces cerevisiae* (yeast β-glucan, yBG, which contains longer β(1,6)-linked glucosyl branches extended by β(1,3)-linked glucose units). No growth was observed on *Alcaligenes faecalis* curdlan (an unbranched β(1,3)-glucan analogous to plant callose), which is poorly water-soluble and forms a gel in aqueous suspension (Fig. 4-2B, Table C-1). Interestingly, *Bu* also grew well on barley MLG (bMLG, which has a ~2.5:1 ratio of β(1,4) to β(1,3) backbone linkages) (Fig. 4-2B, Table C-1), despite lacking a canonical mixed-linkage glucan utilization locus (MLGUL) homologous to the *B. ovatus* MLGUL (211).

We identified in *Bu* a putative β(1,3)-glucan utilization locus that encodes a TonB-dependent transporter (TBDT, a SusC homolog, BACUNI_01489) and a cell-surface glycan-binding protein (SGBP-A, a SusD homolog, BACUNI_01488) as canonical PUL signatures (109), as well as an additional non-homologous SGBP (*Bu*SGBP-B, BACUNI_01487), three glycoside hydrolases (GH16 subfamily 3, GH158, and GH3; BACUNI_01486 to _01484), and a hybrid two-component system (HTCS) transcriptional regulator (BACUNI_01490) (Fig. 4-1). In particular, the GH complement was suggestive of a role in β(1,3)-glucan hydrolysis: GH16 subfamily 3 (GH16_3 (56)) contains known *endo*-laminarinases (among other *endo*-β-glucanases) (211) and GH3 contains *exo*-β-glucosidases (among others) (293). Notably, during the course of this study, GH158 emerged as a new family whose founding member was shown to hydrolyze a chemically derivatized β(1,3)-glucan in a high-throughput screen (294). Concordant with this proposed PUL

specificity, when we probed expression of core genes encoding the TBDT and both SGBPs (BACUNI_01487-01489) we found that they were strongly upregulated in the presence of laminarin, bMLG and yBG as sole carbon sources versus a glucose control (Fig. 4-2E). The syntenic TBDT and SGBP-A (SusC/SusD-homologs) in a partially homologous *B. cellulosilyticus* WH2 PUL (Fig. 4-1) were shown previously to be similarly upregulated in the presence of laminarin and bMLG (123). Notably, *B. ovatus* ATCC 8483, which possess a MLGUL(211), but not a homologous 1,3GUL (Fig. 4-1), does not grow on laminarin (109).

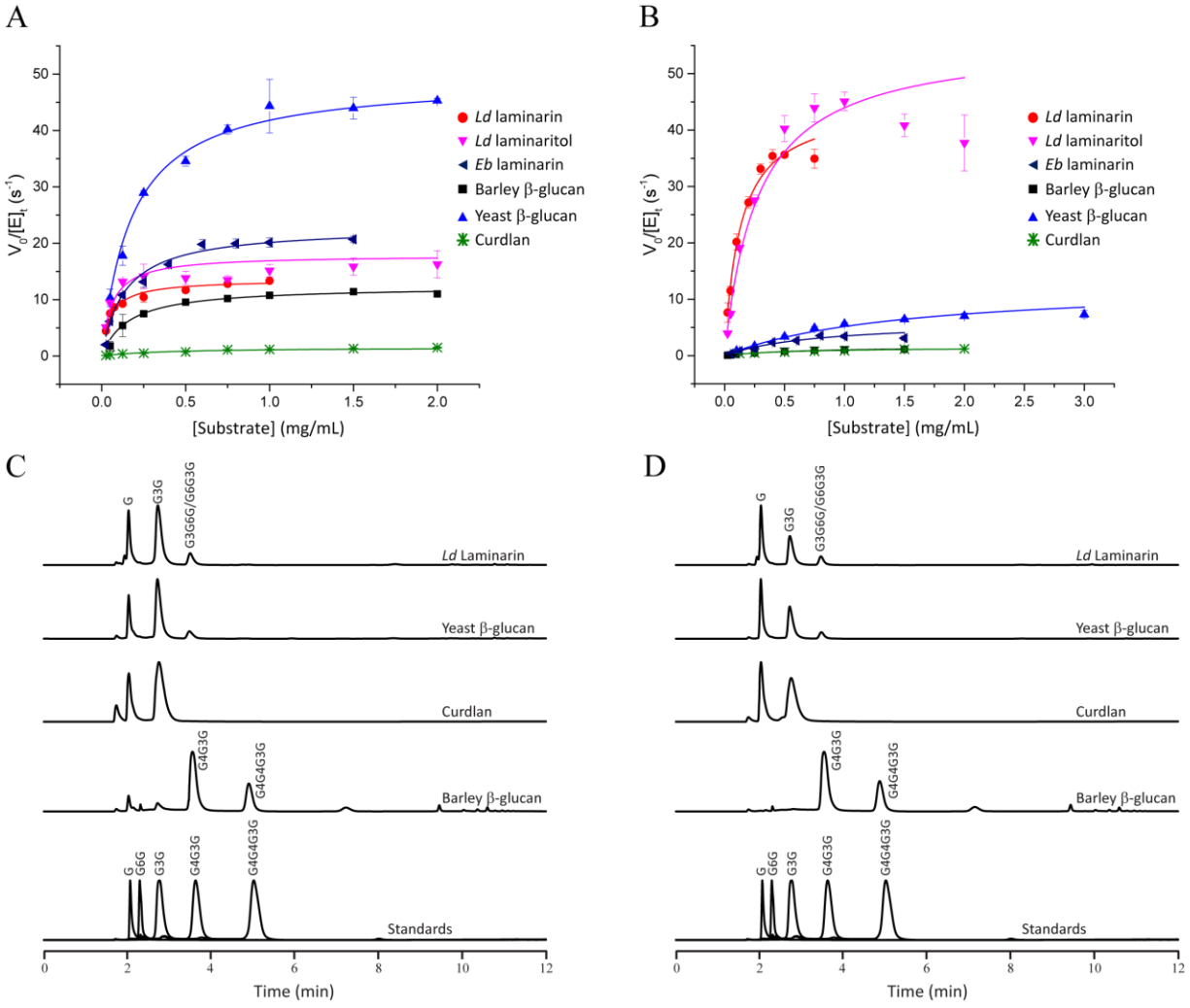

Figure 4-3. Hydrolysis of β-glucans by *Bu*GH16 and *Bu*GH158.

(A) and (B) Initial-rate kinetics analysis of *Bu*GH16 and *Bu*GH158, respectively. Curves represent fits of the Michaelis-Menten equation to the average data points (n = 3); error bars represent standard deviations from the mean. (C) and (D) HPAEC-PAD analysis of limit-digest products of *Bu*GH16 and *Bu*GH158, respectively, at 37 ºC.

4.3.2 Biochemical basis of β(1,3)-glucan recognition and degradation by *B. uniformis*

By analogy with other PUL-encoded systems, we propose a working model of concerted β(1,3)-glucan saccharification and uptake by the proteins of the 1,3GUL (Fig. 4-1). This model involves polysaccharide capture at the cell surface by at least one SGBP, backbone hydrolysis by at least one *endo*-glucanase, transport of oligosaccharide fragments through the outer membrane by the TBDT, and ultimate saccharification by an *exo*-β-glucosidase in the periplasm. Indeed, signal peptide analysis with SignalP, LipoP, and PSORTb predicted that *Bu*GH16 is localized at the cell surface via N-terminal Cys lipidation (PSORTb score 9.7), whereas *Bu*GH3, which also harbors a type-II signal peptide, is predicted to be periplasmic (PSORTb score 9.4). Interestingly, PSORTb was unable to predict the localization of *Bu*SGBP-A, *Bu*SGBP-B, and *Bu*GH158 (score < 2.5), despite the presence of a type-II signal peptide and a +2 serine residue in each, and expected extracellular localization of at least the SGBPs by analogy with the archetypal starch utilization system (10,13).

4.3.2.1 *Bu*GH16 is a broad-specificity endo-β(1,3)/β(1,4)-glucanase

To investigate the catalytic role of *Bu*GH16 in surface polysaccharide break down, hydrolytic activity of the recombinant protein was screened against a library of polysaccharides (Table C-2). *Bu*GH16 displayed activity towards the β(1,3)-glucans *Ld*Lam, *Eisenia bicyclis* laminarin (*Eb*Lam, which contains more frequent β(1,6)-linked glucosyl branches of up to three β(1,6)-linked glucosyl residues (270), Fig. 4-2A), yBG, and curdlan. *Bu*GH16 was also active on bMLG. Using laminarin as substrate, the pH optimum was 6.0 (consistent with extracellular function in the human large intestine (295)) and the maximum temperature of activity was ca. 50 °C (Fig. C-1).

Subsequent Michaelis-Menten kinetics at the optimum pH and 37 °C confirmed that *Bu*GH16 is a broadly specific β(1,3)-glucanase with similar catalytic efficiencies towards laminarins and yBG (Fig. 4-3A, Table C-2). The k_{cat}/K_m value of *Bu*GH16 on bMLG was 4-fold lower than for these all-β(1,3)-backbone glucans (Fig. 4-3A, Table C-2,), and the enzyme was very poorly active on the unbranched β(1,3)-glucan, curdlan (k_{cat}/K_m value ca. 1% of that on the laminarins; Fig. 4-3A, Table C-2). *Ld*Lam and yBG were ultimately hydrolyzed to glucose, laminaribiose and a trisaccharide (G3G6G or G6G3G), as determined by HPAEC-PAD (Fig. 4-3C) and confirmed by MALDI-MS of the per-*O*-acetylated crude mixture (data not shown). The limit-digest products were glucose and laminaribiose from curdlan, and G4G3G and G4G4G3G

from bMLG (Fig. 4-3C). Analyses of the hydrolysis products over time suggest that *Bu*GH16 proceeds by an *endo*-dissociative mode of action (Fig. C-2).

Despite extensive attempts using the full-length protein and the GH16 module only (Fig. C-3), we were not able to obtain a crystal structure of *Bu*GH16 to explain the observed substrate promiscuity. However, protein phylogeny unambiguously places *Bu*GH16 in the "β-bulge laminarinase/MLGase" clade (Fig. C-4), corresponding to GH16 subfamily 3 (56), which is known to contain laminarinases with secondary MLGase activity, as well as *bona fide* MLGases (211). These enzymes require a β(1,3) linkage between the -1 and -2 subsites (30), but beyond the -2 subsite, requirements are less stringent, leading to widespread promiscuity in polysaccharide and bond-cleavage specificity (β(1,3) *vs.* β(1,4)) (211). Notably, structural homology modelling and superposition with an exemplar laminarinase from the marine bacterium *Zobellia galactanivorans* DsijT (296) rationalizes the ability of *Bu*GH16 to accommodate highly branched β(1,3)-glucans, *i.e.* laminarins and yBG (Fig. C-5).

In addition to the catalytic module, *Bu*GH16 comprises a pair of PFAM 13004 domains (Fig. C-3). Despite initial bioinformatics predictions, carbohydrate binding has not been demonstrated for any PFAM 13004 domain to date (138). Likewise, affinity gel electrophoresis (AGE) analysis of a recombinant protein consisting the two PFAM 13004 domains of *Bu*GH16 revealed that they do not bind cognate polysaccharides of the 1,3GUL (Fig. C-6). These domains likely serve a spacer function, as in BoGH5 (138) and analogous to all-β-sheet domains in SGBPs (99,100,102).

4.3.2.2 *Bu*GH158 is a strictly specific, retaining *endo*-β(1,3)glucanase with a TIM-barrel fold

*Bu*GH158 is a member of a newly established GH family, the distantly related founding member of which, Vvad_PD1638 (sequence identity 29%, Fig. C-7) was shown to be active on the artificial proxy substrate, carboxymethyl-curdlan, in a high-throughput screen (294). Hence, we performed detailed kinetic and product analysis to more precisely delineate the specificity of *Bu*GH158 in the context of the 1,3GUL (Fig. 4-3B, Table C-2). In contrast to *Bu*GH16, *Bu*GH158 is highly specific for *Ld*Lam, with a k_{cat}/K_m value ca. 2 orders of magnitude higher than for *Eb*Lam, yBG, MLG, and curdlan (Fig. 4-3B, Table C-2; see also Fig. C-1 and C-2). The corresponding hydrolysis products were identical to those of *Bu*GH16 (Fig. 4-3D).

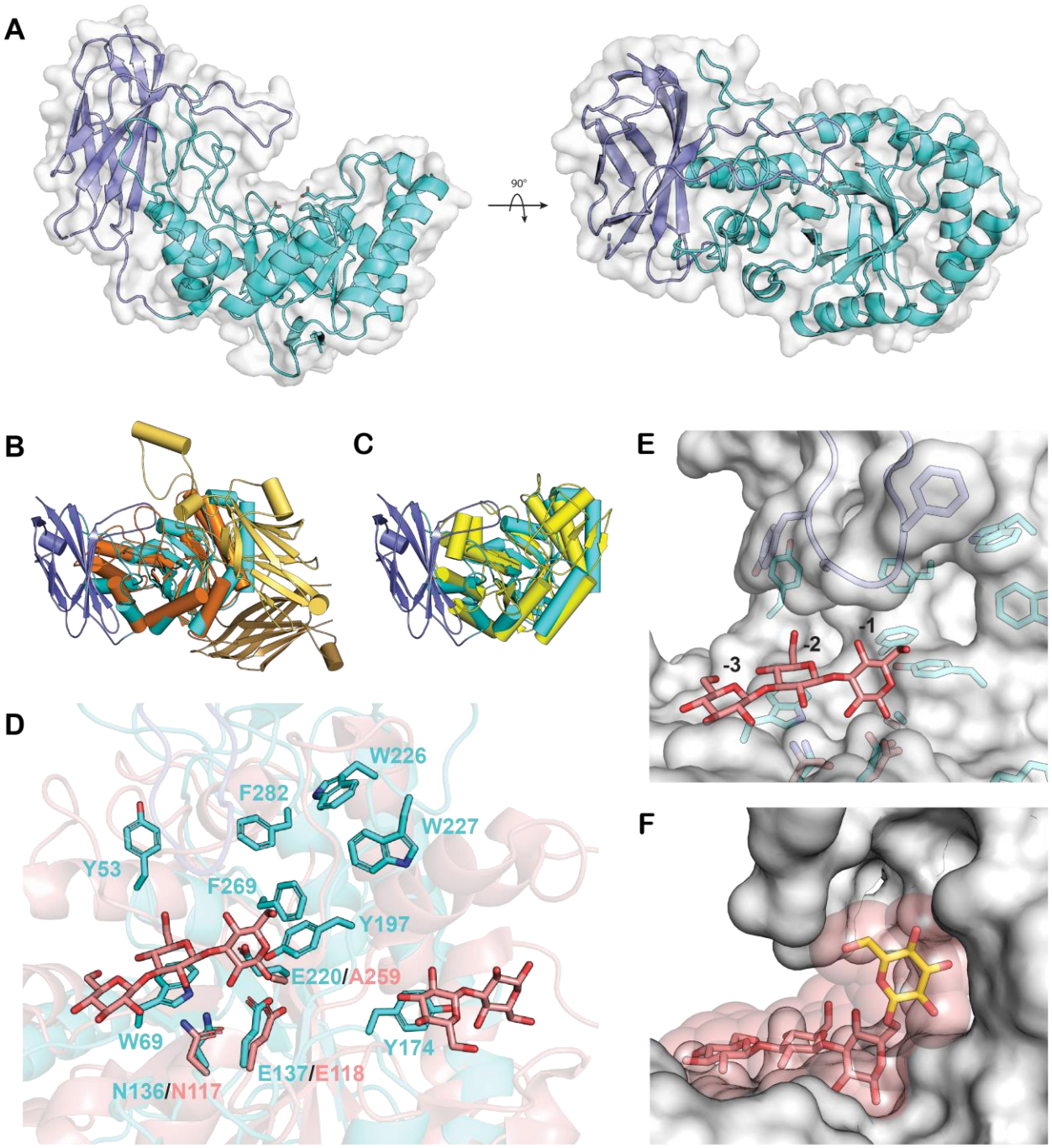

Figure 4-4. *Bu*GH158 tertiary structure.

(A) Overall structure of *Bu*GH158 with the TIM barrel domain colored cyan, the Ig-like domain colored slate, and a semi-transparent surface in white. The catalytic residues E137 and E220 are shown as sticks. (B) Superposition of *Bu*GH158 (colored as in panel A) with the *Bifidobacterium dentium* GH2 β-glucuronidase (RMSD = 4.4 Å across 203 Cα pairs), which has two additional domains N-terminal to the TIM barrel (PDB ID 5Z1B, orange). (C) Superposition of *Bu*GH158 (colored as in panel A) and the *Chrysonilia sitophila* GH5 β(1,4)-mannanase (RMSD = 5.5 Å across 121 Cα pairs), which has no additional domains (PBD ID 4AWE, yellow). (D) Superposition of *Bu*GH158 (colored as in

panel A) with a laminaritriose complex of a GH17 *endo*-β(1,3)-glucanase from *Solanum tuberosum* (StGH17, PDB ID 4GZJ, rose). *Bu*GH158 aromatic residues in the active-site cleft are shown in cyan sticks and homologous catalytic residues in StGH17 are shown in rose sticks. (E) Surface representation of *Bu*GH158 superposed with the StGH17 laminaritriose ligand (rose sticks) occupying the negative subsites as labeled, revealing a pocket near the 6-OH of the glucose in the -1 subsite. (F) A glucose molecule shown as yellow sticks modeled attached to the subsite -1 glucose via a β(1,6)-linkage. The transparent sphere about each atom of the modeled glucose and laminaritriose represent a van der Waals radius of 1.5 Å.

To provide the first three-dimensional insight into substrate specificity and catalysis in GH158, we determined the tertiary structure of the enzyme to 1.8 Å by X-ray crystallography (Table C-3). *Bu*GH158 consists of an N-terminal $(\alpha/\beta)_8$ triose phosphate isomerase (TIM) barrel domain and a C-terminal, eight-stranded immunoglobulin (Ig)-like domain in contact with helices α7 and α8 of the TIM barrel. Additional loops contribute to the extensive contact between domains, with one from the Ig-like domain extending above the TIM barrel to shape the active site cleft (Fig. 4-4A). A Dali structure similarity search returned members of Clan GH-A as the top 20 structural homologs (Table C-4), thus confirming the bioinformatics prediction that GH158 constitutes an evolutionarily distinct family within this Clan (294). Correspondingly, superposition with the top two GH2 and GH5 results show that the TIM barrel core is well conserved and that family-level differences arise from the presence or absence of accessory domains (Fig. 4-4B, 4-4C).

All Clan GH-A enzymes are predicted to use a canonical Koshland double-displacement mechanism employing a covalent glycosyl-enzyme intermediate and resulting in overall retention of the anomeric stereochemistry at the site of polysaccharide backbone hydrolysis (cf. Fig. 1-5) (293). This was confirmed for BuGH158 by NMR analysis of the hydrolysis of 2-chloro-4-nitrophenyl laminaribioside (G3G-CNP), which constitutes the first stereochemical determination for the family (Fig. C-8). Clan GH-A members present the conserved catalytic acid/base and nucleophile residues on loops immediately following strands β4 and β7, corresponding to E137 and E220, respectively, in *Bu*GH158 (Fig. 4-4D). Additionally, a conserved asparagine (N136) precedes the general acid/base (E137), and is anticipated to engage in hydrogen-bonding interactions with the substrate (Fig. 4-4D) (293). Indeed, site-directed mutation of N136, E137, or E220 to alanine completely abolished catalytic activity (data not shown).

GH158 is the fourth clan GH-A family, in addition to GH17, 128, and 148, now known to contain an *endo*-β(1,3)-glucanase activity (293,294). The active-site cleft surrounding the catalytic sidechains is rich in surface-exposed aromatic residues oriented to engage in stacking interactions with the substrate (Fig. 4-4D). Structural alignment with a *Solanum tuberosum* GH17 *endo*-β(1,3)-glucanase:laminaritrose (G3G3G) complex (297) (RMSD = 3.1 Å for 120 Cα pairs, sequence identity = 20.6%) infers the directionality of polysaccharide binding in *Bu*GH158 (Fig. 4-4D). Crucially, this superposition also reveals that the 6-OH of the glucosyl residue in the -1 subsite is oriented toward a pocket lined with aromatic residues that may accommodate a single β-1,6 linked glucose branch in *Bu*GH158 (Fig. 4-4E). Indeed, a glucosyl residue could easily be manually docked into this position without clashes (Fig. 4-4F). This structural feature provides a plausible explanation for the strict *Ld*Lam specificity and poor activity on yBG, the latter of which contains longer β-1,6 linked glucose branches (268) (Fig. 4-2A).

4.3.2.3 Periplasmic saccharification of imported oligosaccharides is mediated by *Bu*GH3, a specific exo-β(1,3)-glucosidase

GH3 is known to contain members with a diversity of exo-β-glycosidases (293), thus warranting detailed kinetic characterization of *Bu*GH3 in the context of the 1,3GUL. Initial substrate screening on chromogenic *p*NP glycosides revealed that *Bu*GH3 is an exo-β-glucosidase (no other activity detected on a panel of *p*NP substrates, Table C-5). Subsequent Michaelis-Menten kinetics on diverse β-gluco-oligosaccharides further established that *Bu*GH3 is a specific β(1,3)-glucosidase poised to efficiently handle the hydrolysis products of the predicted cell-surface GH16 and GH158 enzymes (Table C-5; Fig. C-9). Indeed, incubation with the limit digest products of either *Bu*GH16 or *Bu*GH158 confirmed that *Bu*GH3 is capable of completely degrading all oligosaccharide products of laminarin and bMLG to glucose (Fig. C-2 and C-10).

Despite possessing an apparently broad ability to hydrolyze β(1,3)-, β(1,4)-, and β(1,6)-glucosides, the particular preference of *Bu*GH3 for β(1,3)-linkages is highlighted by a two orders-of-magnitude higher k_{cat}/K_m value for laminaribiose (G3G) over cellobiose (G4G) and gentiobiose (G6G) (Table C-5). Concordant with this finding, catalytic efficiency towards mixed-linkage trisaccharides with a β(1,3)-glucose at the non-reducing end (G3G4G) was two orders of magnitude higher than that with a β(1,4)-glucose at the non-reducing end (G4G3G, Table C-5). Laminari-oligosaccharides of increasing DP were hydrolyzed with comparable k_{cat} values, but K_m

values decreased between the di- and trisaccharide, after which K_m values leveled off, thus suggesting that the *Bu*GH3 likely has two positive subsites of kinetic significance.

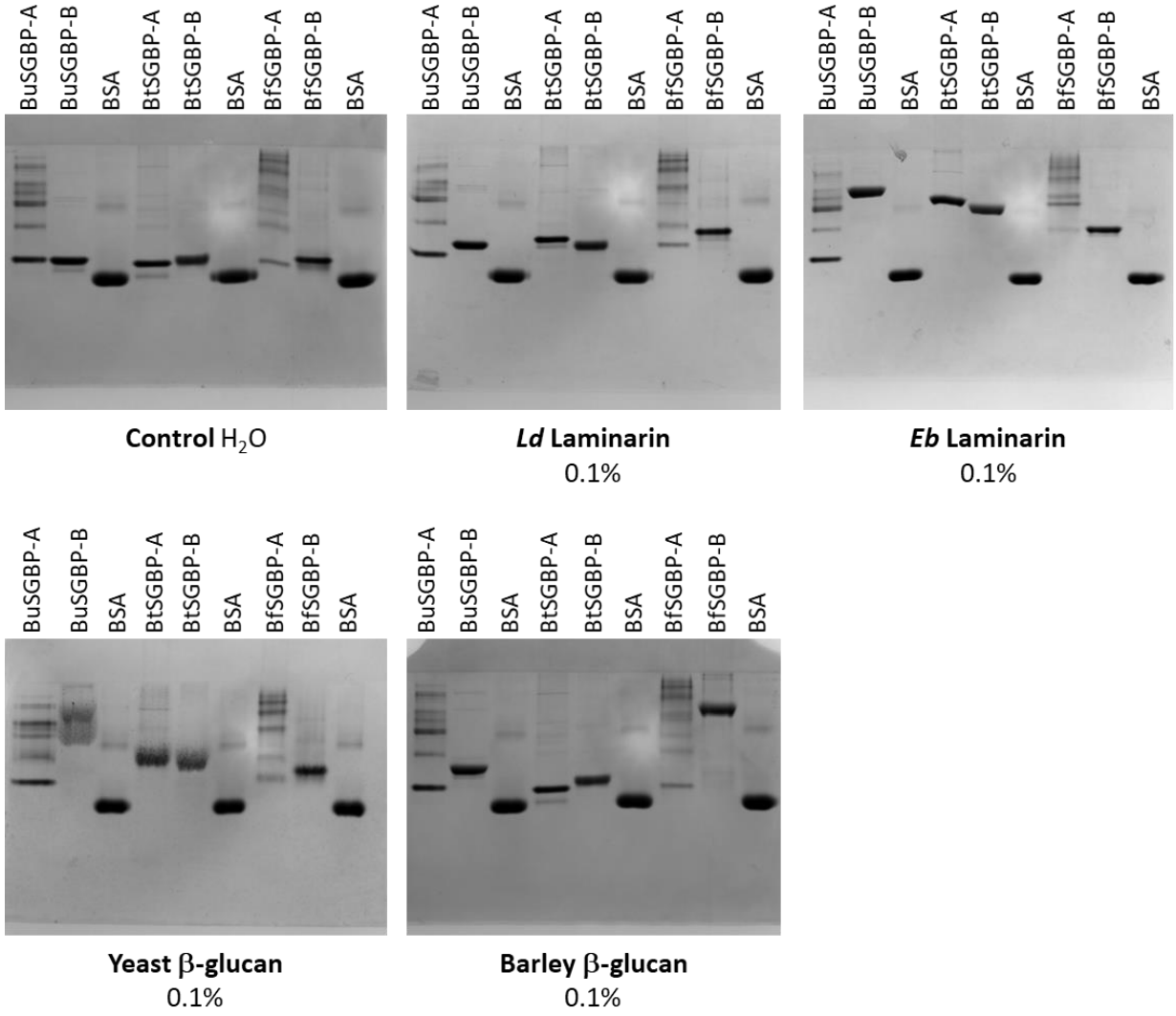

Figure 4-5. Binding of *Bacteroides* SGBPs to β(1,3)-glucans.

Native polyacrylamide (10%) gel electrophoresis containing 0.1% polysaccharide, with bovine serum albumin (BSA) as a control protein.

4.3.3 *Bu*SGBP-B mediates β-glucan specificity

The 1,3GUL encodes two potential SGBPs: The SusD-homolog *Bu*SGBP-A, and the sequence-divergent, "SusE-positioned" (298) *Bu*SGBP-B. Notably, qualitative screening of a library of soluble polysaccharides by affinity gel electrophoresis (AGE), as well as isothermal

titration calorimetry (ITC), indicated that *Bu*SGBP-A does not bind any likely substrates, including *Ld*Lam and bMLG (Figure 4-5, Table C-6, C-7, Fig. C-11, C-12). SGBP-A homologs that do not bind polysaccharide are currently rare, but not entirely unknown (258); indeed the primary role of this *Bacteroides* PUL component appears to be its indispensable structural association with the cognate TBDT (97,99,100,102).

In contrast, AGE demonstrated that *Bu*SGBP-B was able to bind *Ld*Lam, *Eb*Lam, yBG, and bMLG (Fig. 4-5). The smaller shift observed for *Ld*Lam than for *Eb*Lam (Fig. 4-5) recapitulates observations for an SGBP-A from a marine *Bacteroidetes*, *Gramella* sp.(226). Quantitative ITC indicated an order-of-magnitude higher affinity constant (K_a) for *Ld*Lam over bMLG (Table C-6), thus revealing the high specificity of *Bu*SGBP-B for β(1,3)-glucan. The smallest laminari-oligosaccharide bound by *Bu*SGBP-B was the trisaccharide (G3G3G) and affinity increased with degree-of-polymerization (Table C-7; Fig. C-12). Neither SGBP recognized insoluble crystalline cellulose (β(1,4)-glucan) in pull-down assays (data not shown), unlike SGBPs that bind MLG (100), which further underscores the β(1,3)-glucan specificity of *Bu*SGBP-B .

4.3.4 Divergent GH and SGBP specificities collectively dictate the range of β-glucans utilized by *Bacteroides* species

Using the *Bu*1,3GUL as the archetype, we identified several homologous 1,3GULs from the *Bacteroidaceae*, *Porphyromonadaceae*, and *Prevotellaceae* families (Order *Bacteroidales*), of which five representatives are shown in Fig. 4-1. All comprise a syntenic TBDT, SGBP-A, SGBP-B, and GH3 as the core set of conserved genes. However, the predicted *endo*-glucanase (GH16 and GH158) content is notably variable and the SGBPs have particularly low sequence similarity, including among related *Bacteroides* species (Fig. 4-1, Table C-8). With a focus on this genus, we tested the growth of *Bt* and *Bf* vis-à-vis *Bu* to determine how heterogeneity in 1,3GUL gene content might affect the utilization of individual β-glucans. Strikingly, whereas all three species were able to grow on *Ld*Lam as the sole carbon source, only *Bu* and *Bt* grew on yBG, while only *Bu* grew on bMLG (Figure 4-2, Table C-1).

To elucidate the molecular basis of this species-specific β-glucan utilization, we characterized the transcriptional response of the *Bt* and *Bf* 1,3GUL (as a measure of HTCS specificity) and the biochemistry of the predicted surface *endo*-glucanases and SGBPs vis-à-vis the *Bu* system. In *Bt* and *Bf*, the genes encoding the TBDT and SGBP-A were strongly upregulated

by *Ld*Lam and yBG, whereas they were only very weakly activated by bMLG (<10-fold, Fig. 4-2E). These results are generally concordant with the observed growth phenotypes among all three *Bacteroides* species. Although, it was particularly unexpected to observe strong upregulation of the *Bf* 1,3GUL with yBG, on which *Bf* is not able to grow. The *Bf*1,3GUL was still highly activated after dialyzing the substrate to remove any potential small oligosaccharide signals (Fig. C-13), suggesting that this system may be capable of sufficiently cleaving yBG to release an oligosaccharide that activates this PUL but subsequently unable to grow appreciably on the bulk polysaccharide. These distinct regulatory profiles likely involve a combination of surface GH and transporter specificity, combined with the sensing abilities of the respective HTCS regulators. The latter signaling specificity is notable in light of high protein sequence similarity among species (80-85% amino acid identity and 88-91% similarity versus the *Bu*HTCS, Table C-8) and the results shown below that at least *Bt*GH16 can cleave bMLG but does generate an activating cue.

Whereas *Bu* encodes both a GH16_3 member and a GH158 member in its 1,3GUL, *Bt* possesses only a GH16_3 member, while *Bf* possesses only a GH158 member. Enzymology revealed that *Bt*GH16 is a predominant laminarinase with lower, yet comparable, activity on yBG and bMLG, similar to *Bu*GH16 (Figure C-14, Table C-2; the homologous *Pl*GH16 also had broad activity). In contrast, *Bf*GH158 was highly specific for *Ld*Lam, with low activity on yBG and very poor activity on bMLG, similar to *Bu*GH158 (Table C-2).

We also assessed the β-glucan specificities of the SGBPs from the *Bt* and *Bf* 1,3GULs for comparison with those of *Bu*. *Bt*SGBP-A, *Bt*SGBP-B, *Bf*SGBP-A, and *Bf*SGBP-B each bound the all-β(1,3)-linked *Ld*Lam, *Eb*Lam, and yBG, as shown by AGE and ITC, although *Bf*SGBP-A and -B interacted only weakly with yBG (Fig. 4-5; Table C-6; Fig. C-11). Strikingly, *Bf*SGBP-B migration was strongly hindered by the β(1,3)/β(1,4)-linked bMLG in AGE, whereas that of *Bf*SGBP-A, *Bt*SGBP-A and *Bt*SGBP-B was unaffected. Indeed, ITC indicated that *Bf*SGBP-B bound bMLG with a similar affinity as *Bu*SGBP-B (Table C-6).

These biochemical data show that whereas orthologous GH specificity is well-predicted by the family to which it belongs, considerable diversity exists in the specificity of syntenic SGBPs encoded by 1,3GULs of *Bu*, *Bt*, and *Bf*. When combined with microbiological data, a pattern emerges whereby GH content combined with SGBP specificity predicts the range of β-glucan congeners a *Bacteroides* species is able to utilize. In parallel, the signaling specificity or promiscuity through the single HTCS sensor associated with each system must also be able to

respond to the diversity of cues generated from cleavage of related substrates. Of the three species examined, only *Bu* possesses the full complement of synergistic HTCS, GH, and SGBP specificities to enable growth on laminarin, yBG, and bMLG (summarized in Fig. 4-6).

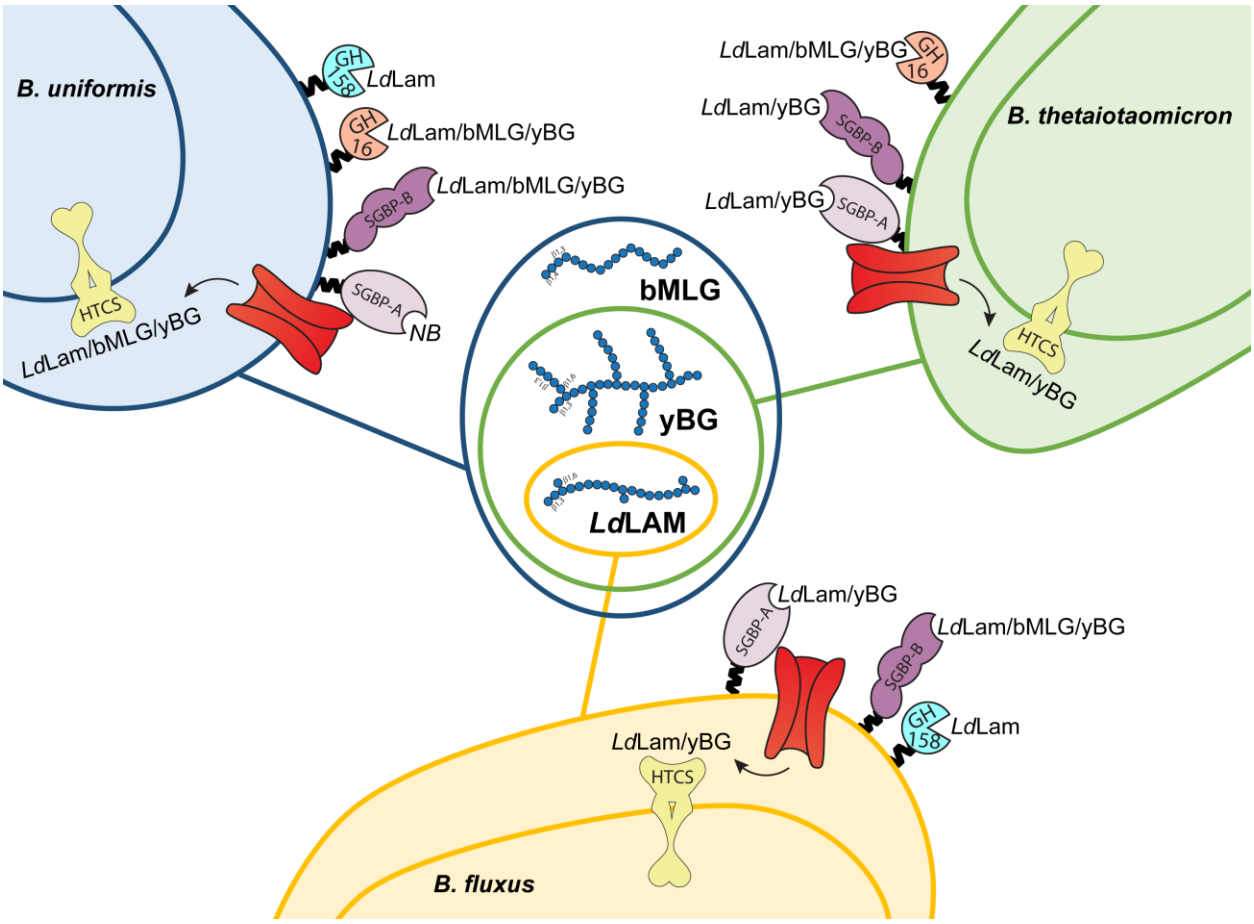

Figure 4-6. Summative diagram of β-glucan utilization by *Bacteroides* of the HGM.

For each species, the binding or catalytic specificity of the proteins encoded by the endogenous 1,3GUL is indicated on the cell surface (Lam, laminarin; bMLG, barley mixed-linkage β-glucan; yBG, yeast β-glucan; NB, no binding). The capacity each species to utilize individual β-glucans for growth, which arises as a combination of these individual specificities, as well as periplasmic transport/sensing, is indicated in the Venn diagram.

4.3.5 Metagenomic survey reveals global 1,3GUL distribution in the human gut

To determine the prevalence of 1,3GULs in the human gut, we surveyed publicly available gut metagenomic data from 2,441 adults across five different continents (North American, South America, Africa, Europe, and Asia; Fig. 4-7). Despite similar genetic synteny, these 1,3GULs have different nucleotide sequences (pairwise identities Bf/Bt, 44%; Bf/Bu, 60%; Bt/Bu, 44%),

allowing us to use the nucleotide sequence to distinguish the presence or absence of a specific 1,3GUL in each metagenomic sample. The *B. uniformis* type 1,3GUL is the most abundant across all samples (48% of the total population), perhaps reflecting the high prevalence of this species in humans from industrialized populations. This is followed by the *B. thetaiotaomicron* type 1,3GUL (26% of samples) and *B. fluxus* (0.53% of samples). The low prevalence of *B. fluxus* may be explained by its low abundance and prevalence in European, North American, and Asian populations, consistent with previous observations for the xyloglucan utilization locus (138). Although, 1,3 GULs are widely distributed throughout human populations (59%), we do not see correlation with any particular geographic region or population. This ubiquity may be explained by the prevalence of β(1,3)-glucans in different diets since edible fungi and yeast fermentation products are common worldwide. Strikingly, we were unable to detect any of the three identified 1,3GULs in the indigenous Hadza and Yanomami tribes, which may be due to a high prevalence of *Prevotella* and not *Bacteroides* in these populations (132,289). This present study, in combination with our previous metagenomic survery of MLG utilization (211), reveals a broadly represented potential for specific β-glucan metabolism to establish niches for individual human gut bacteria.

4.4 Discussion

The human large intestine is a highly competitive ecosystem, in which access to a wide range of carbohydrates confers selective advantage. We outline here a model in which a single PUL enables *Bu* to utilize a range of β-glucan congeners (Fig. 4-1). In this model, *Ld*Lam, yBG, or MLG (Fig. 4-2) is bound by *Bu*SGBP-B on the surface of the bacterium, whereas the SusD homolog *Bu*SGBP-A is passive in this step. Depending on the backbone linkage composition and extent of branching, individual β-glucans are cleaved into oligosaccharides by one or both cell surface-anchored *endo*-glucanases. Whereas *Bu*GH158 is a highly specific laminarinase, *Bu*GH16 is a generalist *endo*-β(1,3)-glucanase which is able to additionally accommodate backbone β(1,4)-glucosyl linkages (as in MLG) and long/frequent β(1,6)-glucosyl branches (as in *Eb*Lam and yBG). Oligosaccharides produced by these endo-glucanases are actively imported by the associated TBDT into the periplasm, where the *exo*-β-glucosidase *Bu*GH3 completes saccharification. Notably, the poor activity of *Bu*GH3 toward β(1,6) and β(1,4) linkages suggests that other periplasmic glucosidases encoded outside of the 1,3GUL may assist with oligosaccharide debranching and complete MLG oligosaccharide hydrolysis. The genome of *B.*

uniformis ATCC 8492 contains three members of GH30, which are known to contain exo-β(1,6)-glucosidase that act on laminarin (299), as well as twenty additional members of GH3, which could contain a potent exo-β(1,4)-glucosidase.

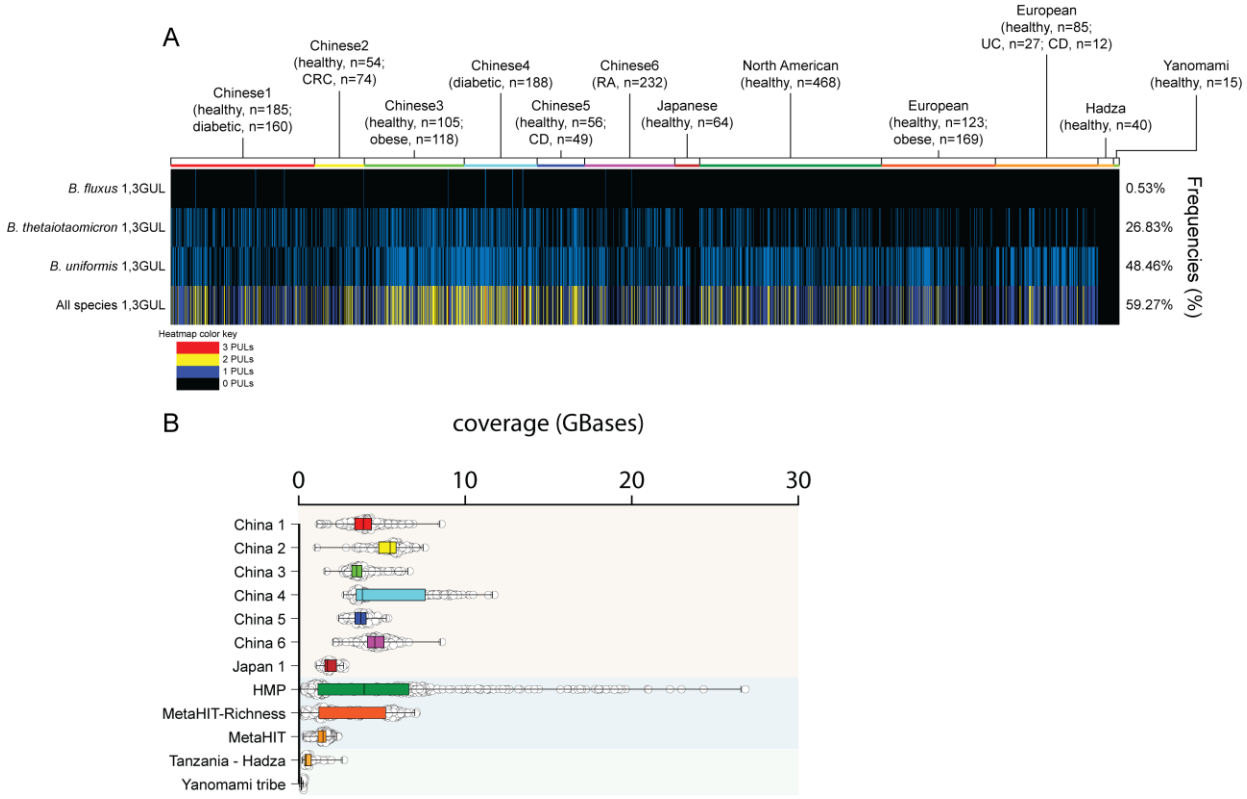

Figure 4-7. Prevalence of 1,3-GULs in 2441 human metagenomes.

(A) Each line represents the presence (blue) or absence (black) of a specific 1,3GUL species related in a single human gut metagenomic sample. The bottom row represents the total number of 1,3GUL that each individual possesses, colored according to the legend. The frequency of 1,3GULs incidence across all 2441 individuals is shown on the right. (B) Coverage variation of each metagenomic sample is indicated by individual samples (open circles) and a box plot with the mean.

The *Bacteroides* 1,3GULs characterized here share partial homology with various laminarin-targeting PULs from marine Bacteroidetes, including *Gramella* (300), *Formosa* (301), and other closely related genera within the family *Flavobacteriaceae* (302), raising questions of evolutionary origin of 1,3GULs in the HGM. Indeed, PULs that target porphyran (a galactan-based seaweed polysaccharide) are thought to have been acquired by the HGM from marine Bacteroidetes *Zobellia galactanivorans* via a horizontal gene transfer event (198). In addition to

syntenic 1,3GULs, recently-characterized *Formosa* species also harbor enlarged laminarin-targeting PULs containing other enzymes. In this system, laminarin must first be debranched by a GH30 exo-β(1,6)-glucosidase before being hydrolyzed by a GH17 endo-β(1,3)-glucanases, which has very low activity on β(1,6)-glucose-branched laminarin (301). In general, GH17 endoglucanases have a very narrow active site cleft that does not effectively accommodate such branches (297,301) compared to BuGH16 and BuGH158 enzymes which allow the 1,3GUL system to bypass the initial debranching requirement. Thus, although GH16 and GH17 laminarin-active enzymes may display similar overall specificities (303), differences in active-site structure may delineate distinct roles in the stepwise total hydrolysis of laminarins.

Comparative genomic and biochemical analysis revealed variations in GH content that, together with complementary HTCS, GH, and SGBP specificity, dictate selective β-glucan utilization. Among the *Bacteroides* species tested, *Bu* is unique in its ability to utilize the trifecta of β-glucans containing contiguous or individual β(1,3) linkages, *viz.* *Ld*Lam, yBG, and bMLG. In contrast, *Bt* can only access *Ld*Lam and yBG, while *Bf* is further restricted to utilizing *Ld*Lam only (summarized in Fig. 4-6). None of the species tested were able to utilize curdlan, perhaps due to its poor solubility in water. Orthologous GHs from syntenic 1,3GUL share similar specificities, yet the GH complement alone is insufficient to confer growth on a particular β-glucan. For example, *Bt* produces a GH16 member that is an approximately equally competent laminarinase, yBGase, and MLGase (Table C-2). However, an inability to capture bMLG at the cell surface, due to a lack of polysaccharide binding by the cognate SGBPs (Fig. 4-5), and an inability to transport or recognize breakdown products in the periplasm (Fig. 4-2) precludes growth on this abundant dietary glycan. It is interesting to ponder to what extent losses- or gains-of-function might be correlated and arise (a)synchronously through the evolution of these distinct protein families more widely among the *Bacteroidetes*.

In this regard, *Bf* is the least versatile of the three *Bacteroides* species explored here. The inability of *Bf* to utilize bMLG clearly is the consequence of three factors (Fig. 4-6): (1) the lack of a polyspecific GH16 ortholog encoded by its 1,3GUL (*cf.* *Bu* and *Bt*), (2) the extremely poor MLGase activity of *Bf*GH158 (Table C-2), and, not least, (3) an inability to transport/sense this cereal polysaccharide in the periplasm (Fig. 4-2). Yet, we note that *Bf* possesses an SGBP-B able to bind both laminarin and bMLG (Fig. 4-5, Table C-6), which is correlated with MLG utilization in the absence of a MLG-binding SGBP-A in *Bu* (Fig. 4-6). Further, the inability of *Bf* to grow on

yBG is consistent with weak binding by *Bf*SGBP-A and *Bf*SGBP-B, poor catalytic efficiency of BfGH158 towards yBG, and the lack of a compensatory GH16 ortholog. This is despite an HTCS which is equally highly responsive to yBG and *Ld*Lam (Fig. 4-2), as well as both SGBP types binding both of these 1,3-β-glucans. Hence, we might anticipate that a gain-of-function mutation to introduce yBG hydrolase activity, for example through acquisition of a versatile GH16 member or broadening the substrate scope of the extant GH158, would result in growth on yBG. Testing this hypothesis will be dependent on the future generation of a transformation system for *Bf* or discovery of a corresponding wild-type strain.

The present study underscores that dietary specificity of related gut commensal strains is gleaned most precisely by systems-based approaches involving genomic, biochemical, and structural biological dissection of PULs (100-103,110,138-140,194,211,256-259), which otherwise could not have been predicted based on sequencing data alone. Whereas the GH complement of PULs has been shown to drive specificity in the levan/fructan system (101), to our knowledge the present study describes the first case in which the interplay of GHs and SGBPs, underpinned by HTCS specificity, collectively dictate glycan utilization among species. Indeed, in most PULs characterized to-date, the specificities of the vanguard GH and SGBPs are concordant (99,100,110,138,140,211,258). On the other hand, the *Bf* case reveals how a limited endo-glycanase in the context of polyspecific SGBPs and a HTCS can restrict nutrient range. Thus, the evolution of both the gene complement and the specificity of individual components within a PUL allows bacteria to access to distinct nutrient niches in the competitive human gut environment. In the context of human nutrition and health, the detailed characterization presented here provides a validated set of molecular markers (Fig. 4-1) to identify β(1,3)-glucan utilization potential among members of the HGM (and other microbiota). This insight may prove especially transformative for our future ability to select strains and dietary formulations in tailoring microbial intervention therapies (14).

Chapter 5: Distinct protein architectures mediate species-specific beta-glucan binding and metabolism in the human gut microbiota

5.1 Introduction

The distal human gastrointestinal tract plays host to a highly dynamic community of microbes, collectively known as the human gut microbiota (HGM). Strong correlational and causal links between composition of the HGM and numerous disease states have been established (6,8,9,304), making HGM manipulation an attractive potential route for therapies (11,12). HGM composition is easily altered on a short timescale (9,10), and is driven mainly by our ingestion of complex polysaccharides (indigestible "dietary fiber") (13,14,253). Bacteroidetes is a dominant phylum within the HGM, members of which owe much of their success in this highly competitive environment to an arsenal of glycan metabolic systems encoded by polysaccharide utilization loci (PULs) (105). By devoting a significant portion of their genome to encoding coordinated sets of carbohydrate-active enzymes (CAZymes), cell-surface glycan-binding proteins (SGBPs), and TonB-dependent transporters (TBDTs) in PULs (Fig. 5-1), individual Bacteroidetes are able to access and grow on broad palettes of complex polysaccharides that nature has to offer (15,109).

Recent efforts to elucidate the pathways by which diverse complex polysaccharides are utilized via PUL systems have provided a wealth of structural insight into the CAZymes (110,135,138,140,211,257-259,305-308). Not least, these studies have revealed many new glycoside hydrolase (GH) families and specificities (reviewed in (106)). In the PUL paradigm, the non-catalytic SGBPs play essential roles in recruiting polysaccharide to the cell surface, as well as facilitating transport of cleavage products across the outer membrane, in concert with cognate TBDTs (95,97). Despite their importance to the function of PUL systems, structural studies of the highly diverse types of SGBPs are currently few (99,100,103,107). As such, this constitutes an important gap in our understanding of the molecular basis of PUL system selectivity and function. Soluble β-glucans are a ubiquitous part of the human diet with numerous health benefits, which we access via fermentation by the HGM (141,208,210). The chemical structures of β-glucans are diverse, but these polysaccharides can be broadly classified based on backbone linkages, each of which is targeted by a corresponding PUL: mixed-linkage β(1,3)/β(1,4)-glucan (MLG; from cereal crops) (100,211), β(1,3)-glucan (from fungi/yeasts, plants) (308), and β(1,6)-glucan (from fungi) (309). In our recent study of homologous β(1,3)-glucan utilization loci (1,3GULs) from

Bacteroides uniformis, *B. thetaiotaomicron*, and *B. fluxus*, we demonstrated that growth on individual branched β(1,3)-glucans and/or MLG is dependent on the combinatorial specificities of cognate GHs and SGBPs encoded by each species. (Fig. 5-1) (308). Here, we explore the structural basis underlying the specificity of the three orthologous SGBPs-A (SusD homologs) and the three sequence-diverse SGBPs-B through X-ray crystallography and isothermal titration calorimetry with defined oligosaccharides. In doing so, we uncover unique structural architectures that allow recognition of distinct classes of β-glucan.

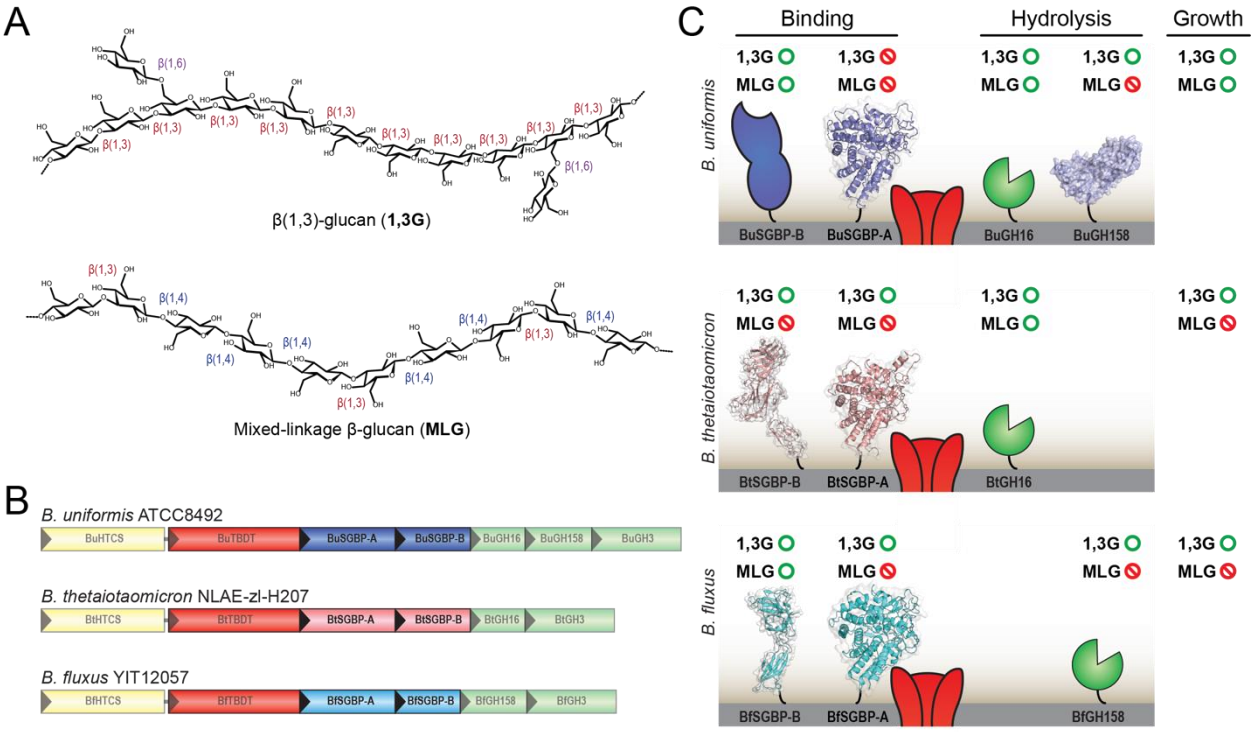

Figure 5-1. β(1,3)-glucan utilization loci (1,3GULs).

(A) Chemical structures of β(1,3)-glucan and mixed-linkage β-glucan (MLG). β(1,3)-glucans from diverse sources vary in the length and frequency of β(1,6)-linked glucose branching. Shown as representative is laminarin from *Laminaria digitata*, which contain single β(1,6)-glucose branching at a frequency of around once per every seven β(1,3)-linked glucose. MLGs from diverse sources have the same general structure but differ in the ratio of cellotriosyl to cellotetraosyl units separated by β(1,3)-linkages. (B) Genetic organization of the 1,3GULs from three different species of Bacteroidetes. HTCS: hybrid two-component system sensor/regulator; TBDT: TonB-dependent transporter (SusC homolog); SGBP: cell surface glycan-binding protein; GH: glycoside hydrolase. (C) Specificities of SGBPs and GHs and resulting ability to support growth on β(1,3)-glucan and MLG. In order to be utilized for growth, the β-glucan must be bound by at least one SGBP and hydrolyzed by at least one GH. The grey bar represents the gram-negative bacterial outer membrane, cognate TBDTs are colored red and GHs without resolved tertiary structure are

shown in green. Polysaccharide specificity data is from Dejean et al. (308). These surface PUL proteins work in concert to capture, hydrolyze, and transport target β-glucan breakdown products at and across the outer membrane. Once sequestered to the periplasmic space, the specific breakdown products of the target polysaccharide bind to the inner membrane spanning regulator proteins to elicit PUL expression. Finally, exo-acting Cazymes eventually hydrolyze oligosaccharides down to their constituent monosaccharides which can cross the inner membrane and enter metabolic pathways.

5.2 Materials and Methods

5.2.1 Substrates

Laminarin from *Laminaria digitata* was purchased from Sigma Aldrich (St. Louis, MO, USA), and laminarin from *Eisenia bicyclis* was purchased from Carbosynth (Compton, UK). Barley beta-glucan (high viscosity), yeast beta-glucan, and curdlan from *Alcaligenes faecalis* were purchased from Megazyme International (Bray, Ireland). All laminarioligosaccharides and mixed-linkage gluco-oligosaccharides used in this study were purchased from Megazyme International (Bray, Ireland).

5.2.2 Cloning and site-directed mutagenesis

All full-length and truncated SGBP constructs were cloned as previously described (308). All site-directed mutants were produced as previously described (100). Primers for cloning and site-directed mutagenesis are listed in Table D-14 and Table D-15, respectively.

5.2.3 Recombinant protein production and purification

Plasmids were transformed into chemically competent *E. coli* BL21 (DE3) for overexpression. All proteins for crystallography were produced in 1 L LB media supplemented with 100 μg/mL ampicillin. Cells were grown at 37 °C with shaking until culture OD_{600} reached ~0.7 at which point isopropyl β-D-thiogalactopyranoside was added to a final concentration of 1 mM. The culture under overexpression condition was incubated over two nights at 16 °C after which the cells were harvested by centrifugation at 4500 *g* for 30 minutes.

Cells were resuspended in buffer A (20 mM HEPES pH 7.4, 500 mM NaCl, 20 mM imidazole) supplemented with 1 mM phenylmethylsulfonyl fluoride (PMSF) and 1 mM ethylenediamine tetraacetic acid (EDTA) then lysed by sonication. Cell debris was pelleted by centrifugation at 24,700 *g* for 45 minutes and the supernatant loaded onto a 5 mL HisTrap™ IMAC FF Crude Ni Sepharose column (GE Life Sciences). Bound SGBPs were washed with buffer A then eluted using a linear gradient of 0 – 100 % buffer B (20 mM HEPES pH 7.4, 500 mM NaCl,

500 mM imidazole). Eluted SGBPs were thoroughly buffer exchanged into buffer C (20 mM HEPES pH 7.8, 100 mM NaCl) using either a 30 kDa or 10 kDa cutoff Vivaspin centrifugal concentrators (Sartorius). His_6-tags were subsequently cleaved from SGBPs by TEV protease (prepared in-house as described in (310)): 1.5 mg TEV protease per 25 – 100 mg SGBP, in 25 – 35 mL reaction volume such that protein concentration does not exceed 5 mg/mL, incubated overnight with gentle rocking at 4 °C. Cleaved His_6-tag, uncleaved SGBP, and His_6-tagged TEV protease were removed by running over freshly charged HisTrap™ column and collecting the flowthrough. SGBPs were further purified by size exclusion chromatography using either Superdex 75 or Superdex 200 resin (GE Life Sciences) packed in an XK 16/100 column (GE Life Sciences) equilibrated and run in buffer D (10 mM HEPES pH 7.0) at 0.8 mL/min. Pure fractions as determined by SDS-PAGE were pooled and concentrated, quickly flash-frozen in liquid nitrogen, and stored at -70 °C until required.

5.2.4 Crystallization and structure determination

For all proteins, initial screening was carried out in 96-well sitting drop format using various commercial screens: Classics Suite, Classics II Suite, JCSG+ Suite, Protein Complex Suite, pH Clear Suite, and PACT Suite (all from Qiagen). Plates for screening were set up using the Phoenix robot (Art Robbin Instruments). Promising hits were subsequently optimized in larger hanging drops in 24-well plates by hand. Final crystallization conditions are presented in Table D-16; all crystals were grown at room temperature. Micro seeding where used was carried out using Seed Beads (Hampton Research) to crush crystals and trialing serial dilutions of seeds in hanging drops in 24-well plates.

For BuSGBP-A and BfSGBP-B, dramatic improvements in crystal morphology and diffraction quality were achieved by using hits from the Additive Screen (Hampton Research). All additives were used at the recommended final concentrations according to the manufacturer's instructions by mixing 1 μL into 5 μL of protein followed by 4 μL of crystallization solution. All additive screening and optimization were set up in hanging drops in 24-well plates. In the case of BuSGBP-A, 10 mM hexamminecobalt(III) chloride improved the diffraction limit from low-resolution (d_{min} > 5.5 Å) to near-atomic (d_{min} = 1.85 Å), and allowed for the determination of experimental phases using the ordered cobalt sites. Hexamminecobalt(III) bound to the slightly negatively charged carboxy end of one of the α-helices that make up TPR3, the exact location where a sodium cation is observed to be bound in BtSGBP-A (Fig. D-1D). A molecule in a

neighboring asymmetric unit also makes contact with hexamminecobalt(III) near its C-terminal end, similarly via δ^- carbonyl oxygens (Fig. D-1C). Overall, these hexamminecobalt(III) ions can be seen to aid in formation of crystal contacts to improve crystal packing (Fig. D-1B), therefore improving diffraction quality of the crystal. In the case of BfSGBP-B, initial flake-like crystals were stacks of very thin plates exhibiting highly anisotropic diffraction ($d_{min} \approx 2.5$ Å in one axis but > 3.5 Å in another). Addition of 100 mM guanidine hydrochloride significantly increased crystal thickness and diffraction isotropy, and improved resolution of the diffraction data (d_{min} = 1.61 – 1.82 Å) (Fig. D-8A). Guanidine molecules were resolved in the electron density and observed to interact with numerous main chain δ^- carbonyl oxygens in sections of polypeptide lacking defined secondary structure (Fig. D-8B). Interactions with aspartates, asparagines and prolines, as well as π-cation interactions with tyrosines are also observed to contribute to the overall improvement in crystal packing (Fig. D-8B).

Soaking with various halides and heavy atom candidates were attempted to obtain experimental phases for BtSGBP-B and BfSGBP-B. Salts of zinc ions produced best results for both structures: sufficient labelling was achieved with a 10-minute soak in crystallization solution supplemented with 100 mM zinc sulfate for BtSGBP-B and 100 mM zinc acetate for BfSGBP-B. All crystals were briefly (10 – 30 seconds) soaked in crystallization solution supplemented with cryoprotectant listed in Table D-14 before flash freezing by plunging in liquid nitrogen.

Diffraction data were collected at the Stanford Synchrotron Radiation Lightsource (SSRL) beamline 9-2, SSRL beamline 12-2, or Advanced Photon Source beamline 23ID-D (Table D-1, D-5, and D-10); all collected at 100 K. Single anomalous dispersion (SAD) experiments were conducted for BuSGBP-A_Co, BtSGBP-B_Zn, and BfSGBP-B_Zn by collecting data at the respective absorption edges determined by fluorescence scans. For BuSGBP-A_Co and BtSGBP-B_Zn (longer wavelength absorption edge and lower resolution diffraction limit, respectively), 720 degrees of data were collected to increase multiplicity.

Data for BtSGBP-A, BuSGBP-A, BfSGBP-B_MLG3, and BtSGBP-B_trunc were processed using autoPROC (311), which utilizes XDS (153), Pointless (278), Aimless (279), and CCP4 (154). Data for BuSGBP-A_Co, BfSGBP-A, BtSGBP-A_lam6, BfSGBP-B_Zn, BfSGBP-B, BfSGBP-B_lam3 were indexed and integrated using XDS (153) and scaled and merged using Aimless (279). Data for BtSGBP-B and BtSGBP-B_Zn were processed and scaled using xia2/DIALS (219,220).

Experimental phasing by SAD for BuSGBP-A_Co and BtSGBP-B were performed in AutoSol (216), subsequent density modification performed using RESOLVE (312), and initial models built using Phenix.autobuild (313). Experimental phasing by SAD for BfSGBP-B was performed in autoSHARP (314), subsequent density modification performed using SOLOMON (315), and an initial model built using ARP/wARP (316). All phasing by molecular replacement were performed using Phaser (157) in the Phenix suite (217) or the CCP4i2 suite (221), and initial models were built using Phenix.autobuild (313), Buccaneer (317), or ARP/wARP (316). All subsequent manual model building and refinement were conducted with Coot (280) and Refmac5 (161), respectively, in the CCP4i2 suite (221). The quality of modelled proteins was monitored using Molprobity (318), and that of modelled sugars was validated using Privateer (164).

5.2.5 Affinity gel electrophoresis

Native polyacrylamide gels consisting of 10 % (w/v) acrylamide and 0.1 % (w/v) polysaccharide (or water for control) were cast as described in (319). SGBPs (6.0 μg) and bovine serum albumin (5.0 μg) were loaded onto gels and subjected to electrophoresis under non-denaturing conditions at 80 V for 3 hours at room temperature. Proteins were visualized by staining with Coomassie Brilliant Blue.

5.2.6 Isothermal titration calorimetry

All ITC experiments were performed using the MicroCal PEAQ-ITC instrument (Malvern Panalytical) calibrated to 25 °C and with reference power of 20.9 μW. The sample cell was loaded with approximately 250 μL of SGBP at 100 μM and the syringe was loaded with approximately 70 μL of oligosaccharide at 1 mM; all in 20 mM HEPES pH 7.0. An initial injection of 0.2 μL was followed by 18 subsequent injections of 2 μL spaced 150 seconds apart and with an injection duration of 4 seconds. All injections were performed with stirring at 750 rpm throughout the run, and the resulting heat of reaction was recorded. Data were analyzed using the MicroCal PEAQ-ITC Analysis Software (Malvern Panalytical).

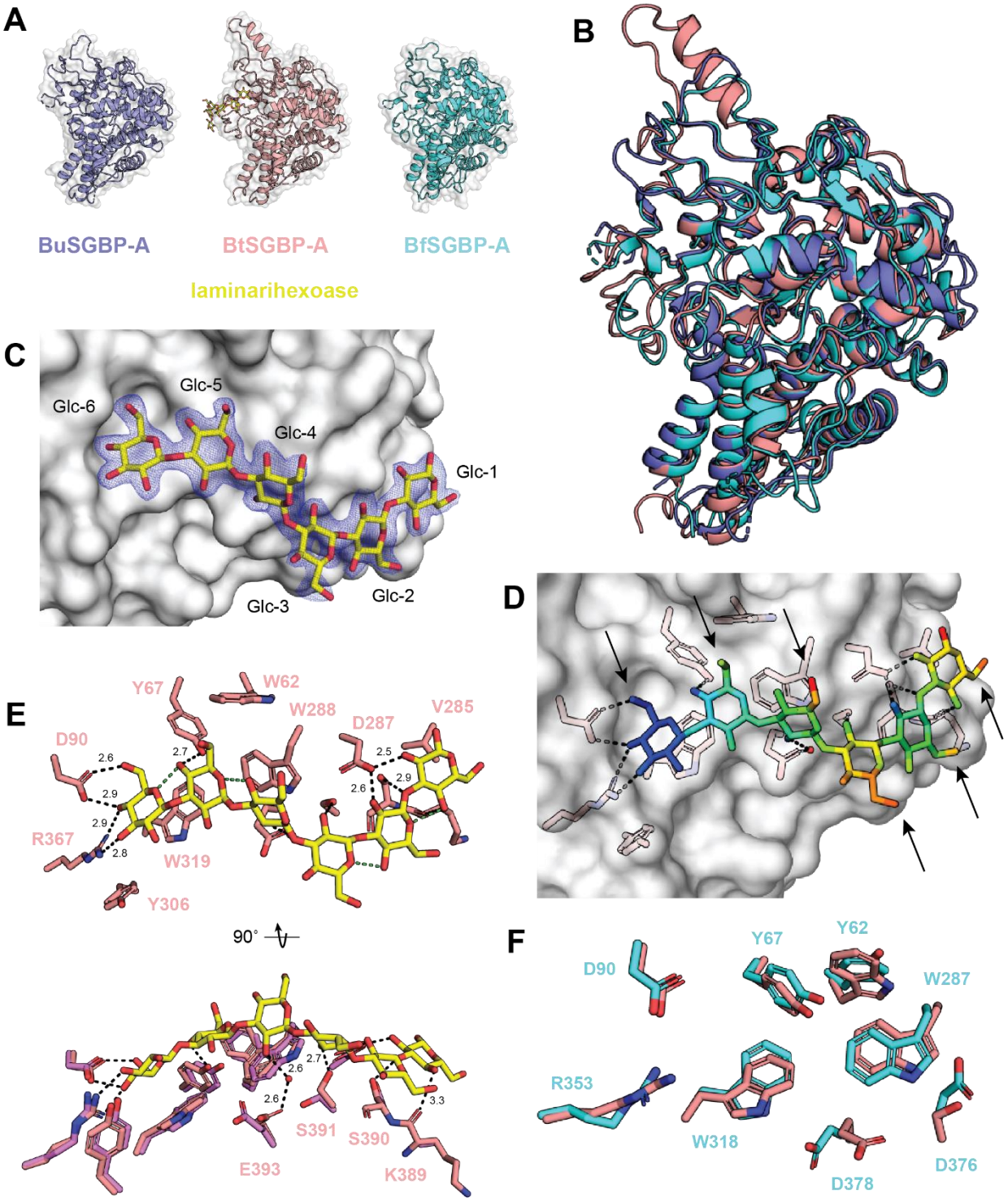

Figure 5-2. 1,3GUL SGBPs-A and BtSGBP-A laminarihexaose complex.

(A) Overall cartoon representation of 1,3GUL SGBPs-A with transparent surface. BuSGBP-A is shown in slate, BtSGBP-A in rose, BfSGBP-A in cyan, and laminarihexaose in yellow throughout all figures. (B) Secondary structure matching (SSM) superposition of the three SGBPs-A. (C) BtSGBP-A in opaque surface representation with bound laminarihexaose. Refined $2F_{obs}$-F_{calc} map contoured at σ = 1.0 about the modelled sugar is shown as blue mesh. (D)

Bound laminarihexaose colored ramped from blue to red representing low to high B-factor. BtSGBP-A is shown with transparent surface with binding site residues shown as sticks and hydrogen bonds as dotted lines. Arrows point to the C6-OH of each monosaccharide and represent space where β(1,6)-linked branches can be accommodated. (E) Details of binding site with key residues and bound laminarihexaose shown as sticks and hydrogen bonds as dotted lines. Hydrogen bond donor-acceptor distances are labelled in Å and intramolecular hydrogen bonds are colored green. The main chain atoms are shown for Lys389 and Ser390 to reveal carbonyl groups; only the side chains are shown for all other residues. Orthogonal view is shown below with additional overlaid binding site residues of unliganded BtSGBP-A included colored magenta. (F) Conserved binding site residues of BtSGBP-A overlaid with those of BfSGBP-A.

5.3 Results

PULs classically encode two distinct type of SGBPs. SGBPs-A (also referred to as SusD-like or SusD-homologs) are highly conserved across PUL systems, in which they form the extracellular lid of an active transport complex with the cognate TBDT (SusC homologs) (95,97). Indeed, the tandem *TBDT/SGBP-A* (*susC*/*susD* homolog) gene pair is a signature feature used to identify PULs in sequenced Bacteroidetes genomes (92,109). On the other hand, PULs often encode one or more additional, often highly sequence-divergent, SGBP(s) immediately downstream of the *SGBP-A* (*susD*) homolog, here denoted as SGBP-B (99,100). These are sometimes referred to a "SusE-positioned" proteins due to this genetic organization but lack of sequence similarity with SusE (107). The 1,3GULs from *Bacteroides uniformis* ATCC8492, *B. thetaiotaomicron* NLAE-zl-H207, and *B. fluxus* YIT12057 each encode one SGBP-A and one SGBP-B (Fig. 5-1B, 5-1C), the tertiary structures of which we determined in free and oligosaccharide-complexed forms.

5.3.1 The three orthologous 1,3GUL SGBPs-A possess the canonical SusD fold

All three SGBPs-A are predicted to be outer membrane surface-anchored via a cysteine lipidation at the N-terminus of their respective mature polypeptides (Cys22 in BuSGBP-A, Cys21 in BtSGBP-A, Cys22 in BfSGBP-A (174,213)). Hence, the recombinant forms of these proteins were produced with the native signal peptide and this cysteine removed. Diffraction quality crystals of BuSGBP-A, BtSGBP-A, and BfSGBP-A were successfully grown following varying amounts of optimization. Crystals of BtSGBP-A required micro seeding to reproduce crystallization screen hits, while the morphology and diffraction quality of BuSGBP-A were significantly improved by the addition of hexamminecobalt(III) chloride (see Experimental procedures and Fig. D-1A).

Experimental phases for BuSGBP-A were obtained by single anomalous dispersion at the cobalt absorption edge (Table D-1). There are two ordered cobalt sites with significant anomalous signal, one per each of two molecules in the asymmetric unit. Both hexamminecobalt(III) complexes were found to mediate crystal contacts to molecules in neighboring asymmetric units, explaining the improved diffraction (Fig. D-1B, D-1C). Coordinates were refined to a final resolution of 1.85 Å against a higher resolution dataset collected at shorter wavelength (Table D-1). The completed model of both molecules in the asymmetric unit comprised residues 43 – 529 with unmodeled gaps at residues 176 – 177, 211 – 219, 293 – 308, and 390 – 394, due to lack of corresponding electron density (Fig. 5-2A).

The crystal structures of BtSGBP-A and BfSGBP-A were solved to 1.80 Å and 1.84 Å, respectively, by molecular replacement using the BuSGBP-A structure as a search model (Table D-1). Both consist of a single molecule in their respective asymmetric units and the resulting electron density allowed near-complete tracing of the protein model comprising residues 38 – 515 for BtSGBP-A and residues 40 – 510 for BfSGBP-A (Fig. 5-2A). As observed for BuSGBP-A, the first ca. 20 amino acids in both proteins were disordered, suggesting that these form flexible tethers from the cell surface in the native lipoproteins. Otherwise, the only unmodeled gaps in the BtSGBP-A and BfSGBP-A structures correspond to residues 172 – 179 and residues 172 – 175, respectively. The C-termini were defined in the electron density of all three SGBPs-A.

The three orthologous SGBPs-A were all observed to possess the canonical SusD fold with tetratricopeptide repeats (TPRs) prominently forming the structural scaffold (Fig. 5-2A) (146). The structures are almost identical and superpose with low pairwise root-mean-square deviation (RMSD) values, which negatively correlate with amino acid sequence identity, as expected (Fig. 5-2B, Table D-2). One prominent difference is the insertion of an α-helix at the periphery of BtSGBP-A (residues 346 – 353), which is not present in BuSGBP-A nor BfSGBP-A. The functional significance of this additional helix, if any, is unknown. Other differences in the observed residues are restricted to minor variations in the positioning of surface loops.

5.3.2 BtSGBP-A and BfSGBP-A bind β(1,3)-glucan via the non-reducing end

Soaking of the native BtSGBP-A with laminarihexaose yielded a 2.05 Å ligand-complexed structure (Table D-1) that clearly revealed an extended β(1,3)-glucan binding platform, notably comprising two key tryptophan residues (Trp288 and Trp318) and specific recognition of the non-reducing end by multiple hydrogen bonds (Fig. 5-2). Electron density was observed for all six

glucosyl residues, which were convincingly modelled in the favored $^4C_1$ conformation (Fig. 5-2C, Table D-2). The bound laminarihexaose molecule adopts an extended helical structure with the binding surface complementing a low-energy conformation of the oligosaccharide (320). As is typical for β(1,3)-glucans, hydrogen bonds between the ring oxygen and the C4-OH of the adjacent glucosyl residue are observed, except between Glc3 and Glc4. Here, there is a ca. 180° flip in the ϕ (O5-C1-O3'-C3') angle of the glycosidic bond between Glc3 and Glc4, imparting an additional twist in the helix.

BtSGBP-A residues involved in binding interactions with the hexasaccharide are borne on a polypeptide connecting the two α-helices comprising TPR1 (residues 59 – 107) and the TPR3 and TPR4 domains (residues 279 – 398). A bis-tris molecule observed in the original native structure was displaced by the non-reducing end glucose (Fig. D-2), which displays a very favorable carbohydrate-aromatic interaction with Trp319 (Fig. 5-2D, 5-2E, 5-2F). In addition, four hydrogen bonds between C3-OH and Arg367-$N^{\eta 1}$, C4-OH and Arg367-$N^{\eta 2}$, C4-OH and Asp90-$O^{\delta 1}$, and C6-OH and Asp90-$O^{\delta 2}$ firmly anchor the non-reducing end glucose in place. The strength of these collective interactions is reflected in the comparatively very low B-factor of Glc6 (Fig. 5-2D). In this context it is also worth noting that Tyr306 is well-positioned to firmly hold Arg367 in place via a π-cation interaction. Trp288 partially stacks against both Glc5 and Glc4, with the former also hydrogen bonding to Tyr67 via the C4-OH, and the latter exhibiting water-mediated hydrogen bonding to Glu393 via the C2-OH (Fig. 5-2E). The only interaction to Glc3 is a hydrogen bond between Ser391 and the C2-OH resulting in a higher B-factor on the C5 side of the ring (Fig. 5-2D). Glc2 and Glc1 both hydrogen-bond to the same O^{δ} of Asp287 via their respective C2-OH groups (Fig. 5-2E). Additional hydrogen bonds were observed between Ser390 and the glycosidic oxygen connecting Glc2 and Glc1, as well as between the main chain carbonyl oxygen of Lys389 and the C4-OH of Glc1. Finally, a possible CH-π interaction between Glc2 and the amide bond between Ser390 and Ser391, as well as hydrophobic interaction between Glc1 and Val285, may also contribute to binding at these sites (Fig. 5-2E). Overall, the native and laminarihexaose-complexed structures of BtSGBP-A are virtually identical, superposing with a RMSD value of 0.19 Å over 402 Cα pairs (Fig. D-2C). The key binding-site residues are also in essentially identical positions except for Ser391 and Glu393, which upon binding laminarihexaose rotate to participate in direct and water-mediated hydrogen bonding, respectively (Fig. 5-2E).

Despite BtSGBP-A binding the non-reducing end of the glycan chain in a manner reminiscent of a type-C carbohydrate-binding module (CBM) (76), the binding pocket is not prominent and rather constitutes a binding platform that is blocked on one end (Fig. 5-2C, 5-2D). Indeed, this extended binding site requires a β(1,3)-glucan substrate with a degree of polymerization (DP) ≥ 5 to effect binding, as reveled by isothermal titration calorimetry (ITC). BtSGBP-A bound laminarihexaose (K_D 26 μM) approximately 10-fold tighter than laminaripentaose (K_D 210 μM), whereas binding to laminaritriose and laminaritetraose was too weak to be quantified (Fig. D-4, Table D-4). Likewise, soaking native BtSGBP-A crystals with laminaritriose did not yield a trisaccharide-complexed structure. In the biological context, our previous affinity gel electrophoresis (AGE) and ITC analyses using polysaccharides demonstrated that BtSGBP-A is responsible for binding β(1,3)-glucans with varying degrees and lengths of β(1,6)-linked glucosyl branching (308). The laminarihexaose complex here reveals the abundance of space around the C6-OH group at every glucose binding subsite, thereby rationalizing this versatility (Fig. 5-2D).

Despite extensive efforts, including soaking and co-crystallization trials, we were not able to obtain a structure of BfSGBP-A in complex with a laminarioligosaccharide. However, sequence alignment combined with structural superposition reveal that many of the key binding site residues are conserved vis-à-vis BtSGBP-A (Fig. 5-2F, D-3). Specifically, the aromatic residues lining the binding platform are in essentially identical positions, as are the key Asp90 and Arg367 residues that block the platform end. As such, BfSGBP-A can be inferred to bind β(1,3)-glucan in a similar fashion to BtSGBP-A (Fig. D-3C). This assumption is supported by both β(1,3)-glucan polysaccharide (308) and laminarioligosaccharide binding data (Fig. D-4, Table D-4), which indicate that BfSGBP-A requires at least a pentasaccharide for binding and is agnostic to branching frequency and branch length.

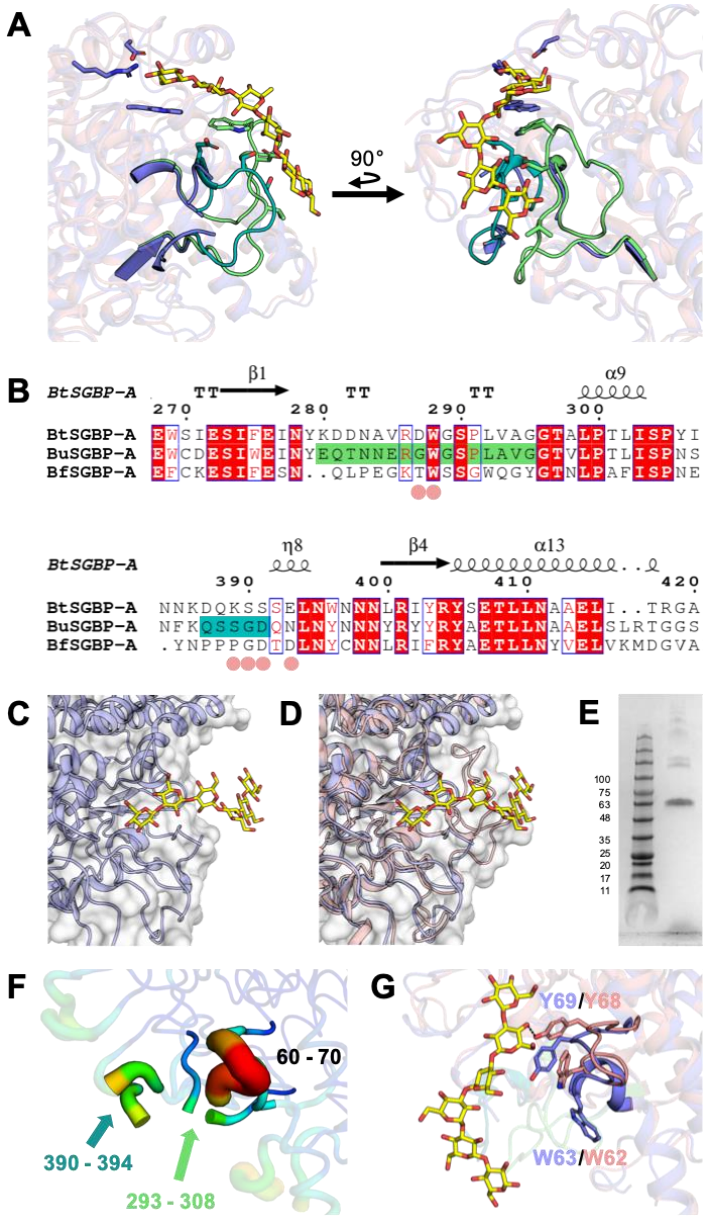

Figure 5-3. BuSGBP-A disrupted key loops.

(A) Binding site loops in BtSGBP-A corresponding to two of the loops missing in BuSGBP-A; loop 293 – 308 in lime and loop 390 – 394 in teal. Sidechains that participate in ligand binding as well as bound laminarihexaose in BtSGBP-A are shown in sticks. (B) Amino acid sequence alignment of 1,3GUL SGBPs-A showing the region containing the missing loops in BuSGBP-A, highlighted in the same colors as panel (A). BtSGBP-A residues involved in binding laminarihexaose are indicated with rose circles. (C) Binding site of BuSGBP-A shown as transparent surface around main chain cartoon with overlaid laminarihexaose from BtSGBP-A. (D) Same as panel (C) with additional overlaid BtSGBP-A main chain cartoon highlighting the binding site architecture to which the missing loops in BuSGBP-A would contribute. (E) SDS-PAGE gel of dissolved BuSGBP-A crystal. (F) The three disrupted BuSGBP-A loops in B-factor putty projection. (G) Conserved aromatic residues in loop 60 – 70 overlaid with those of BtSGBP-A that interact with bound laminarihexaose.

5.3.3 Lack of β(1,3)-glucan binding by BuSGBP-A correlates with structural disorder

Unlike its two orthologs, BuSGBP-A does not display binding to β(1,3)-glucan or other polysaccharides (308). This is particularly intriguing in light of the protein sequence alignment, which indicates that key binding-site residues are conserved vis-à-vis BtSGBP-A and BfSGBP-A (Fig. 5-3, D-3). Furthermore, a bis-tris molecule was observed in the BuSGBP-A structure at the identical position in the native BtSGBP-A structure, coordinated by a conserved Arg/Asp/Trp triad (Fig. D-5A, D-5B). As mentioned above, BuSGBP-A contains four sections of the protein that were not defined in the electron density. Structural superposition with laminarihexaose-bound BtSGBP-A revealed that two of these unmodeled gaps in the polypeptide (residues 293 – 308 and 390 – 394, Fig. D-5C, D-5D) correspond to loops that shape the binding site (Fig. 5-3). Furthermore, loop comprised of residues 293 – 308 contains the conserved tryptophan residue (Trp288) that in BtSGBP-A forms crucial carbohydrate-aromatic stacking interactions with Glc5 and Glc4 (Fig. 5-3B). SDS-PAGE of BuSGBP-A crystals after dissolution in 1 % (v/v) trifluoroacetic acid (observed molar mass ca. 60 kDa, Fig. 5-3E) confirmed that unexpected proteolysis was not the cause of the missing density (Fig. D-5C, D-5D).

It is notable that loop 293 – 308 cannot assume the same conformation and occupy the same space as the corresponding loop in BtSGBP-A, because it would clash with the polypeptide in a symmetry-related molecule (Fig. D-5E). The two other disordered segments of the polypeptide are thought to be inconsequential to binding. Residues 176 – 177 correspond to a loop distant from the binding site that is also disordered in BtSGBP-A and BfSGBP-A, and residues 211 – 219 correspond to an insertion not found in the other two SGBPs-A (Fig. D-3A) but that is also distant from the binding site. On the other hand, another loop that shapes the binding site is loop 60 – 70. In BuSGBP-A, this segment exhibits high B-factor and two conserved aromatic residues, Trp63 and Tyr69, are shifted away from the positions taken by the corresponding residues (Trp62 and Tyr68) in BtSGBP-A (Fig. 5-3F, 5-3G). Trp62 is likely to contribute to the hydrophobicity of the binding surface and Tyr68 makes a hydrogen bond to Glc2 in BtSGBP-A. Thus, three loops critical to shaping the binding site are structurally disordered in BuSGBP-A, which is remarkably coincident with the inability of this protein to bind β(1,3)-glucans (Fig. 5-3F).

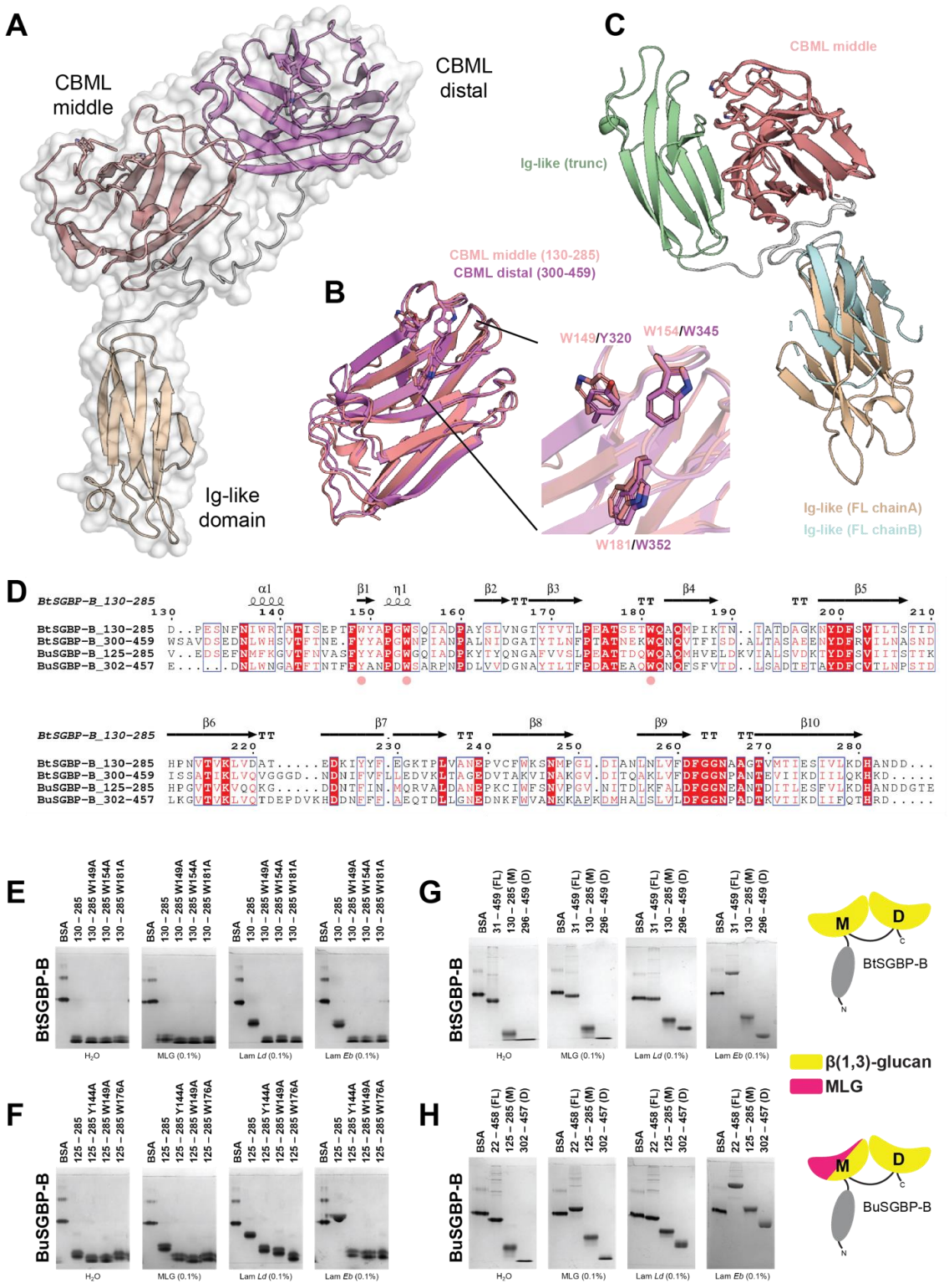

Figure 5-4. BtSGBP-B structure and biochemistry.

(A) Overall structure of BtSGBP-B in cartoon representation with transparent surface. Individual ordered domains are colored in different tints and interdomain linkers are colored white. (B) SSM superposition of CBML-middle (rose) and CBML-distal (magenta). A close-up of the binding site aromatic residues in stick representation is shown to the

right. (C) Different orientations of the Ig-like domain (colored in different tints) relative to the CBML-middle domain (rose) observed in crystal structures with the linker colored white. Respective CBML-middle domains were aligned by SSM. (D) Amino acid sequence alignment of CBML domains from BtSGBP-B and BuSGBP-B. Conserved/similar binding site aromatic residues are indicated by rose circles. (E and F) Affinity gel electrophoresis (AGE) results of BtSGBP-B and BuSGBP-B binding site mutants, respectively. Lam *Ld* – laminarin from *Laminaria digitate*, Lam *Eb* – laminarin from *Eisenia bicyclis*. (G and H) AGE results of BtSGBP-B and BuSGBP-B domain dissections, respectively. Binding specificity cartoon is shown to the right.

5.3.4 β(1,3)-glucan-specific SGBPs-B comprise unique multi-domain architectures

Across the three syntenic 1,3GULs, BuSGBP-B and BfSGBP-B bind both β(1,3)-glucan and MLG, whereas BtSGBP-B is specific for β(1,3)-glucan (Fig. 5-1) (308). Of these proteins, BtSGBP-B and BuSGBP-B share considerable sequence identity (53 %), while the primary structure of BfSGBP-B is notably distinct (< 20 % identity the other two SGBPs-B). To provide insight into the structural features responsible for these different binding properties, we obtained crystal structures for BtSGBP-B and BfSGBP-B in several free and oligosaccharide-complexed forms.

We obtained diffraction quality crystals of full-length, native BtSGBP-B, but the corresponding selenomethionine-labelled crystals diffracted poorly (d_{min} > 6.5 Å). Therefore, experimental phases were obtained by single anomalous dispersion from crystals soaked in 100 mM zinc sulfate following a screening trial. Twenty zinc sites with significant anomalous signal were identified in the structure, *viz.* ten in each of two molecules comprising the asymmetric unit (Fig. D-6). The final 2.60 Å native structure of BtSGBP-B was obtained by molecular replacement using the zinc-labelled structure as the search model (Table D-5).

The completed model reveals that BtSGBP-B has an extended multimodular architecture, which is typical of SGBPs-B, although this modularity is generally highly diverse (99,100,103,107) (reviewed in (106)). In this case, BtSGBP-B is comprised of three independent domains: an N-terminal immunoglobulin-like (Ig-like) domain and two β-sandwich, carbohydrate binding module-like (CBML) domains (Fig. 5-4A). The CBML-middle and CBML-distal domains both comprise two β-sheets of five anti-parallel β-strands each, which superpose very closely (RMSD = 0.87 Å for 143 Cα pairs) (Fig. 5-4B). DALI analysis (321) using CBML-middle as the search model indicates that the fold of these CBML domains are most similar to that of CBM family 4, including a β(1,3)-glucan-binding module from *Thermotoga maritima* (PDB ID 1GUI,

(322)) (Fig. D-7A, Table D-6). However, the poor sequence similarity of the individual CBML sequences with *bona fide* CBM4 members precludes their inclusion in this family (B. Henrissat, AFMB-CNRS (184), personal communication).

Also notable, the polypeptide linkers connecting the discrete domains are very long. Over 10 amino acids separate the N-terminal Ig-like and CBML-middle domains, and over 20 amino acids separate the CBML-middle and CBML-distal domains. The flexibility of these linkers may allow different relative conformations of the domains, observed for the two different molecules in the asymmetric unit (Fig. 5-4C). A truncated form of BtSGBP-B, comprising only the Ig-like and CBML-middle domains (residues 28-285), assumed an additional divergent conformation, further indicating the articulation of the domains relative to one another (Fig. 5-4C, Table D-5). Despite extensive efforts, we were unable to determine the structure of full-length, truncation variants, nor surface entropy reduction variants of BuSGBP-B. However, the individual BuSGBP-B domains could successfully be dissected and produced in soluble form, guided by sequence alignment with BtSGBP-B (Fig. 5-4D, D-7A, D-7B, Table D-7).

Inspection of the CBML domains in BtSGBP-B revealed a cluster of three aromatic residues at one edge of the concave side of the β-sandwich (three tryptophans in CBML-middle and two tryptophans and a tyrosine in CBML-distal), suggesting the location of the binding site (Fig. 5-4B). Although we were unsuccessful in obtaining an oligosaccharide-complexed structure, site-directed alanine mutants of each of these aromatic residues had a deleterious effect on binding capacity, which confirmed this hypothesis. Furthermore, these aromatic residues are conserved in BuSGBP-B, in which they are also critical for ligand binding in that protein (Fig. 5-4E, 5-4F).

5.3.5 Complementary domain specificities dictate β-glucan recognition by BtSGBP-B and BuSGBP-B

To understand the distinct specificities of BtSGBP-B and BuSGBP-B for β(1,3)-glucan and MLG despite their similar protein architectures, we dissected the four individual CMBL domains and assessed their specificities (Fig. 5-4G, 5-4H). AGE indicate that, while both CBML domains of BuSGBP-B bind β(1,3)-glucan, CBML-middle is exclusively responsible for the binding of MLG (Fig. 5-4H). ITC using defined laminarioligosaccharides revealed similar affinities for both CBML domains (K_d values in the 10^{-5} M range), with the difference that CBML-distal bound laminarioligosaccharides of DP 3 or greater, while CBML-middle required laminaritetraose or longer oligosaccharides (Fig. D-9, Table D-8, D-9). These chain-length

dependences and affinity values were precisely recapitulated by BtSGBP-B (Fig. D-8, Table D-8). The defining difference is that, unlike its counterpart, CBML-middle domain of BtSGBP cannot bind MLG (Fig. 5-4G).

5.3.6 BfSGBP-B comprises a distinct two-domain architecture

Like BuSGBP-B, BfSGBP-B binds both β(1,3)-glucan and MLG (308). However, low sequence similarity, including a significantly different polypeptide length, suggested a different structural architecture versus BtSGBP-B and BuSGBP-B. Following optimization, the crystal morphology and diffraction quality of full-length BfSGBP-B were significantly improved by the addition of 100 mM guanidine hydrochloride, which mediates crystal contact between molecules in adjacent asymmetric units (Fig. D-10; loosely analogous to hexamminecobalt(III) chloride in the case of BuSGBP-A, Fig. D-1). Experimental phases for BfSGBP-B were obtained by single anomalous dispersion at the absorption edge of zinc, ions of which were incorporated into the crystal by soaking in 100 mM zinc acetate. Four zinc sites with significant anomalous signal were identified in the structure, coordinated by surface-exposed aspartate, glutamate, and histidine residues (Fig. D-11). The unliganded structure was determined at 1.82 Å resolution by rigid body refinement using coordinates obtained from the phasing model (Table D-10).

The overall structure of the final model is comprised of two domains: an N-terminal polycystic kidney disease (PKD) domain (residues 38 – 118) and a C-terminal β-barrel domain (residues 119 – 290) (Fig. 5-5A). The PKD domain consists of a β-sandwich with one sheet containing three anti-parallel strands and the other containing four. The β-barrel domain consists of eight anti-parallel β-strands comprising the core. Additional pairs of β-strands are present at the top face of the barrel, in addition to short α-helices that abut the top and bottom faces of the barrel (Fig. 5-5A). A β-barrel fold involved in carbohydrate binding is unique among SGBPs-B, which are typified by β-sandwich folds. A DALI search (321) of the β-barrel domain returned structures of uncharacterized or functionally unrelated proteins with poor Z scores (<8.1) as top results, although, notably, many of the uncharacterized proteins originate from *Bacteroides* species (data not shown).

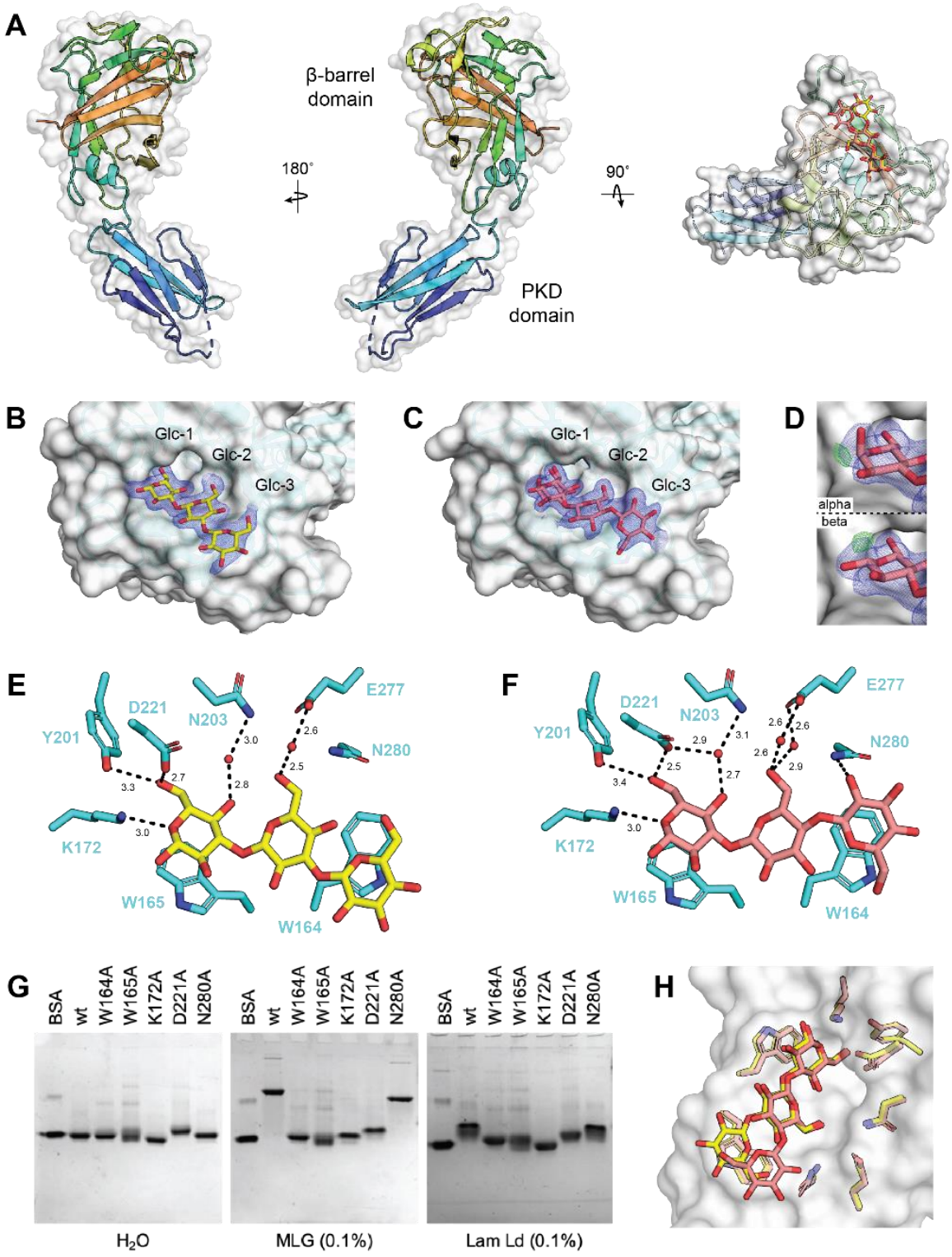

Figure 5-5. BfSGBP-B laminaritriose and MLG triose (G4G3G) complexes.

(A) Overall structure of BfSGBP-B in cartoon representation color ramped from blue (N-terminus) to red (C-terminus) with transparent surface. Two side views and one top view is presented with the laminaritriose in yellow and G4G3G in salmon shown as sticks in the top view only. (B) Close-up of bound laminaritriose with F_{obs}-F_{calc} omit map contoured

at σ = 3.0 about the modelled sugar shown as blue mesh. (C) Close-up of bound G4G3G with F_{obs}-F_{calc} omit map contoured at σ = 3.0 about the modelled sugar shown as blue mesh. (D) Refined $2F_{obs}$-F_{calc} map contoured at σ = 1.0 and positive F_{obs}-F_{calc} map contoured at σ = 3.0 about the reducing-end glucose modelled in the α and β anomers. (E) Details of binding site interaction with bound laminaritriose with ligand and key residues shown as sticks and hydrogen bonds as dotted lines. Hydrogen bond donor-acceptor distances are labelled in Å. (F) Same as (E) with bound G4G3G; only the β anomer is shown for clarity. (G) AGE results of binding site mutants (Lam *Ld* – laminarin from *Laminaria digitata*) H, Overlay of bound laminaritriose and G4G3G ligands with binding site residues in respective complexes in the same color as the ligand (only the β anomer of G4G3G is shown for clarity).

5.3.7 BfSGBP-B binds β(1,3)-glucan via the reducing-end

To illuminate the structural features responsible for the dual β(1,3)-glucan/MLG binding ability of BfSGBP-B, structures in complex with laminaritriose (G3G3G) and the mixed-linkage trisaccharide G4G3G were solved to 1.76 and 1.61 Å resolution, respectively. Partial enzyme digest products of laminarin or MLG, comprising a mixture of oligosaccharides, were soaked into BfSGBP-B crystals, and phases were obtained from the unliganded BfSGBP-B coordinates by molecular replacement (Table D-10). In both cases, three glucose residues in the favored $^4C_1$ conformation were convincingly modelled (Table D-11, D-12) at the same position on the top face of the β-barrel domain (Fig. 5-5A, 5-5B, 5-5C). Strikingly, these complexes revealed that BfSGBP-B binds β-glucans via the reducing end, as unambiguously indicated by the observation of electron density corresponding to both the α- and β-anomers of Glc1 (Fig. 5-5D).

A series of CH-π and hydrogen bonding interactions effect binding of the two oligosaccharides within defined binding subsites. At the terminal subsite, the reducing-end glucose stacks against Trp165 and also makes key hydrogen-bonding interactions with Lys172 via the ring oxygen and with Asp221 via the C6-OH (Fig. 5-5E, 5-5F). The indispensable nature of not just the carbohydrate-aromatic interaction but these two hydrogen bonds is highlighted by a total loss of binding when any of these residues is mutated to an alanine (Fig. 5-5G). Water-mediated hydrogen-bonding is also observed between the C4-OH and Asn203. The glucose residue occupying the second subsite is relatively bereft of interactions other than water-mediated hydrogen bonds between the C6-OH and Glu277 (Fig. 5-5E, 5-5F). In both trisaccharides, the reducing-end Glc1 and Glc2 are linked via a β(1,3)-bond and, as such, are posed identically in the G4G3G and G3G3G complexes. The wide third subsite, delineated by a stacking interaction with Trp164, is able to accommodate either a β(1,3)-linked or a β(1,4)-linked glucosyl residue (Fig. 5-

5E, 5-5F, 5-5H). Like that observed with Trp165, this carbohydrate-aromatic interaction is indispensable (Fig. 5-5G). A hydrogen bond between Asn280 and the C2-OH is observed only for the β(1,4)-linked glucose in G4G3G.

Although the BfSGBP-B crystals were soaked with solutions containing longer oligosaccharides, there is a lack of convincing electron density beyond Glc3, despite the presence of minor positive F_{obs}-F_{calc} density around the terminal C3 hydroxyl group in the G3G3G complex and the C4 hydroxyl group in the G4G3G complex (data not shown). The observation of three well-defined binding subsites is consistent with ITC, which indicated binding to MLG oligosaccharides and laminarioligosaccharides of DP 3 and higher, but not laminaribiose (Table D-13). Further corroborating the crystallography, ITC data also showed that a β(1,3)-glucosyl linkage is strictly required at the reducing end, as evidenced by binding to the MLG oligosaccharides G4G3G and G4G4G3G, but not G3G4G, G3G4G4G, and G4G3G4G (Fig. D-13, Table D-13).

5.4 Discussion

5.4.1 Homologous SGBPs-A feed and/or cap the TBDT

The structurally diverse SGBPs play critical or accessory roles in Bacteroidetes of the HGM and other ecosystems by facilitating the attachment to, and uptake of, complex carbohydrates (99,100,102). SGBP-A orthologs, which are syntenic across PULs, have high tertiary structural conservation vis-à-vis the archetype, SusD from the *B. thetaiotaomicron* starch utilization system (146). This conservation is exemplified here by the structures of BuSGBP-A, BtSGBP-A and BfSGBP-A.

Apart from carbohydrate binding, SGBPs-A (SusD homologs) play an indispensable role in transport by forming a functional complex with their cognate TBDT (SusC homolog), in which they comprise the extracellular lid of a "pedal-bin" arrangement (97) (Fig. 5-6A). Targeted gene deletion and site-directed mutagenesis studies have shown that whereas the role as a lid is indispensable for complex formation, the ability of SGBPs to bind the target glycan is not required for bacterial growth in some cases (99,108). In other cases, both the presence of the SGBP-A at the cell surface and its ability to bind target glycan are indispensable (100). In this context, superposition of the three SGBP-A structures with a fructan SGBP-A/TBDT complex (PDB ID 5T3R (97)) highlights that the homologous glycan-binding surfaces are poised over the top of the entrance to the β-barrel (Fig. 5-6B). In the superposition with BtSGBP-A, the laminarihexaose in

the binding site is observed to fit within the diameter of the barrel with the non-reducing end positioned more centrally and the reducing end extending towards the edge (Fig. 5-6C).

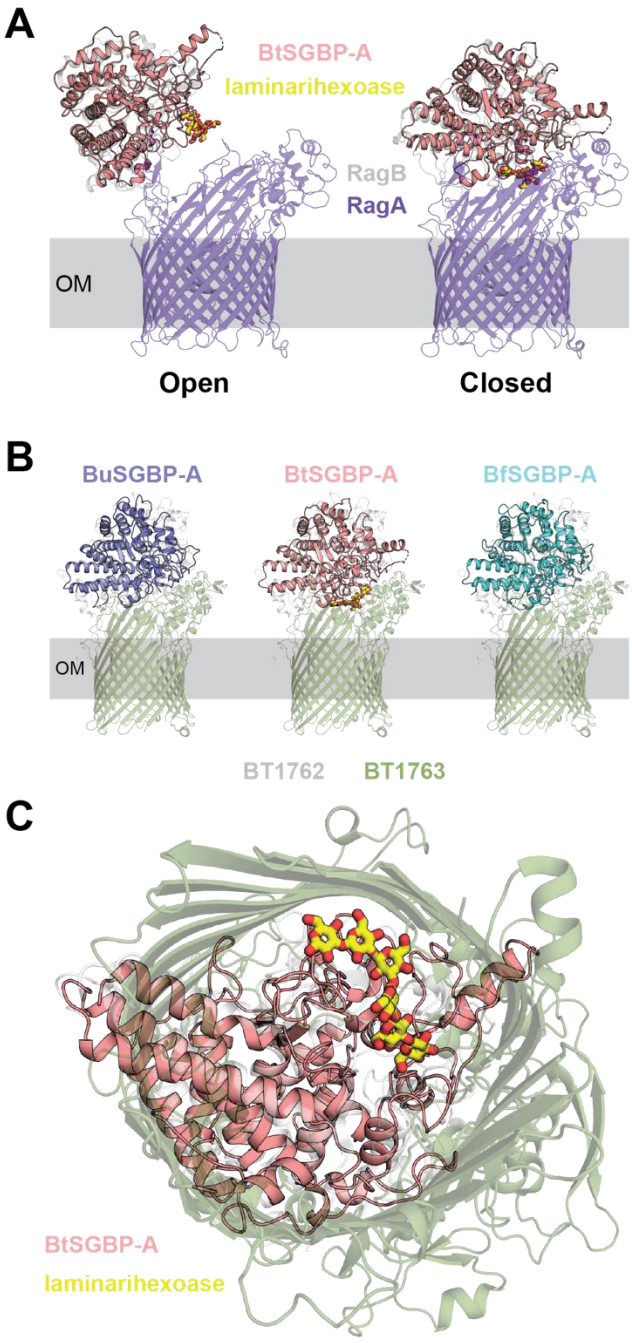

Figure 5-6. 1,3GUL SGBPs-A overlays with known SusCD-homolog complexes.

(A) SSM superposition of BtSGBP-A (rose) with RagB (transparent white) in complex with RagA (transparent purple) in the open (PDB ID: 6SML) and closed (PDB ID: 6SM3) conformations. Laminarihexaose bound to BtSGBP-A is shown as thick yellow sticks throughout figure. (B) SSM superposition of BuSGBP-A (slate), BtSGBP-B (rose), and

BfSGBP-B (cyan) with BT1762 (transparent white) in complex with BT1763 (transparent green) in the closed conformation (PDB ID: 5T3R). (C) A view through the barrel from the bottom of the BT1762/1763 complex (the plug domain of the TBDT was not resolved in 5T3R) with BtSGBP-A overlaid on BT1763.

In this structural context, it is notable that the syntenic orthologs BtSGBP-A and BfSGBP-A both bind β(1,3)-glucans by specifically recognizing the non-reducing end of the polysaccharide. This *exo*-binding mode contrasts the *endo*-binding mode observed, for example, with the MLG-specific homolog BoSGBP$_{MLG}$-A from the *Bacteroides ovatus* MLG utilization locus (MLGUL) (100). On the other hand, a *B. thetaiotaomicron* SGBP-A (BT1043) from a mucin utilization locus was found to bind the disaccharide *N*-acetyl lactosamine (LacNAc, the enzymatic hydrolysis product of poly-LacNAc) via the reducing end (224). Inspection of the modelled BtSGBP-A:laminarihexose/TBDT complex suggests that glycan chain-end binding may play a specific functional role in capturing short oligosaccharides in the "lid" of the "pedal-bin" in a concerted cycle that includes opening and closing of the opposite end of the TBDT by movement of the plug domain.

Intriguingly, the free BtSGBP-A can bind longer β(1,3)-glucan polysaccharides, yet these superposed *models* suggest that binding laminarioligosaccharides with DP > 6 would create a steric clash with the side of the TBDT in the fully closed conformation. Yet, in the absence of an *experimental* structure of BtSGBP-A in complex with its cognate TBDT, the molecular details of the transport cycle of β(1,3)-glucans remain unresolved. The extent to which individual SGBPs-A function to initially capture and thread longer oligosaccharide and polysaccharides into the β-barrel, and then remain in an open conformation, will require the direct observation of active ternary complexes. This latter mode can be likened to feeding a length of rope through an open hand, which may be especially relevant for *endo*-type SGBPs-A.

We also note that the targeting of chain ends may also be a consequence of adaption to the often short β(1,3)-glucan molecules found in nature from microbial sources (320). Plant MLGs, in contrast, are typically much longer polymers (235). Overall, the binding sites of BtSGBP-A and BfSGBP-A are considerably more round when compared to the flat binding platform of the MLG-specific *B. ovatus* SGBP$_{MLG}$-A (100). This difference is concordant with the conformations assumed by their respective targets: β(1,3)-glucans have an extended helical structure in solution, whereas MLGs have regular kinks in an otherwise linear and flat glucan chain (Fig. 5-1) (235,320).

Overall, tailoring of the binding site to distinct β-glucan types underscores the evolutionary plasticity of the variable region of SGBPs-A relative to the TPR repeats that forms the scaffold of these SusD homologs. In this context, we previously showed that BuSGBP-A of the *B. uniformis* 1,3GUL does not bind the cognate polysaccharide (308). Here, crystallography of BuSGBP-A leads us to speculate that accumulated mutations causing disorder of the binding site loops may have disrupted glycan binding, in the absence of selective pressure. In this 1,3GUL system, the corresponding SGBP-B, the structure of which we describe below, appears to compensate for this loss-of-function (Fig. 5-1) (308).

Our tertiary structural analysis also underscores how the PUL paradigm transcends environmental niches by revealing the commonality between human-gut and marine Bacteroidetes glycan capture systems. A homologous SGBP-A, GMSusD, from the marine Bacteroidetes *Gramella* sp. MAR_2010_102, was recently structurally characterized, although a ligand-complexed structure is not available (226). However, three tryptophan sidechains determined to be critical to binding in GMSusD are positioned homologously in the BtSGBP-A:laminarihexaose complex (all-Cα RMSD 1.26 Å, sequence identity 27.9 %) (Fig. D-12). Likewise, key binding site residues such as Asp90 and Arg367, which make multiple, critical interactions to the reducing end glucosyl residue in BtSGBP-A, are conserved, suggesting that GMSusD binds β(1,3)-glucans in the same orientation. The gene encoding GMSusD is part of a predicted laminarin utilization locus (226) that is partially syntenic to the human gut *Bacteroides* 1,3GULs. This observation underscores the evolutionary connection of not just the proteins, but also the PULs to which they belong, despite operating in distinctly different environments. In this context, there is notable precedent for the transfer of carbohydrate utilization systems from marine to human gut bacteria (198,323).

5.4.2 Structurally diverse SGBPs-B provide complement glycan recruitment

PULs very often encode additional SGBP(s), immediately downstream of the *SGBP-A* (*SusD*) homolog, to aid in recruitment of target polysaccharide to the cell surface. Such is the case for the three syntenic Bacteroides 1,3GULs (Fig. 5-1). However, these SGBPs cannot be confidently identified by bioinformatic approaches due to their poor primary and tertiary structural similarity. Yet despite this lack of conservation, the few (seven) known SGBPs-B are typified by multi-modular architectures (99,100,103,107) (see also unpublished PDB IDs 3ORJ and 6D2Y). The SGBP-B structures presented here further add to this diversity with unique tandem

arrangements and distinct structures of the individual modules (Figs. 5-4, 5-5). Variability even within syntenic PULs highlights the significant degree of convergent evolution of SGBPs-B in the PUL paradigm, in stark contrast to the rigorously conserved SGBPs-A.

Our full-length structures of BuSGBP-B and BtSGBP-B reveal that the individual tandem CBM-like domains are nearly identical and thus likely arose as a result of an intra-gene duplication event. A further distinguishing feature of BuSGBP-B and BtSGBP-B is the comparatively long linkers separating the discrete CBM-like domains following the N-terminal Ig-like membrane spacer domain. These linkers tantalizingly suggest mobility of these domains with respect to one another. Indeed, full-length and truncation variant structures captured in various conformations suggest the proteins behave like beads on a loose string.

Among the handful of SGBP-B structures determined to date, only the starch-targeting SusE and SusF were found to have multiple domains that each possess glycan binding capacity (107). In contrast, those targeting xyloglucan (99), heparin/heparan sulfate (103), and MLG (100) all only possess a single binding site at the distal C-terminal domain. Therefore, BuSGBP-B and BtSGBP-B presented here represent newly discovered members of the former category. We also show that the CBML-middle domain is solely responsible for enabling BuSGBP-B to bind both β(1,3)-glucan and MLG. The ability of BuSGBP-B to bind these structurally distinct β-glucans is an important factor allowing *B. uniformis* to grow on both polysaccharides (Fig. 5-1) (308). In contrast, *B. thetaiotaomicron* cannot grow on MLG, despite possessing an enzyme that can hydrolyze it in its 1,3GUL, due to a lack of an SGBP that binds the cereal glucan (Fig. 5-1) (308). Unfortunately, the structural basis by which the single domain of BuSGBP-B recognizes both β(1,3)-glucan and MLG remains unclear due to our inability to obtain a glycan complex.

A single, promiscuous binding site that can accommodate different glycans is not uncommon among glycan-binding proteins, but generally involves polysaccharides with similar solution structures. In this context, reports of proteins that can bind glycans with distinctly different physicochemical properties like β(1,3)-glucan and MLG is unusual. Indeed, a survey of the current CBM families (51,293) reveals that four families that contain β(1,3)-glucan-binders (CBM39, 43, 52, and 56) and six contain MLG-binders (CBM11, 72, 78, 79, 80, and 81),but only two families contain mixed β(1,3)-glucan/MLG-binders (CBM4 and 6), e.g. *Thermotoga maritima* CBM4-2 attached to a GH16 laminarinase (322,324) and *Cellvibrio mixtus* CBM6-2 attached to a GH5 endo-glucanase (325,326). However, structural insights that explain the ability to bind both types

of β-glucans are scarce. CmCBM6-2 does so through two binding clefts with individual specificities (325,326). Thus, binding sites optimized for β(1,3)-glucans generally possess features that are incompatible with binding MLG and *vice versa* (324,327). The CBML-middle module of BuSGBP-B appears to be one of few exceptions.

BfSGBP-B represents an additional, distinct architecture that supports binding of both β(1,3)-glucans and MLGs. Whereas BuSGBP-B comprises an Ig-like::CBML::CBML trimodular structure, BfSGBP-B consists of a PKD::β-barrel bimodular structure, which presents the binding site on the C-terminal module. We surmise that the PKD domain acts as a spacer to distance the binding end of the protein further away from the membrane, analogous to the Ig-like domain of BuSGBP-B. Such "passive" spacer domains are also observed in cell-surface GHs encoded by PULs (e.g. PFAM PF13004) (138,308). Notably, our individual complexes of BfSGBP-B with laminaritriose and G4G3G represent, to our knowledge, the first structures of a single binding site that accommodates β(1,3)-glucans and MLGs with equal affinity. In this case, binding MLGs by BfSGBP-B would appear to be a non-deleterious, off-target activity: The 1,3GUL of *B. fluxus* lacks a corresponding outer-membrane GH to hydrolyze MLG and therefore the bacterium does not grow on this cereal polysaccharide (Fig. 5-1) (308).

5.4.3 Binding chain ends as a strategy for β(1,3)-glucan recognition

CBMs have been delineated into three classes based their modes of binding, which is influenced by binding-site topology. Type-A CBMs bind crystalline polysaccharides via complementary flat faces, Type-B CBMs bind in an *endo*-mode along the chain of amorphous/soluble glycans (often with cleft-shaped binding sites), and Type-C CBMs bind in an *exo*-mode at the termini of glycan chains (often with pocket-shaped binding sites) (76). In this context, the distinction of SGBPs, CBMs, and lectins, which are united by their functions as glycan-binding proteins, is largely semantic (328). The majority of SGBPs characterized to-date bind their substrates in a Type-A or Type-B mode (99,100,103,107,146). Thus, the binding of β-glucan at the chain termini in a Type-C or *exo*-mode at the non-reducing end, in the case of BtSGBP-A, and the reducing end, in the case of BfSGBP-B, is noteworthy.

Recognition of chain ends may be a general strategy for targeting β(1,3)-glucans, which has precedent outside of the PUL paradigm. That is, diverse CBMs, lectins, receptors and pattern-recognition proteins also specifically bind β(1,3)-glucan chain ends (329). The structure of BhCBM6 from a *Bacillus halodurans* laminarinase was determined with a laminarihexaose

molecule bound via the reducing end (330). Similar to BtSGBP-A and BfSGBP-A, BhCBM6 bound to laminarioligosaccharides with DP ≥ 5, suggesting the presence of a comparable number of subsites. In contrast however, the reducing-end glucose is sandwiched in BhCBM6 by two tryptophan side chains making CH-π interactions to both faces of the sugar ring. The binding site of BhCBM6 is therefore closer to a *bona fide* CBM Type-C binding "pocket" (330). Similarly, in CmCBM6-2 (*vide supra*) the higher affinity laminarin binding site resembles a pocket lined by a tyrosine and tryptophan on opposing sides (326). CsCBM6-3 from a putative *Clostridium stercorarium* xylanase likewise employs a phenylalanine and tyrosine to sandwich the reducing end of laminaribiose, despite β(1,3)-glucan not being the main target of this CBM (331). In contrast, the binding sites of BtSGBP-A and BfSGBP-A are better described as shallow binding clefts or platforms that are blocked at one end; this topology is reminiscent of some *exo*-acting β(1,4)-glucanases (332,333). Altogether, these structural data point to convergent evolution of chain-end recognition as an effective strategy for binding β(1,3)-glucans in diverse niches.

5.5 Conclusion

Within the Bacteroidetes PUL paradigm, non-catalytic, cell surface-anchored glycan-binding proteins – SGBPs – play cornerstone roles in glycan recognition and transport. The suite of SGBP structures from three syntenic 1,3GULs presented here significantly expands our structural insight into this important class of PUL components. As underscored here, SGBPs-A have a highly conserved SusD-like scaffold, which is central to their essential role in pairing with a cognate TBDT. Yet, a tunable glycan-binding site allows tailoring to specific glycans, in this case β(1,3)-glucan. In stark contrast to SGBPs-A, SGBPs-B are structurally diverse, a theme that we elaborate here by revealing novel domain architectures and binding strategies. Understanding how SGBPs-B might work in concert (104,334) with the TBDT/SGBP-A complex (94,97) will be key to fully illuminating the glycan catch-and-sequester scheme of PULs.

Chapter 6: Conclusion

Scientific as well as public interest in the HGM has flourished in recent decades, in concert with growing appreciation of its influence on almost all aspects of human health. In particular, the field has greatly benefited from the rapid advances in sequencing technologies and bioinformatics capabilities, and we now have a much better understanding of this symbiotic community of microbes than we did even a single decade ago (171,287,335). Despite the advances, to say that we have only scratched the surface of comprehensive knowledge of the HGM is almost an understatement. The sheer complexity of the system makes the task of gaining enough insight to harness it as a therapeutic route to treat diseases and improve health, a mammoth undertaking. Indeed, the task necessarily requires a multi-faceted endeavor bringing together many different scientific disciplines and expertise.

One area of critical interest is the impact of diet-microbiota interactions on human health. Diet constantly and dramatically alters the HGM, the overall composition and population dynamics of which is shaped by glycan metabolism by the HGM (9,10,13). Albeit complex and personalized, significant associations between HGM composition and various disease states have been shown. In this context, it is essential to gain functional insights into glycan utilization by gut bacteria, the root process behind taxa-specific promotion of proliferation.

Bacteroidetes are a major phylum of bacteria that dominates the HGM (336), alongside the Firmicutes, thanks in large part to their consolidated PUL strategy. By devoting a significant portion of their genome to unique PULs (109), Bacteroidetes are able to grow on a huge diversity of glycans that may be available at any given time, earning them the moniker of "glycan generalists".

MLG and β(1,3)-glucan are two such glycans from the β-glucan category that prominently feature in the human diet worldwide. Chapters 2 and 3 together outline the molecular pathway of MLG degradation and sequestration by *B. ovatus* with focus on the GHs and SGBPs, respectively. MLGULs were shown to be genetic markers of MLG metabolism, and the prevalence of MLGUL-harboring Bacteroidetes in the HGM of populations across nationalities and disease states points to general relevance of potential MLG-based therapeutic strategies. Chapters 4 and 5 outline the molecular pathway of variable β(1,3)-glucan utilization capabilities by *B. uniformis*, *B. thetaiotaomicron*, and *B. fluxus*. These 1,3GULs were interestingly found to have differential

abilities to effect utilization of β(1,3)-glucan congeners as well as even MLG in the case of that from *B. uniformis*. Therefore, in these chapters, I make extensive use of X-ray crystallography to reveal structural bases of specificities towards β(1,3)-glucan and MLG. This includes the determination of the first structural representative of GH158 as well as a series of unique SGBPs, the proteins which together combine to dictate parent 1,3GUL specificity.

These two prominent health-promoting categories of β-glucan in the broader context of the human diet is but a portion of the plethora of glycans presented to the HGM *de jour*. The works within this thesis thus represent a few important pieces of the puzzle, many more pieces of which must be uncovered to gain holistic understanding of glycan metabolism by the HGM. In this context, tremendous concomitant progress has been made in functionally characterizing the pathways by which the HGM mediates deconstruction of diverse storage polysaccharides (101,134,337), hemicelluloses (99,100,110,135,138,211,338-341), marine polysaccharides (198,306,323), pectins (35,111), fungal/microbial glycans (194,308,309), host/eukaryotic glycans (103,257,259), and human milk oligosaccharides (49). In contrast to the wealth of new recent insights focused heavily on the soluble proteins (CAZymes and SGBPs), insights into the transport mechanism as well as regulation are lagging. Valiant efforts are certainly being made in these areas, but the road ahead is certainly significantly longer than that behind.

We now have a much better grasp of how individual PULs operate in isolation; however in many cases these apparently "selfish" PUL systems may be "leaky" with particular benefit to the community; partial breakdown products (PBPs) released by the action of certain PULs can be shared with neighboring bacteria and support the dynamic response of microbial communities (196,234). One example of this distributive mechanism is observed in xylan utilization by *B. ovatus*, where PBP produced at the cell surface diffuse into the extracellular environment and are utilized by *Bifidobacterium adolescentis*, a species lacking the enzymatic machinery to catalyze the initial depolymerisation step (Fig. 6-1A) (110). Interestingly, this form of syntrophy was observed during the utilization of relatively simple glucuronoxylan but not during utilization of more complex glucuronoarabinoxylan. These synergistic interactions therefore appear to be glycan- and species-specific, and may reflect hierarchies in selective metabolism of substrates. In this regard, the preferential degradation of some glycans over others is likely to play a central role in shaping the complex microbial relationships of the microbiota (194,342).

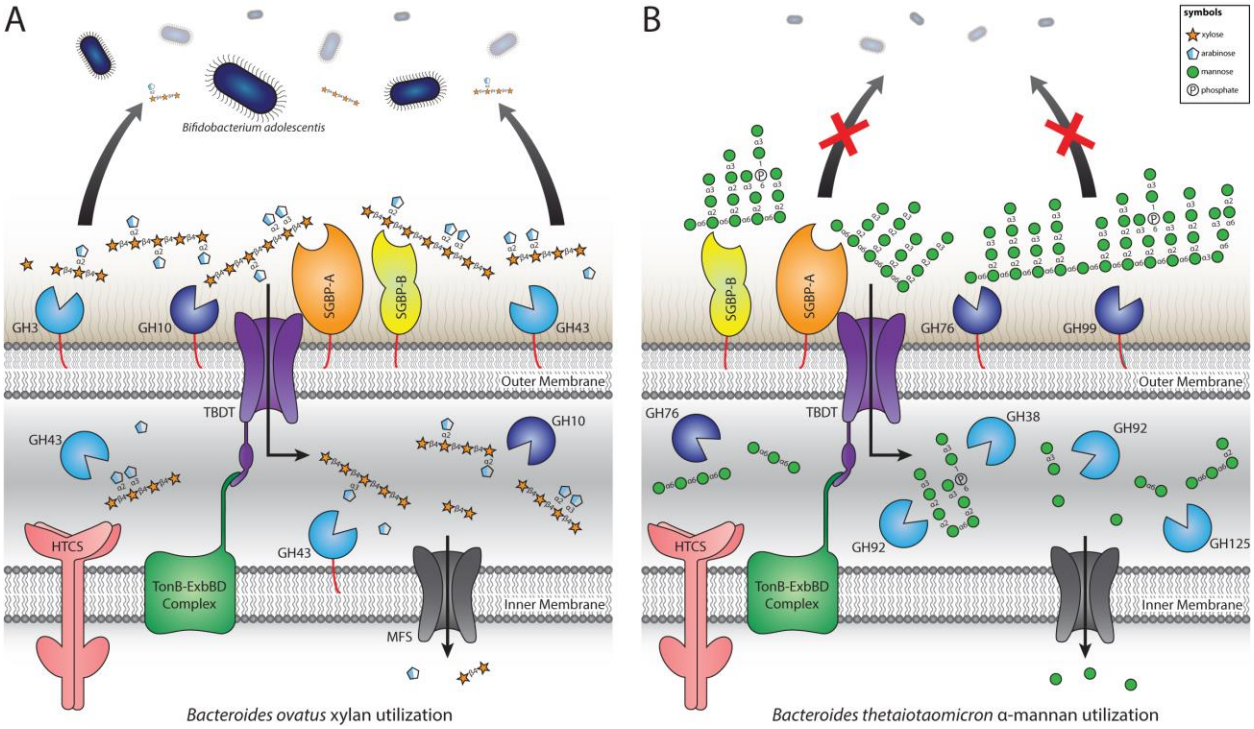

Figure 6-1. Nutritional foraging strategies encoded by PULs and their roles in microbial ecological interactions. (A) In a distributive mechanism, utilization of wheat arabinoxylan by *B. ovatus* releases PBPs that diffuse into the extracellular environment and support the growth of *Bifidobacterium adolescentis* (343). (B) In contrast, the selfish mechanism employed by *B. thetaiotaomicron* in the digestion of yeast α-mannans results in rapid import of extracellular products into the periplasm where saccharification is culminated (344). The concerted actions of these two PUL models drive syntrophic and cooperative networks in the context of the complex microbial environment of the gut microbiota. Proteins are coloured as in Figure 1. Monosaccharides are represented by Consortium for Functional Glycomics symbols (26).

The PUL-mediated liberation of PBPs contributes to the complex metabolic web of cross-feeding interactions that has been mapped between several Bacteroidales type-strains (234), although in the context of the entire gut microbiota these relationships are likely to be much more complex. For example, in some species CAZymes are selectively packaged into outer membrane vesicles and released into the extracellular environment where they are thought to mediate the production of free glycan fragments for use by the greater gut community (176,234). Remarkably, certain species, such as *B. ovatus*, secrete enzymes that are not required for the utilization of glycans such as inulin by the bacterium itself or by its clonemates; rather, this effort appears to benefit other species in the gut community (196). This seemingly altruistic act results in significant fitness benefits for *B. ovatus* that are only realized in the context of a complex microbiota.

In contrast to the extensive and complex dynamic relationships that exist between cohorts of bacteria in the gut, certain species such as *B. thetaiotaomicron* exhibit relatively little collaboration during the digestion of complex glycans. This truly "selfish" mechanism is deployed by *B. thetaiotaomicron* during the utilization of yeast α-mannans, in which manno-oligosaccharides generated at the cell surface are rapidly imported into the periplasm for further breakdown, conferring no direct benefits to neighboring species (Fig. 6-1B) (194).

Thus, progress is also being made in revealing how these individual puzzle pieces fit together, which is critical to obtaining a clear picture of glycan metabolism by the HGM in its entirety. Disentangling cross-feeding networks and community interactions in the greater context of the HGM will be essential to informing future prebiotic and other novel therapeutic strategies based on precise manipulation of this microbial community.

Though of great prominence in the context of the HGM, Bacteroidetes and other adept utilizers of diverse glycans are widespread across ecological niches, including in the gastrointestinal tracts of other animals such as ruminants and insects, as well as in marine, fresh water, and terrestrial habitats (Fig. 6-2). By extension, PULs fuel microbial communities all around us, not just within us. Therefore, advances in functional understanding of individual (gp)PULs and their role in the broader context of the HGM will inform microbial community dynamics in other ecological niches, and vice versa. Furthermore, applications beyond the realm of human health may spring from PUL research; for instance, efficient degradation of biomass feedstock is of particular interest in the renewable energy space. Their broad relevance in glycobiology, microbiology, and ecology combined with continued appreciation of the importance of the HGM to human health means that strategies for complex glycan utilization will remain a flourishing area of research for the foreseeable future.

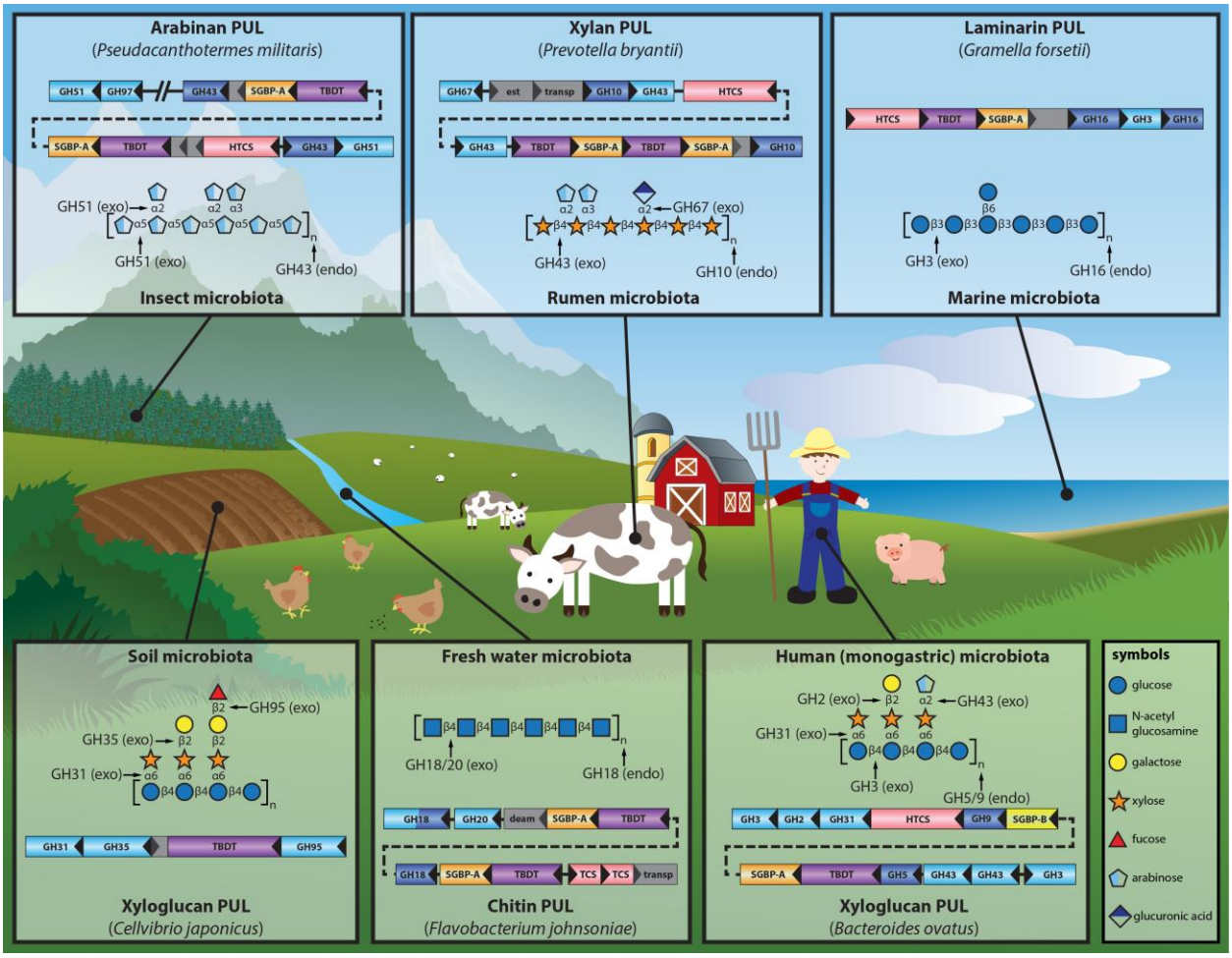

Figure 6-2 The ecological distribution of PULs in nature.

PULs are found in a variety of microbial communities, highlighting the global role of this polysaccharide utilization strategy. Each semitransparent box contains a representative bacterial PUL from a distinct microbial ecosystem, as well as the schematic structure of the target glycan. Clockwise from top left: arabinan PUL from *Pseudocanthotermes militaris* (345), xylan PUL from *Prevotella bryantii* (346), laminarin PUL from *Gramella forsetii* (347), xyloglucan PUL from *Bacteroides ovatus* (138), chitin PUL from *Flavobacterium johnsoniae* (348), xyloglucan PUL from *Cellvibrio japonicus* (349). Genes are colored according to protein function as follows: blue, *endo*-glycoside hydrolase (GH); cyan, *exo*-GH; orange, SusD-homologous surface glycan binding protein (SGBP); yellow, other SGBP; purple, TonB-dependent transporter (TBDT); pink, (hybrid) two-component sensor (HTCS/TCS); grey, unknown or other function (est, esterase; transp, transporter; deam, deaminase). Monosaccharides are represented by Consortium for Functional Glycomics symbols (26).

References

1. Sender, R., Fuchs, S., and Milo, R. (2016) Revised Estimates for the Number of Human and Bacteria Cells in the Body. *Plos Biology* **14**
2. Thomas, S., Izard, J., Walsh, E., Batich, K., Chongsathidkiet, P., Clarke, G., Sela, D. A., Muller, A. J., Mullin, J. M., Albert, K., Gilligan, J. P., DiGuilio, K., Dilbarova, R., Alexander, W., and Prendergast, G. C. (2017) The Host Microbiome Regulates and Maintains Human Health: A Primer and Perspective for Non-Microbiologists. *Cancer Research* **77**, 1783-1812
3. Zhang, X., Chen, B. D., Zhao, L. D., and Li, H. (2020) The Gut Microbiota: Emerging Evidence in Autoimmune Diseases. *Trends Mol Med*
4. Turnbaugh, P. J., Ley, R. E., Mahowald, M. a., Magrini, V., Mardis, E. R., and Gordon, J. I. (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027-1031
5. Arrieta, M.-c., Stiemsma, L. T., Dimitriu, P. A., Thorson, L., Russell, S., Yurist-doutsch, S., Kuzeljevic, B., Gold, M. J., Britton, H. M., Lefebvre, D. L., Subbarao, P., Mandhane, P., Becker, A., McNagny, K. M., Sears, M. R., Kollmann, T., Investigators, S., Mohn, W. W., Turvey, S. E., and Finlay, B. B. (2015) Early infancy microbial and metabolic alterations affect risk of childhood asthma. **7**
6. Fujimura, K. E., and Lynch, S. V. (2015) Microbiota in Allergy and Asthma and the Emerging Relationship with the Gut Microbiome. *Cell Host & Microbe* **17**, 592-602
7. Schwabe, R. F., and Jobin, C. (2013) The microbiome and cancer. *Nature Reviews Cancer* **13**, 800-812
8. Ridaura, V. K., Faith, J. J., Rey, F. E., Cheng, J. Y., Duncan, A. E., Kau, A. L., Griffin, N. W., Lombard, V., Henrissat, B., Bain, J. R., Muehlbauer, M. J., Ilkayeva, O., Semenkovich, C. F., Funai, K., Hayashi, D. K., Lyle, B. J., Martini, M. C., Ursell, L. K., Clemente, J. C., Van Treuren, W., Walters, W. A., Knight, R., Newgard, C. B., Heath, A. C., and Gordon, J. I. (2013) Gut Microbiota from Twins Discordant for Obesity Modulate Metabolism in Mice. *Science* **341**, 1079
9. 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., Nunez, G., and Martens, E. C. (2016) A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell* **167**, 1339-1353
10. David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., Ling, A. V., Devlin, A. S., Varma, Y., Fischbach, M. A., Biddinger, S. B., Dutton, R. J., and Turnbaugh, P. J. (2014) Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559-563
11. Zmora, N., Soffer, E., and Elinav, E. (2019) Transforming medicine with the microbiome. *Science Translational Medicine* **11**, 3
12. Kolodziejczyk, A. A., Zheng, D. P., and Elinav, E. (2019) Diet-microbiota interactions and personalized nutrition. *Nature Reviews Microbiology* **17**, 742-753
13. Koropatkin, N. M., Cameron, E. A., and Martens, E. C. (2012) How glycan metabolism shapes the human gut microbiota. *Nature Reviews Microbiology* **10**, 323-335

14. Zmora, N., Suez, J., and Elinav, E. (2019) You are what you eat: diet, health and the gut microbiota. *Nature Reviews Gastroenterology & Hepatology* **16**, 35-56
15. El Kaoutari, A., Armougom, F., Gordon, J. I., Raoult, D., and Henrissat, B. (2013) The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nature Reviews Microbiology* **11**, 497-504
16. Goodrich, J. K., Davenport, E. R., Clark, A. G., and Ley, R. E. (2017) The Relationship Between the Human Genome and Microbiome Comes into View. in *Annual Review of Genetics, Vol 51* (Bonini, N. M. ed.), Annual Reviews, Palo Alto. pp 413-433
17. O'Hara, A. M., and Shanahan, F. (2006) The gut flora as a forgotten organ. *Embo Reports* **7**, 688-693
18. Grosskopf, A., and Simm, A. (2020) Carbohydrates in nutrition: friend or foe? *Zeitschrift Fur Gerontologie Und Geriatrie* **53**, 290-294
19. Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A. G., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H. (2015) Essentials of Glycobiology. pp
20. Zakowski, J. J., and Bruns, D. E. (1985) BIOCHEMISTRY OF HUMAN ALPHA-AMYLASE ISOENZYMES. *Crc Critical Reviews in Clinical Laboratory Sciences* **21**, 283-322
21. Cerqueira, F. M., Photenhauer, A. L., Pollet, R. M., Brown, H. A., and Koropatkin, N. M. (2020) Starch Digestion by Gut Bacteria: Crowdsourcing for Carbs. *Trends in Microbiology* **28**, 95-108
22. Keegstra, K. (2010) Plant Cell Walls. *Plant Physiology* **154**, 483-486
23. Doblin, M. S., Pettolino, F., and Bacic, A. (2010) Plant cell walls: the skeleton of the plant world. *Functional Plant Biology* **37**, 357-381
24. Scheller, H. V., and Ulvskov, P. (2010) Hemicelluloses. in *Annual Review of Plant Biology, Vol 61* (Merchant, S., Briggs, W. R., and Ort, D. eds.), Annual Reviews, Palo Alto. pp 263-289
25. Rennie, E. A., and Scheller, H. V. (2014) Xylan biosynthesis. *Current Opinion in Biotechnology* **26**, 100-107
26. Varki, A., Cummings, R. D., Aebi, M., Packer, N. H., Seeberger, P. H., Esko, J. D., Stanley, P., Hart, G., Darvill, A., Kinoshita, T., Prestegard, J. J., Schnaar, R. L., Freeze, H. H., Marth, J. D., Bertozzi, C. R., Etzler, M. E., Frank, M., Vliegenthart, J. F., Lutteke, T., Perez, S., Bolton, E., Rudd, P., Paulson, J., Kanehisa, M., Toukach, P., Aoki-Kinoshita, K. F., Dell, A., Narimatsu, H., York, W., Taniguchi, N., and Kornfeld, S. (2015) Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* **25**, 1323-1324
27. Vincken, J. P., York, W. S., Beldman, G., and Voragen, A. G. J. (1997) Two general branching patterns of xyloglucan, XXXG and XXGG. *Plant Physiology* **114**, 9-13
28. Schultink, A., Liu, L., Zhu, L., and Pauly, M. (2014) Structural Diversity and Function of Xyloglucan Sidechain Substituents. *Plants (Basel)* **3**, 526-542
29. Moreira, L. R. S., and Filho, E. X. F. (2008) An overview of mannan structure and mannan-degrading enzyme systems. *Applied Microbiology and Biotechnology* **79**, 165-178
30. Planas, A. (2000) Bacterial 1,3-1,4-beta-glucanases: structure, function and protein engineering. *Biochim Biophys Acta* **1543**, 361-382

31. Woodward, J. R., Fincher, G. B., and Stone, B. A. (1983) Water-soluble (1→3), (1→4)-β-D-glucans from barley (Hordeum vulgare) endosperm. II. Fine structure. *Carbohydrate Polymers* **3**, 207-225
32. Xue, X. X., and Fry, S. C. (2012) Evolution of mixed-linkage (13,14)--d-glucan (MLG) and xyloglucan in Equisetum (horsetails) and other monilophytes. *Annals of Botany* **109**, 873-886
33. Mohnen, D. (2008) Pectin structure and biosynthesis. *Current Opinion in Plant Biology* **11**, 266-277
34. Thakur, B. R., Singh, R. K., and Handa, A. K. (1997) Chemistry and uses of pectin - A review. *Critical Reviews in Food Science and Nutrition* **37**, 47-73
35. Ndeh, D., Rogowski, A., Cartmell, A., Luis, A. S., Baslé, A., Gray, J., Venditto, I., Briggs, J., Zhang, X., Labourel, A., Terrapon, N., Buffetto, F., Nepogodiev, S., Xiao, Y., Field, R. A., Zhu, Y., O'Neill, M. A., Urbanowicz, B. R., York, W. S., Davies, G. J., Abbott, D. W., Ralet, M.-C., Martens, E. C., Henrissat, B., and Gilbert, H. J. (2017) Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature*
36. Luna, E., Pastor, V., Robert, J., Flors, V., Mauch-Mani, B., and Ton, J. (2011) Callose Deposition: A Multifaceted Plant Defense Response. *Molecular Plant-Microbe Interactions* **24**, 183-193
37. Silva, J., Ferraz, R., Dupree, P., Showalter, A. M., and Coimbra, S. (2020) Three Decades of Advances in Arabinogalactan-Protein Biosynthesis. *Frontiers in Plant Science* **11**, 18
38. Le Bastard, Q., Chapelet, G., Javaudin, F., Lepelletier, D., Batard, E., and Montassier, E. (2020) The effects of inulin on gut microbial composition: a systematic review of evidence from human studies. *European Journal of Clinical Microbiology & Infectious Diseases* **39**, 403-413
39. Joshi, S., S., E., and V., J. (2019) Marine Polysaccharides: Biomedical and Tissue Engineering Applications. in *Marine-Derived Biomaterials for Tissue Engineering Applications. Springer Series in Biomaterials Science and Engineering* (A., C., and B., B.-N. eds.), Springer, Singapore. pp 443-487
40. Usov, A. I., and Zelinsky, N. D. (2014) Chemical structures of algal polysaccharides. in *Functional Ingredients from Algae for Foods and Nutraceuticals*, Woodhead Publishing Series in Food Science, Technology and Nutrition. pp 23-86
41. Wang, X. Y., and Quinn, P. J. (2010) Lipopolysaccharide: Biosynthetic pathway and structure modification. *Progress in Lipid Research* **49**, 97-107
42. Kwiatkowski, S., and Kwiatkowski, S. E. (2012) *Yeast (Saccharomyces cerevisiae) Glucan Polysaccharides – Occurrence, Separation and Application in Food, Feed and Health Industries*, InTech
43. Friedman, M. (2016) Mushroom Polysaccharides: Chemistry and Antiobesity, Antidiabetes, Anticancer, and Antibiotic Properties in Cells, Rodents, and Humans. *Foods* **5**
44. McGuckin, M. A., Linden, S. K., Sutton, P., and Florin, T. H. (2011) Mucin dynamics and enteric pathogens. *Nature Reviews Microbiology* **9**, 265-278
45. Morla, S. (2019) Glycosaminoglycans and Glycosaminoglycan Mimetics in Cancer and Inflammation. *International Journal of Molecular Sciences* **20**, 19
46. Sarrazin, S., Lamanna, W. C., and Esko, J. D. (2011) Heparan Sulfate Proteoglycans. *Cold Spring Harbor Perspectives in Biology* **3**, 33

47. Mikami, T., and Kitagawa, H. (2013) Biosynthesis and function of chondroitin sulfate. *Biochimica Et Biophysica Acta-General Subjects* **1830**, 4719-4733
48. Caterson, B., and Melrose, J. (2018) Keratan sulfate, a complex glycosaminoglycan with unique functional capability. *Glycobiology* **28**, 182-206
49. Sakanaka, M., Gotoh, A., Yoshida, K., Odamaki, T., Koguchi, H., Xiao, J., Kitaoka, M., and Katayama, T. (2020) Varied Pathways of Infant Gut-Associated Bifidobacterium to Assimilate Human Milk Oligosaccharides: Prevalence of the Gene Set and Its Correlation with Bifidobacteria-Rich Microbiota Formation. *Nutrients* **12**, 21
50. Cantarel, B. I., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Research* **37**, 233-238
51. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**, D490-495
52. Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P., and Davies, G. (1995) CONSERVED CATALYTIC MACHINERY AND THE PREDICTION OF A COMMON FOLD FOR SEVERAL FAMILIES OF GLYCOSYL HYDROLASES. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 7090-7094
53. Aspeborg, H., Coutinho, P. M., Wang, Y., Brumer, H., and Henrissat, B. (2012) Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evolutionary Biology* **12**, 186-186
54. Stam, M. R., Danchin, E. G. J., Rancurel, C., Coutinho, P. M., and Henrissat, B. (2006) Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of alpha-amylase-related proteins. *Protein Engineering Design & Selection* **19**, 555-562
55. St John, F. J., Gonzalez, J. M., and Pozharski, E. (2010) Consolidation of glycosyl hydrolase family 30: A dual domain 4/7 hydrolase family consisting of two structurally distinct groups. *Febs Letters* **584**, 4435-4441
56. Viborg, A. H., Terrapon, N., Lombard, V., Michel, G., Czjzek, M., Henrissat, B., and Brumer, H. (2019) A subfamily roadmap of the evolutionarily diverse glycoside hydrolase family 16 (GH16). *Journal of Biological Chemistry* **294**, 15973-15986
57. Mewis, K., Lenfant, N., Lombard, V., and Henrissat, B. (2016) Dividing the Large Glycoside Hydrolase Family 43 into Subfamilies: a Motivation for Detailed Enzyme Characterization. *Applied and Environmental Microbiology* **82**, 1686-1692
58. Abbott, W., Alber, O., Bayer, E., Berrin, J. G., Boraston, A., Brumer, H., Brzezinski, R., Clarke, A., Cobucci-Ponzano, B., Cockburn, D., Coutinho, P., Czjzek, M., Dassa, B., Davies, G. J., Eijsink, V., Eklof, J., Felice, A., Ficko-Blean, E., Pincher, G., Fontaine, T., Fujimoto, Z., Fujita, K., Fushinobu, S., Gilbert, H., Gloster, T., Goddard-Borger, E., Greig, I., Hehemann, J. H., Hemsworth, G., Henrissat, B., Hidaka, M., Hurtado-Guerrero, R., Igarashi, K., Ishida, T., Janecek, S., Jongkees, S., Juge, N., Kaneko, S., Katayama, T., Kitaoka, M., Konno, N., Kracher, D., Kulminskaya, A., van Bueren, A. L., Larsen, S., Lee, J., Linder, M., LoLeggio, L., Ludwig, R., Luis, A., Maksimainen, M., Mark, B., McLean, R., Michel, G., Montanier, C., Moracci, M., Mori, H., Nakai, H., Nerinckx, W., Ohnuma, T., Pickersgill, R., Piens, K., Pons, T., Rebuffet, E., Reilly, P., Remaud-Simeon, M., Rempel, B., Robinson, K., Rose, D., Rouvinen, J., Saburi, W., Sakamoto,

Y., Sandgren, M., Shaikh, F., Shoham, Y., St John, F., Stahlberg, J., Suits, M., Sulzenbacher, G., Sumida, T., Suzuki, R., Svensson, B., Taira, T., Taylor, E., Tonozuka, T., Urbanowicz, B., Vaaje-Kolstad, G., Van den Ende, W., Varrot, A., Versluys, M., Vincent, F., Vocadlo, D., Wakarchuk, W., Wennekes, T., Williams, R., Williams, S., Wilson, D., Withers, S., Yaoi, K., Yip, V., Zhang, R., and Consortium, C. A. (2018) Ten years of CAZypedia: a living encyclopedia of carbohydrate-active enzymes. *Glycobiology* **28**, 3-8

59. Koshland, D. E. (1953) STEREOCHEMISTRY AND THE MECHANISM OF ENZYMATIC REACTIONS. *Biological Reviews of the Cambridge Philosophical Society* **28**, 416-436
60. Davies, G., and Henrissat, B. (1995) STRUCTURES AND MECHANISMS OF GLYCOSYL HYDROLASES. *Structure* **3**, 853-859
61. McCarter, J. D., and Withers, S. G. (1994) MECHANISMS OF ENZYMATIC GLYCOSIDE HYDROLYSIS. *Current Opinion in Structural Biology* **4**, 885-892
62. Ren, W. W., Farren-Dai, M., Sannikova, N., Swiderek, K., Wang, Y., Akintola, O., Britton, R., Moliner, V., and Bennet, A. J. (2020) Glycoside hydrolase stabilization of transition state charge: new directions for inhibitor design. *Chemical Science* **11**, 10488-10495
63. Davies, G. J., and Williams, S. J. (2016) Carbohydrate-active enzymes: sequences, shapes, contortions and cells. *Biochemical Society Transactions* **44**, 79-87
64. Sinnott, M. L. (1990) CATALYTIC MECHANISMS OF ENZYMATIC GLYCOSYL TRANSFER. *Chemical Reviews* **90**, 1171-1202
65. Bissaro, B., Monsan, P., Faure, R., and O'Donohue, M. J. (2015) Glycosynthesis in a waterworld: new insight into the molecular basis of transglycosylation in retaining glycoside hydrolases. *Biochemical Journal* **467**, 17-35
66. Arnal, G., Stogios, P. J., Asohan, J., Attia, M. A., Skarina, T., Viborg, A. H., Henrissat, B., Savchenko, A., and Brumer, H. (2019) Substrate specificity, regiospecificity, and processivity in glycoside hydrolase family 74. *Journal of Biological Chemistry* **294**, 13233-13247
67. Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008) Glycosyltransferases: Structures, functions, and mechanisms. in *Annual Review of Biochemistry*, Annual Reviews, Palo Alto. pp 521-555
68. Garron, M. L., and Cygler, M. (2010) Structural and mechanistic classification of uronic acid-containing polysaccharide lyases. *Glycobiology* **20**, 1547-1573
69. Lombard, V., Bernard, T., Rancurel, C., Brumer, H., Coutinho, P. M., and Henrissat, B. (2010) A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochemical Journal* **432**, 437-444
70. Biely, P. (2012) Microbial carbohydrate esterases deacetylating plant polysaccharides. *Biotechnology Advances* **30**, 1575-1588
71. Nakamura, A. M., Nascimento, A. S., and Polikarpov, I. (2017) Structural diversity of carbohydrate esterases. *Biotechnology Research and Innovation* **1**, 35-51
72. Polgar, L. (2005) The catalytic triad of serine peptidases. *Cellular and Molecular Life Sciences* **62**, 2161-2172
73. Montanier, C., Money, V. A., Pires, V. M. R., Flint, J. E., Pinheiro, B. A., Goyal, A., Prates, J. A. M., Izumi, A., Stalbrand, H., Morland, C., Cartmell, A., Kolenova, K., Topakas, E., Dodson, E. J., Bolam, D. N., Davies, G. J., Fontes, C., and Gilbert, H. J.

(2009) The Active Site of a Carbohydrate Esterase Displays Divergent Catalytic and Noncatalytic Binding Functions. *Plos Biology* **7**, 687-697
74. Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M., and Henrissat, B. (2013) Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnology for Biofuels* **6**, 14
75. Hemsworth, G. R., Johnston, E. M., Davies, G. J., and Walton, P. H. (2015) Lytic Polysaccharide Monooxygenases in Biomass Conversion. *Trends in Biotechnology* **33**, 747-761
76. Gilbert, H. J., Knox, J. P., and Boraston, A. B. (2013) Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. *Current Opinion in Structural Biology* **23**, 669-677
77. Abbott, D. W., and van Bueren, A. L. (2014) Using structure to inform carbohydrate binding module function. *Current Opinion in Structural Biology* **28**, 32-40
78. Hamaker, B. R., and Tuncil, Y. E. (2014) A perspective on the complexity of dietary fiber structures and their potential effect on the gut microbiota. *Journal of Molecular Biology* **426**, 3838-3850
79. Martens, E. C., Kelly, A. G., Tauzin, A. S., and Brumer, H. (2014) The Devil Lies in the Details: How Variations in Polysaccharide Fine-Structure Impact the Physiology and Evolution of Gut Microbes. *Journal of Molecular Biology* **426**, 3851-3865
80. Anderson, K. L., and Salyers, A. A. (1989) Biochemical-Evidence that Starch Breakdown by Bacteroides-Thetaiotaomicron Involves Outer-Membrane Starch-Binding Sites and Periplasmic Starch-Degrading Enzymes. *Journal of Bacteriology* **171**, 3192-3198
81. Anderson, K. L., and Salyers, A. A. (1989) Genetic evidence that outer membrane binding of starch is required for starch utilization by Bacteroides thetaiotaomicron. *Journal of Bacteriology* **171**, 3199-3204
82. Tancula, E., Feldhaus, M. J., Bedzyk, L. A., and Salyers, A. A. (1992) Location and characterization of genes involved in binding of starch to the surface of Bacteroides thetaiotaomicron. *Journal of Bacteriology* **174**, 5609-5616
83. Reeves, A. R., D'Elia, J. N., Frias, J., and Salyers, A. A. (1996) A Bacteroides thetaiotaomicron outer membrane protein that is essential for utilization of maltooligosaccharides and starch . These include : A Bacteroides thetaiotaomicron Outer Membrane Protein That Is Essential for Utilization of Maltooligosaccharides. *Journal of Bacteriology* **178**, 823-830
84. D'Elia, J. N., and Salyers, A. A. (1996) Contribution of a neopullulanase, a pullulanase, and an ??-glucosidase to growth of Bacteroides thetaiotaomicron on starch. *Journal of Bacteriology* **178**, 7173-7179
85. D'Elia, J. N., and Salyers, A. A. (1996) Effect of regulatory protein levels on utilization of starch by Bacteroides thetaiotaomicron. *Journal of Bacteriology* **178**, 7180-7186
86. Reeves, a. R., Wang, G. R., Salyers, a. a., Reeves, A. R., Wang, G.-r., and Salyers, A. a. (1997) Characterization of four outer membrane proteins that play a role in utilization of starch by Bacteroides thetaiotaomicron . Characterization of Four Outer Membrane Proteins That Play a Role in Utilization of Starch by Bacteroides thetaiotaomicron. *Journal of Bacteriology* **179**, 643-649
87. Shipman, J. a., Berleman, J. E., and Salyers, a. a. (2000) Characterization of four outer membrane proteins involved in binding starch to the cell surface of Bacteroides thetaiotaomicron. *Journal of Bacteriology* **182**, 5365-5372

88. Cho, K. H., Cho, D., Wang, G. R., and Salyers, A. A. (2001) New regulatory gene that contributes to control of Bacteroides thetaiotaomicron starch utilization genes. *Journal of Bacteriology* **183**, 7198-7205
89. Cho, K. H., and Salyers, A. A. (2001) Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by Bacteroides thetaiotaomicron. *Journal of Bacteriology* **183**, 7224-7230
90. Shipman, J. A., Cho, K. H., Siegel, H. A., and Salyers, A. A. (1999) Physiological characterization of SusG, an outer membrane protein essential for starch utilization by Bacteroides thetaiotaomicron. *Journal of Bacteriology* **181**, 7206-7211
91. Terrapon, N., Lombard, V., Gilbert, H. J., and Henrissat, B. (2015) Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. *Bioinformatics* **31**, 647-655
92. Terrapon, N., Lombard, V., Lap, P., Al-masaudi, S., Gilbert, H. J., and Henrissat, B. (2018) PULDB : the expanded database of Polysaccharide Utilization Loci. **46**, 677-683
93. Bolam, D. N., and van den Berg, B. (2018) TonB-dependent transport by the gut microbiota: novel aspects of an old problem. *Current Opinion in Structural Biology* **51**, 35-43
94. Gray, D. A., White, J. B. R., Oluwole, A. O., Rath, P., Glenwright, A. J., Mazur, A., Zahn, M., Basle, A., Morland, C., Evans, S. L., Cartmell, A., Robinson, C. V., Hiller, S., Ranson, N. A., Bolam, D. N., and van den Berg, B. (2020) Insights into glycan import by a prominent gut symbiont. *bioRxiv*
95. Bolam, D. N., and Koropatkin, N. M. (2012) Glycan recognition by the Bacteroidetes Sus-like systems. *Current Opinion in Structural Biology* **22**, 563-569
96. Bakolitsa, C., Xu, Q. P., Rife, C. L., Abdubek, P., Astakhova, T., Axelrod, H. L., Carlton, D., Chen, C., Chiu, H. J., Clayton, T., Das, D., Deller, M. C., Duan, L. A., Ellrott, K., Farr, C. L., Feuerhelm, J., Grant, J. C., Grzechnik, A., Han, G. W., Jaroszewski, L., Jin, K. K., Klock, H. E., Knuth, M. W., Kozbial, P., Krishna, S. S., Kumar, A., Lam, W. W., Marciano, D., McMullan, D., Miller, M. D., Morse, A. T., Nigoghossian, E., Nopakun, A., Okach, L., Puckett, C., Reyes, R., Tien, H. J., Trame, C. B., van den Bedem, H., Weekes, D., Hodgson, K. O., Wooley, J., Elsliger, M. A., Deacon, A. M., Godzik, A., Lesley, S. A., and Wilson, I. A. (2010) Structure of BT_3984, a member of the SusD/RagB family of nutrient-binding molecules. *Acta Crystallographica Section F-Structural Biology Communications* **66**, 1274-1280
97. Glenwright, A. J., Pothula, K. R., Bhamidimarri, S. P., Chorev, D. S., Basle, A., Firbank, S. J., Zheng, H. J., Robinson, C. V., Winterhalter, M., Kleinekathofer, U., Bolam, D. N., and van den Berg, B. (2017) Structural basis for nutrient acquisition by dominant members of the human gut microbiota. *Nature* **541**, 407-411
98. Madej, M., White, J. B. R., Nowakowska, Z., Rawson, S., Scavenius, C., Enghild, J. J., Bereta, G. P., Pothula, K., Kleinekathoefer, U., Basle, A., Ranson, N. A., Potempa, J., and van den Berg, B. Structural and functional insights into oligopeptide acquisition by the RagAB transporter from Porphyromonas gingivalis. *Nature Microbiology*, 20
99. Tauzin, A. S., Kwiatkowski, K. J., Orlovsky, N. I., Smith, C. J., Creagh, A. L., Haynes, C. A., Wawrzak, Z., Brumer, H., and Koropatkin, N. M. (2016) Molecular Dissection of Xyloglucan Recognition in a Prominent Human Gut Symbiont. *Mbio* **7**, 15
100. Tamura, K., Foley, M. H., Gardill, B. R., Dejean, G., Schnizlein, M., Bahr, C. M. E., Louise Creagh, A., van Petegem, F., Koropatkin, N. M., and Brumer, H. (2019) Surface

glycan-binding proteins are essential for cereal beta-glucan utilization by the human gut symbiont Bacteroides ovatus. *Cell Mol Life Sci* **76**, 4319–4340
101. Sonnenburg, E. D., Zheng, H., Joglekar, P., Higginbottom, S. K., Firbank, S. J., Bolam, D. N., and Sonnenburg, J. L. (2010) Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. *Cell* **141**, 1241-1252
102. Cameron, E. A., Kwiatkowski, K. J., Lee, B. H., Hamaker, B. R., Koropatkin, N. M., and Martens, E. C. (2014) Multifunctional Nutrient-Binding Proteins Adapt Human Symbiotic Bacteria for Glycan Competition in the Gut by Separately Promoting Enhanced Sensing and Catalysis. *Mbio* **5**, 1-12
103. Cartmell, A., Lowe, E. C., Basle, A., Firbank, S. J., Ndeh, D. A., Murray, H., Terrapon, N., Lombard, V., Henrissat, B., Turnbull, J. E., Czjzek, M., Gilbert, H. J., and 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**, 7037-7042
104. Karunatilaka, K., Cameron, E., Coupland, B., Koropatkin, N., Martens, E., and Biteen, J. (2013) Super-Resolution Imaging of Live Anaerobic Bacteria: Assembly and Dynamics of the B. Theta Starch Utilization System. *Biophysical Journal* **104**, 526
105. Grondin, J. M., Tamura, K., Dejean, G., Abbott, D. W., and Brumer, H. (2017) Polysaccharide Utilization Loci: Fueling Microbial Communities. *Journal of Bacteriology* **199**, 1-15
106. Tamura, K., and Brumer, H. (2021) Glycan utilization systems in the human gut microbiota: a gold mine for structural discoveries. *Current Opinion in Structural Biology* **68**, 26-40
107. Cameron, E. A., Maynard, M. A., Smith, C. J., Smith, T. J., Koropatkin, N. M., and Martens, E. C. (2012) Multidomain Carbohydrate-binding Proteins Involved in Bacteroides thetaiotaomicron Starch Metabolism. *Journal of Biological Chemistry* **287**, 34614-34625
108. Foley, M. H., Martens, E. C., and Koropatkin, N. M. (2018) SusE facilitates starch uptake independent of starch binding in B . thetaiotaomicron. **108**, 551-566
109. 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., and Gordon, J. I. (2011) Recognition and Degradation of Plant Cell Wall Polysaccharides by Two Human Gut Symbionts. *Plos Biology* **9**, 1-16
110. Rogowski, A., Briggs, J. A., Mortimer, J. C., Tryfona, T., Terrapon, N., Lowe, E. C., Basle, A., Morland, C., Day, A. M., Zheng, H. J., Rogers, T. E., Thompson, P., Hawkins, A. R., Yadav, M. P., Henrissat, B., Martens, E. C., Dupree, P., Gilbert, H. J., and Bolam, D. N. (2015) Glycan complexity dictates microbial resource allocation in the large intestine. *Nature Communications* **6**, 15
111. Luis, A. S., Briggs, J., Zhang, X., Farnell, B., Ndeh, D., Labourel, A., Baslé, A., Cartmell, A., Terrapon, N., Stott, K., Lowe, E. C., McLean, R., Shearer, K., Schückel, J., Venditto, I., Ralet, M.-c., Henrissat, B., Martens, E. C., Mosimann, S. C., Abbott, D. W., and Gilbert, H. J. (2018) Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic Bacteroides. *Nature Microbiology* **3**
112. Groisman, E. A. (2016) Feedback Control of Two-Component Regulatory Systems. in *Annual Review of Microbiology, Vol 70* (Gottesman, S. ed.), Annual Reviews, Palo Alto. pp 103-124

113. Capra, E. J., and Laub, M. T. (2012) Evolution of Two-Component Signal Transduction Systems. in *Annual Review of Microbiology, Vol 66* (Gottesman, S., Harwood, C. S., and Schneewind, O. eds.), Annual Reviews, Palo Alto. pp 325-347
114. Jacob-Dubuisson, F., Mechaly, A., Betton, J. M., and Antoine, R. (2018) Structural insights into the signalling mechanisms of two-component systems. *Nature Reviews Microbiology* **16**, 585-593
115. Raghavan, V., and Groisman, E. A. (2010) Orphan and hybrid two-component system proteins in health and disease. *Current Opinion in Microbiology* **13**, 226-231
116. Sineva, E., Savkina, M., and Ades, S. E. (2017) Themes and variations in gene regulation by extracytoplasmic function (ECF) sigma factors. *Current Opinion in Microbiology* **36**, 128-137
117. Martens, E. C., Chiang, H. C., and Gordon, J. I. (2008) Mucosal Glycan Foraging Enhances Fitness and Transmission of a Saccharolytic Human Gut Bacterial Symbiont. *Cell Host & Microbe* **4**, 447-457
118. Sheridan, P. O., Martin, J. C., Lawley, T. D., Browne, H. P., Harris, H. M. B., Bernalier-Donadille, A., Duncan, S. H., O'Toole, P. W., Scott, K. P., and Flint, H. J. (2016) Polysaccharide utilization loci and nutritional specialization in a dominant group of butyrate-producing human colonic Firmicutes. *Microbial Genomics* **2**, 16
119. Ndeh, D., and Gilbert, H. J. (2018) Biochemistry of complex glycan depolymerisation by the human gut microbiota. *Fems Microbiology Reviews* **42**, 146-164
120. Ze, X. L., Ben David, Y., Laverde-Gomez, J. A., Dassa, B., Sheridan, P. O., Duncan, S. H., Louis, P., Henrissat, B., Juge, N., Koropatkin, N. M., Bayer, E. A., and Flint, H. J. (2015) Unique Organization of Extracellular Amylases into Amylosomes in the Resistant Starch-Utilizing Human Colonic Firmicutes Bacterium Ruminococcus bromii. *Mbio* **6**, 11
121. Artzi, L., Bayer, E. A., and Morais, S. (2017) Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. *Nature Reviews Microbiology* **15**, 83-95
122. Bayer, E. A., Belaich, J. P., Shoham, Y., and Lamed, R. (2004) The cellulosomes: Multienzyme machines for degradation of plant cell wall polysaccharides. *Annual Review of Microbiology* **58**, 521-554
123. McNulty, N. P., Wu, M., Erickson, A. R., Pan, C. L., Erickson, B. K., Martens, E. C., Pudlo, N. A., Muegge, B. D., Henrissat, B., Hettich, R. L., and 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**, 1-20
124. Littman, D. R., and Pamer, E. G. (2011) Role of the commensal microbiota in normal and pathogenic host immune responses. *Cell Host and Microbe* **10**, 311-323
125. Fujimura, K. E., Slusher, N. A., Cabana, M. D., and Lynch, S. V. (2010) Role of the gut microbiota in defining human health. *Expert Review of Anti-Infective Therapy* **8**, 435-454
126. Kau, A. L., Ahern, P. P., Griffin, N. W., Goodman, A. L., and Gordon, J. I. (2011) Human nutrition, the gut microbiome and the immune system. *Nature* **474**, 327-336
127. Biedermann, L., and Rogler, G. (2015) The intestinal microbiota: its role in health and disease. *European Journal of Pediatrics* **174**, 151-167
128. Kootte, R. S., Vrieze, A., Holleman, F., Dallinga-Thie, G. M., Zoetendal, E. G., de Vos, W. M., Groen, A. K., Hoekstra, J. B. L., Stroes, E. S., and Nieuwdorp, M. (2012) The therapeutic potential of manipulating gut microbiota in obesity and type 2 diabetes mellitus. *Diabetes, Obesity and Metabolism* **14**, 112-120

129. Subramanian, S., Blanton, L. V., Frese, S. A., Charbonneau, M., Mills, D. A., and Gordon, J. I. (2015) Cultivating healthy growth and nutrition through the gut microbiota. *Cell* **161**, 36-48
130. Blanton, L. V., Barratt, M. J., Charbonneau, M. R., Ahmed, T., and Gordon, J. I. (2016) Childhood undernutrition, the gut microbiota, and microbiota-directed therapeutics. *Science* **352**, 1533-1533
131. Haak, B. W., Levi, M., and Wiersinga, W. J. (2017) Microbiota-targeted therapies on the intensive care unit. *Current Opinions in Critical Care* **23**, 1-8
132. De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J. B., Massart, S., Collini, S., Pieraccini, G., and Lionetti, P. (2010) Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 14691-14696
133. Sonnenburg, E. D., and Sonnenburg, J. L. (2014) Starving our Microbial Self: The Deleterious Consequences of a Diet Deficient in Microbiota-Accessible Carbohydrates. *Cell Metabolism* **20**, 779-786
134. Martens, E. C., Koropatkin, N. M., Smith, T. J., and Gordon, J. I. (2009) Complex Glycan Catabolism by the Human Gut Microbiota: The Bacteroidetes Sus-like Paradigm. *Journal of Biological Chemistry* **284**, 24673-24677
135. Hemsworth, G. R., Thompson, A. J., Stepper, J., Sobala, Ł., Coyle, T., Larsbrink, J., Spadiut, O., Goddard-Borger, E. D., Stubbs, K. A., Brumer, H., and Davies, G. J. (2016) Structural dissection of a complex Bacteroides ovatus gene locus conferring xyloglucan metabolism in the human gut. *Open Biol* **6**
136. Ciorba, M. (2012) A Gastroenterologist's Guide to Probiotics. *Clinical gastroenterology and hepatology* **10**, 960-968
137. Slavin, J. (2013) Fiber and prebiotics: Mechanisms and health benefits. *Nutrients* **5**, 1417-1435
138. Larsbrink, J., Rogers, T. E., Hemsworth, G. R., McKee, L. S., Tauzin, A. S., Spadiut, O., Klinter, S., Pudlo, N. A., Urs, K., Koropatkin, N. M., Creagh, A. L., Haynes, C. A., Kelly, A. G., Cederholm, S. N., Davies, G. J., Martens, E. C., and Brumer, H. (2014) A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature* **506**, 498-502
139. Bågenholm, V., Reddy, S. K., Bouraoui, H., Morrill, J., Kulcinskaja, E., Bahr, C. M., Aurelius, O., Rogers, T., Xiao, Y., Logan, D. T., Martens, E. C., Koropatkin, N. M., and Stålbrand, H. (2017) Galactomannan Catabolism Conferred by a Polysaccharide Utilization Locus of Bacteroides ovatus. **292**, 229-243
140. Ndeh, D., Rogowski, A., Cartmell, A., Luis, A. S., Basle, A., Gray, J., Venditto, I., Briggs, J., Zhang, X. Y., Labourel, A., Terrapon, N., Buffetto, F., Nepogodiev, S., Xiao, Y., Field, R. A., Zhu, Y. P., O'Neill, M. A., Urbanowicz, B. R., York, W. S., Davies, G. J., Abbott, D. W., Ralet, M. C., Martens, E. C., Henrissat, B., and Gilbert, H. J. (2017) Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature* **544**, 65-70
141. El Khoury, D., Cuda, C., Luhovyy, B. L., and Anderson, G. H. (2012) Beta glucan: health benefits in obesity and metabolic syndrome. *J Nutr Metab* **2012**, 851362
142. McNeil, N. I. (1984) The contribution of the large-intestine to energy supplies in man. *American Journal of Clinical Nutrition* **39**, 338-342

143. Cummings, J. H., and Macfarlane, G. T. (1997) Role of intestinal bacteria in nutrient metabolism. *Journal of Parenteral and Enteral Nutrition* **21**, 357-365
144. Othman, R. A., Moghadasian, M. H., and Jones, P. J. H. (2011) Cholesterol-lowering effects of oat β-glucan. *Nutrition Reviews* **69**, 299-309
145. Lazaridou, A., Biliaderis, C. G., Micha-Screttas, M., and Steele, B. R. (2004) A comparative study on structure-function relations of mixed-linkage (1→3), (1→4) linear β-D-glucans. *Food Hydrocolloids* **18**, 837-855
146. Koropatkin, N. M., Martens, E. C., Gordon, J. I., and Smith, T. J. (2008) Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure* **16**, 1105-1115
147. Eschenfeldt, W. H. S. L. M. C. S. J. A. D. M. I. (2009) A Family of LIC Vectors for High-Throughput Cloning and Purificaiton of Proteins. *Methods in Molecular Biology: High Throughput Protein Expression and Purification* **498**, 105-115
148. Sundqvist, G., Stenvall, M., Berglund, H., Ottosson, J., and Brumer, H. (2007) A general, robust method for the quality control of intact proteins using LC-ESI-MS. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **852**, 188-194
149. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005) Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Protocols Handbook*, 571-607
150. Abdel-Akher, M., and Smith, F. (1951) The Reduction of Sugars with Sodium Borohydride. *Journal of the American Chemical Society* **73**, 4691-4692
151. Viladot, J. L., Moreau, V., Planas, A., and Driguez, H. (1997) Transglycosylation activity of Bacillus 1,3-1,4-β-D-glucan 4-glucanohydrolases. Enzymic snthesis of alternate 1,3-1,4-β-D-glucooligosaccharides. *Journal of the Chemical Society, Perkin Transactions 1* **338**, 2383-2387
152. Ibatullin, F. M., Baumann, M. J., Greffe, L., and Brumer, H. (2008) Kinetic analyses of retaining endo-(xylo)glucanases from plant and microbial sources using new chromogenic xylogluco-oligosaccharide aryl glycosides. *Biochemistry* **47**, 7762-7769
153. Kabsch, W. (2010) XDS. *Acta Crystallographica Section D-Biological Crystallography* **66**, 125-132
154. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D-Biological Crystallography* **67**, 235-242
155. Labourel, A., Jam, M., Legentil, L., Sylla, B., Hehemann, J. H., Ferrières, V., Czjzek, M., and Michel, G. (2015) Structural and biochemical characterization of the laminarinase ZgLamCGH16 from Zobellia galactanivorans suggests preferred recognition of branched laminarin. *Acta Crystallographica Section D: Biological Crystallography* **71**, 173-184
156. Stein, N. (2008) CHAINSAW: A program for mutating pdb files used as templates in molecular replacement. *Journal of Applied Crystallography* **41**, 641-643
157. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658-674

158. Cowtan, K. (2010) Recent developments in classical density modification. *Acta Crystallographica Section D: Biological Crystallography* **66**, 470-478
159. Cowtan, K. (2007) Fitting molecular fragments into electron density. *Acta Crystallographica Section D: Biological Crystallography* **64**, 83-89
160. Debreczeni, J. E., and Emsley, P. (2012) Handling ligands with Coot. *Acta Crystallographica Section D-Biological Crystallography* **68**, 425-430
161. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D-Biological Crystallography* **67**, 355-367
162. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D-Biological Crystallography* **60**, 2126-2132
163. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research* **35**, W375-W383
164. Agirre, J., Iglesias-fernández, J., Rovira, C., Davies, G. J., Wilson, K. S., and Cowtan, K. D. (2015) Privateer : software for the conformational validation of carbohydrate structures. *Nature Structural and Molecular Biology* **22**, 833-834
165. Edgar, R. C. (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-1797
166. Larsson, A. (2014) AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276-3278
167. Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., Keduas, V., and Notredame, C. (2006) Expresso: Automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Research* **34**, 604-608
168. Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30**, 2725-2729
169. Kurokawa, K., Itoh, T., Kuwahara, T., Oshima, K., Toh, H., Toyoda, A., Takami, H., Morita, H., Sharma, V. K., Srivastava, T. P., Taylor, T. D., Noguchi, H., Mori, H., Ogura, Y., Ehrlich, D. S., Itoh, K., Takagi, T., Sakaki, Y., Hayashi, T., and Hattori, M. (2007) Comparative Metagenomics Revealed Commonly Enriched Gene Sets in Human Gut Microbiomes. *DNA Research* **14**, 169-181
170. Qin, J. J., Li, R. Q., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J. H., Xu, J. M., Li, S. C., Li, D. F., Cao, J. J., Wang, B., Liang, H. Q., Zheng, H. S., Xie, Y. L., Tap, J., Lepage, P., Bertalan, M., Batto, J. M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H. B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H. M., Yu, C., Li, S. T., Jian, M., Zhou, Y., Li, Y. R., Zhang, X. Q., Li, S. G., Qin, N., Yang, H. M., Wang, J., Brunak, S., Dore, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Bork, P., Ehrlich, S. D., and Meta, H. I. T. C. (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59-U70
171. Huttenhower, C., Fah Sathirapongsasuti, J., Segata, N., Gevers, D., Earl, A. M., Fitzgerald, M. G., Young, S. K., Zeng, Q., Alm, E. J., Alvarado, L., Anderson, S., Arachchi, H. M., Bloom, T., Ciulla, D. M., Erlich, R. L., Feldgarden, M., Fisher, S.,

Friedrich, D. C., Giannoukos, G., Goldberg, J. M., Griggs, A., Gujja, S., Haas, B. J., Hepburn, T. A., Howarth, C., Huang, K. H., Kells, C., Lennon, N., Mehta, T., Nusbaum, C., Pearson, M., Priest, M. E., Russ, C., Shenoy, N., Sykes, S. M., Tabbaa, D. G., Ward, D. V., Yandava, C., Zucker, J. D., Birren, B. W., Knight, R., Clemente, J. C., Lozupone, C. A., McDonald, D., Abubucker, S., Chinwalla, A. T., Fulton, R. S., Hallsworth-Pepin, K., Lobos, E. A., Magrini, V., Martin, J. C., Mitreva, M., Sodergren, E. J., Wollam, A. M., Appelbaum, E., Bhonagiri, V., Chen, L., Clifton, S. W., Delehaunty, K. D., Dooling, D. J., Farmer, C. N., Fronick, C. C., Fulton, L. L., Gao, H., Herter, B., Kota, K. C., Mardis, E. R., Mihindukulasuriya, K. A., Minx, P. J., Oglaughlin, M., Pohl, C., Tomlinson, C. M., Walker, J., Wang, Z., Warren, W., Wylie, K. M., Wylie, T., Ye, L., Zhou, Y., Weinstock, G. M., Wilson, R. K., Badger, J. H., Madupu, R., Bihan, M., Busam, D. A., Scott Durkin, A., Foster, L., Goll, J., Li, K., McCorrison, J. M., Miller, J. R., Rogers, Y.-H. H., Sanka, R. K., Singh, I., Sutton, G. G., Thiagarajan, M., Torralba, M., Methé, B. A., Nelson, K. E., Creasy, H. H., Giglio, M. G., Wortman, J. R., Abolude, O. O., Arze, C. A., Cantarel, B. L., Crabtree, J., Davidovics, N. J., Felix, V. M., Jordan, C., Mahurkar, A. A., Orvis, J., Ravel, J., Schriml, L., White, J. R., White, O., Muzny, D. M., Worley, K. C., Buhay, C. J., Ding, Y., Dugan, S. P., Holder, M. E., Jiang, H., Joshi, V., Kovar, C. L., Lee, S. L., Lewis, L., Liu, Y., Newsham, I., Qin, X., Reid, J. G., Wilczek-Boney, K., Wu, Y., Zhang, L., Zhu, Y., Gibbs, R. A., Highlander, S. K., Petrosino, J. F., Versalovic, J., Aagaard, K. M., Allen-Vercoe, E., Andersen, G. L., Armitage, G., Ayvaz, T., Keitel, W. A., Ross, M. C., Youmans, B. P., Baker, C. C., Begg, L., Belachew, T., Campbell, J. L., Deal, C., Di Francesco, V., Giblin, C., Giovanni, M. Y., Blaser, M. J., Bonazzi, V., Chhibba, S., McEwen, J., Peterson, J., Proctor, L. M., Schloss, J. A., Wang, L., Wellington, C., Wetterstrand, K. A., Paul Brooks, J., Buck, G. A., Rivera, M. C., Sheth, N. U., Canon, S. R., Chain, P. S. G., Lo, C.-C. C., Scholz, M., Kyrpides, N. C., Liolios, K., Markowitz, V. M., Mavromatis, K., Pagani, I., Chen, I. M. A. M. A., Chu, K., Palaniappan, K., Cutting, M. A., Hamilton, H. A., Harris, E. L., Dwayne Lunsford, R., McInnes, P., Davis, C. C., Desantis, T. Z., Dewhirst, F. E., Izard, J., Lemon, K. P., Deych, E., La Rosa, P. S., Shannon, W. D., Michael Dunne, W., Watson, M. A., Edgar, R. C., Farrell, R. M., Sharp, R. R., Faust, K., Raes, J., Fodor, A. A., Forney, L. J., Friedman, J., Smillie, C. S., Garcia, N., Gonzalez, A., Knights, D., Kinder Haake, S., Hoffmann, D. E., Huse, S. M., Jansson, J. K., Katancik, J. A., Kelley, S. T., Rodriguez-Mueller, B., King, N. B., Kong, H. H., Koren, O., Ley, R. E., Koren, S., Liu, B., Pop, M., Sommer, D. D., Lewis, C. M., Spicer, P., Madden, T., Mannon, P. J., McGuire, A. L., Patel, S. M., Podar, M., Vishnivetskaya, T. A., Pollard, K. S., Sharpton, T. J., Truty, R. M., Rho, M., Ye, Y., Rhodes, R., Riehle, K. P., Sankar, P., Schloss, P. D., Schubert, A. M., Schmidt, T. M., Simone, G. A., Sobel, J. D., Treangen, T. J., Yooseph, S., Zoloth, L., Conlan, S., Segre, J. A., and O'Laughlin, M. (2012) Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207-214

172. Junjie Qin, Y. L. Z. C. S. L. J. Z. F. Z. S. L. W. Z. Y. G. D. S. Y. P. D. Z. Z. J. W. W. Y. Q. W. (2012) A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55-60
173. Planas, A. (2000) Bacterial 1,3-1,4-β-glucanases: structure, function and protein engineering. *Biochimica et biophysica acta* **1543**, 361-382
174. Juncker, A. S., and Willenbrock, H. (2003) Prediction of lipoprotein signal peptides in Gram negative bacteria. *Protein science* **12**, 1652-1662

175. Paetzel, M., Karla, A., Strynadka, N. C. J., and Dalbey, R. E. (2002) Signal peptidases. *Chemical Reviews* **102**, 4549-4579
176. Elhenawy, W., Debelyy, M. O., and Feldman, M. F. (2014) Preferential packing of acidic glycosidases and proteases into Bacteroides outer membrane vesicles. *mBio* **5**, 1-12
177. Martin, K., McDougall, B. M., McIlroy, S., Jayus, Chen, J. Z., and Seviour, R. J. (2007) Biochemistry and molecular biology of exocellular fungal beta-(1,3)- and beta-(1,6)-glucanases. *Fems Microbiology Reviews* **31**, 168-192
178. Lowman, D. W., West, L. J., Bearden, D. W., Wempe, M. F., Power, T. D., Ensley, H. E., Haynes, K., Williams, D. L., and Kruppa, M. D. (2011) New insights into the structure of (1-3,1-6)-β-D-glucan side chains in the Candida glabrata cell wall. *PLoS ONE* **6**, 1-10
179. Harada, T., Misaki, A., and Saito, H. (1968) Curdlan: A bacterial gel-forming β-1, 3-glucan. *Archives of Biochemistry and Biophysics* **124**, 292-298
180. Gaiser, O. J., Piotukh, K., Ponnuswamy, M. N., Planas, A., Borriss, R., and Heinemann, U. (2006) Structural basis for the substrate specificity of a Bacillus 1,3-1,4-β-glucanase. *Journal of Molecular Biology* **357**, 1211-1225
181. McGregor, N., Yin, V., Tung, C.-c., Petegem, F. V., and Brumer, H. (2017) Crystallographic insight into the evolutionary origins of xyloglucan endotransglycosylases and endohydrolases. *The Plant Journal* **89**, 651-670
182. Davies, G. J., Wilson, K. S., and Henrissat, B. (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases. *The Biochemical journal* **321**, 557-559
183. McGregor, N., Morar, M., Fenger, T. H., Stogios, P., Lenfant, N., Yin, V., Xu, X., Evdokimova, E., Cui, H., Henrissat, B., Savchenko, A., and Brumer, H. (2016) Structure-function analysis of a mixed-linkage β-glucanase/xyloglucanase from the key ruminal bacteroidetes prevotella bryantii B14. *Journal of Biological Chemistry* **291**, 1175-1197
184. Davies, G. J., and Sinnot, M. L. (2008) Sorting the diverse: The sequence based classifications of carbohydrate active enzymes. *The Biochemist* **30**, 26-32
185. Barbeyron, T., Gerard, A., Potin, P., Henrissat, B., and Kloareg, B. (1998) The Kappa-Carrageenase of the Marine Bacterium Cytophaga drobachiensis. Structural and Phylogenetic Relationships Within Family-16 Glycoside Hydrolases. *Mol. Biol. Evol* **15**, 528-537
186. Michel, G., Chantalat, L., Duee, E., Barbeyron, T., Henrissat, B., Kloareg, B., and Dideberg, O. (2001) The κ-carrageenase of P. carrageenovora features a tunnel-shaped active site: A novel insight in the evolution of clan-B glycoside hydrolases. *Structure* **9**, 513-525
187. Holm, L., and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Research* **38**, W545-W549
188. Hahn, M., Pons, J., Planas, a., Querol, E., and Heinemann, U. (1995) Crystal structure of Bacillus licheniformis 1,3-1,4-beta-D-glucan 4-glucanohydrolase at 1.8 A resolution. *FEBS letters* **374**, 221-224
189. Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**, 785-786
190. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015) The Phyre2 web portal for protein modeling , prediction and analysis. *Nature Protocols* **10**, 845-858

191. Pozzo, T., Pasten, J. L., Karlsson, E. N., and Logan, D. T. (2010) Structural and Functional Analyses of β-Glucosidase 3B from Thermotoga neapolitana: A Thermostable Three-Domain Representative of Glycoside Hydrolase 3. *Journal of Molecular Biology* **397**, 724-739
192. Karkehabadi, S., Helmich, K. E., Kaper, T., Hansson, H., Mikkelsen, N.-e., Gudmundsson, M., Piens, K., Fujdala, M., Banerjee, G., Scott-craig, J. S., Walton, J. D., Phillips, G. N., and Sandgren, M. (2014) Biochemical Characterization and Crystal Structures of a Fungal Family 3 β-Glucosidase , Cel3A from Hypocrea jecorina. **289**, 31624-31637
193. Dotsenko, G. S., Sinitsyna, O. A., Hinz, S. W. A., Wery, J., and Sinitsyn, A. P. (2012) Bioresource Technology Characterization of a GH family 3 β-glycoside hydrolase from Chrysosporium lucknowense and its application to the hydrolysis of β-glucan and xylan. *Bioresource Technology* **112**, 345-349
194. Cuskin, F., Lowe, E. C., Temple, M. J., Zhu, Y. P., Cameron, E. A., Pudlo, N. A., Porter, N. T., Urs, K., Thompson, A. J., Cartmell, A., Rogowski, A., Hamilton, B. S., Chen, R., Tolbert, T. J., Piens, K., Bracke, D., Vervecken, W., Hakki, Z., Speciale, G., Munoz-Munoz, J. L., Day, A., Pena, M. J., McLean, R., Suits, M. D., Boraston, A. B., Atherly, T., Ziemer, C. J., Williams, S. J., Davies, G. J., Abbott, D. W., Martens, E. C., and Gilbert, H. J. (2015) Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature* **517**, 165-169
195. Zheng, X., Li, L., and Wang, Q. (2011) Distribution and molecular characterization of β-glucans from hull-less barley bran, shorts and flour. *International Journal of Molecular Sciences* **12**, 1563-1574
196. Rakoff-Nahoum, S., Foster, K. R., and Comstock, L. E. (2016) The evolution of cooperation within the gut microbiota. *Nature* **533**, 255-259
197. Ilari, A., Fiorillo, A., Angelaccio, S., Florio, R., Chiaraluce, R., Van Der Oost, J., and Consalvi, V. (2009) Crystal structure of a family 16 endoglucanase from the hyperthermophile Pyrococcus furiosus - Structural basis of substrate recognition. *FEBS Journal* **276**, 1048-1058
198. Hehemann, J. H., Correc, G., Barbeyron, T., Helbert, W., Czjzek, M., and Michel, G. (2010) Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464**, 908-912
199. Zhong, Y., Marungruang, N., Fak, F., and Nyman, M. (2015) Effects of two whole-grain barley varieties on caecal SCFA, gut microbiota and plasma inflammatory markers in rats consuming low- and high-fat diets. *British Journal of Nutrition* **113**, 1558-1570
200. Patrascu, O., Béguet-Crespel, F., Marinelli, L., Chatelier, E. L., Abraham, A.-l., Leclerc, M., Klopp, C., Terrapon, N., Henrissat, B., Blottière, H. M., Doré, J., and Béra-maillet, C. (2017) A fibrolytic potential in the human ileum mucosal microbiota revealed by functional metagenomic. *Scientific Reports* **7**, 1-15
201. Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., and Gordon, J. I. (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027-1031
202. Arrieta, M. C., Stiemsma, L. T., Dimitriu, P. A., Thorson, L., Russell, S., Yurist-Doutsch, S., Kuzeljevic, B., Gold, M. J., Britton, H. M., Lefebvre, D. L., Subbarao, P., Mandhane, P., Becker, A., McNagny, K. M., Sears, M. R., Kollmann, T., Mohn, W. W., Turvey, S.

E., Finlay, B. B., and Investigators, C. S. (2015) Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Science Translational Medicine* **7**, 14
203. Claesson, M. J., Jeffery, I. B., Conde, S., Power, S. E., O'Connor, E. M., Cusack, S., Harris, H. M. B., Coakley, M., Lakshminarayanan, B., O'Sullivan, O., Fitzgerald, G. F., Deane, J., O'Connor, M., Harnedy, N., O'Connor, K., O'Mahony, D., van Sinderen, D., Wallace, M., Brennan, L., Stanton, C., Marchesi, J. R., Fitzgerald, A. P., Shanahan, F., Hill, C., Ross, R. P., and O'Toole, P. W. (2012) Gut microbiota composition correlates with diet and health in the elderly. *Nature* **488**, 178-184
204. Gorham, J. B., Kang, S., Williams, B. A., Grant, L. J., McSweeney, C. S., Gidley, M. J., and Mikkelsen, D. (2017) Addition of arabinoxylan and mixed linkage glucans in porcine diets affects the large intestinal bacterial populations. *Eur J Nutr* **56**, 2193-2206
205. Williams, B. A., Grant, L. J., Gidley, M. J., and Mikkelsen, D. (2017) Gut Fermentation of Dietary Fibres: Physico-Chemistry of Plant Cell Walls and Implications for Health. *Int J Mol Sci* **18**
206. Ding, T., and Schloss, P. D. (2014) Dynamics and associations of microbial community types across the human body. *Nature* **509**, 357-360
207. Othman, R. A., Moghadasian, M. H., and Jones, P. J. H. (2011) Cholesterol-lowering effects of oat beta-glucan. *Nutrition Reviews* **69**, 299-309
208. Gunness, P., Michiels, J., Vanhaecke, L., De Smet, S., Kravchuk, O., Van de Meene, A., and Gidley, M. J. (2016) Reduction in circulating bile acid and restricted diffusion across the intestinal epithelium are associated with a decrease in blood cholesterol in the presence of oat β-glucan. *FASEB J* **30**, 4227-4238
209. Fehlbaum, S., Prudence, K., Kieboom, J., Heerikhuisen, M., van den Broek, T., Schuren, F. H. J., Steinert, R. E., and Raederstorff, D. (2018) In Vitro Fermentation of Selected Prebiotics and Their Effects on the Composition and Activity of the Adult Gut Microbiota. *Int J Mol Sci* **19**
210. Nilsson, U., Johansson, M., Nilsson, A., Björck, I., and Nyman, M. (2008) Dietary supplementation with beta-glucan enriched oat bran increases faecal concentration of carboxylic acids in healthy subjects. *Eur J Clin Nutr* **62**, 978-984
211. Tamura, K., Hemsworth, G. R., DeJean, G., Rogers, T. E., Pudlo, N. A., Urs, K., Jain, N., Davies, G. J., Martens, E. C., and Brumer, H. (2017) Molecular Mechanism by which Prominent Human Gut Bacteroidetes Utilize Mixed-Linkage Beta-Glucans, Major Health-Promoting Cereal Polysaccharides. *Cell Reports* **21**, 417-430
212. McGregor, N., Morar, M., Fenger, T. H., Stogios, P., Lenfant, N., Yin, V., Xu, X. H., Evdokimova, E., Cui, H., Henrissat, B., Savchenko, A., and Brumer, H. (2016) Structure-Function Analysis of a Mixed-linkage beta-Glucanase/Xyloglucanase from the Key Ruminal Bacteroidetes Prevotella bryantii B(1)4. *Journal of Biological Chemistry* **291**, 1175-1197
213. Nielsen, H. (2017) Predicting Secretory Proteins with SignalP. *Methods Mol Biol* **1611**, 59-73
214. Doublié, S. (2007) Production of selenomethionyl proteins in prokaryotic and eukaryotic expression systems. *Methods Mol Biol* **363**, 91-108
215. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. in *Methods in Enzymology, Macromolecular Crystallography, Pt A*. pp 307-326

216. Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H., and Hung, L. W. (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallographica Section D-Biological Crystallography* **65**, 582-601
217. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D-Biological Crystallography* **66**, 213-221
218. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D-Structural Biology* **68**, 352-367
219. Winter, G., Waterman, D. G., Parkhurst, J. M., Brewster, A. S., Gildea, R. J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., Michels-Clark, T., Young, I. D., Sauter, N. K., and Evans, G. (2018) DIALS: implementation and evaluation of a new integration package. *Acta Crystallographica Section D-Structural Biology* **74**, 85-97
220. Winter, G. (2010) xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* **43**, 186-190
221. Potterton, L., Agirre, J., Ballard, C., Cowtan, K., Dodson, E., Evans, P. R., Jenkins, H. T., Keegan, R., Krissinel, E., Stevenson, K., Lebedev, A., McNicholas, S. J., Nicholls, R. A., Noble, M., Pannu, N. S., Roth, C., Sheldrick, G., Skubak, P., Turkenburg, J., Uski, V., von Delft, F., Waterman, D., Wilson, K., Winn, M., and Wojdyr, M. (2018) CCP4i2: the new graphical user interface to the CCP4 program suite. *Acta Crystallographica Section D-Structural Biology* **74**, 68-84
222. Vagin, A., and Teplyakov, A. (2010) Molecular replacement with MOLREP. *Acta Crystallographica Section D-Biological Crystallography* **66**, 22-25
223. Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S. M., and Bricogne, G. (2004) Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallographica Section D-Structural Biology* **60**, 2210-2221
224. Koropatkin, N., Martens, E. C., Gordon, J. I., and Smith, T. J. (2009) Structure of a SusD Homologue , BT1043 , Involved in Mucin O -Glycan Utilization in a Prominent Human Gut Symbiont Structure of a SusD Homologue , BT1043 , Involved in Mucin O -Glycan Utilization in a Prominent Human Gut Symbiont †,‡. 1532-1542
225. Phansopa, C., Roy, S., Rafferty, J. B., Douglas, C. W. I., Pandhal, J., Wright, P. C., Kelly, D. J., and Stafford, G. P. (2014) Structural and functional characterization of NanU, a novel high-affinity sialic acid-inducible binding protein of oral and gut-dwelling Bacteroidetes species. *Biochemical Journal* **458**, 499-511
226. Mystkowska, A. A., Robb, C., Vidal-Melgosa, S., Vanni, C., Fernandez-Guerra, A., Hohne, M., and Hehemann, J. H. (2018) Molecular recognition of the beta-glucans laminarin and pustulan by a SusD-like glycan-binding protein of a marine Bacteroidetes. *Febs Journal* **285**, 4465-4481
227. Larsbrink, J., Zhu, Y., Kharade, S. S., Kwiatkowski, K. J., Eijsink, V. G. H., Koropatkin, N. M., McBride, M. J., and Pope, P. B. (2016) A polysaccharide utilization locus from

Flavobacterium johnsoniae enables conversion of recalcitrant chitin. *Biotechnology for Biofuels* **9**, 16
228. Hudson, K. L., Bartlett, G. J., Diehl, R. C., Agirre, J., Gallagher, T., Kiessling, L. L., and Woolfson, D. N. (2015) Carbohydrate-Aromatic Interactions in Proteins. *Journal of the American Chemical Society* **137**, 15152-15160
229. Asensio, J. L., Ardá, A., Cañada, F. J., and Jiménez-Barbero, J. (2013) Carbohydrate-aromatic interactions. *Acc Chem Res* **46**, 946-954
230. Jeffrey, G. A. (1997) *An introduction to hydrogen bonding*, Oxford University Press, New York
231. Holm, L., and Laakso, L. M. (2016) Dali server update. *Nucleic Acids Research* **44**, W351-W355
232. Barsanti, L., Passarelli, V., Evangelista, V., Frassanito, A. M., and Gualtieri, P. (2011) Chemistry, physico-chemistry and applications linked to biological activities of beta-glucans. *Natural Product Reports* **28**, 457-466
233. Wood, P. J., Weisz, J., and Blackwell, B. A. (1994) Structural studies of (1-3), (1-4)-beta-D-glucans by C(13)-nuclear magnetic-resonance spectroscopy and by rapid analysis of cellulose-like regions using high-performance anion-exchange chromatography of oligosaccharides released by lichenase. *Cereal Chemistry* **71**, 301-307
234. Rakoff-Nahoum, S., Coyne, M. J., and Comstock, L. E. (2014) An Ecological Network of Polysaccharide Utilization among Human Intestinal Symbionts. *Current Biology* **24**, 40-49
235. Lazaridou, A., Biliaderis, C. G., Micha-Screttas, M., and Steele, B. R. (2004) A comparative study on structure-function relations of mixed-linkage (1 -> 3), (1 -> 4) linear beta-D-glucans. *Food Hydrocolloids* **18**, 837-855
236. Mackenzie, A. K., Pope, P. B., Pedersen, H. L., Gupta, R., Morrison, M., Willats, W. G., and Eijsink, V. G. (2012) Two SusD-like proteins encoded within a polysaccharide utilization locus of an uncultured ruminant Bacteroidetes phylotype bind strongly to cellulose. *Appl Environ Microbiol* **78**, 5935-5937
237. Mackenzie, A. K., Naas, A. E., Kracun, S. K., Sch??ckel, J., Fangel, J. U., Agger, J. W., Willats, W. G. T., Eijsink, V. G. H., and Pope, P. B. (2015) A polysaccharide utilization locus from an uncultured Bacteroidetes phylotype suggests ecological adaptation and substrate versatility. *Applied and Environmental Microbiology* **81**, 187-195
238. Naas, A. E., Mackenzie, A. K., Mravec, J., Schückel, J., Willats, W. G., Eijsink, V. G., and Pope, P. B. (2014) Do rumen Bacteroidetes utilize an alternative mechanism for cellulose degradation? *MBio* **5**, e01401-01414
239. Kiemle, S. N., Zhang, X., Esker, A. R., Toriz, G., Gatenholm, P., and Cosgrove, D. J. (2014) Role of (1,3)(1,4)-beta-Glucan in Cell Walls: Interaction with Cellulose. *Biomacromolecules* **15**, 1727-1736
240. Cann, I., Bernardi, R. C., and Mackie, R. I. (2016) Cellulose degradation in the human gut: Ruminococcus champanellensis expands the cellulosome paradigm. *Environmental Microbiology* **18**, 307-310
241. Haskey, N., and Gibson, D. L. (2017) An Examination of Diet for the Maintenance of Remission in Inflammatory Bowel Disease. *Nutrients* **9**
242. Armstrong, Z., Mewis, K., Liu, F., Morgan-Lang, C., Scofield, M., Durno, E., Chen, H. M., Mehr, K., Withers, S. G., and Hallam, S. J. (2018) Metagenomics reveals functional

synergy and novel polysaccharide utilization loci in the Castor canadensis fecal microbiome. *Isme Journal* **12**, 2757-2769
243. Shepherd, E. S., DeLoache, W. C., Pruss, K. M., Whitaker, W. R., and Sonnenburg, J. L. (2018) An exclusive metabolic niche enables strain engraftment in the gut microbiota. *Nature* **557**, 434-438
244. Joglekar, P., Sonnenburg, E. D., Higginbottom, S. K., Earle, K. A., Morland, C., Shapiro-Ward, S., Bolam, D. N., and Sonnenburg, J. L. (2018) Genetic Variation of the SusC/SusD Homologs from a Polysaccharide Utilization Locus Underlies Divergent Fructan Specificities and Functional Adaptation in Strains. *mSphere* **3**
245. Farrar, M. D., Whitehead, T. R., Lan, J., Dilger, P., Thorpe, R., Holland, K. T., and Carding, S. R. (2005) Engineering of the gut commensal bacterium Bacteroides ovatus to produce and secrete biologically active murine interleukin-2 in response to xylan. *J Appl Microbiol* **98**, 1191-1197
246. Backhed, F. (2005) Host-Bacterial Mutualism in the Human Intestine. *Science* **307**, 1915-1920
247. Clemente, J. C., Ursell, L. K., Parfrey, L. W., and Knight, R. (2012) The impact of the gut microbiota on human health: an integrative view. *Cell* **148**, 1258-1270
248. Frank, D. N., St Amand, A. L., Feldman, R. A., Boedeker, E. C., Harpaz, N., and Pace, N. R. (2007) Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci U S A* **104**, 13780-13785
249. Hooper, L. V., Littman, D. R., and Macpherson, A. J. (2012) Interactions between the microbiota and the immune system. *Science* **336**, 1268-1273
250. Sonnenburg, J. L., and Backhed, F. (2016) Diet-microbiota interactions as moderators of human metabolism. *Nature* **535**, 56-64
251. Costello, E. K., Stagaman, K., Dethlefsen, L., Bohannan, B. J., and Relman, D. A. (2012) The application of ecological theory toward an understanding of the human microbiome. *Science* **336**, 1255-1262
252. Heintz-Buschart, A., and Wilmes, P. (2018) Human Gut Microbiome: Function Matters. *Trends Microbiol* **26**, 563-574
253. Porter, N. T., and Martens, E. C. (2017) The Critical Roles of Polysaccharides in Gut Microbial Ecology and Physiology. *Annu Rev Microbiol* **71**, 349-369
254. Lapebie, P., Lombard, V., Drula, E., Terrapon, N., and Henrissat, B. (2019) Bacteroidetes use thousands of enzyme combinations to break down glycans. *Nature Communications* **10**, 7
255. Larsbrink, J., Thompson, A. J., Lundqvist, M., Gardner, J. G., Davies, G. J., and Brumer, H. (2014) A complex gene locus enables xyloglucan utilization in the model saprophyte Cellvibrio japonicus. *Molecular Microbiology* **94**, 418-433
256. Temple, M. J., Cuskin, F., Basle, A., Hickey, N., Speciale, G., Williams, S. J., Gilbert, H. J., and Lowe, E. C. (2017) A Bacteroidetes locus dedicated to fungal 1,6-beta-glucan degradation: Unique substrate conformation drives specificity of the key endo-1,6-beta-glucanase. *J Biol Chem* **292**, 10639-10650
257. Cartmell, A., Munoz-Munoz, J., Briggs, J. A., Ndeh, D. A., Lowe, E. C., Basle, 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., and Gilbert, H. (2018) A

surface endogalactanase in Bacteroides thetaiotaomicron confers keystone status for arabinogalactan degradation. *Nature Microbiology* **3**, 1314-1326
258. Luis, A. S., Briggs, J., Zhang, X. Y., Farnell, B., Ndeh, D., Labourel, A., Basle, A., Cartmell, A., Terrapon, N., Stott, K., Lowe, E. C., McLean, R., Shearer, K., Schuckel, J., Venditto, I., Ralet, M. C., Henrissat, B., Martens, E. C., Mosimann, S. C., Abbott, D. W., and Gilbert, H. J. (2018) Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic Bacteroides. *Nature Microbiology* **3**, 210-219
259. Briliute, J., Urbanowicz, P. A., Luis, A. S., Basle, A., Paterson, N., Rebello, O., Hendel, J., Ndeh, D. A., Lowe, E. C., Martens, E. C., Spencer, D. I. R., Bolam, D. N., and Crouch, L. I. (2019) Complex N-glycan breakdown by gut Bacteroides involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci. *Nature Microbiology* **4**, 1571-1581
260. Stone, B. A., and Clarke, A. E. (1992) *Chemistry and biology of (1-3)-beta-glucans*, LaTrobe University Press, Australia
261. O'Sullivan, L., Murphy, B., McLoughlin, P., Duggan, P., Lawlor, P. G., Hughes, H., and Gardiner, G. E. (2010) Prebiotics from Marine Macroalgae for Human and Animal Health Applications. *Marine Drugs* **8**, 2038-2064
262. Salyers, A. A., Palmer, J. K., and Wilkins, T. D. (1977) Laminarinase (beta-glucanase) activity in Bacteroides from the human colon. *Appl Environ Microbiol* **33**, 1118-1124
263. Zhao, J., and Cheung, P. C. (2011) Fermentation of beta-glucans derived from different sources by bifidobacteria: evaluation of their bifidogenic effect. *J Agric Food Chem* **59**, 5986-5992
264. Cloetens, L., Ulmius, M., Johansson-Persson, A., Akesson, B., and Onning, G. (2012) Role of dietary beta-glucans in the prevention of the metabolic syndrome. *Nutr Rev* **70**, 444-458
265. Driscoll, M., Hansen, R., Ding, C., Cramer, D. E., and Yan, J. (2009) Therapeutic potential of various beta-glucan sources in conjunction with anti-tumor monoclonal antibody in cancer therapy. *Cancer Biol Ther* **8**, 218-225
266. Stier, H., Ebbeskotte, V., and Gruenwald, J. (2014) Immune-modulatory effects of dietary Yeast Beta-1,3/1,6-D-glucan. *Nutr J* **13**, 38
267. Lazaridou, A., and Biliaderis, C. G. (2007) Molecular aspects of cereal beta-glucan functionality: Physical properties, technological applications and physiological effects. *Journal of Cereal Science* **46**, 101-118
268. Manners, D. J., Masson, A. J., and Patterson, J. C. (1973) Structure of a beta-(1-3)-D-glucan from yeast-cell walls. *Biochemical Journal* **135**, 19-30
269. Read, S. M., Currie, G., and Bacic, A. (1996) Analysis of the structural heterogeneity of laminarin by electrospray-ionisation-mass spectrometry. *Carbohydr Res* **281**, 187-201
270. Menshova, R. V., Ermakova, S. P., Anastyuk, S. D., Isakov, V. V., Dubrovskaya, Y. V., Kusaykin, M. I., Um, B. H., and Zvyagintseva, T. N. (2014) Structure, enzymatic transformation and anticancer activity of branched high molecular weight laminaran from brown alga Eisenia bicyclis. *Carbohydr Polym* **99**, 101-109
271. Abdek-Akher, M., Hamilton J.K., and Smith F. (1951) The Reduction of Sugars with Sodium Borohydride. *Journal of The American Chemical Society* **73**, 4691-4692
272. Viladot, J. L., Moreau, V., Planas, A., and Driguez, H. (1997) Transglycosylation activity of Bacillus 1,3-1,4-beta-D-glucan 4-glucanohydrolases. Enzymic synthesis of alternate

1,3,-1,4-beta-D-glucooligosaccharides. *Journal of the Chemical Society-Perkin Transactions 1*, 2383-2387
273. Chen, I. A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., Huntemann, M., Varghese, N., White, J. R., Seshadri, R., Smirnova, T., Kirton, E., Jungbluth, S. P., Woyke, T., Eloe-Fadrosh, E. A., Ivanova, N. N., and Kyrpides, N. C. (2019) IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res* **47**, D666-d677
274. Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S. C., Ester, M., Foster, L. J., and Brinkman, F. S. L. (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26**, 1608-1615
275. Eschenfeldt, W. H., Makowska-Grzyska, M., Stols, L., Donnelly, M. I., Jedrzejczak, R., and Joachimiak, A. (2013) New LIC vectors for production of proteins from genes containing rare codons. *J Struct Funct Genomics* **14**, 135-144
276. Hutchison, C. A., Phillips, S., Edgell, M. H., Gillam, S., Jahnke, P., and Smith, M. (1978) MUTAGENESIS AT A SPECIFIC POSITION IN A DNA-SEQUENCE. *Journal of Biological Chemistry* **253**, 6551-6560
277. Arnal, G., Attia, M. A., Asohan, J., and Brumer, H. (2017) A Low-Volume, Parallel Copper-Bicinchoninic Acid (BCA) Assay for Glycoside Hydrolases. *Methods Mol Biol* **1588**, 3-14
278. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallographica Section D-Structural Biology* **62**, 72-82
279. Evans, P. R., and Murshudov, G. N. (2013) How good are my data and what is the resolution? *Acta Crystallographica Section D-Biological Crystallography* **69**, 1204-1214
280. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallographica Section D-Biological Crystallography* **66**, 486-501
281. Yu, J., Feng, Q., Wong, S. H., Zhang, D., Liang, Q. Y., Qin, Y. W., Tang, L. Q., Zhao, H., Stenvang, J., Li, Y. L., Wang, X. K., Xu, X. Q., Chen, N., Wu, W. K. K., Al-Aama, J., Nielsen, H. J., Kiilerich, P., Jensen, B. A. H., Yau, T. O., Lan, Z., Jia, H. J., Li, J. H., Xiao, L., Lam, T. Y. T., Ng, S. C., Cheng, A. S. L., Wong, V. W. S., Chan, F. K. L., Xu, X., Yang, H., Madsen, L., Datz, C., Tilg, H., Wang, J., Brunner, N., Kristiansen, K., Arumugam, M., and Sung, J. J. Y. (2017) Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut* **66**, 70-+
282. Liu, R. X., Hong, J., Xu, X. Q., Feng, Q., Zhang, D. Y., Gu, Y. Y., Shi, J., Zhao, S. Q., Liu, W., Wang, X. K., Xia, H. H., Liu, Z. P., Cui, B., Liang, P. W., Xi, L. Q., Jin, J. B., Ying, X. Y., Wang, X. L., Zhao, X. J., Li, W. Y., Jia, H. J., Lan, Z., Li, F. Y., Wang, R., Sun, Y. K., Yang, M. L., Shen, Y. X., Jie, Z. Y., Li, J. H., Chen, X. M., Zhong, H. Z., Xie, H. L., Zhang, Y. F., Gu, W. Q., Deng, X. X., Shen, B. Y., Xu, X., Yang, H. M., Xu, G. W., Bi, Y. F., Lai, S. H., Wang, J., Qi, L., Madsen, L., Wang, J. Q., Ning, G., Kristiansen, K., and Wang, W. Q. (2017) Gut microbiome and serum metabolome alterations in obesity and after weight-loss intervention. *Nature Medicine* **23**, 859-+
283. Gu, Y. Y., Wang, X. K., Li, J. H., Zhang, Y. F., Zhong, H. Z., Liu, R. X., Zhang, D. Y., Feng, Q., Xie, X. Y., Hong, J., Ren, H. H., Liu, W., Ma, J., Su, Q., Zhang, H. M., Yang, J. L., Wang, X. L., Zhao, X. J., Gu, W. Q., Bi, Y. F., Peng, Y. D., Xu, X. Q., Xia, H. H., Li, F., Xu, X., Yang, H. M., Xu, G. W., Madsen, L., Kristiansen, K., Ning, G., and Wang,

W. Q. (2017) Analyses of gut microbiota and plasma bile acids enable stratification of patients for antidiabetic treatment. *Nature Communications* **8**, 12

284. He, Q., Gao, Y., Jie, Z. Y., Yu, X. L., Laursen, J. M., Xiao, L., Li, Y., Li, L. L., Zhang, F. M., Feng, Q., Li, X. P., Yu, J. H., Liu, C., Lan, P., Yan, T., Liu, X., Xu, X., Yang, H. M., Wang, J., Madsen, L., Brix, S., Wang, J. P., Kristiansen, K., and Jia, H. J. (2017) Two distinct metacommunities characterize the gut microbiota in Crohn's disease patients. *Gigascience* **6**, 43
285. Zhang, X., Zhang, D. Y., Jia, H. J., Feng, Q., Wang, D. H., Liang, D., Wu, X. N., Li, J. H., Tang, L. Q., Li, Y., Lan, Z., Chen, B., Li, Y. L., Zhong, H. Z., Xie, H. L., Jie, Z. Y., Chen, W. N., Tang, S. M., Xu, X. Q., Wang, X. K., Cai, X. H., Liu, S., Xia, Y., Li, J. Y., Qiao, X. Y., Al-Aama, J. Y., Chen, H., Wang, L., Wu, Q. J., Zhang, F. C., Zheng, W. J., Li, Y. Z., Zhang, M. R., Luo, G. W., Xue, W. B., Xiao, L., Li, J., Chen, W. T., Xu, X., Yin, Y., Yang, H. M., Wang, J., Kristiansen, K., Liu, L., Li, T., Huang, Q. C., and Li, Y. R. (2015) The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nature Medicine* **21**, 895-905
286. Nishijima, S., Suda, W., Oshima, K., Kim, S. W., Hirose, Y., Morita, H., and Hattori, M. (2016) The gut microbiome of healthy Japanese and its microbial and functional uniqueness. *DNA Research* **23**, 125-133
287. Lloyd-Price, J., Mahurkar, A., Ahnavard, G. R., Rabtree, J. C., Rvis, J. O., Hall, A. B. R., Rady, A. B., Reasy, H. H. C., McCracken, C., Giglio, M. G., McDonald, D., Franzosa, E. A., Knight, R., White, O., and Huttenhower, C. (2017) Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* **550**, 61-+
288. Le Chatelier, E., Nielsen, T., Qin, J. J., Prifti, E., Hildebrand, F., Falony, G., Almeida, M., Arumugam, M., Batto, J. M., Kennedy, S., Leonard, P., Li, J. H., Burgdorf, K., Grarup, N., Jorgensen, T., Brandslund, I., Nielsen, H. B., Juncker, A. S., Bertalan, M., Levenez, F., Pons, N., Rasmussen, S., Sunagawa, S., Tap, J., Tims, S., Zoetendal, E. G., Brunak, S., Clement, K., Dore, J., Kleerebezem, M., Kristiansen, K., Renault, P., Sicheritz-Ponten, T., de Vos, W. M., Zucker, J. D., Raes, J., Hansen, T., Bork, P., Wang, J., Ehrlich, S. D., Pedersen, O., and Meta, H. I. T. C. (2013) Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541-+
289. Smits, S. A., Leach, J., Sonnenburg, E. D., Gonzalez, C. G., Lichtman, J. S., Reid, G., Knight, R., Manjurano, A., Changalucha, J., Elias, J. E., Dominguez-Bello, M. G., and Sonnenburg, J. L. (2017) Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. *Science* **357**, 802-+
290. Conteville, L. C., Oliveira-Ferreira, J., and Vicente, A. C. P. (2019) Gut Microbiome Biomarkers and Functional Diversity Within an Amazonian Semi-Nomadic Hunter Gatherer Group. *Frontiers in Microbiology* **10**, 10
291. Boratyn, G. M., Thierry-Mieg, J., Thierry-Mieg, D., Busby, B., and Madden, T. L. (2019) Magic-BLAST, an accurate RNA-seq aligner for long and short reads. *Bmc Bioinformatics* **20**, 19
292. Quinlan, A. R., and Hall, I. M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842
293. CAZypedia Consortium. (2018) Ten years of CAZypedia: a living encyclopedia of carbohydrate-active enzymes. *Glycobiology* **28**, 3-8
294. Helbert, W., Poulet, L., Drouillard, S., Mathieu, S., Loiodice, M., Couturier, M., Lombard, V., Terrapon, N., Turchetto, J., Vincentelli, R., and 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**, 6063-6068
295. Evans, D. F., Pye, G., Bramley, R., Clark, A. G., Dyson, T. J., and Hardcastle, J. D. (1988) Measurement of gastrointestinal pH profiles in normal ambulant human subjects. *Gut* **29**, 1035-1041
296. Labourel, A., Jam, M., Legentil, L., Sylla, B., Hehemann, J. H., Ferrieres, V., Czjzek, M., and Michel, G. (2015) Structural and biochemical characterization of the laminarinase ZgLamC(GH16) from Zobellia galactanivorans suggests preferred recognition of branched laminarin. *Acta Crystallographica Section D-Structural Biology* **71**, 173-184
297. Wojtkowiak, A., Witek, K., Hennig, J., and Jaskolski, M. (2013) Structures of an active-site mutant of a plant 1,3-beta-glucanase in complex with oligosaccharide products of hydrolysis. *Acta Crystallogr D Biol Crystallogr* **69**, 52-62
298. Grondin, J. M., Tamura, K., Dejean, G., Abbott, D. W., and Brumer, H. (2017) Polysaccharide Utilization Loci: Fuelling microbial communities. *J Bacteriol*
299. Becker, S., Scheffel, A., Polz, M. F., and Hehemann, J. H. (2017) Accurate Quantification of Laminarin in Marine Organic Matter with Enzymes from Marine Microbes. *Applied and Environmental Microbiology* **83**, 14
300. Kabisch, A., Otto, A., Konig, S., Becher, D., Albrecht, D., Schuler, M., Teeling, H., Amann, R. I., and Schweder, T. (2014) Functional characterization of polysaccharide utilization loci in the marine Bacteroidetes 'Gramella forsetii' KT0803. *Isme Journal* **8**, 1492-1502
301. Unfried, F., Becker, S., Robb, C. S., Hehemann, J. H., Markert, S., Heiden, S. E., Hinzke, T., Becher, D., Reintjes, G., Kruger, K., Avci, B., Kappelmann, L., Hahnke, R. L., Fischer, T., Harder, J., Teeling, H., Fuchs, B., Barbeyron, T., Amann, R. I., and Schweder, T. (2018) Adaptive mechanisms that provide competitive advantages to marine bacteroidetes during microalgal blooms. *Isme Journal* **12**, 2894-2906
302. Panschin, I., Huang, S., Meier-Kolthoff, J. P., Tindall, B. J., Rohde, M., Verbarg, S., Lapidus, A., Han, J., Trong, S., Haynes, M., Reddy, T. B. K., Huntemann, M., Pati, A., Ivanova, N. N., Mavromatis, K., Markowitz, V., Woyke, T., Goker, M., Klenk, H. P., Kyrpides, N. C., and Hahnke, R. L. (2016) Comparing polysaccharide decomposition between the type strains Gramella echinicola KMM 6050(T) (DSM 19838(T)) and Gramella portivictoriae UST040801-001(T) (DSM 23547(T)), and emended description of Gramella echinicola Nedashkovskaya et al. 2005 emend. Shahina et al. 2014 and Gramella portivictoriae Lau et al. 2005. *Standards in Genomic Sciences* **11**, 16
303. Badur, A. H., Ammar, E. M., Yalamanchili, G., Hehemann, J. H., and Rao, C. V. (2020) Characterization of the GH16 and GH17 laminarinases from Vibrio breoganii 1C10. *Applied Microbiology and Biotechnology* **104**, 161-171
304. Zhao, L. P., Zhang, F., Ding, X. Y., Wu, G. J., Lam, Y. Y., Wang, X. J., Fu, H. Q., Xue, X. H., Lu, C. H., Ma, J. L., Yu, L. H., Xu, C. M., Ren, Z. Y., Xu, Y., Xu, S. M., Shen, H. L., Zhu, X. L., Shi, Y., Shen, Q. Y., Dong, W. P., Liu, R., Ling, Y. X., Zeng, Y., Wang, X. P., Zhang, Q. P., Wang, J., Wang, L. H., Wu, Y. Q., Zeng, B. H., Wei, H., Zhang, M. H., Peng, Y. D., and Zhang, C. H. (2018) Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes. *Science* **359**, 1151-1156
305. Ficko-Blean, E., Prechoux, A., Thomas, F., Rochat, T., Larocque, R., Zhu, Y. T., Stam, M., Genicot, S., Jam, M., Calteau, A., Viart, B., Ropartz, D., Perez-Pascual, D., Correc,

G., Matard-Mann, M., Stubbs, K. A., Rogniaux, H., Jeudy, A., Barbeyron, T., Medigue, C., Czjzek, M., Vallenet, D., McBride, M. J., Duchaud, E., and Michel, G. (2017) Carrageenan catabolism is encoded by a complex regulon in marine heterotrophic bacteria. *Nature Communications* **8**, 17

306. Pluvinage, B., Grondin, J. M., Amundsen, C., Klassen, L., Moote, P. E., Xiao, Y., Thomas, D., Pudlo, N. A., Anele, A., Martens, E. C., Inglis, D., Uwiera, R. E. R., Boraston, A. B., and Abbott, D. W. (2018) Molecular basis of an agarose metabolic pathway acquired by a human intestinal symbiont. *Nature Communications* **9**, 14
307. Hettle, A. G., Hobbs, J. K., Pluvinage, B., Vickers, C., Abe, K. T., Salama-Alber, O., McGuire, B. E., Hehemann, J. H., Hui, J. P. M., Berrue, F., Banskota, A., Zhang, J., Bottos, E. M., Van Hamme, J., and Boraston, A. B. (2019) Insights into the κ/ι-carrageenan metabolism pathway of some marine *Pseudoalteromanas* species. *Commun Biol* **2**, 474
308. Déjean, G., Tamura, K., Cabrera, A., Jain, N., Pudlo, N. A., Pereira, G., Viborg, A. H., Van Petegem, F., Martens, E. C., and Brumer, H. (2020) Synergy between Cell Surface Glycosidases and Glycan-Binding Proteins Dictates the Utilization of Specific Beta(1,3)-Glucans by Human Gut. *mBio* **11**
309. Temple, M. J., Cuskin, F., Basle, A., Hickey, N., Speciale, G., Williams, S. J., Gilbert, H. J., and Lowe, E. C. (2017) A Bacteroidetes locus dedicated to fungal 1,6--glucan degradation: Unique substrate conformation drives specificity of the key endo-1,6--glucanase. *Journal of Biological Chemistry* **292**, 10639-10650
310. Tropea, J. E., Cherry, S., and Waugh, D. S. (2009) Expression and purification of soluble His(6)-tagged TEV protease. *Methods Mol Biol* **498**, 297-307
311. Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T., and Bricogne, G. (2011) Data processing and analysis with the autoPROC toolbox. *Acta Crystallographica Section D-Structural Biology* **67**, 293-302
312. Terwilliger, T. (2004) SOLVE and RESOLVE: automated structure solution, density modification, and model building. *Journal of Synchrotron Radiation* **11**, 49-52
313. Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J., and Adams, P. D. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallographica Section D-Biological Crystallography* **64**, 61-69
314. Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007) Automated structure solution with autoSHARP. *Methods Mol Biol* **364**, 215-230
315. Abrahams, J. P., and Leslie, A. G. W. (1996) Methods used in the structure determination of bovine mitochondrial F-1 ATPase. *Acta Crystallographica Section D-Biological Crystallography* **52**, 30-42
316. Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nature Protocols* **3**, 1171-1179
317. Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D-Biological Crystallography* **62**, 1002-1011
318. Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall, W. B., Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S., and Richardson, D. C.

(2018) MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* **27**, 293-315
319. Cockburn, D., Wilkens, C., and Svensson, B. (2017) Affinity Electrophoresis for Analysis of Catalytic Module-Carbohydrate Interactions. *Methods Mol Biol* **1588**, 119-127
320. Bacic, A., Fincher, G. B., and Stone, B. A. (2009) *Chemistry, Biochemistry, and Biology of 1-3 Beta Glucans and Related Polysaccharides*, Academic Press
321. Holm, L. (2020) DALI and the persistence of protein shape. *Protein Science* **29**, 128-140
322. Boraston, A. B., Warren, R. A., and Kilburn, D. G. (2001) beta-1,3-Glucan binding by a thermostable carbohydrate-binding module from Thermotoga maritima. *Biochemistry* **40**, 14679-14685
323. Hehemann, J. H., Kelly, A. G., Pudlo, N. A., Martens, E. C., and Boraston, A. B. (2012) Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 19786-19791
324. Boraston, A. B., Nurizzo, D., Notenboom, V., Ducros, V., Rose, D. R., Kilburn, D. G., and Davies, G. J. (2002) Differential oligosaccharide recognition by evolutionarily-related beta-1,4 and beta-1,3 glucan-binding modules. *J Mol Biol* **319**, 1143-1156
325. Henshaw, J. L., Bolam, D. N., Pires, V. M. R., Czjzek, M., Henrissat, B., Ferreira, L. M. A., Fontes, C., and Gilbert, H. J. (2004) The family 6 carbohydrate binding module CmCBM6-2 contains two ligand-binding sites with distinct specificities. *Journal of Biological Chemistry* **279**, 21552-21559
326. Pires, V. M. R., Henshaw, J. L., Prates, J. A. M., Bolam, D. N., Ferreira, L. M. A., Fontes, C., Henrissat, B., Planas, A., Gilbert, H. J., and Czjzek, M. (2004) The crystal structure of the family 6 carbohydrate binding module from Cellvibrio mixtus endoglucanase 5A in complex with oligosaccharides reveals two distinct binding sites with different ligand Specificities. *Journal of Biological Chemistry* **279**, 21560-21568
327. Ribeiro, D. O., Viegas, A., Pires, V. M. R., Medeiros-Silva, J., Bule, P., Chai, W. G., Marcelo, F., Fontes, C., Cabrita, E. J., Palma, A. S., and Carvalho, A. L. (2019) Molecular basis for the preferential recognition of beta 1,3-1,4-glucans by the family 11 carbohydrate-binding module from Clostridium thermocellum. *Febs Journal*, 21
328. Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochemical Journal* **382**, 769-781
329. Legentil, L., Paris, F., Ballet, C., Trouvelot, S., Daire, X., Vetvicka, V., and Ferrières, V. (2015) Molecular Interactions of β-(1→3)-Glucans with Their Receptors. *Molecules* **20**, 9745-9766
330. van Bueren, A. L., Morland, C., Gilbert, H. J., and Boraston, A. B. (2005) Family 6 carbohydrate binding modules recognize the non-reducing end of beta-1,3-linked glucans by presenting a unique ligand binding surface. *Journal of Biological Chemistry* **280**, 530-537
331. Boraston, A. B., Notenboom, V., Warren, R. A. J., Kilburn, D. G., Rose, D. R., and Davies, G. (2003) Structure and ligand binding of carbohydrate-binding module CsCBM6-3 reveals similarities with fucose-specific lectins and "galactose-binding" domains. *Journal of Molecular Biology* **327**, 659-669

332. Sakon, J., Irwin, D., Wilson, D. B., and Karplus, P. A. (1997) Structure and mechanism of endo/exocellulase E4 from Thermomonospora fusca. *Nature Structural Biology* **4**, 810-818
333. Yaoi, K., Kondo, H., Noro, N., Suzuki, M., Tsuda, S., and Mitsuishi, Y. (2004) Tandem repeat of a seven-bladed beta-propeller domain in oligoxyloglucan reducing-end-specific cellobiohydrolase. *Structure* **12**, 1209-1217
334. Tuson, H. H., Foley, M. H., Koropatkin, N. M., and Biteen, J. S. (2018) The Starch Utilization System Assembles around Stationary Starch-Binding Proteins. *Biophysical Journal* **115**, 242-250
335. Proctor, L. M., Creasy, H. H., Fettweis, J. M., Lloyd-Price, J., Mahurkar, A., Zhou, W. Y., Buck, G. A., Snyder, M. P., Strauss, J. F., Weinstock, G. M., White, O., Huttenhower, C., and Integrative, H. M. P. i. R. N. (2019) The Integrative Human Microbiome Project. *Nature* **569**, 641-648
336. Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., Fernandes, G. R., Tap, J., Bruls, T., Batto, J. M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, H. B., Nielsen, T., Pons, N., Poulain, J., Qin, J. J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, E. G., Wang, J., Guarner, F., Pedersen, O., de Vos, W. M., Brunak, S., Dore, J., Weissenbach, J., Ehrlich, S. D., Bork, P., and Meta, H. I. T. C. (2011) Enterotypes of the human gut microbiome. *Nature* **473**, 174-180
337. Foley, M. H., Cockburn, D. W., and Koropatkin, N. M. (2016) The Sus operon: a model system for starch uptake by the human gut Bacteroidetes. *Cellular and Molecular Life Sciences* **73**, 2603-2617
338. Bagenholm, V., Reddy, S. K., Bouraoui, H., Morrill, J., Kulcinskaja, E., Bahr, C. M., Aurelius, O., Rogers, T., Xiao, Y., Logan, D. T., Martens, E. C., Koropatkin, N. M., and Stalbrand, H. (2017) Galactomannan Catabolism Conferred by a Polysaccharide Utilization Locus of Bacteroides ovatus: enzyme synergy and crystal structure of a β-mannanase. *Journal of Biological Chemistry* **292**, 229-243
339. Bagenholm, V., Wiemann, M., Reddy, S. K., Bhattacharya, A., Rosengren, A., Logan, D. T., and Stalbrand, H. (2019) A surface-exposed GH26 β-mannanase from Bacteroides ovatus: Structure, role, and phylogenetic analysis of BoMan26B. *Journal of Biological Chemistry* **294**, 9100-9117
340. Leth, M. L., Ejby, M., Workman, C., Ewald, D. A., Pedersen, S. S., Sternberg, C., Bahl, M. I., Licht, T. R., Aachmann, F. L., Westereng, B., and Abou Hachem, M. (2018) Differential bacterial capture and transport preferences facilitate co-growth on dietary xylan in the human gut. *Nature Microbiology* **3**, 570-580
341. La Rosa, S. L., Leth, M. L., Michalak, L., Hansen, M. E., Pudlo, N. A., Glowacki, R., Pereira, G., Workman, C. T., Arntzen, M. O., Pope, P. B., Martens, E. C., Abou Hachem, M., and Westereng, B. (2019) The human gut Firmicute Roseburia intestinalis is a primary degrader of dietary beta-mannans. *Nature Communications* **10**, 14
342. Rogers, T. E., Pudlo, N. A., Koropatkin, N. M., Bell, J. S. K., Moya Balasch, M., Jasker, K., and Martens, E. C. (2013) Dynamic responses of Bacteroides thetaiotaomicron during growth on glycan mixtures. *Molecular Microbiology* **88**, 876-890
343. Rogowski, A., Briggs, J. A., Mortimer, J. C., Tryfona, T., Terrapon, N., Lowe, E. C., Basle, A., Morland, C., Day, A. M., Zheng, H., Rogers, T. E., Thompson, P., Hawkins,

A. R., Yadav, M. P., Henrissat, B., Martens, E. C., Dupree, P., Gilbert, H. J., and Bolam, D. N. (2015) Glycan complexity dictates microbial resource allocation in the large intestine. *Nat Commun* **6**, 7481

344. Cuskin, F., Lowe, E. C., Temple, M. J., Zhu, Y., Cameron, E. A., Pudlo, N. A., Porter, N. T., Urs, K., Thompson, A. J., Cartmell, A., Rogowski, A., Hamilton, B. S., Chen, R., Tolbert, T. J., Piens, K., Bracke, D., Vervecken, W., Hakki, Z., Speciale, G., Munoz-Munoz, J. L., Day, A., Pena, M. J., McLean, R., Suits, M. D., Boraston, A. B., Atherly, T., Ziemer, C. J., Williams, S. J., Davies, G. J., Abbott, D. W., Martens, E. C., and Gilbert, H. J. (2015) Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature* **517**, 165-169
345. Arnal, G., Bastien, G., Monties, N., Abot, A., Anton Leberre, V., Bozonnet, S., O'Donohue, M., and Dumon, C. (2015) Investigating the function of an arabinan utilization locus isolated from a termite gut community. *Applied and environmental microbiology* **81**, 31-39
346. Dodd, D., Moon, Y. H., Swaminathan, K., Mackie, R. I., and Cann, I. K. (2010) Transcriptomic analyses of xylan degradation by *Prevotella bryantii* and insights into energy acquisition by xylanolytic bacteroidetes. *The Journal of biological chemistry* **285**, 30261-30273
347. Kabisch, A., Otto, A., Konig, S., Becher, D., Albrecht, D., Schuler, M., Teeling, H., Amann, R. I., and Schweder, T. (2014) Functional characterization of polysaccharide utilization loci in the marine Bacteroidetes '*Gramella forsetii*' KT0803. *ISME J* **8**, 1492-1502
348. Larsbrink, J., Zhu, Y., Kharade, S. S., Kwiatkowski, K. J., Eijsink, V. G., Koropatkin, N. M., McBride, M. J., and Pope, P. B. (2016) A polysaccharide utilization locus from *Flavobacterium johnsoniae* enables conversion of recalcitrant chitin. *Biotechnology for biofuels* **9**, 260
349. Larsbrink, J., Thompson, A. J., Lundqvist, M., Gardner, J. G., Davies, G. J., and Brumer, H. (2014) A complex gene locus enables xyloglucan utilization in the model saprophyte *Cellvibrio japonicus*. *Mol Microbiol* **94**, 418-433
350. Geoghegan, K. F., Dixon, H. B., Rosner, P. J., Hoth, L. R., Lanzetti, a. J., Borzilleri, K. a., Marr, E. S., Pezzullo, L. H., Martin, L. B., LeMotte, P. K., McColl, a. S., Kamath, a. V., and Stroh, J. G. (1999) Spontaneous alpha-N-6-phosphogluconoylation of a "His tag" in Escherichia coli: the cause of extra mass of 258 or 178 Da in fusion proteins. *Analytical biochemistry* **267**, 169-184
351. Gouet, P., Robert, X., and Courcelle, E. (2003) ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Research* **31**, 3320-3323
352. Labourel, A., Jam, M., Jeudy, A., Hehemann, J. H., Czjzek, M., and Michel, G. (2014) The β-glucanase ZgLamA from zobellia galactanivorans evolved a bent active site adapted for efficient degradation of algal laminarin. *Journal of Biological Chemistry* **289**, 2027-2042
353. Kitamura, E., Myouga, H., and Kamei, Y. (2002) Polysaccharolytic activities of bacterial enzymes that degrade the cell walls of Pythium porphyrae, a causative fungus of red rot disease in Porphyra yezoensis. *Fisheries Science* **68**, 436-445

354. Kobayashi, T., Uchimura, K., Kubota, T., Nunoura, T., and Deguchi, S. (2016) Biochemical and genetic characterization of β-1,3 glucanase from a deep subseafloor Laceyella putida. *Applied Microbiology and Biotechnology* **100**, 203-214
355. Hong, T. Y., Hsiao, Y. Y., Meng, M., and Li, T. T. (2008) The 1.5 A˚structure of endo-1,3-β-glucanase from Streptomyces sioyaensis: Evolution of the active-site structure for 1,3-β-glucan-binding specificity and hydrolysis. *Acta Crystallographica Section D: Biological Crystallography* **64**, 964-970
356. Li, J., Xu, X., Shi, P., Liu, B., Zhang, Y., and Zhang, W. (2015) Overexpression and characterization of a novel endo-β-1,3(4)-glucanase from thermophilic fungus Humicola insolens Y1. *Protein Expression and Purification* **138**, 63-68
357. Masuda, S., Endo, K., Koizumi, N., Hayami, T., Fukazawa, T., Yatsunami, R., Fukui, T., and Nakamura, S. (2006) Molecular identification of a novel beta-1,3-glucanase from alkaliphilic Nocardiopsis sp strain F96. *Extremophiles* **10**, 251-255
358. Krah, M., Misselwitz, R., Politz, O., Thomsen, K. K., Welfle, H., and Borriss, R. (1998) The laminarinase from thermophilic eubacterium Rhodothermus marinus--conformation, stability, and identification of active site carboxylic residues by site-directed mutagenesis. *European journal of biochemistry / FEBS* **257**, 101-111
359. Fuchs, K. P., Zverlov, V. V., Velikodvorskaya, G. A., Lottspeich, F., and Schwarz, W. H. (2003) Lic16A of Clostridium thermocellum, a non-cellulosomal, highly complex endo-β-1,3-glucanase bound to the outer cell surface. *Microbiology* **149**, 1021-1031
360. Zverlov, V. V., Volkov, I. Y., Velikodvorskayal, T. V., and Schwarz, W. H. (1997) Highly thermostable endo-1,3-β-glucanase (laminarinase) LamA from Thermotoga neapolitana: nucleotide sequence of the gene and characterization of the recombinant gene product. *Microbiology* **143**, 1701-1708
361. Gorlach, J. M., Van Der Knaap, E., and Walton, J. D. (1998) Cloning and targeted disruption of MLG1, a gene encoding two of three extracellular mixed-linked glucanases of Cochliobolus carbonum. *Applied and Environmental Microbiology* **64**, 385-391
362. Nakajima, M., Yamashita, T., Takahashi, M., Nakano, Y., and Takeda, T. (2012) A novel glycosylphosphatidylinositol-anchored glycoside hydrolase from Ustilago esculenta functions in β-1,3-glucan degradation. *Applied and Environmental Microbiology* **78**, 5682-5689
363. Kawai, R., Igarashi, K., Yoshida, M., Kitaoka, M., and Samejima, M. (2006) Hydrolysis of β-1,3/1,6-glucan by glycoside hydrolase family 16 endo-1,3(4)-β-glucanase from the basidiomycete Phanerochaete chrysosporium. *Applied Microbiology and Biotechnology* **71**, 898-906
364. Hong, T. Y., and Meng, M. (2003) Biochemical characterization and antifungal activity of an endo-1,3-beta-glucanase of Paenibacillus sp. isolated from garden soil. *Applied microbiology and biotechnology* **61**, 472-478
365. Pauchet, Y., Freitak, D., Heidel-Fischer, H. M., Heckel, D. G., and Vogel, H. (2009) Immunity or digestion: Glucanase activity in a glucan-binding protein family from lepidoptera. *Journal of Biological Chemistry* **284**, 2214-2224
366. Genta, F. A., Bragatto, I., Terra, W. R., and Ferreira, C. (2009) Purification, characterization and sequencing of the major β-1,3-glucanase from the midgut of Tenebrio molitor larvae. *Insect Biochemistry and Molecular Biology* **39**, 861-874

367. Kumagai, Y., and Ojima, T. (2009) Enzymatic properties and the primary structure of a β-1,3-glucanase from the digestive fluid of the Pacific abalone Haliotis discus hannai. *Comparative Biochemistry and Physiology, Part B* **154**, 113-120
368. Kovalchuk, S. N., Sundukova, E. V., Kusaykin, M. I., Guzev, K. V., Anastiuk, S. D., Likhatskaya, G. N., Trifonov, E. V., Nurminski, E. A., Kozhemyako, V. B., Zvyagintseva, T. N., and Rasskazov, V. A. (2006) Purification, cDNA cloning and homology modeling of endo-1,3-β-D- glucanase from scallop Mizuhopecten yessoensis. *Comparative Biochemistry and Physiology, Part B* **143**, 473-485
369. Ferrer, P. (2006) Revisiting the Cellulosimicrobium cellulans yeast-lytic β-1,3-glucanases toolbox: a review. *Microbial cell factories* **5**, 10-10
370. Jeng, W. Y., Wang, N. C., Lin, C. T., Shyur, L. F., and Wang, A. H. J. (2011) Crystal structures of the laminarinase catalytic domain from Thermotoga maritima MSB8 in complex with inhibitors: Essential residues for β-1,3- and β-1,4-glucan selection. *Journal of Biological Chemistry* **286**, 45030-45040
371. Cota, J., Alvarez, T. M., Citadini, A. P., Santos, C. R., de Oliveira Neto, M., Oliveira, R. R., Pastore, G. M., Ruller, R., Prade, R. A., Murakami, M. T., and Squina, F. M. (2011) Mode of operation and low-resolution structure of a multi-domain and hyperthermophilic endo-β-1,3-glucanase from Thermotoga petrophila. *Biochemical and Biophysical Research Communications* **406**, 590-594
372. Hahn, M., Olsen, O., Politz, O., Borriss, R., and Heinemann, U. (1995) Crystal structure and site-directed mutagenesis of Bacillus macerans endo- 1,3-1,4-β-glucanase.
373. Furtado, G. P., Ribeiro, L. F., Santos, C. R., Tonoli, C. C., De Souza, A. R., Oliveira, R. R., Murakami, M. T., and Ward, R. J. (2011) Biochemical and structural characterization of a β-1,3-1,4-glucanase from Bacillus subtilis 168. *Process Biochemistry* **46**, 1202-1206

Appendices

Appendix A Supporting Information for Chapter 2

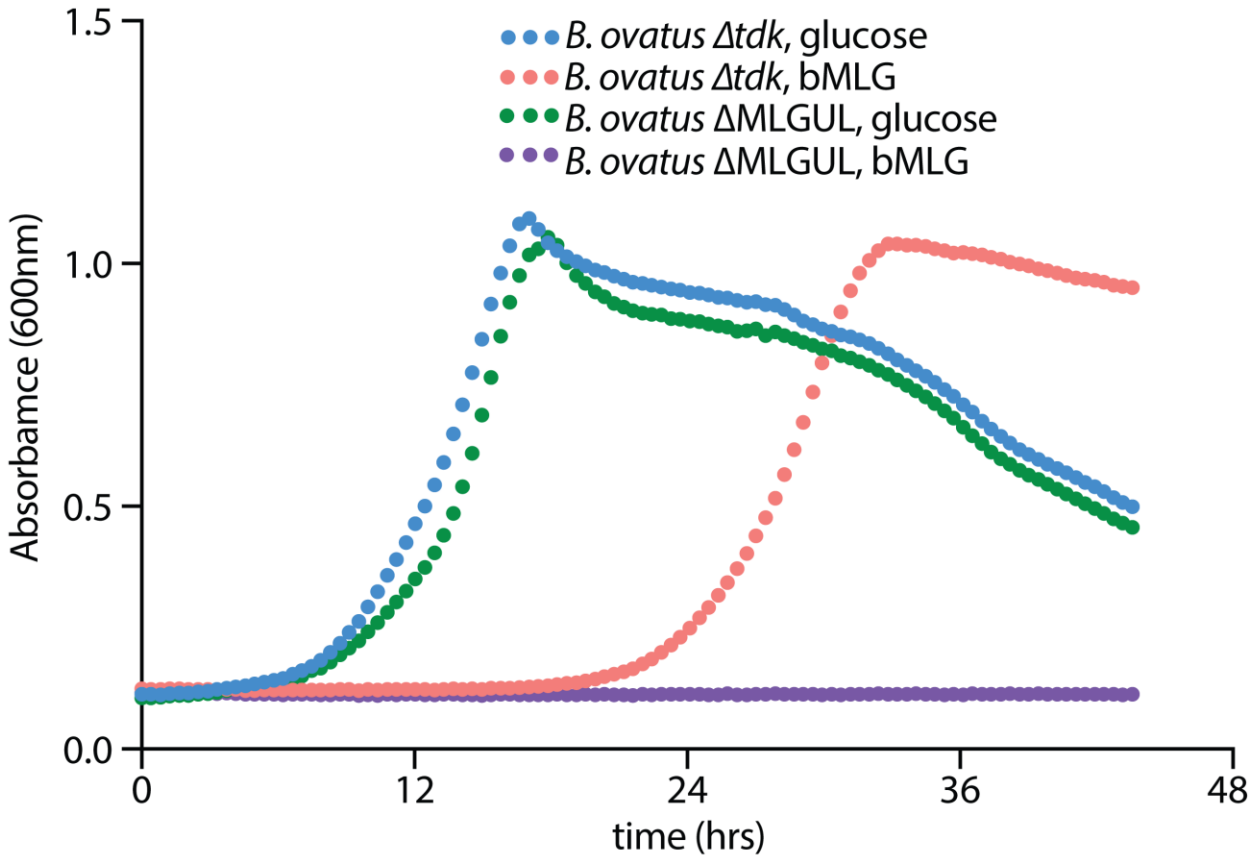

Figure A-1. Growth profile of B. ovatus strains.
B. ovatus Δtdk (wt) and *B. ovatus ΔMLGUL* (whole-PUL knockout) were cultured in minimal medium containing either glucose or bMLG as the sole carbon source (average of n = 2 growths per strain).

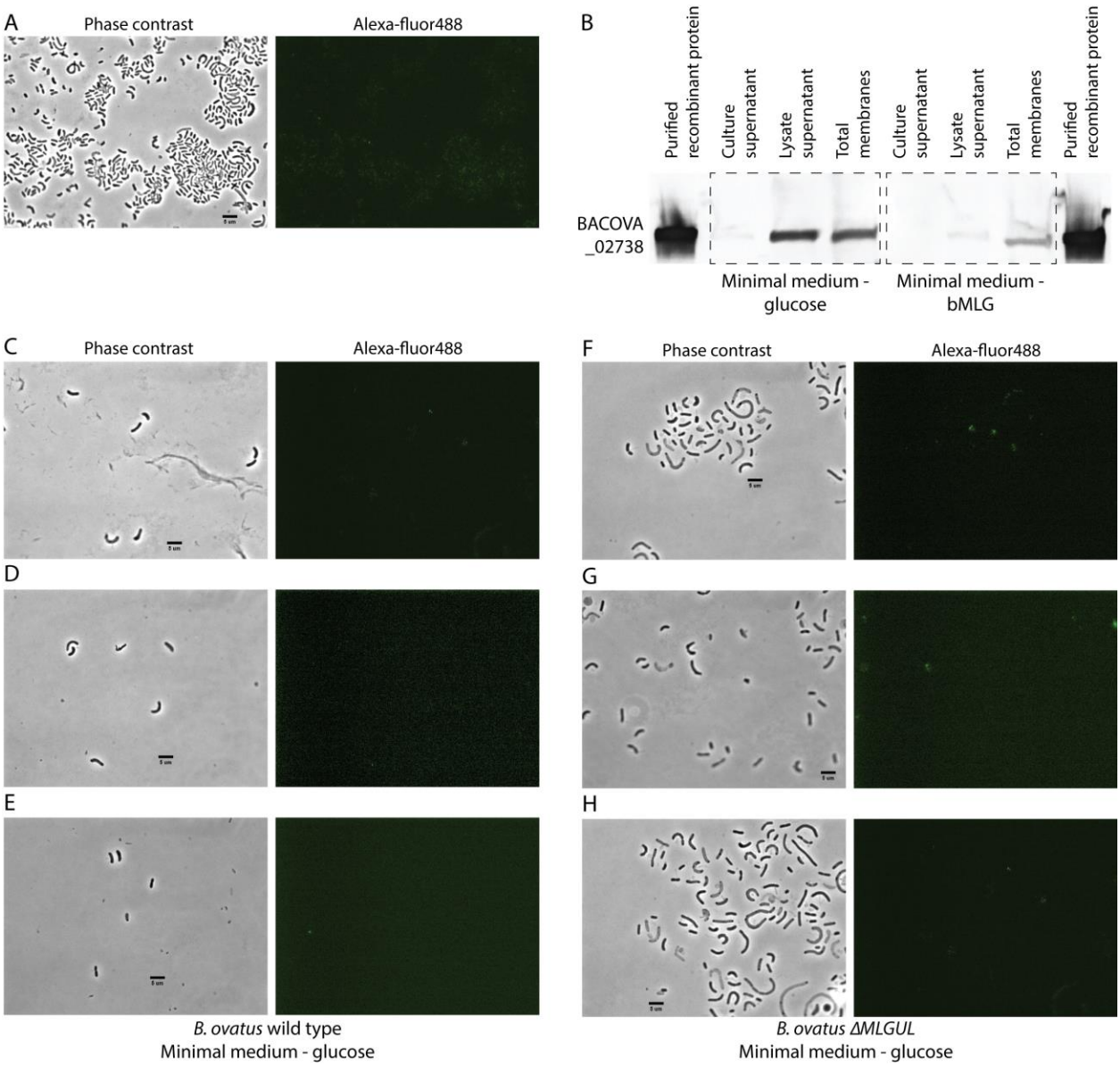

Figure A-2. Enzyme localization analysis.

A: phase contrast and corresponding fluorescence microscope images of *B. ovatus Δtdk* cells grown in minimal medium with bMLG as the sole carbon source probed with custom polyclonal antibodies against recBACOVA_02738(GH3). B: Western blot of protein collected from the culture supernatant, cell lysate supernatant, and cell lysate membrane fraction of *B. ovatus* Δtdk cells grown in minimal medium with glucose or bMLG as a sole carbon source. Phase contrast and corresponding fluorescence microscope images of wild type *B. ovatus* cells grown in minimal medium with glucose as the sole carbon source probed with custom polyclonal antibodies against recBoGH16$_{MLG}$ (C), recBoGH3$_{MLG}$ (D), and recBACOVA_02738(GH3) (E). Phase contrast and corresponding fluorescence microscope images of *B. ovatus ΔMLGUL* cells grown in minimal medium with glucose as the sole carbon source probed with custom polyclonal antibodies against recBoGH16$_{MLG}$ (F), recBoGH3$_{MLG}$ (G), and recBACOVA_02738(GH3) (H).

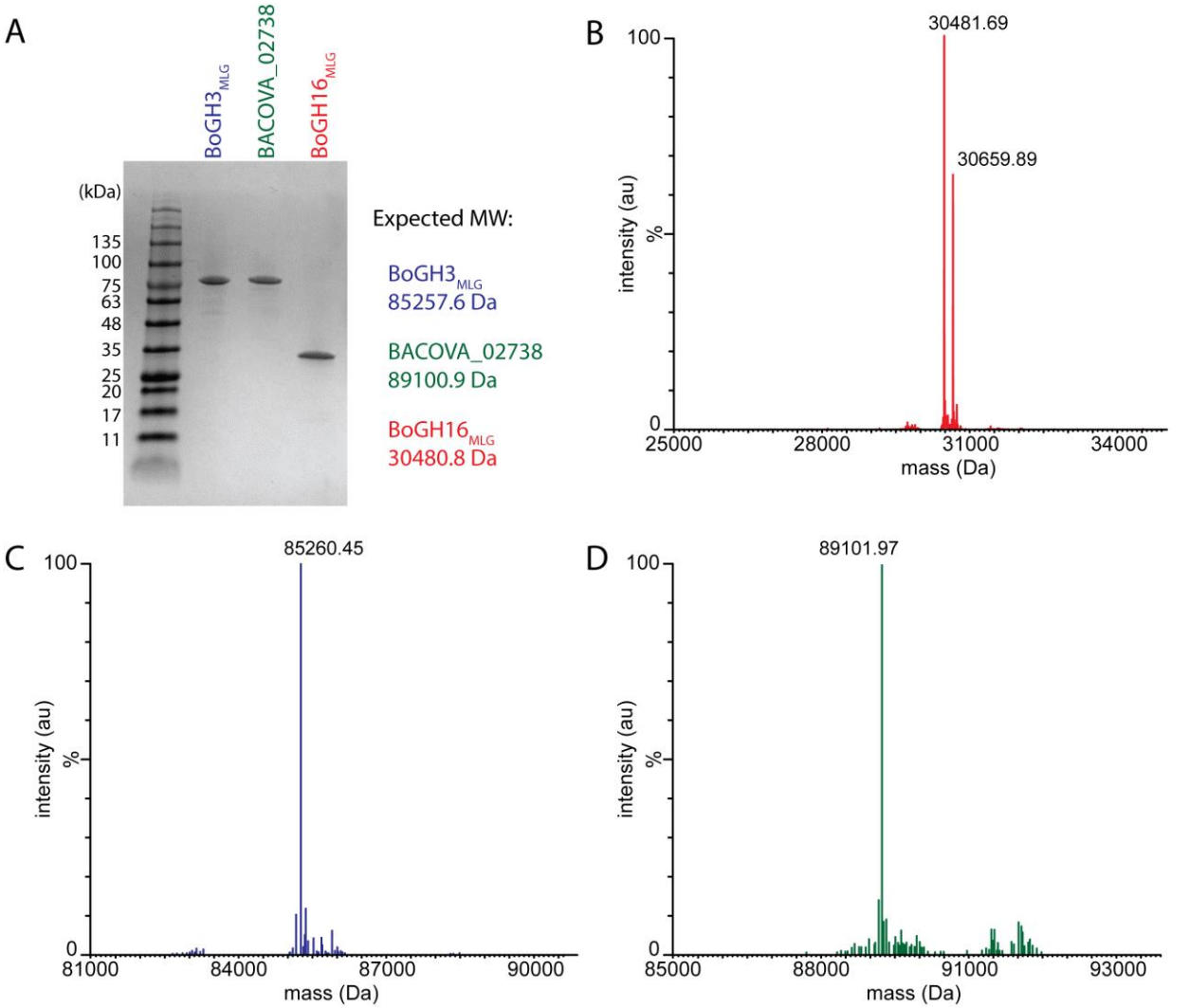

Figure A-3. Purity and molecular mass of recombinant MLGUL proteins.
A: SDS-PAGE of recBoGH16$_{MLG}$, recBoGH3$_{MLG}$, and recBACOVA_02738(GH3). B: Reconstructed mass spectrum of intact recBoGH16$_{MLG}$. The second peak at +178.2 Da from the main peak is a species that has been spontaneously phosphogluconoylated at the N-terminal his-tag during production in *E. coli* (350). C: Reconstructed mass spectrum of intact recBoGH3$_{MLG}$. D: Reconstructed mass spectrum of intact recBACOVA_02738(GH3).

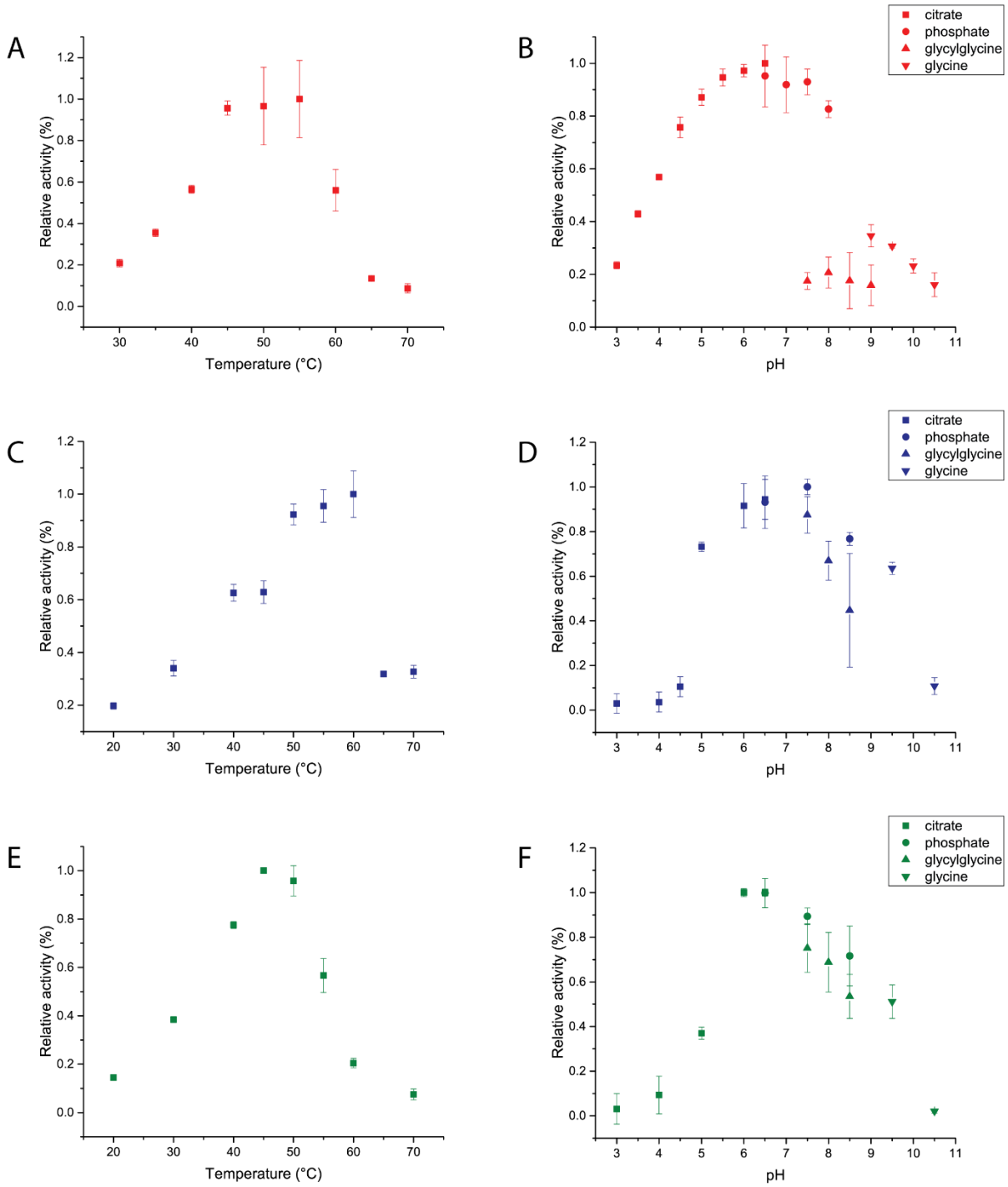

Figure A-4. Optimum conditions for recombinant MLGUL enzymes.
A: Optimum temperature profile of recBoGH16. B: Optimum pH profile of recBoGH16. C: Optimum temperature profile of recBoGH3A. D: Optimum pH profile of recBoGH3A. E: Optimum temperature profile of recBoGH3B. F: Optimum pH profile of recBoGH3B.

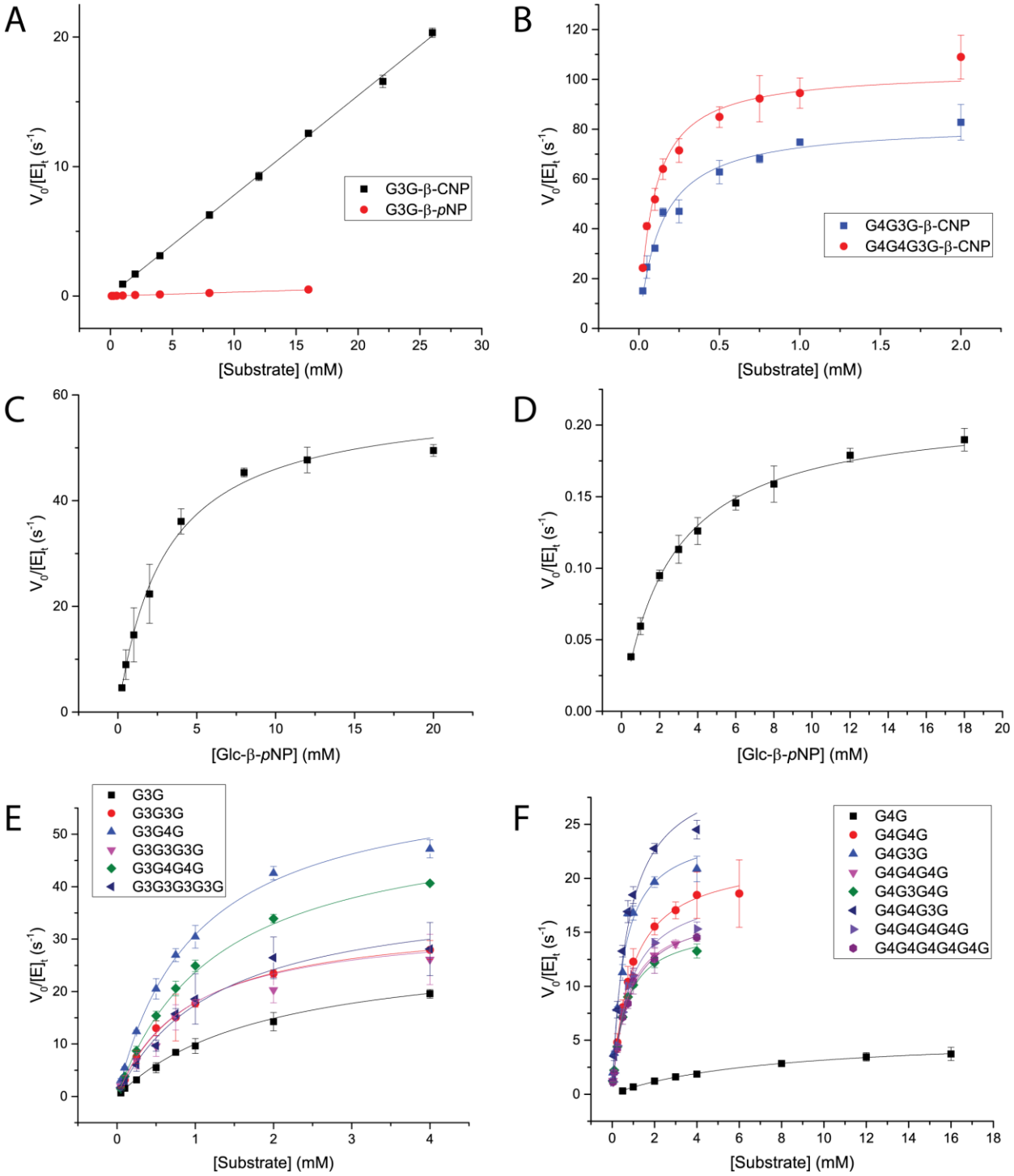

Figure A-5. Initial-rate kinetic analysis of MLGUL GHs.

A: BoGH16$_{MLG}$ against G3G-β-CNP and G3G-β-*p*NP fitted to a linear equation. B: BoGH16$_{MLG}$ against G4G3G-β-CNP and G4G4G3G-β-CNP fitted to the Michealis-Menten equation. C: BoGH3$_{MLG}$ against glucose-β-*p*NP fitted to the Michealis-Menten equation. D: BACOVA_02738(GH3) against glucose-β-*p*NP fitted to the Michealis-Menten equation. E: BoGH3$_{MLG}$ against oligosaccharides with a β(1,3) bond at the non-reducing end. D: BoGH3$_{MLG}$ with oligosaccharides with a β(1,4) bond at the non-reducing end. Curve fitting was done on OriginPro 2015 and error bars represent standard deviations from the mean.

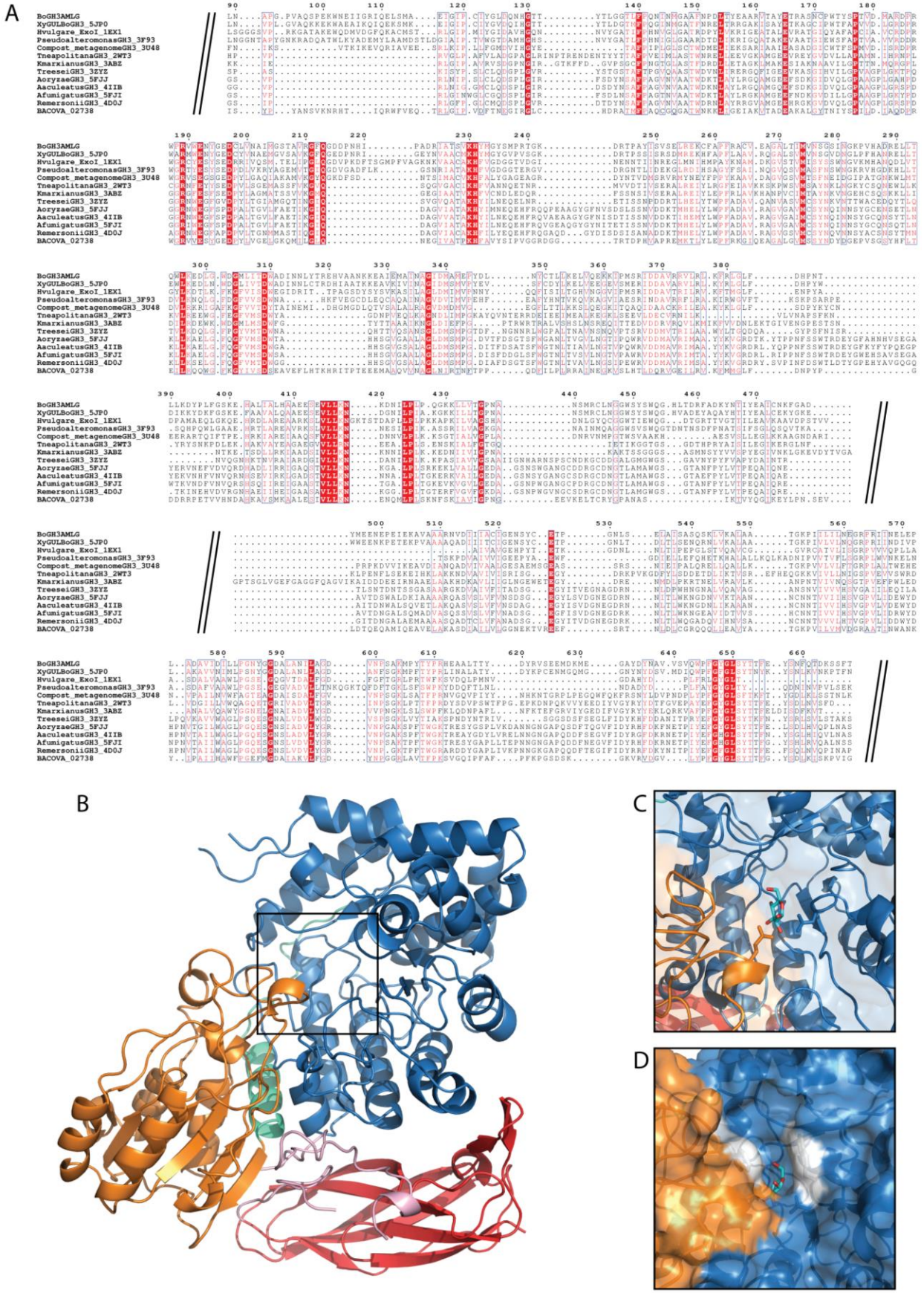
A
B
C
D

Figure A-6. Primary and tertiary structure analysis of BoGH3$_{MLG}$.
A: Structure-based sequence aligmnment of BoGH3$_{MLG}$ and BACOVA_02738(GH3) with structurally characterized GH3 β-glucosidases. Portions of the alignment have been removed for brevity and breaks are indicated by double hash lines. Red highlights indicate invariant positions, blue outlines indicate similar positions and green arrows indicate catalytic residues. Alignment illustration created with ESPript (351). B: Homology model of BoGH3$_{MLG}$ generated by Phyre2. The $(\alpha/\beta)_8$ TIM barrel (blue) is connected by a linker (teal) to a central α/β sandwich (orange), in turn connected by a linker (pink) to a C-terminal fibronectin type-III (FN-III) domain (red). The black box indicates the location of the active site. C: Catalytic site of BoGH3$_{MLG}$ with a bound glucose from the XyGUL BoGH3B overlay in the active site pocket. The catalytic nucleophile (Asp-309) and catalytic acid/base (Glu-527) are shown as sticks. D: Surface representation of the entrance to the BoGH3$_{MLG}$ active site with the two tryptophan residues that line the positive subsite shown in white.

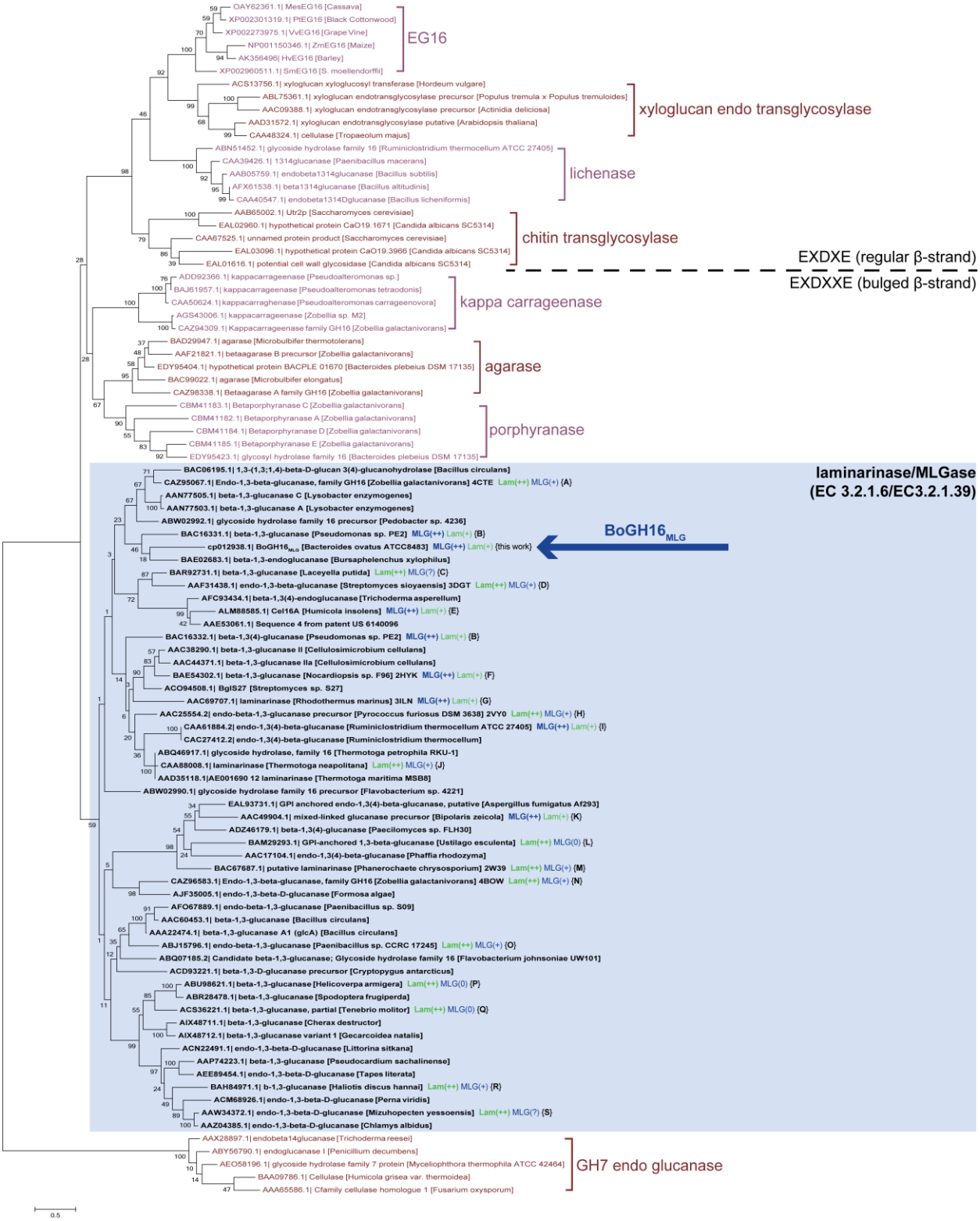

Figure A-7. Phylogenetic tree of characterized Glycoside Hydrolase Family 16 sequences.
Clades distinguishing major substrate specificities (see https://www.cazypedia.org/index.php/GH16) are represented by at least 5 members. Leaf names contain GenBank accession number, enzyme name, organism of origin, and a PDB code where available. Bootstrap values are shown for each node. The dotted horizontal line separates GH

Family 16 enzymes with active-site residues on a regular β-strand from those with a bulged β-strand. The clade highlighted with a blue background is traditionally referred to as a "laminarinase" (EC 3.2.1.39) group, yet our current literature analysis also indicates the presence of predominant MLGases (EC 3.2.1.6). Where known, biochemically determined activities are shown next to select sequences, with the predominant activity displayed in bold type; (++) indicates better activity; (+) indicates poorer activity, (0) indicates no activity, and (?) indicates that specific activity data is unavailable. Reported activities are based on the following references, A: (352), B: (353), C: (354), D: (355), E: (356), F: (357), G: (358), H: (197) , I: (359), J: (360), K: (361), L: (362), M: (363), N: (155), O: (364), P: (365), Q: (366), R: (367), S: (368)

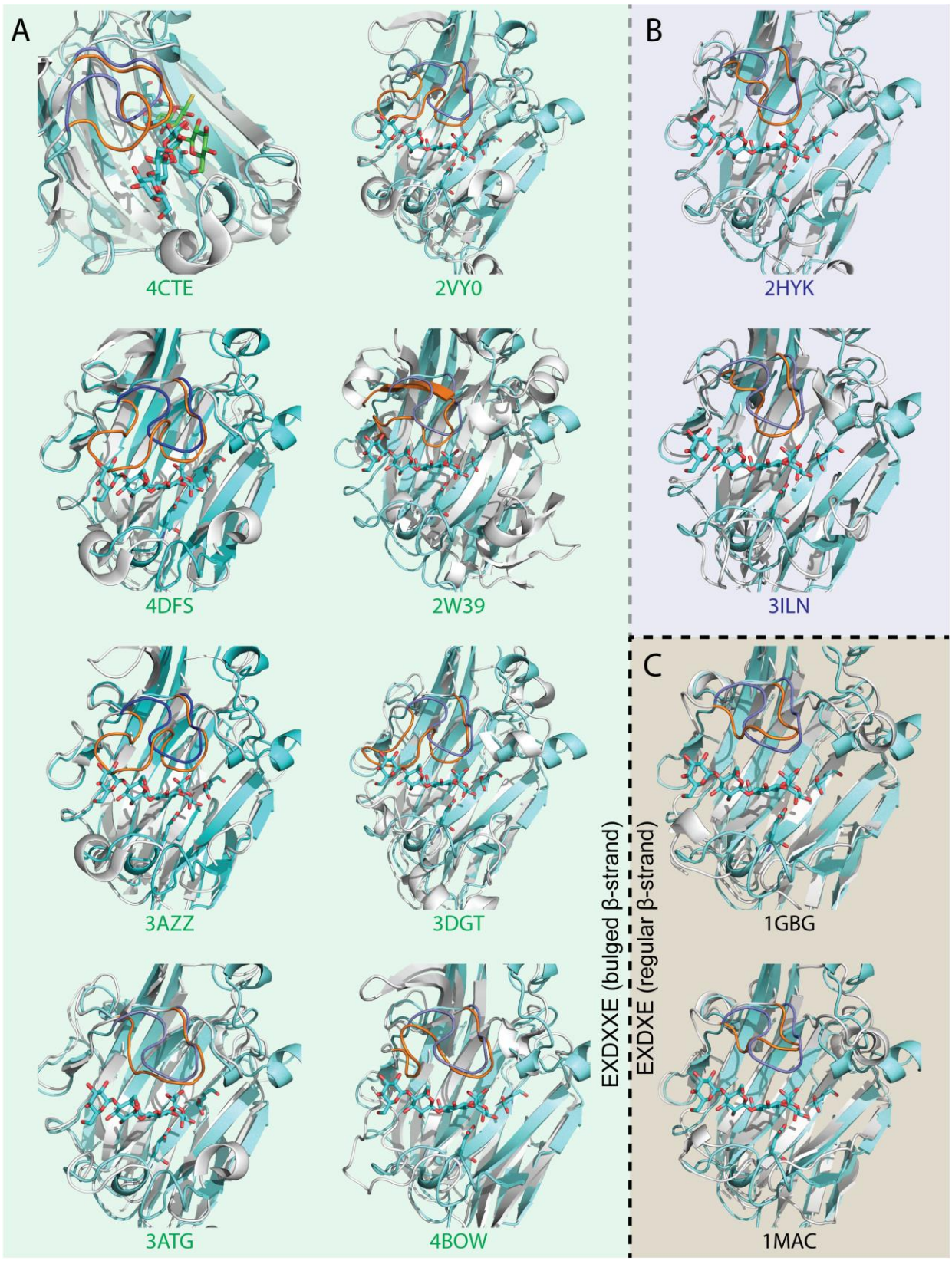

Figure A-8. BoGH16$_{MLG}$ loop comparison with laminarinases and MLGases.
BoGH16$_{MLG}$ (cyan) is structurally aligned with all 10 available β-bulge-containing laminarinases/MLGases and two representative regular β-stranded canonical lichenases (white); PDB ID of the compared structure shown below each

alignment. The $BoGH16_{MLG}$ loop is colored blue and the loop of the compared structure is colored orange. The $BoGH16_{MLG}$ catalytic residues and mixed-linkage oligosaccharide G4G4G3G in complex with $BoGH16_{MLG}$ are also shown in cyan. The thio-β-1,3-trisaccharide in complex with $ZgLamC_{GH16\text{-}E142S}$ (PDB code 4CTE) is shown in green A: β-bulge-containing laminarinases, B: β-bulge-containing MLGases, C: regular β-stranded MLGases (canonical lichenases).

Table A-1. List of primers used for cloning.

Primer Name	Primer sequence (5' → 3')	Vector
$BoGH16_{MLG}$_F	GACGACCATATGTCGGATTCTGTTGGAACG	pET28
$BoGH16_{MLG}$_R	GACGACCTCGAGCTATAATATTTTCACCCA	pET28
$BoGH3_{MLG}$_F	TACTTCCAATCCAATGCCATGGTTCCCACTGCCATTCCTGAA	pMCSG53
$BoGH3_{MLG}$_R	TTATCCACTTCCAATGTTATTACTTGCATATAATATTCAGTGTTTGA	pMCSG53
BACOVA_02738_F	TACTTCCAATCCAATGCCATGAACAACAAACCTACTGATAACA	pMCSG53
BACOVA_02738_R	TTATCCACTTCCAATGTTATTATTGGACCTCAAAACTCCCCT	pMCSG53

Restriction sites are underlined and pMCSG LIC vector complementary sequences are double underlined.

Table A-2. Transcriptional expression of MLGUL and neighboring genes.[a]

Locus tag	Putative protein ID	bMLG	MM-Glc	bMLG/MM-Glc
BACOVA_02736		879.4 ± 29.4	511.6 ± 94.3	1.7
BACOVA_02737		1072.3 ± 23.0	590.4 ± 147.6	1.8
BACOVA_02738	GH3	355.7 ± 46.9	229.0 ± 55.5	1.6
BACOVA_02739	Sigma 70, region 4	22.7 ± 0.2	37.2 ± 8.9	0.6
BACOVA_02740	**HTCS**	**376.8 ± 4.4**	**197.6 ± 0.1**	**1.9**
BACOVA_02741	**GH16$_{MLG}$**	**3657.6 ± 77.7**	**16.0 ± 16.5**	**228.8**
BACOVA_02742	**TBDT**	**3401.1 ± 375.5**	**13.1 ± 5.7**	**259.7**
BACOVA_02743	**SGBP-A**	**1837.5 ± 111.1**	**14.7 ± 6.3**	**124.9**
BACOVA_02744	**SGBP-B**	**2748.9 ± 103.9**	**9.2 ± 1.0**	**298.3**
BACOVA_02745	**GH3$_{MLG}$**	**2988.3 ± 53.7**	**16.5 ± 0.4**	**180.7**
BACOVA_02746[b]	Transposase	2763.9 ± 111.5	11.9 ± 9.7	232.8
BACOVA_02747[b]	Helicase	2821.0 ± 91.9	16.8 ± 5.8	167.6
BACOVA_02748		54.5 ± 10.5	25.8 ± 9.1	2.1
BACOVA_02749		62.3 ± 6.0	30.1 ± 6.8	2.1

[a] Microarray results of upregulation under bMLG induction normalized to a glucose background are shown with MLGUL genes in bold type. Values reported as averages and standard deviations of two biological replicates. Data are from (109).

[b] Loci BACOVA_02746 and BACOVA_02747 are predicted to encoded proteins of only 63 and 44 amino acids, respectively, and were therefore likely to have been originally mis-annotated based on limited regional sequence similarity. Despite apparently high transcript levels, which may result from read-through downstream of highly active operons whether there is protein coding function or not, these loci are not considered to be part of the MLGUL.

Table A-3. Kinetic parameters of BoGH16$_{MLG}$ on polysaccharide substrates.

Substrate	k_{cat} (s^{-1})	K_m (mg mL^{-1})	$kcat/K_m$ (s^{-1} mg^{-1} mL)	Assay
barley MLG	85.8 ± 4.6	0.364 ± 0.051	238	BCA
laminarin	18.1 ± 5.25	2.54 ± 0.97	7.12	BCA
yeast β-glucan	0.875 ± 0.101	0.541 ± 0.169	1.61	BCA
curdlan	ND	ND	0.042	BCA

Data is only presented for substrates on which BoGH16$_{MLG}$ showed activity (no detectable activity on tamarind xyloglucan, beechwood xylan, wheat arabinoxylan, carob galactomannan, konjac glucomannan, carboxymethyl cellulose, hydroxyethylcellulose, xanthan gum, and ulvan as determined by BCA and HPLC analyses). ND: not determined (in cases where Michealis-Menten curve fitting was not feasible, individual k_{cat} and K_m values are not reported and k_{cat}/K_m value was determined from linear curve fit to initial rate data in the [S] << $K_{m(apparent)}$ range). Data are represented as mean ± standard deviation.

Table A-4. Kinetic parameters of BoGH16$_{MLG}$ on chromogenic substrates.

Substrate	k_{cat} (s^{-1})	K_m (mM)	$kcat/K_m$ (s^{-1} mM^{-1})	Assay
G-*p*NP	NA	NA	NA	*p*NP
G-CNP	NA	NA	NA	CNP
G4G-CNP	NA	NA	NA	CNP
G4G4G-CNP	NA	NA	NA	CNP
G3G-*p*NP	ND	ND	0.0298	*p*NP
G3G-CNP	ND	ND	0.768	CNP
G4G3G-CNP	82.3 ± 3.8	0.134 ± 0.018	614	CNP
G4G4G3G-CNP	103.7 ± 3.4	0.0895 ± 0.0097	1160	CNP

NA: no detectable activity. ND: not determined (in cases where Michealis-Menten curve fitting was not feasible, individual k_{cat} and K_m values are not reported and k_{cat}/K_m value was determined from linear curve fit to initial rate data in the [S] << $K_{m(apparent)}$ range). Data are represented as mean ± standard deviation.

Table A-5. Data collection and refinement statistics for BoGH16$_{MLG}$ structures.

	Apo-BoGH16$_{MLG}$ (5NBO)	**G4G4G3G-BoGH16$_{MLG}$ (5NBP)**
Data collection		
Space group	C2	C2
Cell dimensions		
a, *b*, *c* (Å)	167.5, 60.8, 49.4	167.3, 60.2 49.7
α, β, γ (°)	90.0, 94.5, 90.0	90.0, 93.5, 90.0
Wavelength (Å)	0.976	0.979
Resolution (Å)	49.23–1.80 (1.84–1.80)	49.57-1.80 (1.84-1.80)
R_{merge}	0.067 (0.446)	0.094 (0.735)
I / σI	10.0 (2.0)	10.8 (1.7)
Completeness (%)	98.9 (99.9)	99.0 (99.0)
Redundancy	3.0 (3.0)	3.9 (3.9)
Half-set correlation CC(1/2)	0.996 (0.669)	0.994 (0.516)
$R_{p.i.m.}$	0.053 (0.365)	0.066 (0.499)
Refinement		
Resolution (Å)	49.23–1.80	49.57-1.80
No. reflections (Work/Free)	43,072/2,392	45,346/2,384
R_{work} / R_{free}	0.161/0.204	0.175/0.212
No. atoms		
Protein	3,766	3,767
Ligand/solvent/ion	8	93
Water	353	262
Average B-factors (Å^2)		
Protein	18.6	21.6
Ligand/ion	25.4	29.9
Water	29.2	28.6
R.m.s deviations		
Bond lengths (Å)	0.018	0.013
Bond angles (°)	1.83	1.62

Values in parentheses represent data in the highest resolution shell.

Table A-6. Dali search results.

PDB ID	Organism	Protein name	Z score	RMSD (Å)	% ID	Active site	Predominant activity
4CTE	*Zobellia galactanivorans*	ZgLamC	29.3	2.0	38	β-bulge	laminarinase[a]
2HYK	*Nocardiopsis sp. F96*	BglF	28.8	1.9	32	β-bulge	MLGase[b]
3ATG	*Cellulosimicrobium cellulans*	BglII	28.3	1.9	34	β-bulge	laminarinase[c]
3AZY	*Thermotoga maritima*	Lam16	27.8	1.9	32	β-bulge	laminarinase[d]
2VY0	*Pyrococcus furiosus*	LamA	27.3	1.8	31	β-bulge	laminarinase[e]
4DFS	*Thermotoga petrophila*	TpLam	27.3	2.1	31	β-bulge	laminarinase[f]
4BOW	*Zobellia galactanivorans*	ZgLamA	26.9	2.0	30	β-bulge	laminarinase[g]
3ILN	*Rhodothermus marinus*	LamR	26.4	2.2	29	β-bulge	MLGase[h]
1MAC	*Paenibacillus macerans*	Bgi	25.1	2.0	22	regular β-strand	MLGase[i]
3DGT	*Streptomyces sioyaensis*	Curd1	25.0	2.4	28	β-bulge	laminarinase[j]
1GBG	*Bacillus licheniformis*	Bg1	25.0	2.1	23	regular β-strand	MLGase[k]
3O5S	*Bacillus subtilis*	BglS	25.0	2.0	24	regular β-strand	MLGase[l]

(352)
[b] (357)
[c] (369)
[d] (370)
[e] (197)
[f] (371)
[g] (155)
[h] (358)
[i] (372)
[j] (355)
[k] (188)
[l] (373)

Appendix B Supporting Information for Chapter 3

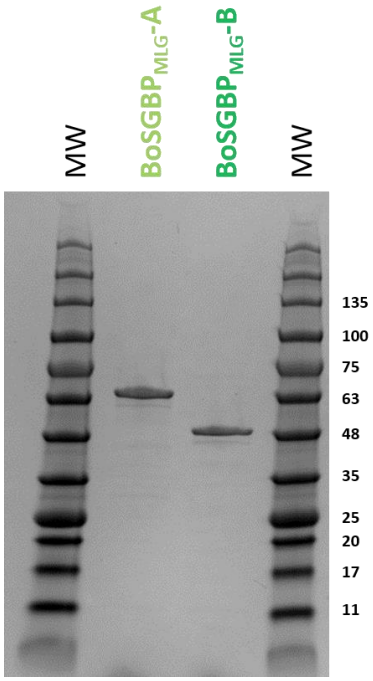

Figure B-1. Recombinant proteins.
SDS-PAGE of purified SGBPs with molecular weight (MW) ladders on either side; ladder band sizes are on the right in kilodaltons (kDa). The calculated expected molecular weights are 62.5 kDa for BoSGBP$_{MLG}$-A and 45.5 kDa for BoSGBP$_{MLG}$-B.

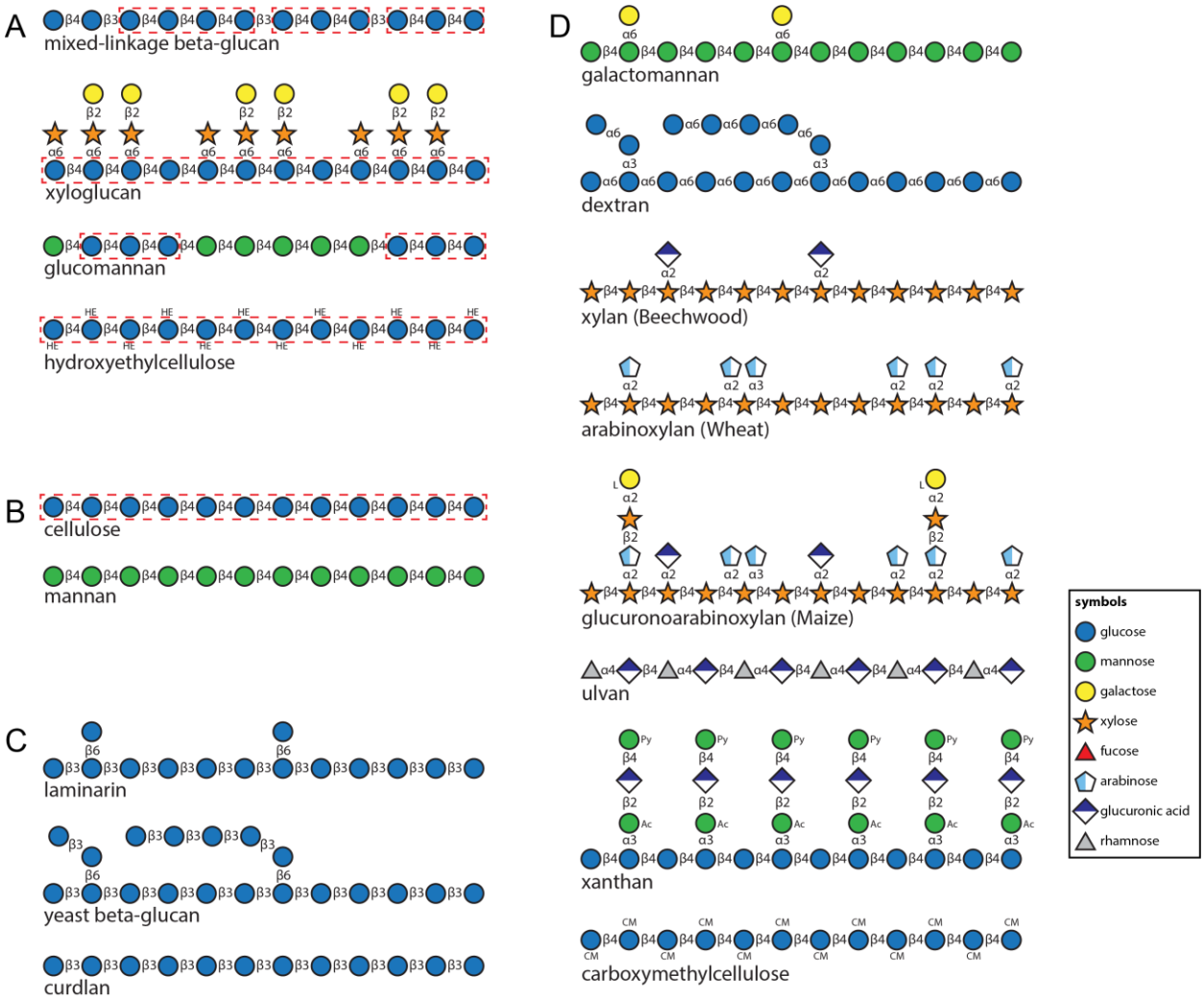

Figure B-2. Structures of polysaccharides.

(A) Polysaccharides that BoSGBP$_{MLG}$-A and BoSGBP$_{MLG}$-B bind (B) Insoluble polysaccharides. (C) Other β-glucans tested that BoSGBP$_{MLG}$-A and BoSGBP$_{MLG}$-B did not bind. (D) Other polysaccharides tested that BoSGBP$_{MLG}$-A and BoSGBP$_{MLG}$-B did not bind. The dashed red boxes highlight the β(1,4)-linked glucose units in the backbone, which are common to all polysaccharides that BoSGBP$_{MLG}$-A and BoSGBP$_{MLG}$-B bind.

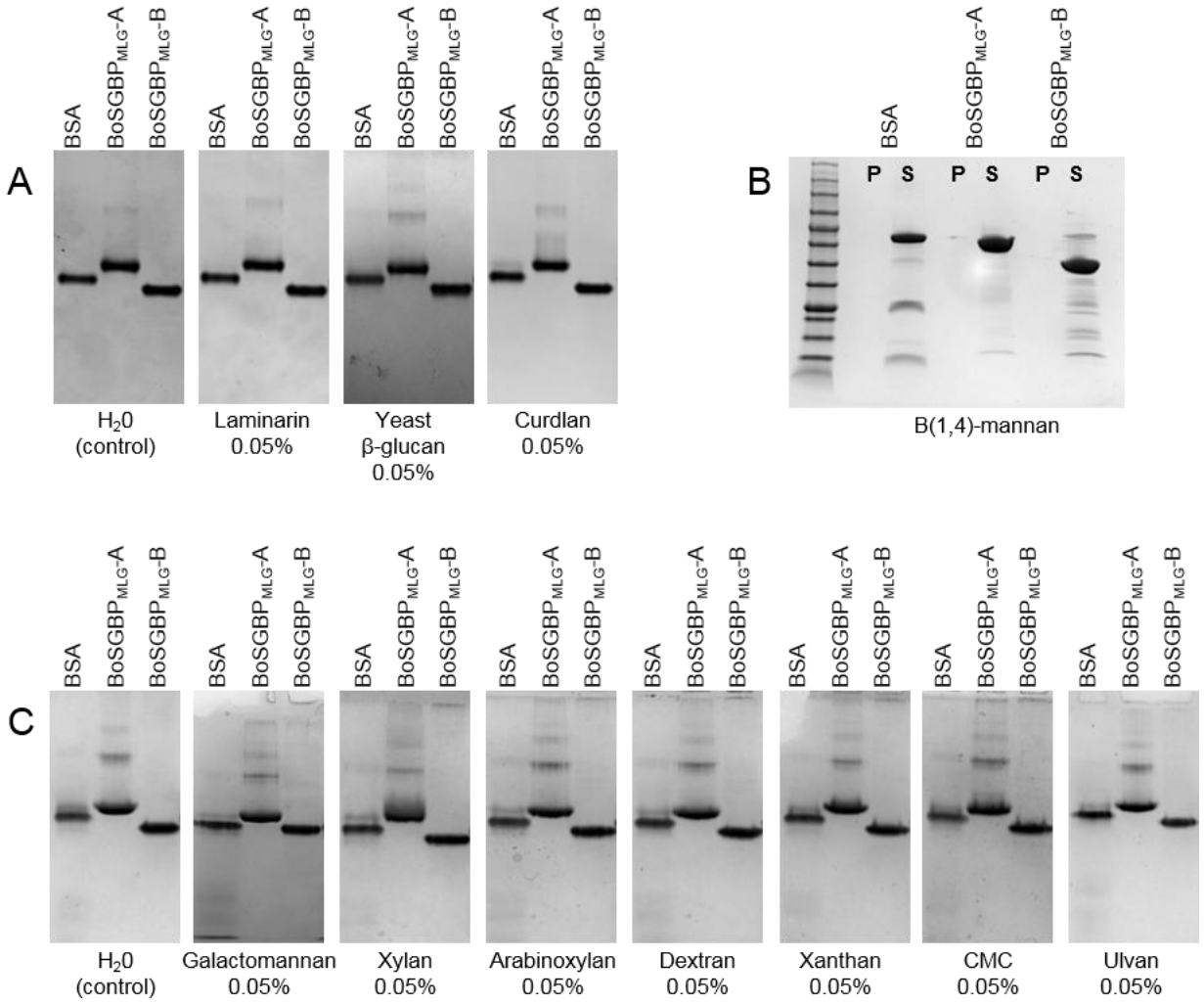

Figure B-3. BoSGBP$_{MLG}$-A and BoSGBP$_{MLG}$-B non-binding polysaccharides.
(A) Affinity electrophoresis gels against β(1,3)-glucans. (B) SDS-PAGE of pull-down assay against insoluble mannan. (C) Affinity electrophoresis gels against various other polysaccharides.

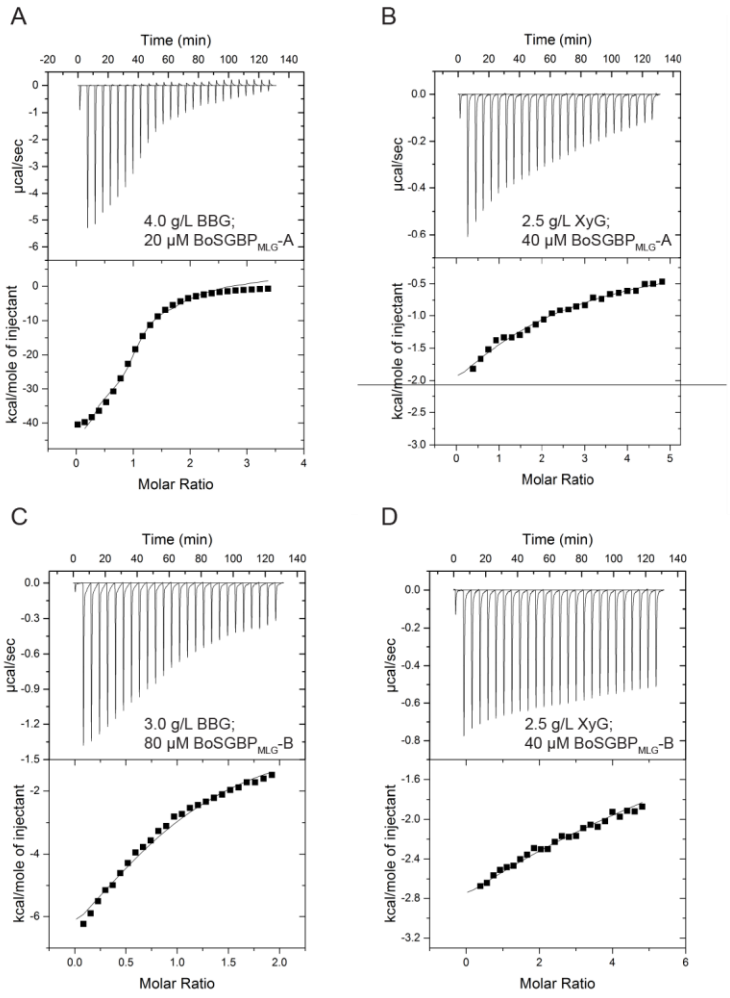

Figure B-4. Representative isothermal titration calorimetry (ITC) results for BoSGBP$_{MLG}$-A and BoSGBP$_{MLG}$-B with bMLG and XyG.

All titrations were performed in 50 mM Sodium Phosphate pH 7.0 with the exception of BoSGBP$_{MLG}$-B with bMLG (performed in 10 mM HEPES pH 7.0) and at 25 °C. In each case, the upper graph shows the raw heat signal for the 10 µL injections of carbohydrate into protein; the bottom graph shows the integrated heats and, where appropriate, fits to a 1:1 binding model. Concentrations of the protein and glycan are indicated in the figure. (A) BoSGBP$_{MLG}$-A with bMLG, (B) BoSGBP$_{MLG}$-A with XyG, (C) BoSGBP$_{MLG}$-B with bMLG, (D) BoSGBP$_{MLG}$-B with XyG.

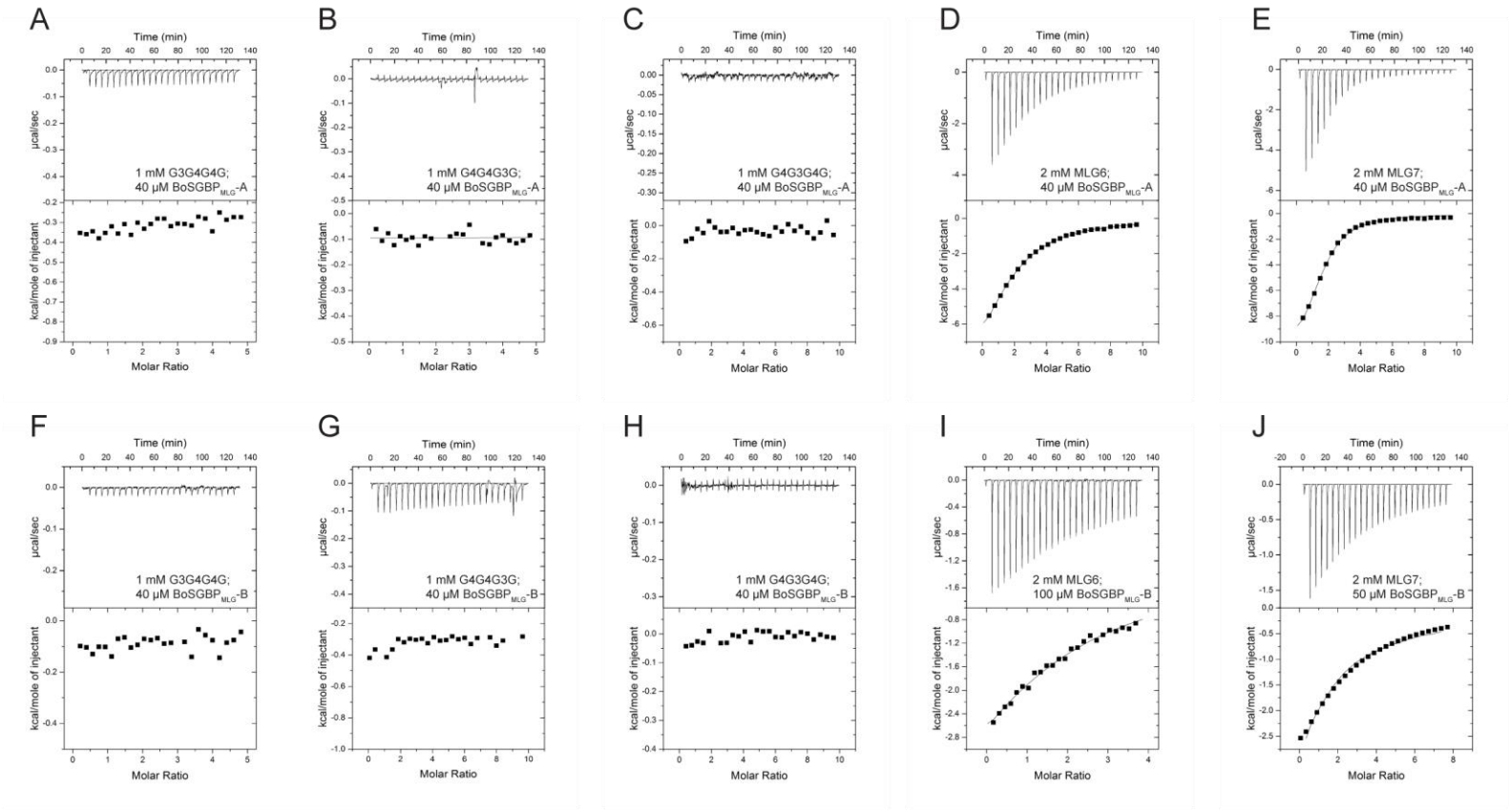

Figure B-5. Representative ITC results for BoSGBP$_{MLG}$-A and BoSGBP$_{MLG}$-B with MLG oligosaccharides.
All titrations were performed in 50 mM Sodium Phosphate pH 7.0 and at 25 °C. In each case, the upper graph shows the raw heat signal for the 10 µL injections of carbohydrate into protein; the bottom graph shows the integrated heats and, where appropriate, fits to a 1:1 binding model. Concentrations of the protein and glycan are indicated in the figure. (A) BoSGBP$_{MLG}$-A with G3G4G4G, (B) BoSGBP$_{MLG}$-A with G4G4G3G, (C) BoSGBP$_{MLG}$-A with G4G3G4G, (D) BoSGBP$_{MLG}$-A with MLG6, (E) BoSGBP$_{MLG}$-A with MLG7, (F) BoSGBP$_{MLG}$-B with G3G4G4G, (G) BoSGBP$_{MLG}$-B with G4G4G3G, (H) BoSGBP$_{MLG}$-B with G4G3G4G, (I) BoSGBP$_{MLG}$-B with MLG6, (J) BoSGBP$_{MLG}$-B with MLG7.

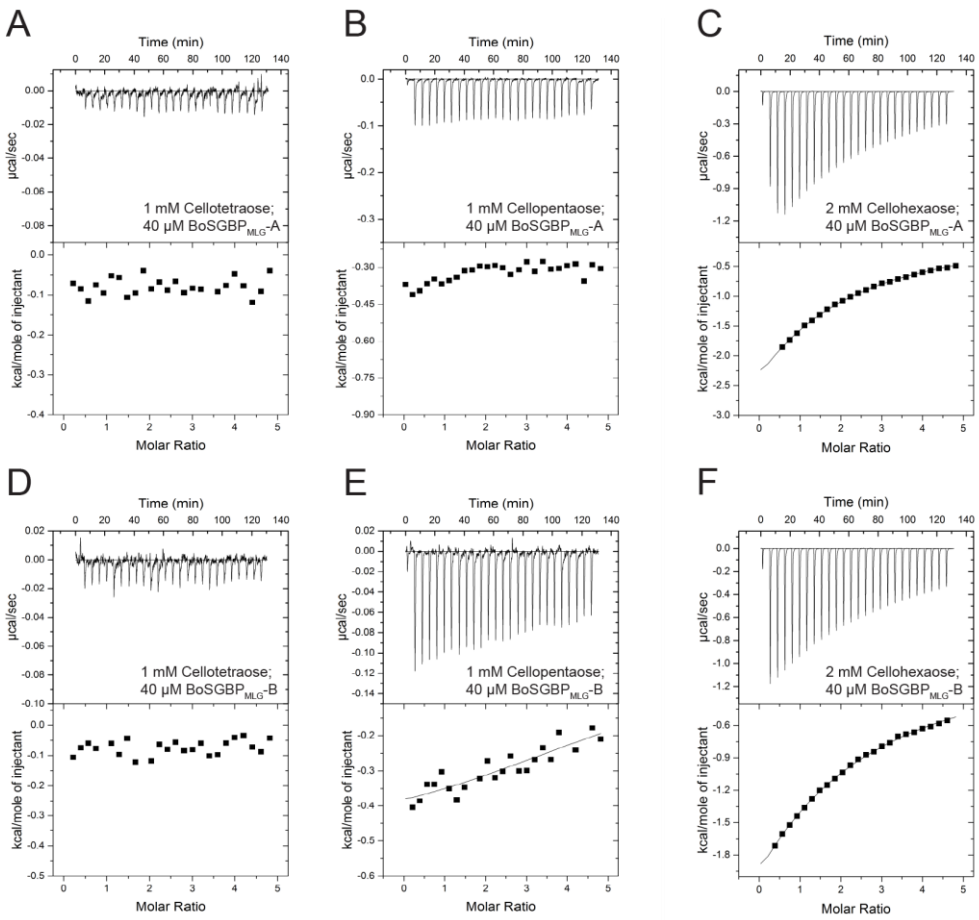

Figure B-6. Representative ITC results for BoSGBP$_{MLG}$-A and BoSGBP$_{MLG}$-B with cellooligosaccharides.
All titrations were performed in 50 mM Sodium Phosphate pH 7.0 and at 25 °C. In each case, the upper graph shows the raw heat signal for the 10 µL injections of carbohydrate into protein; the bottom graph shows the integrated heats and, where appropriate, fits to a 1:1 binding model. Concentrations of the protein and glycan are indicated in the figure. (A) BoSGBP$_{MLG}$-A with cellotetraose, (B) BoSGBP$_{MLG}$-A with cellopentaose, (C) BoSGBP$_{MLG}$-A with cellohexaose, (D) BoSGBP$_{MLG}$-B with cellotetraose, (E) BoSGBP$_{MLG}$-B with cellopentaose (F) BoSGBP$_{MLG}$-B with cellohexaose.

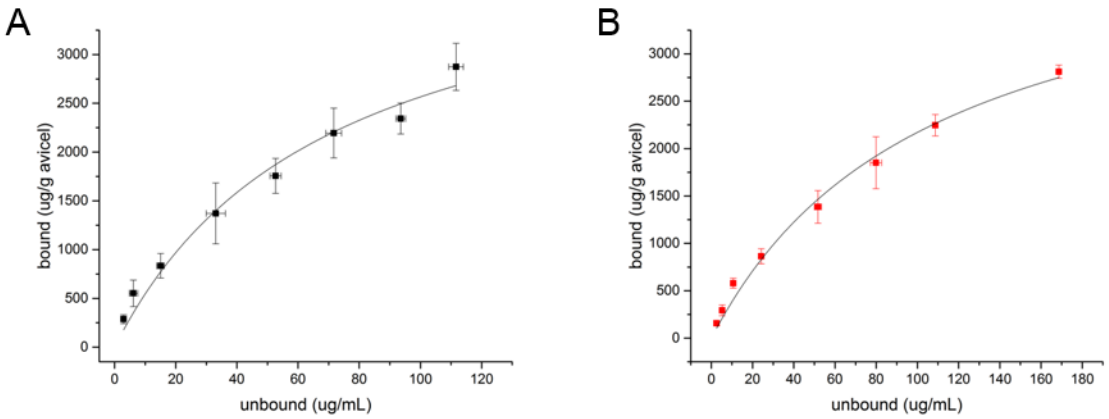

Figure B-7. Avicel depletion isotherm for (A) GFP-BoSGBP$_{MLG}$-A and (B) GFP-BoSGBP$_{MLG}$-B.
The dissociation constants (K_d) extracted from curve fitting (see methods for details) are 49.1 ± 13.1 μM and 117.4 ± 21.0 μM for GFP-BoSGBP$_{MLG}$-A and GFP-BoSGBP$_{MLG}$-B, respectively. Corresponding association constants (K_a) are 2.04 (± 0.54) x 10^4 M^{-1} and 8.52 (± 1.50) x 10^3 M^{-1} respectively for GFP-BoSGBP$_{MLG}$-A and GFP-BoSGBP$_{MLG}$-B.

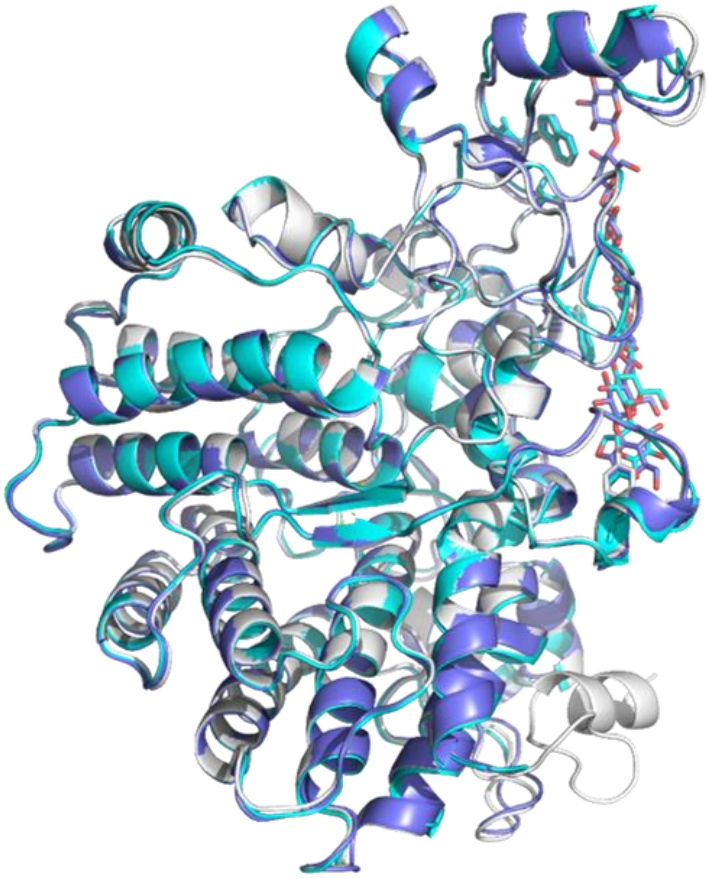

Figure B-8. Overlay of BoSGBP$_{MLG}$-A secondary structures.
Unliganded structure in white, cellohexaose complex in cyan, and MLG7 complex in slate show no major conformational difference from one another. Electron density for an extra 19 N-terminal residues were resolved only in the high resolution (1.50 Å, 6E60) unliganded structure, a portion of which folds into a short α-helix. Because this extra resolved density is involved in crystal contact with the binding platform of a symmetry-related molecule not present in either of the complex structures (6DMF and 6E61), includes a part of the non-native recombinant sequence, and is a part of the sequence that is believed to form a disordered linker between the globular fold of the protein and the phospholipid to which it is anchored, we interpret this extra ordered structure (in particular the α-helix) to be a crystal artifact.

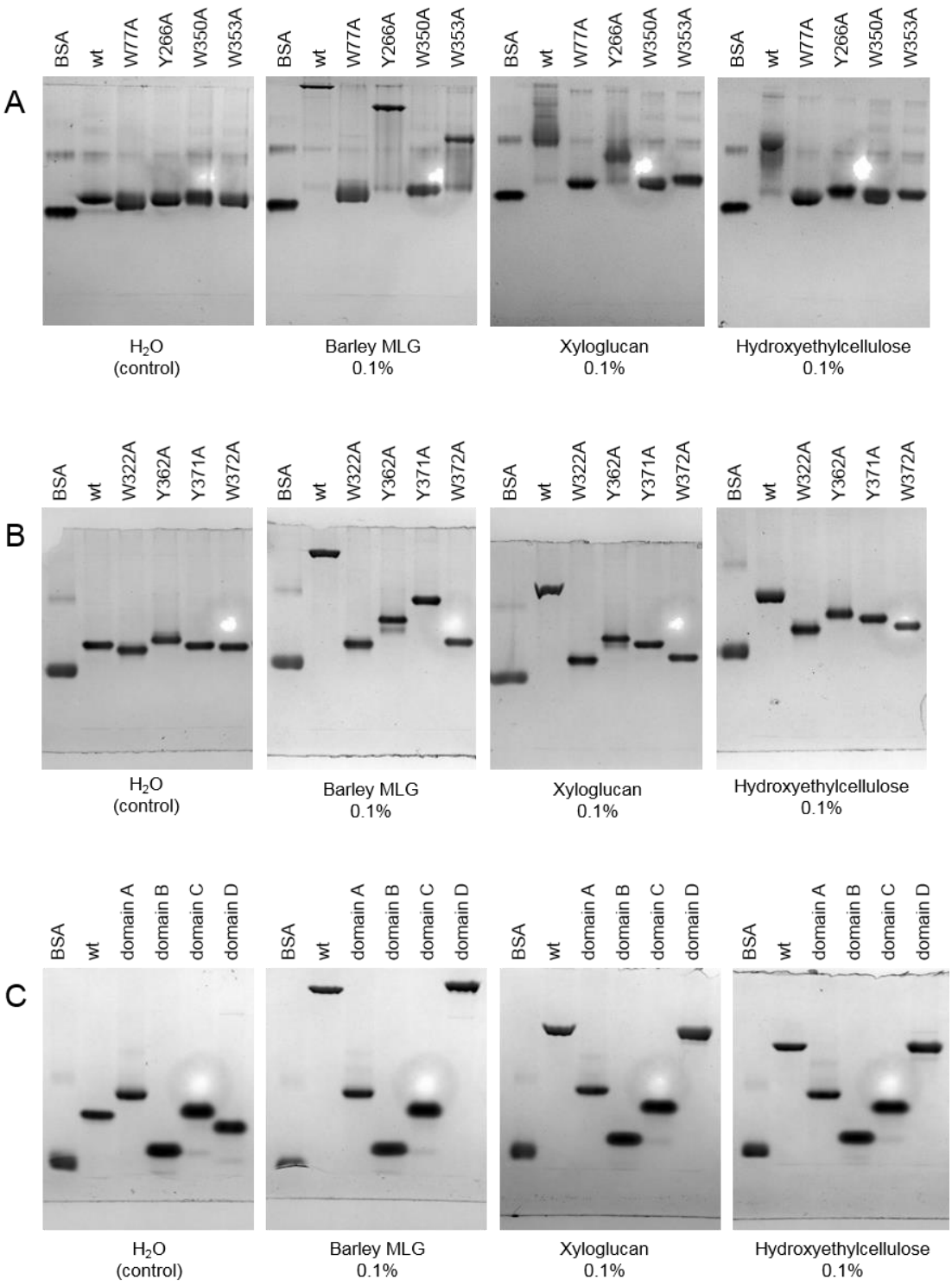

Figure B-9. Affinity electrophoresis gels of site directed mutants and domain dissections on various polysaccharides.

(A) $BoSGBP_{MLG}$-A binding platform site-directed mutants, (B) $BoSGBP_{MLG}$-B binding platform site-directed mutants, (C) $BoSGBP_{MLG}$-B domains. To facilitate comparison, the water and bMLG gels in panel A is identical to that shown in Fig. 3F, while those in panels B and C are identical to those shown in Fig. 4E.

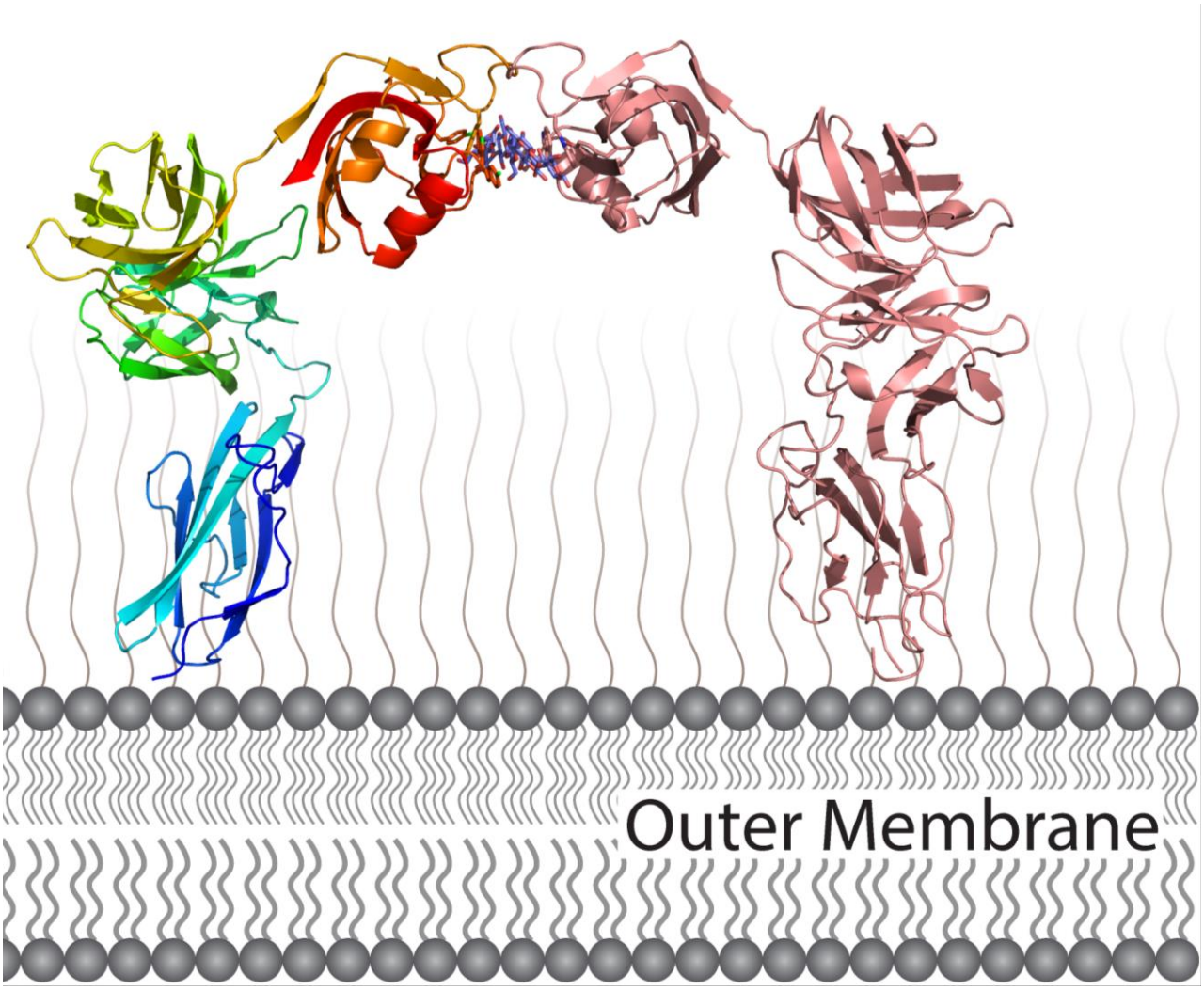

Figure B-10. Representation of a single MLG7 ligand bound by two molecules of BoSGBP$_{MLG}$-B at the cell surface, based on crystallographic data (PDB ID 6E9B).
The molecule on the left is color ramped from blue (N-terminus) to red (C-terminus), the symmetry related molecule on the right is colored salmon and the shared MLG7 ligand in the middle is colored slate. An analogous binding mode was observed for a cellohexaose complex (PDB ID 6E57).

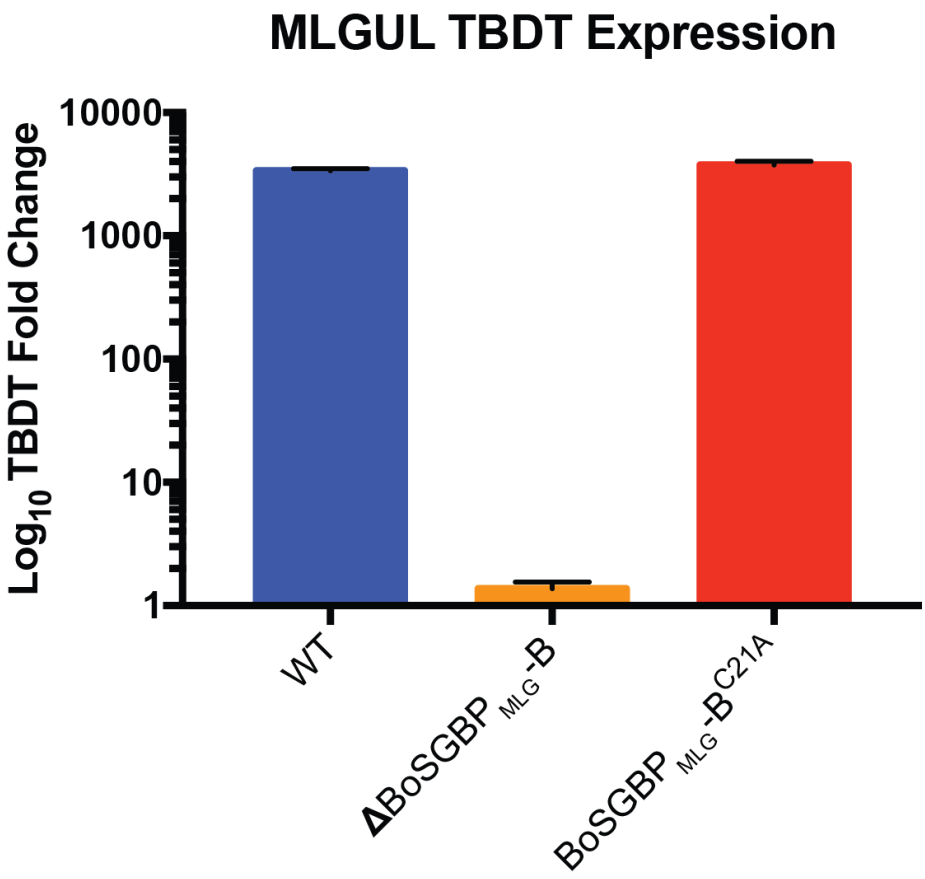

Figure B-11. Transcriptional analysis of MLGUL TBDT in *BoSGBP$_{MLG}$-B* mutants.
Fold change transcriptional response of the MLGUL TBDT from cells grown on MLG (5 mg/mL) vs. glucose (5 mg/mL).

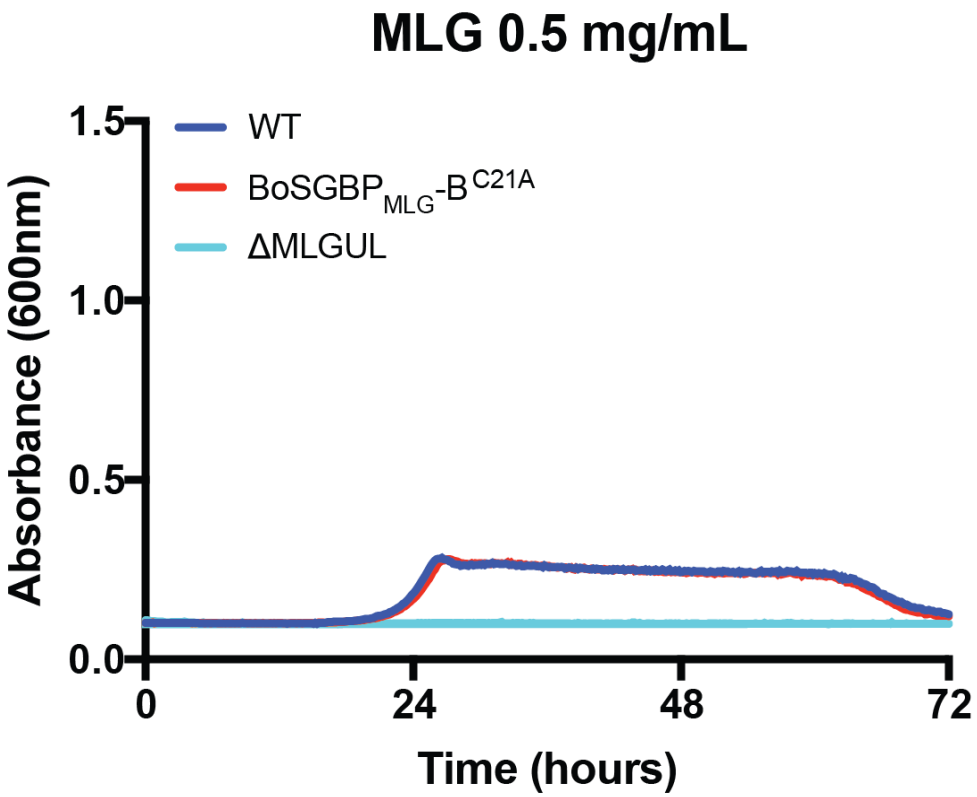

Figure B-12. Average growth of *B. ovatus* on 0.5 mg/mL MLG.
Limited (low maximum O.D.) or no growth of strains on 0.5 mg/mL MLG.

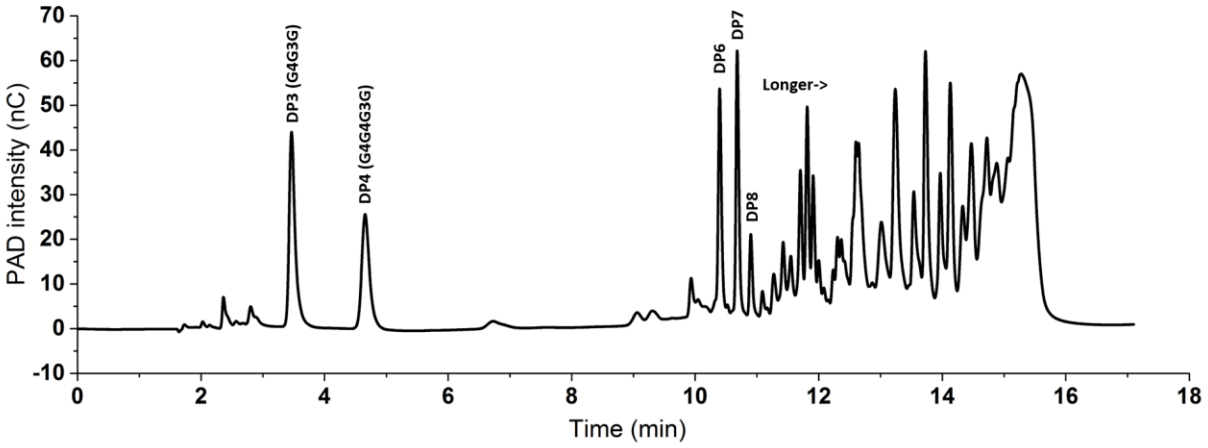

Figure B-13. MLG partial digest HPLC profile.
Controlled hydrolysis of MLG by recombinant $BoGH16_{MLG}$ produced a mixture containing predominantly oligosaccharides longer than the limit digest trisaccharide and tetrasaccharide.

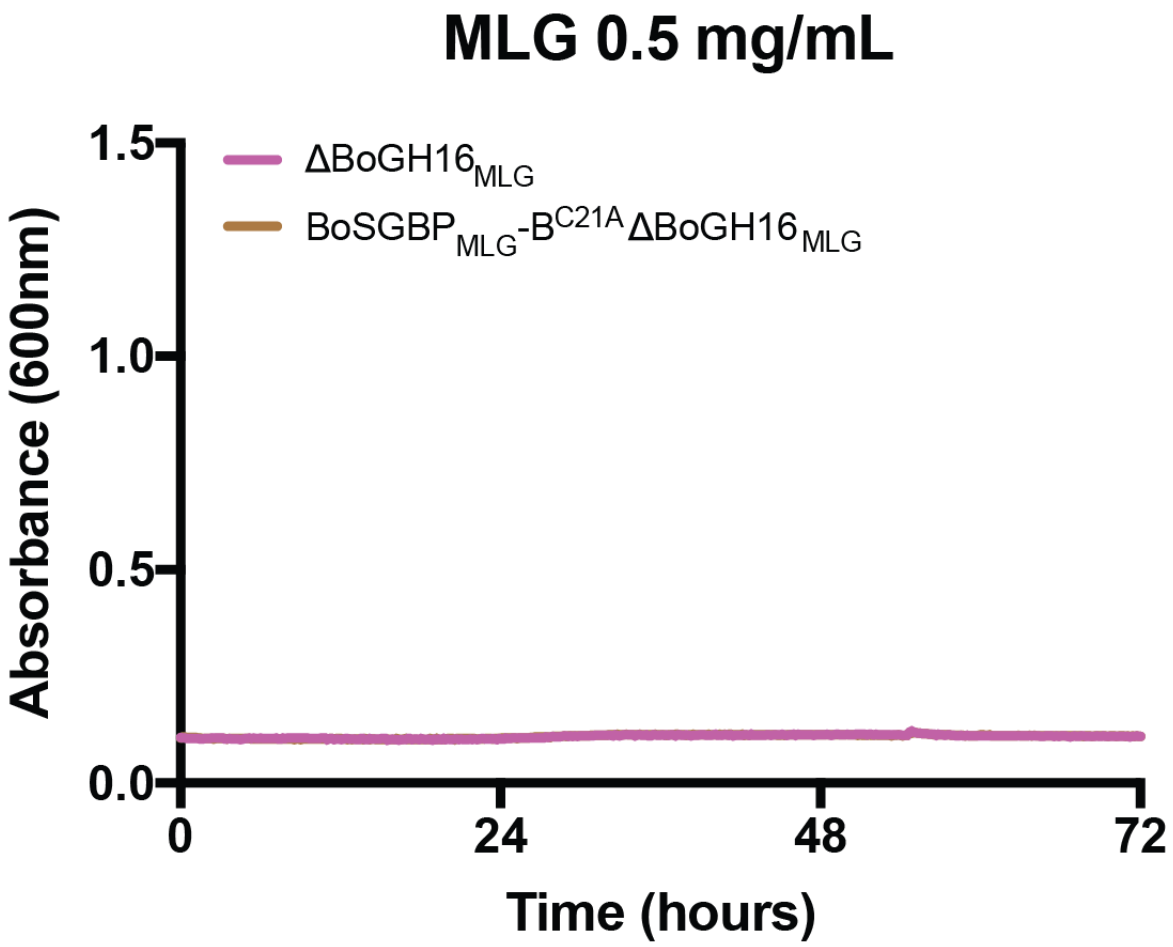

Figure B-14. Average growth of *B. ovatus* strains lacking vanguard BoGH16 on 5 mg/mL MLG. BoGH16 is required for growth on MLG.

Table B-1. Oligonucleotides used in this study.

Primer Name	Sequence (5'-3')	Use
SGBP-A F	GCGCGCCATATGGATGAATACATGGAAAAC	Cloning
SGBP-A R	GACGACCTCGAGTTAGTTTTCAGTATCCCA	Cloning
SGBP-A W77A F	CCGTAGTTGGTGGTGTTCGCACATCCCATCAGTTGTTG	Site-directed mutagenesis
SGBP-A W77A R	CAACAACTGATGGGATGTGCGAACACCACCAACTACGG	Site-directed mutagenesis
SGBP-A Y266A F	ATGAATTTCCGCGATAATCAGAGGCAGCTTCTTGTCCGAAACTAAATG	Site-directed mutagenesis
SGBP-A Y266A R	CATTTAGTTTCGGACAAGAAGCTGCCTCTGATTATCGCGGAAATTCAT	Site-directed mutagenesis
SGBP-A W350A F	GTAGGCCATGGTTCCGCCGAGTAGGCACCTGG	Site-directed mutagenesis
SGBP-A W350A R	CCAGGTGCCTACTCGGCGGAACCATGGCCTAC	Site-directed mutagenesis
SGBP-A W353A F	CCCTGTAGGCGCTGGTTCCCACGAGTAGGC	Site-directed mutagenesis
SGBP-A W353A R	GCCTACTCGTGGGAACCAGCGCCTACAGGG	Site-directed mutagenesis
SGBP-B F	GAAGAACATATGACCGAGGAAGAACCGTTC	Cloning
SGBP-B R	GACGACCTCGAGTTATTTCACGGTTACCAA	Cloning
SGBP-B W301A F	CACCGTCCGGCTTATCCGCCGAAACGTAATGGTGTC	Site-directed mutagenesis
SGBP-B W301A R	GACACCATTACGTTTCGGCGGATAAGCCGGACGGTG	Site-directed mutagenesis
SGBP-B Y341A F	CAAGCTGCATTTTATGAGCTTCGGCAGTAGGTTCTATGGAGTAG	Site-directed mutagenesis
SGBP-B Y341A R	CTACTCCATAGAACCTACTGCCGAAGCTCATAAAATGCAGCTTG	Site-directed mutagenesis
SGBP-B Y350A F	GCCAAGCCTGTCCAAGCGCCTGTGGCAAGCTG	Site-directed mutagenesis
SGBP-B Y350A R	CAGCTTGCCACAGGCGCTTGGACAGGCTTGGC	Site-directed mutagenesis
SGBP-B W351A F	CTTGCCAAGCCTGTCGCATAGCCTGTGGCAAGC	Site-directed mutagenesis
SGBP-B W351A R	GCTTGCCACAGGCTATGCGACAGGCTTGGCAAG	Site-directed mutagenesis
SGBP-B domainA R	TTATTACTCGACATCAGCCAACGGATTGAC	Cloning
SGBP-B domainB F	TACTGCCATATGGACCCTCAATCGAAAGAA	Cloning
SGBP-B domainB R	TAGTCGCTCGAGATTGGTTACTTTCACCAT	Cloning
SGBP-B domainC F	TAGTAGCATATGGCTTCCCTTGTCATTTCG	Cloning
SGBP-B domainC R	ATAATACTCGAGGGTAGATACGGTCACAGT	Cloning
SGBP-B domainD F	TCGACGCATATGGAAATAACGCTTTGGTCA	Cloning
GFP-fusion F	GTCAGCTAGCATGGTTAGCAAAGGTGAAGAA	Cloning
GFP-fusion R	GATGATGGATCCGCTGCCTTTATACAGTTCATC	Cloning

SGBP-A_GFP F	GACGACGGATCCATGGATGAATACATGGAAAAC	Cloning
SGBP-A_GFP R	GACGATCTCGAGTTAGTTTTCAGTATCCCACCA	Cloning
SGBP-B_GFP F	TATTATGGATCCATGACCGAGGAAGAACCGTTC	Cloning
SGBP-B_GFP R	GACGACCTCGAGTTATTTCACGGTTACCAAATC	Cloning
dMLGUL-A UpF	GAAGATAACATTCGAgtcgacGGTGGCCGCTAAGGTAGGCGAG	Deleting SGBP-A
dMLGUL-A UpR	GGATGATTGTTTAATAGGATATATCTTTTTAGAATTTCACATTTACATTAAAACCATAGCTTC	Deleting SGBP-A
dMLGUL-A DownF	GTGAAATTCTAAAAAGATATATCCTATTAAACAATCATCCATTATGAAAAAGATATATATTGCACTATTTGC	Deleting SGBP-A
dMLGUL-A DownR	GGCGGCCGCTCTAGAACTACATACGAGATTGATGCCGGAAATGGTCC	Deleting SGBP-A
dMLGUL-B UpF	GAAGATAACATTCGAgtcgacGGCTGATGAGTGCCACCAGTC	Deleting SGBP-B
dMLGUL-B UpR	GAATGATTTTTAAACTTACTTAGGAACAAATGGATGATTGTTTAATAGTTAGTTTTCAGTATCCC	Deleting SGBP-B
dMLGUL-B DownF	CTATTAAACAATCATCCATTTGTTCCTAAGTAAGTTTAAAAATCATTCATACCATAAGTATCTGTC	Deleting SGBP-B
dMLGUL-B DownR	GGCGGCCGCTCTAGAACTACTTGGAAGCCTCGGACAGCAG	Deleting SGBP-B
MLGUL-A* UpR	GGTATTCATGTTTTCCATGTATTCATCGCTGCAAGAAGCGAAAAACAGGGCAC	Replacing SGBP-A with SGBP-A*
MLGUL-A* F	AGCGATGAATACATGGAAAACATGAATACC	Replacing SGBP-A with SGBP-A*
MLGUL-A* R	TTAGTTTTCAGTATCCCACCACAGAAGGC	Replacing SGBP-A with SGBP-A*
MLGUL-A* DownF	GCCTTCTGTGGTGGGATACTGAAAACTAACTATTAAACAATCATCCATTATGAAAAAG	Replacing SGBP-A with SGBP-A*
dGH16 UpF	GAAAGAAGATAACATTCGAgtcgacGCTGCCGAACTGCTGAAAGAAGG	Deleting GH16
dGH16 UpR	GCATGTATATTTAAGATTCCGTTTTAGCATTTTAAAGGTTAAACGTAATATGTGC	Deleting GH16
dGH16 DownF2	CCTTTAAAATGCTAAAACGGAATCTTAAATATACATGCTATGAATATGGAAAAATGTAAGTATCTACTG	Deleting GH16
dGH16 DownR2	GGCGGCCGCTCTAGAACTACTGGCACGGCTACCATAAAGTGC	Deleting GH16
MLGUL-B C21A UpF	GAAAGAAGATAACATTCGAgtcgacGGCTGATGAGTGCCACCAGTC	Replacing SGBP-B with C21A allele
MLGUL-B C21A UpR	CTTCCTCGGTGCATGCAGTAAATGTGGAG	Replacing SGBP-B with C21A allele
MLGUL-B C21A DownF	CTCCACATTTACTGCATGCACCGAGGAAG	Replacing SGBP-B with C21A allele
MLGUL-B C21A DownR	GGCGGCCGCTCTAGAACTACTTGGAAGCCTCGGACAGCAG	Replacing SGBP-B with C21A allele
MLGUL TBDT qPCR F	CTATGTCTGCCCGTGCTGCTTAC	Probing TBDT transcription
MLGUL TBDT qPCR R	CCGGCTGCCAATCTTTCTTCT	Probing TBDT transcription
B ovatus 16s F	GGTAGTCCACACAGTAAACGATGAA	Normalizing transcription to 16s gene
B ovatus 16s F	CCCGTCAATTCCTTTGAGTTTC	Normalizing transcription to 16s gene

Table B-2. Privateer validation results

Residue Name	Conformation	Average B-factor	RSCC	Diagnostic
$BoSGBP_{MLG}$-A MLG7 complex				
BGC-1	$^4C_1$	50.39	0.81	Ok
BGC-2	$^4C_1$	35.58	0.9	Ok
BGC-3	$^4C_1$	25.89	0.91	Ok
BGC-4	$^4C_1$	23.09	0.92	Ok
BGC-5	$^4C_1$	23.09	0.91	Ok
BGC-6	$^4C_1$	25.22	0.91	Ok
BGC-7	$^4C_1$	29.61	0.84	Ok
BGC-8	$^4C_1$	46.71	0.78	Ok
BGC-9	$^4C_1$	34.56	0.86	Ok
BGC-10	$^4C_1$	25.81	0.9	Ok
BGC-11	$^4C_1$	24.13	0.89	Ok
BGC-12	$^4C_1$	25.42	0.89	Ok
BGC-13	$^4C_1$	25.73	0.9	Ok
BGC-14	$^4C_1$	28.2	0.83	Ok
$BoSGBP_{MLG}$-B MLG7 complex				
BGC-1	$^4C_1$	105.22	0.36	Ok
BGC-2	$^4C_1$	65.6	0.78	Ok
BGC-3	$^4C_1$	58.94	0.79	Ok
BGC-4	$^4C_1$	62.14	0.73	Ok
BGC-5	$^4C_1$	63.7	0.71	Ok
BGC-6	$^4C_1$	74.5	0.63	Ok
BGC-7	$^4C_1$	94.15	0.34	Ok
$BoSGBP_{MLG}$-A cellohexaose complex				
BGC-1	$^4C_1$	60.1742	0.85	Ok
BGC-2	$^4C_1$	50.3745	0.93	Ok
BGC-3	$^4C_1$	46.6764	0.89	Ok
BGC-4	$^4C_1$	37.11	0.93	Ok
BGC-5	$^4C_1$	43.7827	0.87	Ok
BGC-6	$^4C_1$	58.41	0.79	Ok
BGC-7	$^4C_1$	46.1691	0.91	Ok
BGC-8	$^4C_1$	37.44	0.9	Ok
BGC-9	$^4C_1$	33.8336	0.93	Ok
BGC-10	$^4C_1$	39.5718	0.94	Ok

BGC-11	$^4C_1$	47.8636	0.83	Ok
BGC-12	$^4C_1$	67.5033	0.82	Ok
BGC-13	$^4C_1$	51.0927	0.9	Ok
BGC-14	$^4C_1$	43.9373	0.92	Ok
BGC-15	$^4C_1$	39.58	0.93	Ok
BGC-16	$^4C_1$	42.3864	0.95	Ok
BGC-17	$^4C_1$	49.8864	0.9	Ok
BGC-18	$^4C_1$	62.215	0.86	Ok
BGC-19	$^4C_1$	56.0618	0.89	Ok
BGC-20	$^4C_1$	51.0864	0.91	Ok
BGC-21	$^4C_1$	52.7482	0.91	Ok
BGC-22	$^4C_1$	56.2327	0.91	Ok
BGC-23	$^4C_1$	59.2864	0.89	Ok
BGC-24	$^4C_1$	65.1833	0.78	Ok
BGC-25	$^4C_1$	52.58	0.88	Ok
BGC-26	$^4C_1$	42.8782	0.92	Ok
BGC-27	$^4C_1$	37.2736	0.93	Ok
BGC-28	$^4C_1$	41.7655	0.89	Ok
BGC-29	$^4C_1$	41.3182	0.9	Ok
BGC-30	$^4C_1$	68.6325	0.8	Ok
BGC-31	$^4C_1$	57.5191	0.87	Ok
BGC-32	$^4C_1$	63.8582	0.83	Ok
BGC-33	$^4C_1$	56.5745	0.89	Ok
BGC-34	$^4C_1$	62.7182	0.88	Ok
BGC-35	$^4C_1$	70.1945	0.86	Ok
BGC-36	$^4C_1$	59.9958	0.85	Ok
BGC-37	$^4C_1$	47.8464	0.92	Ok
BGC-38	$^4C_1$	44.1164	0.92	Ok
BGC-39	$^4C_1$	38.3709	0.91	Ok
BGC-40	$^4C_1$	41.2236	0.94	Ok
BGC-41	$^4C_1$	45.3373	0.85	Ok
BGC-42	$^4C_1$	62.9917	0.83	Ok
BGC-43	$^4C_1$	54.4318	0.83	Ok
BGC-44	$^4C_1$	50.1036	0.84	Ok
BGC-45	$^4C_1$	50.4491	0.91	Ok
BGC-46	$^4C_1$	50.4491	0.91	Ok

BGC-47	$^4C_1$	55.8391	0.87	Ok
BGC-48	$^4C_1$	63.435	0.8	Ok
BGC-49	$^4C_1$	47.2318	0.91	Ok
BGC-50	$^4C_1$	43.8173	0.9	Ok
BGC-51	$^4C_1$	39.7791	0.92	Ok
BGC-52	$^4C_1$	43.8336	0.91	Ok
BGC-53	$^4C_1$	45.1427	0.89	Ok
BGC-54	$^4C_1$	58.5008	0.8	Ok
BGC-55	$^4C_1$	48.9773	0.9	Ok
BGC-56	$^4C_1$	40.7064	0.9	Ok
BGC-57	$^4C_1$	35.0218	0.91	Ok
BGC-58	$^4C_1$	41.4518	0.92	Ok
BGC-59	$^4C_1$	52.3018	0.87	Ok
	$BoSGBP_{MLG}$-B cellohexaose complex			
BGC-1	$^4C_1$	47.1325	0.83	Ok
BGC-2	$^4C_1$	40.6309	0.88	Ok
BGC-3	$^4C_1$	40.0718	0.86	Ok
BGC-4	$^4C_1$	39.8464	0.86	Ok
BGC-5	$^4C_1$	46.6291	0.81	Ok
BGC-6	$^4C_1$	105.583	0.16	Ok
BGC-7	$^4C_1$	93.8809	0.54	Ok
BGC-8	$^4C_1$	86.8264	0.61	Ok
BGC-9	$^4C_1$	79.6155	0.73	Ok
BGC-10	$^4C_1$	80.81	0.67	Ok
BGC-11	$^4C_1$	110.487	0.25	Ok
BGC-12	$^4C_1$	93.9518	0.4	Ok
BGC-13	$^4C_1$	85.1655	0.55	Ok
BGC-14	$^4C_1$	77.7655	0.7	Ok
BGC-15	$^4C_1$	83.9127	0.62	Ok

Appendix C Supporting Information for Chapter 4

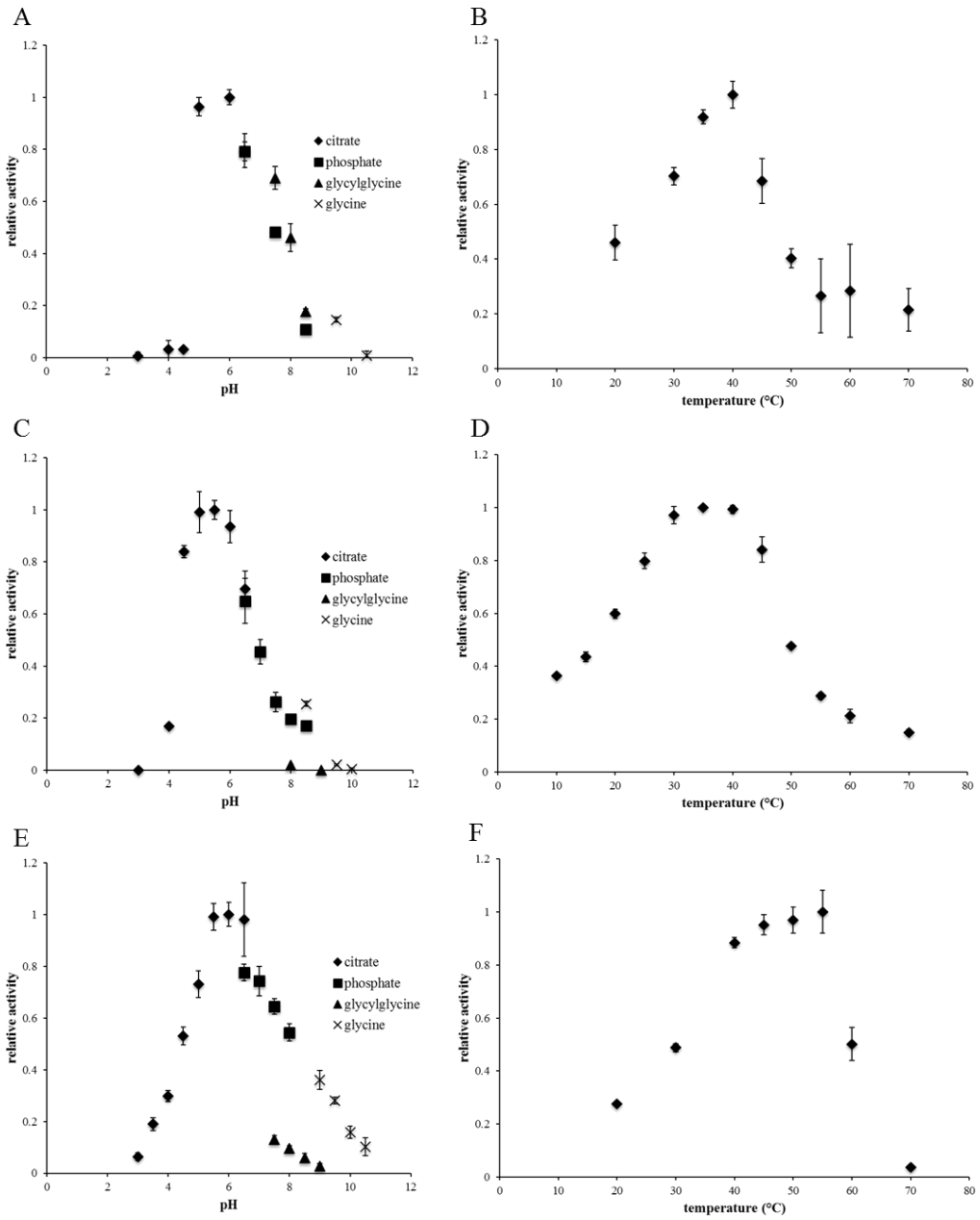

Figure C-1. Temperature and pH optima of recombinant *Bu* 1,3GUL GH proteins.
A. BuGH3 pH profile **B.** BuGH3 temperature profile **C.** BuGH158 pH profile **D.** BuGH158 temperature profile **E.** BuGH16 pH profile **F.** BuGH16 temperature profile for glucose-β-pNP (BuGH3) and laminarin (BuGH158 and BuGH16). Error bars represent standard errors of the mean for three replicates.

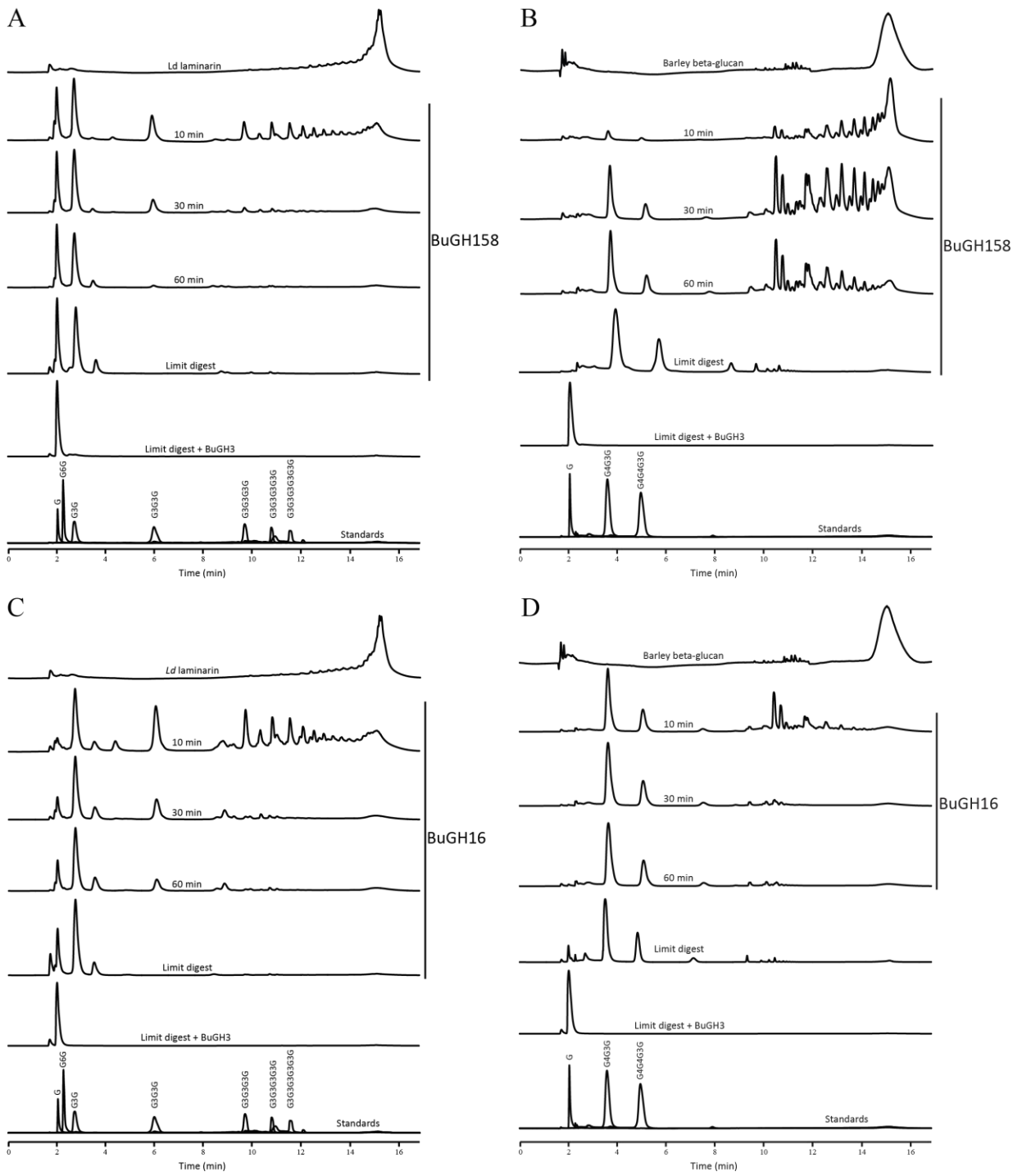

Figure C-2. *Bu* 1,3GUL GHs Product Analysis.
Chromatograms of reaction progress time course and limit digest of laminarin (A and C) and barley β-glucan (B and D) and its hydrolysis products by BuGH158 and BuGH3 (A and B), and by BuGH16 and BuGH3 (C and D) separated by HPAEC-PAD.

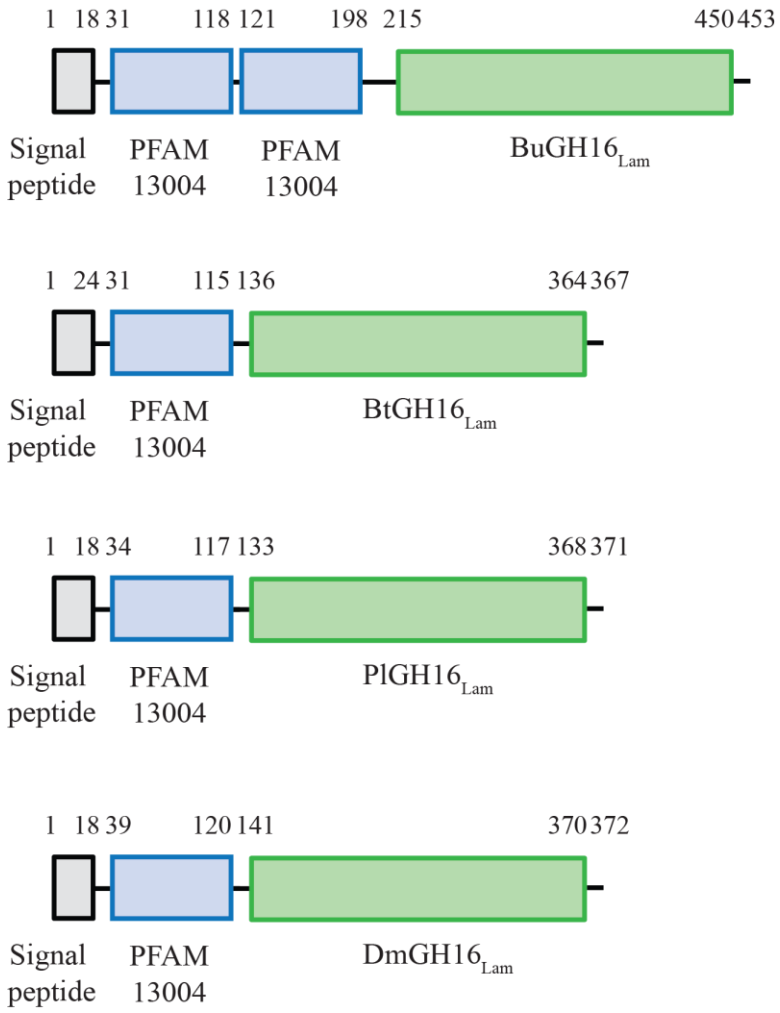

Figure C-3. Modular architecture of the GH16 gene products.

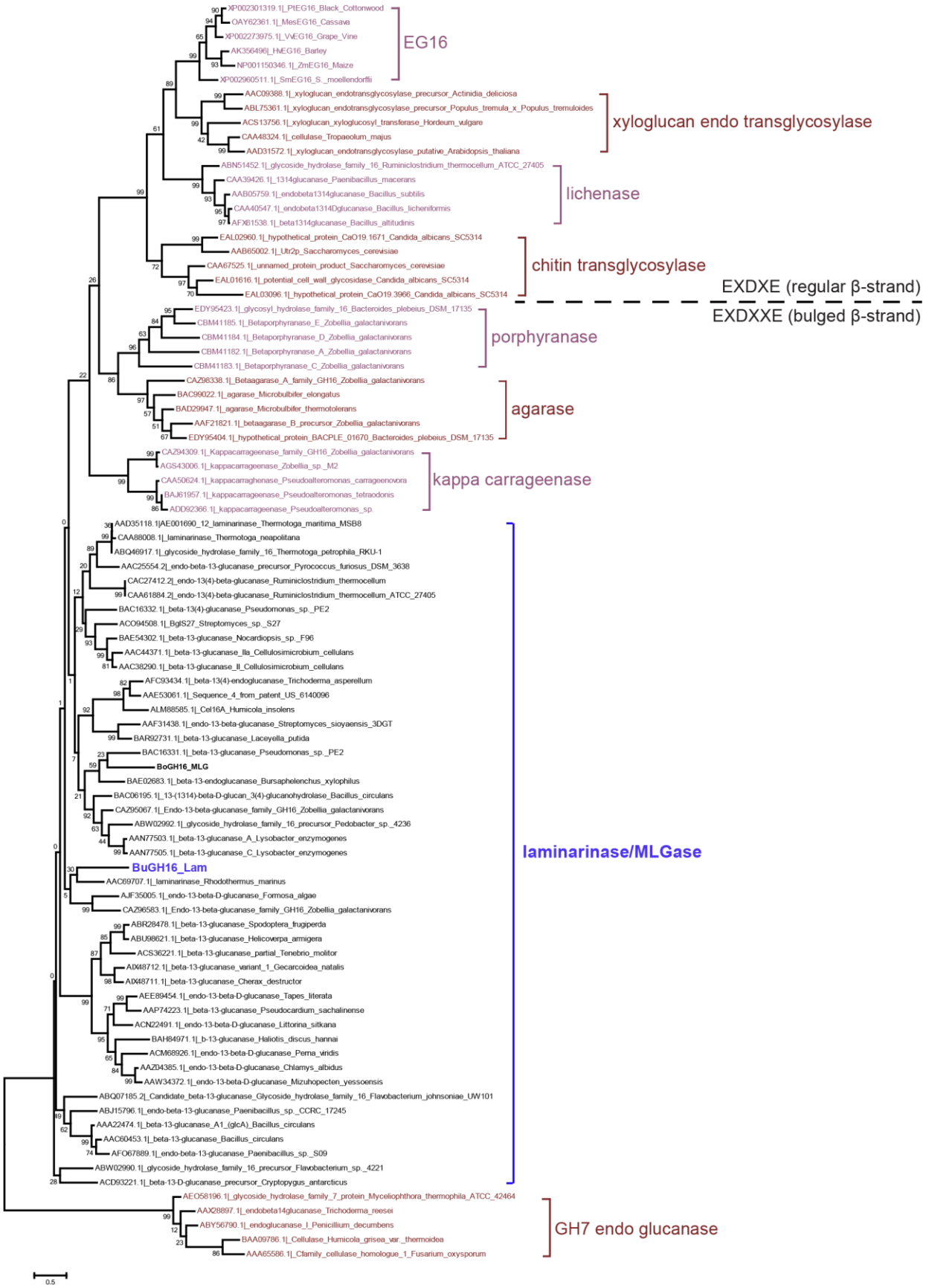
XP002301319.1|_PtEG16_Black_Cottonwood
OAY62361.1|_MesEG16_Cassava
XP002273975.1|_VvEG16_Grape_Vine
AK356496|_HvEG16_Barley
NP001150346.1|_ZmEG16_Maize
XP002960511.1|_SmEG16_S._moellendorffii
EG16
AAC09388.1|_xyloglucan_endotransglycosylase_precursor_Actinidia_deliciosa
ABL75361.1|_xyloglucan_endotransglycosylase_precursor_Populus_tremula_x_Populus_tremuloides
ACS13756.1|_xyloglucan_xyloglucosyl_transferase_Hordeum_vulgare
CAA48324.1|_cellulase_Tropaeolum_majus
AAD31572.1|_xyloglucan_endotransglycosylase_putative_Arabidopsis_thaliana
xyloglucan endo transglycosylase
ABN51452.1|_glycoside_hydrolase_family_16_Ruminiclostridium_thermocellum_ATCC_27405
CAA39426.1|_1314glucanase_Paenibacillus_macerans
AAB05759.1|_endobeta1314glucanase_Bacillus_subtilis
CAA40547.1|_endobeta1314Dglucanase_Bacillus_licheniformis
AFX61538.1|_beta1314glucanase_Bacillus_altitudinis
lichenase
EAL02960.1|_hypothetical_protein_CaO19.1671_Candida_albicans_SC5314
AAB65002.1|_Utr2p_Saccharomyces_cerevisiae
CAA67525.1|_unnamed_protein_product_Saccharomyces_cerevisiae
EAL01616.1|_potential_cell_wall_glycosidase_Candida_albicans_SC5314
EAL03096.1|_hypothetical_protein_CaO19.3966_Candida_albicans_SC5314
chitin transglycosylase
EXDXE (regular β-strand)
EXDXXE (bulged β-strand)
EDY95423.1|_glycosyl_hydrolase_family_16_Bacteroides_plebeius_DSM_17135
CBM41185.1|_Betaporphyranase_E_Zobellia_galactanivorans
CBM41184.1|_Betaporphyranase_D_Zobellia_galactanivorans
CBM41182.1|_Betaporphyranase_A_Zobellia_galactanivorans
CBM41183.1|_Betaporphyranase_C_Zobellia_galactanivorans
porphyranase
CAZ98338.1|_Betaagarase_A_family_GH16_Zobellia_galactanivorans
BAC99022.1|_agarase_Microbulbifer_elongatus
BAD29947.1|_agarase_Microbulbifer_thermotolerans
AAF21821.1|_betaagarase_B_precursor_Zobellia_galactanivorans
EDY95404.1|_hypothetical_protein_BACPLE_01670_Bacteroides_plebeius_DSM_17135
agarase
CAZ94309.1|_Kappacarrageenase_family_GH16_Zobellia_galactanivorans
AGS43006.1|_kappacarrageenase_Zobellia_sp._M2
CAA50624.1|_kappacarraghenase_Pseudoalteromonas_carrageenovora
BAJ61957.1|_kappacarrageenase_Pseudoalteromonas_tetraodonis
ADD92366.1|_kappacarrageenase_Pseudoalteromonas_sp.
kappa carrageenase
AAD35118.1|AE001690_12_laminarinase_Thermotoga_maritima_MSB8
CAA88008.1|_laminarinase_Thermotoga_neapolitana
ABQ46917.1|_glycoside_hydrolase_family_16_Thermotoga_petrophila_RKU-1
AAC25554.2|_endo-beta-13-glucanase_precursor_Pyrococcus_furiosus_DSM_3638
CAC27412.2|_endo-13(4)-beta-glucanase_Ruminiclostridium_thermocellum
CAA61884.2|_endo-13(4)-beta-glucanase_Ruminiclostridium_thermocellum_ATCC_27405
BAC16332.1|_beta-13(4)-glucanase_Pseudomonas_sp._PE2
ACO94508.1|_BglS27_Streptomyces_sp._S27
BAE54302.1|_beta-13-glucanase_Nocardiopsis_sp._F96
AAC44371.1|_beta-13-glucanase_IIa_Cellulosimicrobium_cellulans
AAC38290.1|_beta-13-glucanase_II_Cellulosimicrobium_cellulans
AFC93434.1|_beta-13(4)-endoglucanase_Trichoderma_asperellum
AAE53061.1|_Sequence_4_from_patent_US_6140096
ALM88585.1|_Cel16A_Humicola_insolens
AAF31438.1|_endo-13-beta-glucanase_Streptomyces_sioyaensis_3DGT
BAR92731.1|_beta-13-glucanase_Laceyella_putida
BAC16331.1|_beta-13-glucanase_Pseudomonas_sp._PE2
BoGH16_MLG
BAE02683.1|_beta-13-endoglucanase_Bursaphelenchus_xylophilus
BAC06195.1|_13-(1314)-beta-D-glucan_3(4)-glucanohydrolase_Bacillus_circulans
CAZ95067.1|_Endo-13-beta-glucanase_family_GH16_Zobellia_galactanivorans
ABW02992.1|_glycoside_hydrolase_family_16_precursor_Pedobacter_sp._4236
AAN77503.1|_beta-13-glucanase_A_Lysobacter_enzymogenes
AAN77505.1|_beta-13-glucanase_C_Lysobacter_enzymogenes
BuGH16_Lam
AAC69707.1|_laminarinase_Rhodothermus_marinus
AJF35005.1|_endo-13-beta-D-glucanase_Formosa_algae
CAZ96583.1|_Endo-13-beta-glucanase_family_GH16_Zobellia_galactanivorans
ABR28478.1|_beta-13-glucanase_Spodoptera_frugiperda
ABU98621.1|_beta-13-glucanase_Helicoverpa_armigera
ACS36221.1|_beta-13-glucanase_partial_Tenebrio_molitor
AIX48712.1|_beta-13-glucanase_variant_1_Gecarcoidea_natalis
AIX48711.1|_beta-13-glucanase_Cherax_destructor
AEE89454.1|_endo-13-beta-D-glucanase_Tapes_literata
AAP74223.1|_beta-13-glucanase_Pseudocardium_sachalinense
ACN22491.1|_endo-13-beta-D-glucanase_Littorina_sitkana
BAH84971.1|_b-13-glucanase_Haliotis_discus_hannai
ACM68926.1|_endo-13-beta-D-glucanase_Perna_viridis
AAZ04385.1|_endo-13-beta-D-glucanase_Chlamys_albidus
AAW34372.1|_endo-13-beta-D-glucanase_Mizuhopecten_yessoensis
ABQ07185.2|_Candidate_beta-13-glucanase_Glycoside_hydrolase_family_16_Flavobacterium_johnsoniae_UW101
ABJ15796.1|_endo-beta-13-glucanase_Paenibacillus_sp._CCRC_17245
AAA22474.1|_beta-13-glucanase_A1_(glcA)_Bacillus_circulans
AAC60453.1|_beta-13-glucanase_Bacillus_circulans
AFO67889.1|_endo-beta-13-glucanase_Paenibacillus_sp._S09
ABW02990.1|_glycoside_hydrolase_family_16_precursor_Flavobacterium_sp._4221
ACD93221.1|_beta-13-D-glucanase_precursor_Cryptopygus_antarcticus
laminarinase/MLGase
AEO58196.1|_glycoside_hydrolase_family_7_protein_Myceliophthora_thermophila_ATCC_42464
AAX28897.1|_endobeta14glucanase_Trichoderma_reesei
ABY56790.1|_endoglucanase_I_Penicillium_decumbens
BAA09786.1|_Cellulase_Humicola_grisea_var._thermoidea
AAA65586.1|_Cfamily_cellulase_homologue_1_Fusarium_oxysporum
GH7 endo glucanase
0.5

Figure C-4. **Phylogeny of GH16.**
Maximum Likelihood phylogenetic tree of characterized GH16 members, updated from the phylogeny of (Tamura, K., Hemsworth, G. R., Dejean, G., Rogers, T. E., Pudlo, N. A., Urs, K., Jain, N., Davies, G. J., Martens, E. C., and Brumer, H. 21: 417-430, 2017, doi:10.1093/molbev/mst197) Sequences were initially aligned with MUSCLE (Edgar, R. C. 32: 1792-1797, 2004, doi:10.1093/nar/gkh340) , manually trimmed to remove amino acids outside of the GH16 catalytic domain, and realigned with T-Coffee Expresso (Notredame, C., Higgins, D. G., and Heringa, J. 302: 205-217, 2000, doi:10.1006/jmbi.2000.4042) . The tree was estimated using MEGA6 (Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. 30: 2725-2729, 2013, doi:10.1093/molbev/mst197) and reliability of nodes was tested by bootstrap analysis using 100 resamplings. Five cellulases from GH7 were used as an outgroup to root the tree.

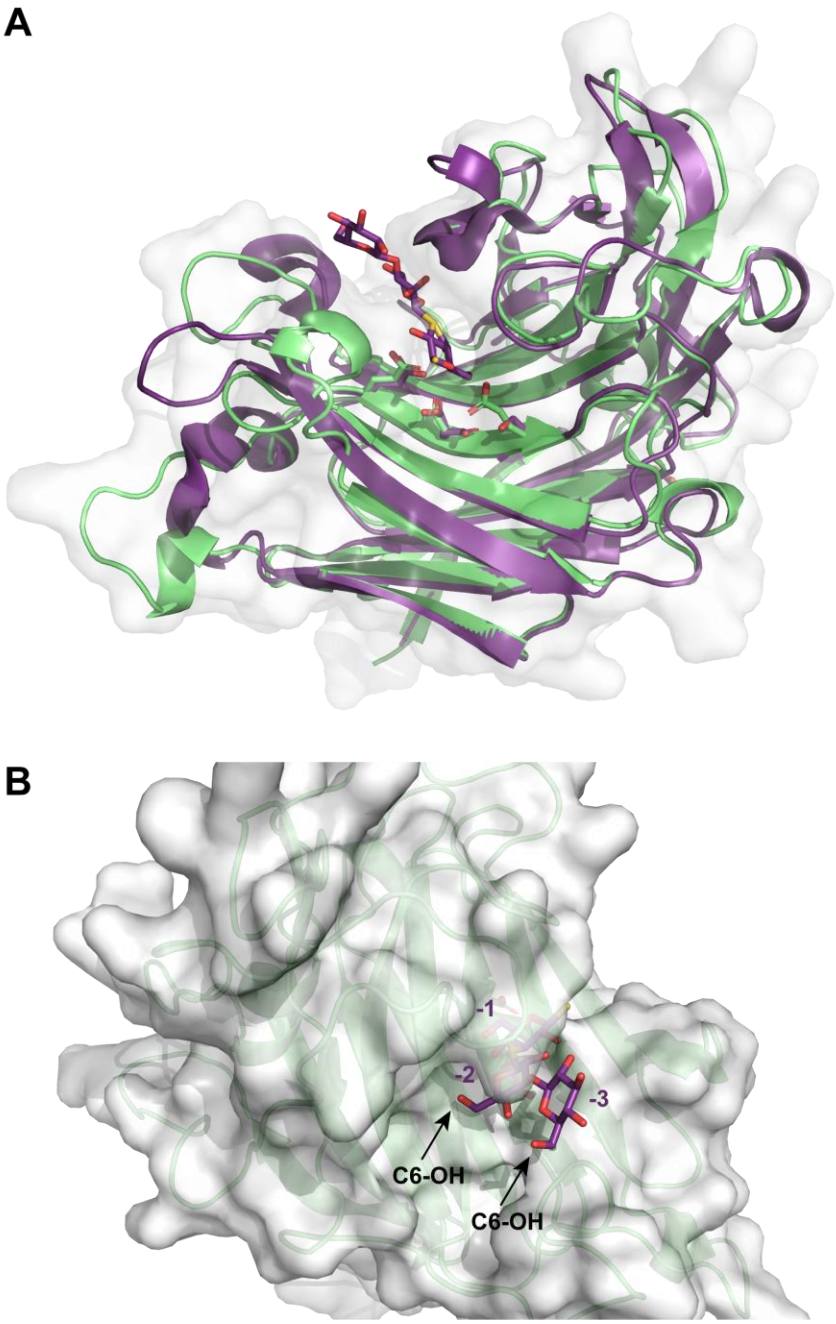

Figure C-5. Tertiary structure homology model of *Bu*GH16.
(A) Superposition of a *Bu*GH16 Phyre2 homology model (Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. 10: 845-858, 2015, doi:10.1038/nprot.2015.053) (teal) with a catalytically inactive, β(1,3)-thio-oligosaccharide ligand complex structure of the *Zobellia galactanivorans* GH16 laminarinase *Zg*LamC$_{GH16\text{-}E142S}$ (PDB ID 4CTE (Labourel, A., Jam, M., Legentil, L., Sylla, B., Hehemann, J. H., Ferrieres, V., Czjzek, M., and Michel, G. 71: 173-184, 2015, doi:10.1107/S139900471402450X) , purple). The conserved catalytic residues and ligand of *Zg*LamC$_{GH16\text{-}E142S}$ are shown as sticks. (B) Semi-transparent surface representation of the superposition, revealing an open space where a single or extended β(1,6)-glucosyl branch could be accommodated in the -2 subsite and beyond.

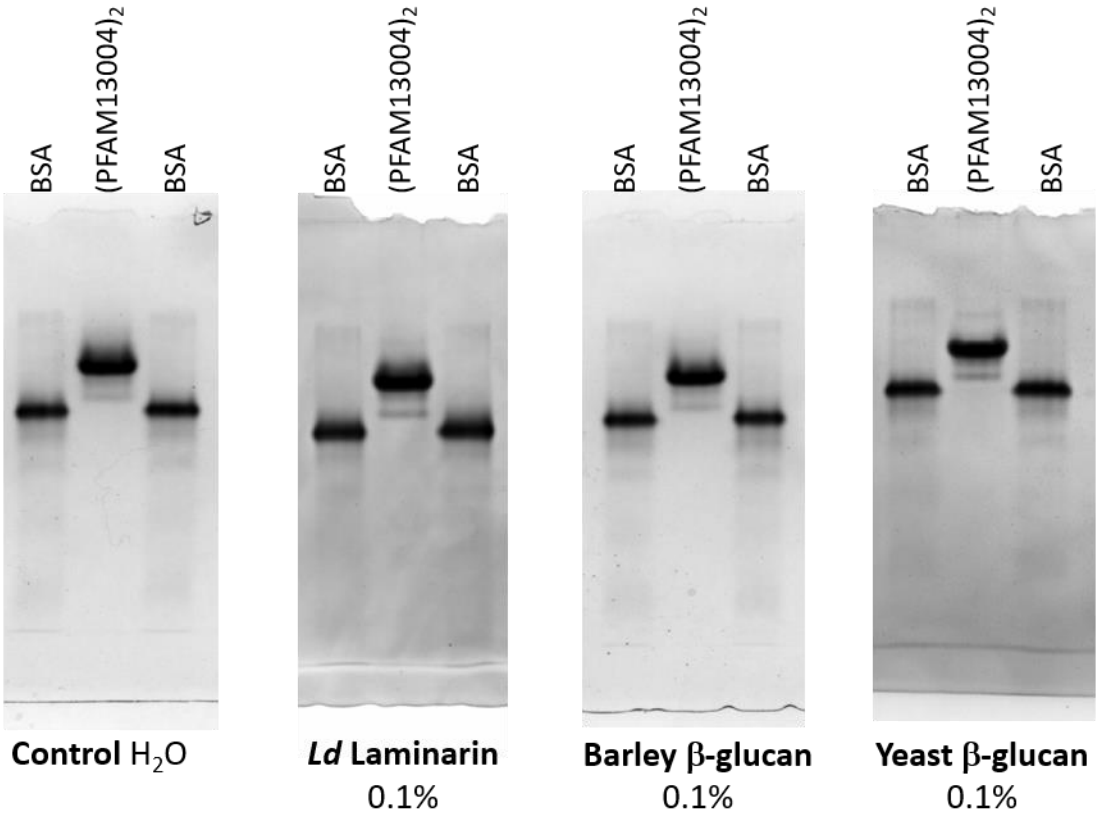

Figure C-6. Native polyacrylamide gel electrophoresis of the tandem PFAM 13004 domains from BuGH16 with β-glucans.
10 % polyacrylamide gels electrophoresis containing 0.1 % polysaccharide, with bovine serum albumin (BSA) as a control protein.

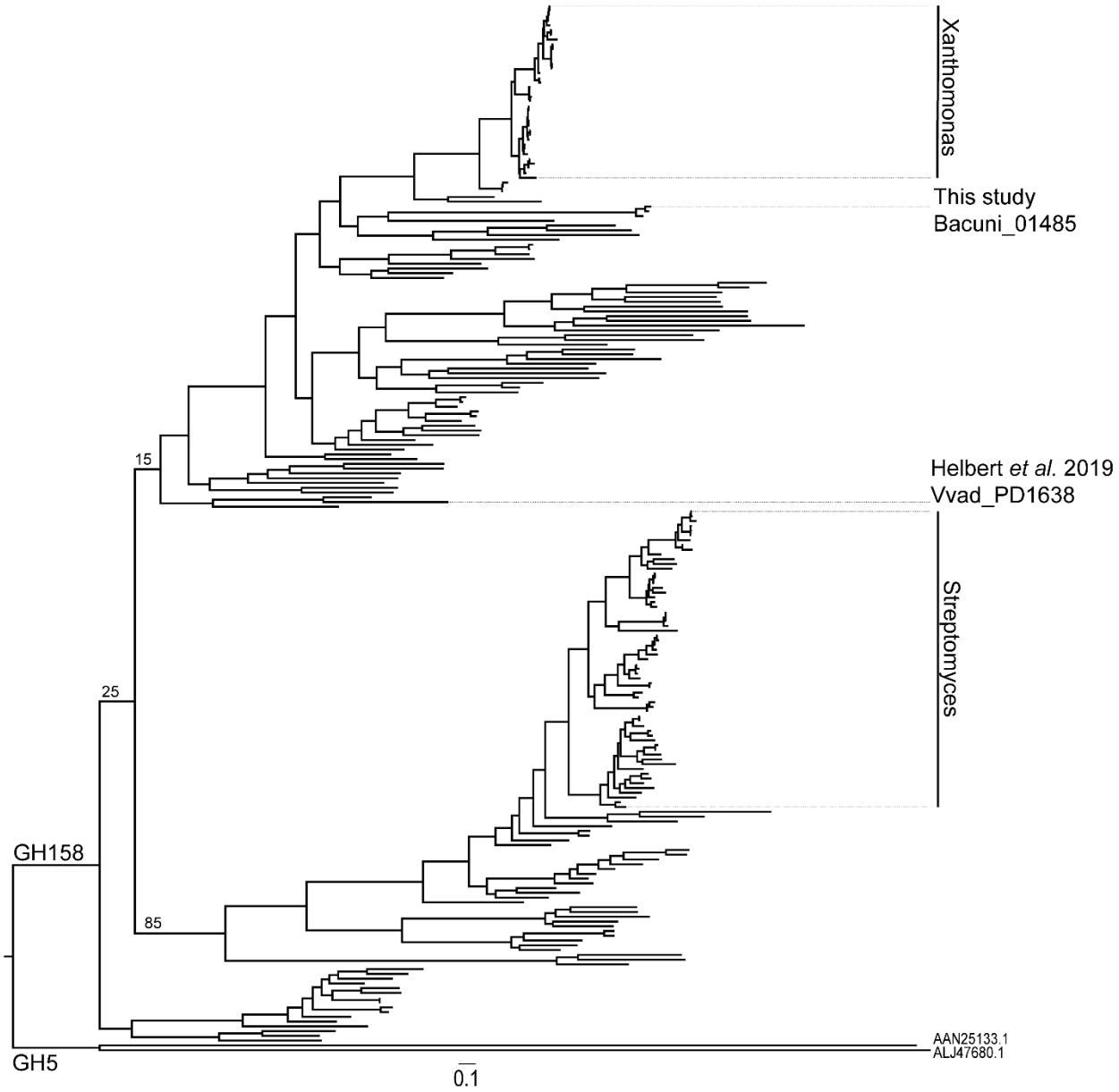

Figure C-7. Phylogeny of GH158.

308 GH158 sequences were retrieved from the CAZy database and redundant sequences were removed using UCLUST (Edgar, R. C. 26: 2460-2461, 2010, doi:10.1093/bioinformatics/btq461). The resulting 218 sequences were initially aligned with MAFFT and trimmed based on the identified boundaries of the GH158 catalytic domain. The trimmed sequences were subsequently realigned with MAFFT using the G-INS-I strategy (Katoh, K., and Standley, D. M. 30: 772-780, 2013, doi:10.1093/molbev/mst010). Two GH5 sequences (GenBank accessions AAN25133.1 and ALJ47680.1) were included to establish an outgroup. The phylogenetic tree was estimated using RAxML (Stamatakis, A. 30: 1312-1313, 2014, doi:10.1093/bioinformatics/btu033) based on 100 bootstrap replicates, and visualised with FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

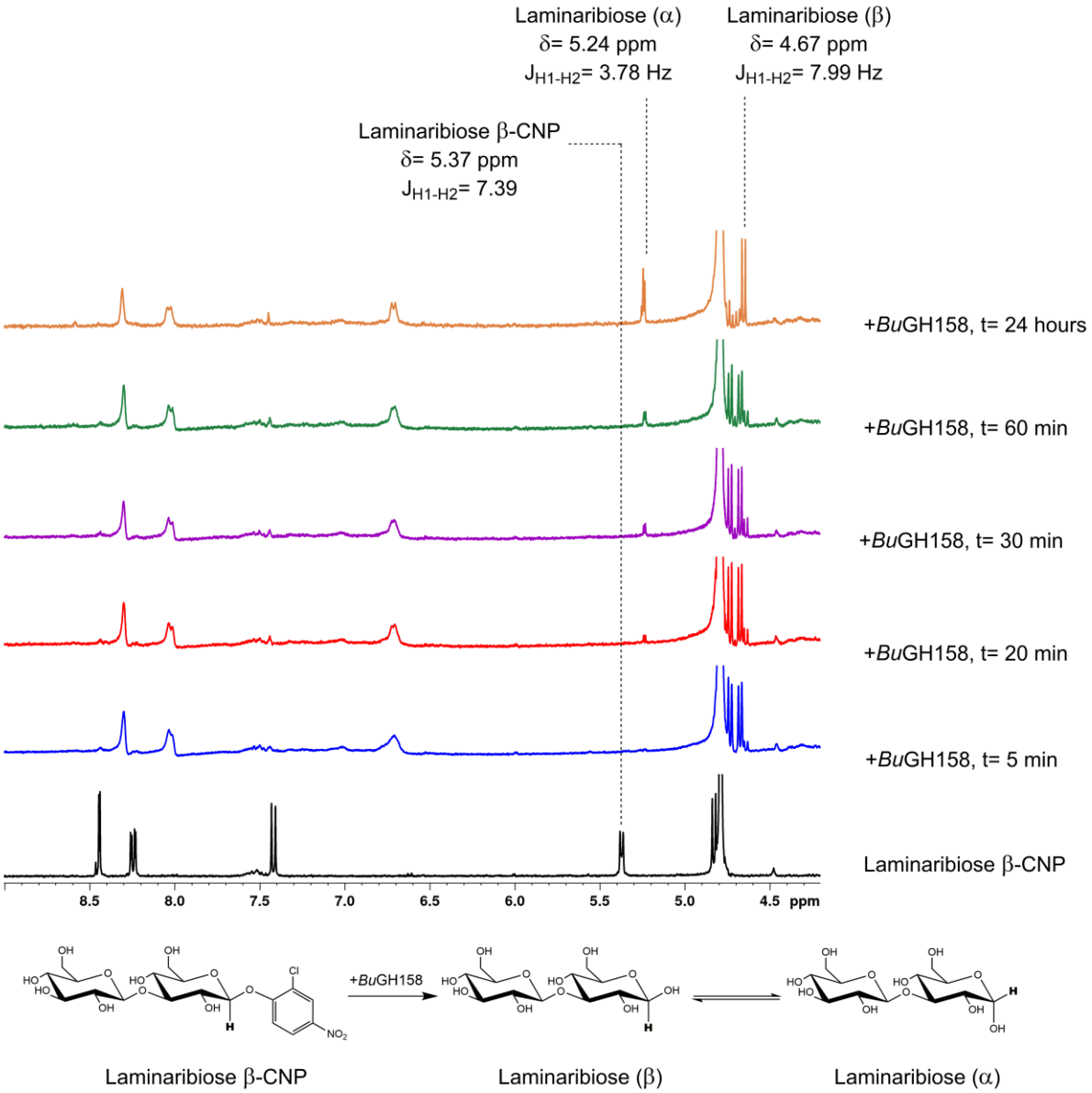

Figure C-8. ^{1}H-NMR analysis of BuGH158 mechanism of hydrolysis.
The change in anomeric configuration of the product of 10 mM laminaribiose-β-CNP hydrolysis by 20 μM BuGH158 was monitored over time. Spectra recorded at the indicated time points suggest an immediate release of β-laminaribiose by enzymatic action (within 5 minutes), and a gradual mutarotation of the product into a mixture of α and β anomers over the subsequent 24 hours. The off-chart peak at 4.8 ppm is HOD.

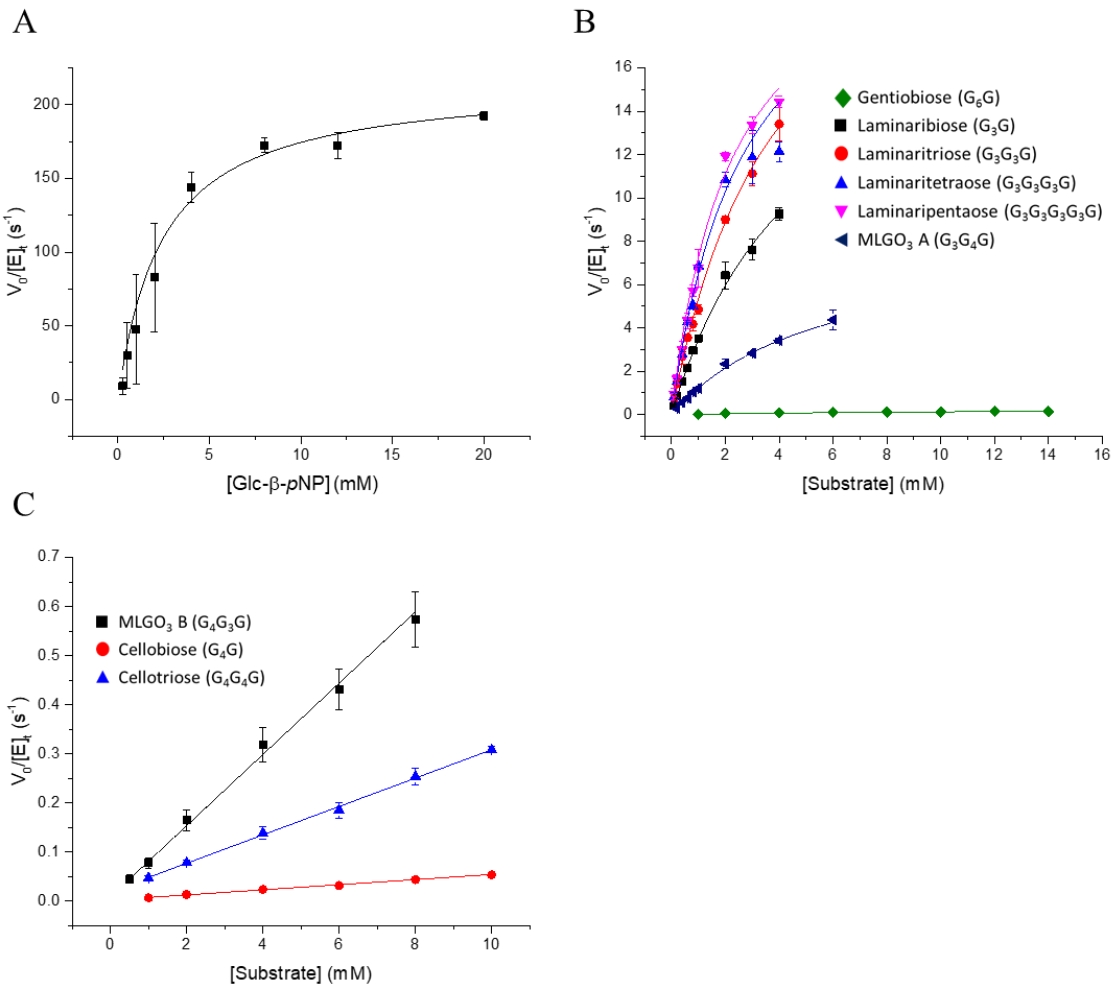

Figure C-9. Initial-rate kinetic analysis of BuGH3.
(A) BuGH3 against glucose-β-*p*NP (B) BuGH3 against oligosaccharides with a β(1,3) bond or a β(1,6) bond at the non-reducing end fitted to the Michealis-Menten equation (C) BuGH3 against oligosaccharides with a β(1,4) bond at the non-reducing end fitted to a linear equation. Curve fitting was done on OriginPro 2015 and error bars represent standard deviations from the mean.

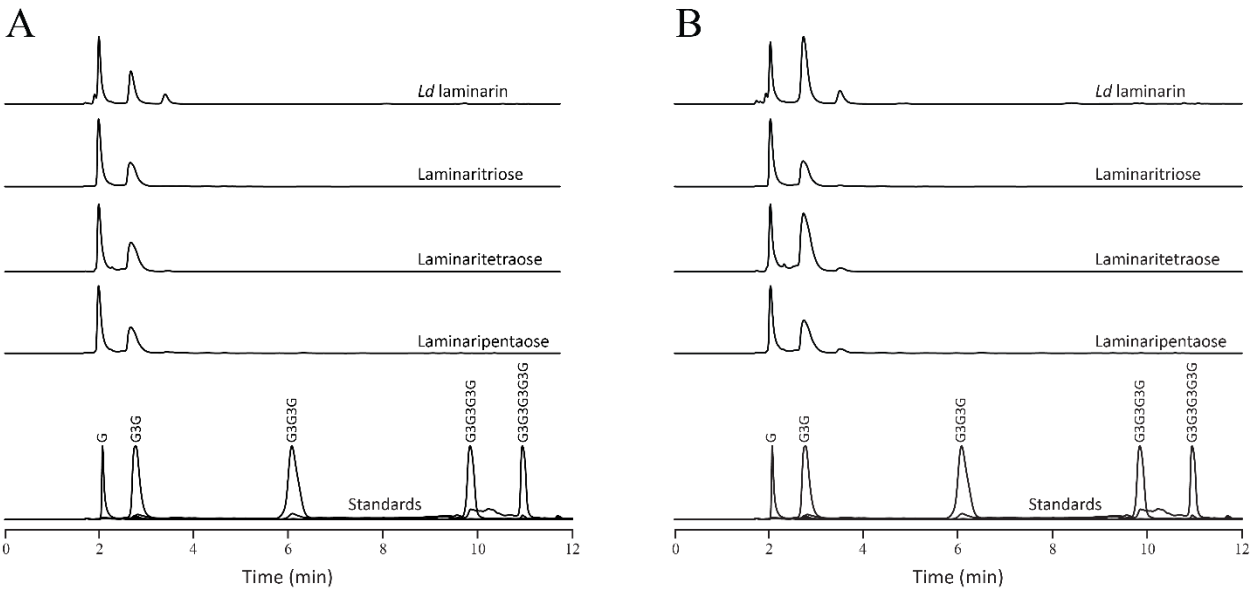

Figure C-10. *Bu* 1,3GUL GHs Product Analysis.
Chromatograms of limit digest products of laminarioligosaccharides by BuGH158 (A) and BuGH16 (B), separated by HPAEC-PAD.

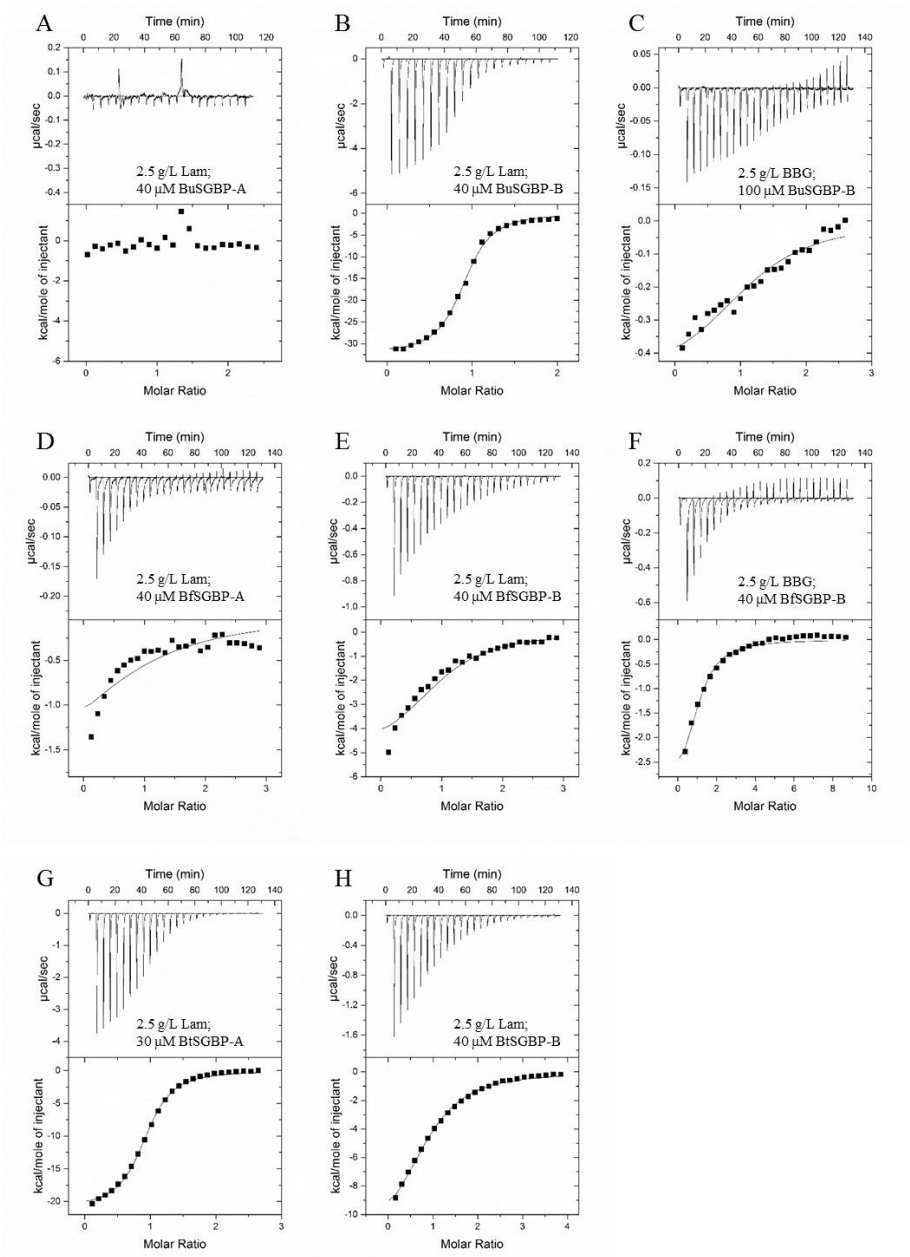

Figure C-11. Representative isothermal titration calorimetry (ITC) results for BuSGBP-A and -B, BfSGBP-A and -B, BtSGBP-A and -B titrations with Laminarin (Lam) and Barley β-glucan (BBG).
All titrations were performed in 50 mM Sodium Phosphate (pH 7.0) at 25 °C. In each case, the upper graph shows the raw injection heat signal, and the bottom graph shows the integrated data and, where appropriate, fits to a 1:1 binding model. Concentrations of the protein and glycan are indicated on the upper panel. (A) BuSGBP-A with Lam; (B) BuSGBP-B with Lam; (C) BuSGBP-B with BBG; (D) BfSGBP-A with Lam; (E) BfSGBP-B with Lam; (F) BfSGBP-B with BBG; (G) BtSGBP-A with Lam; (H) BtSGBP-B with Lam.

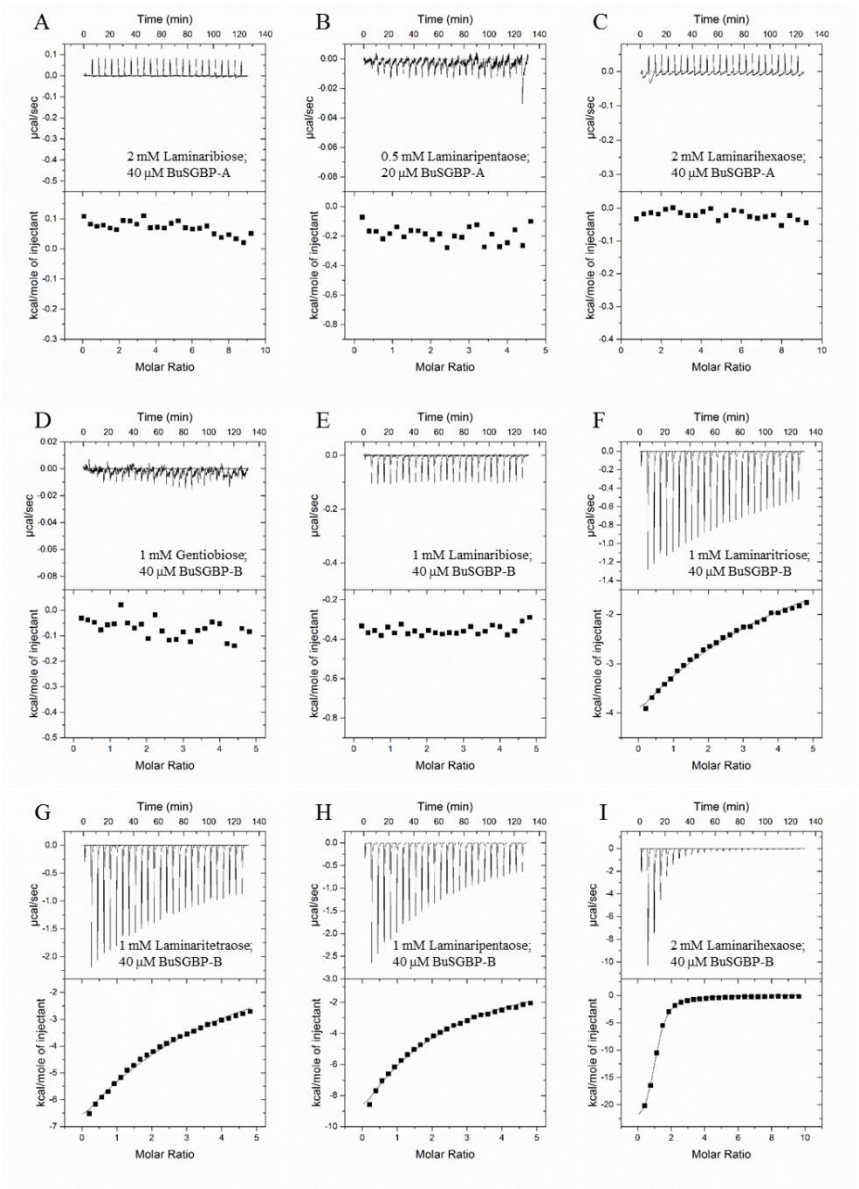

Figure C-12. Representative isothermal titration calorimetry (ITC) results for BuSGBP-A and –B with laminarioligosaccharides.

All titrations were performed in 50 mM Sodium Phosphate (pH 7.0) at 25 °C. In each case, the upper graph shows the raw injection heat signal, and the bottom graph shows the integrated data and, where appropriate, fits to a 1:1 binding model. Concentrations of the protein and glycan are indicated on the upper panel. (A) BuSGBP-A with laminaribiose; (B) BuSGBP-A with laminaripentaose; (C) BuSGBP-A with laminarihexaose; (D) BuSGBP-B with gentiobiose; (E) BuSGBP-B with laminaribiose; (F) BuSGBP-B with laminaritriose; (G) BuSGBP-B with laminaritetraose; (H) BuSGBP-B with laminaripentaose; (I) BuSGBP-B with laminarihexaose.

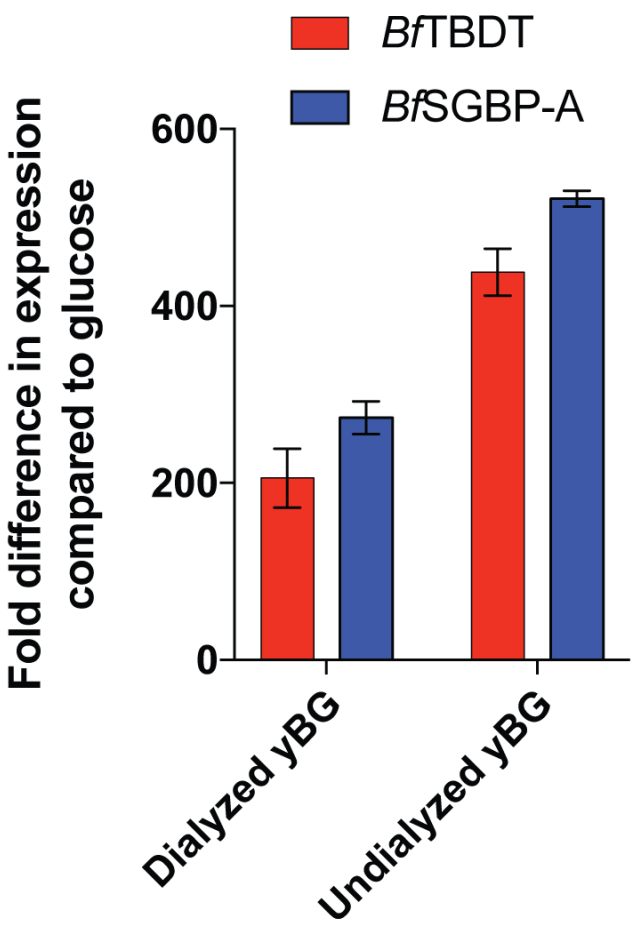

Figure C-13. Transcriptional response of *Bf*1,3GUL to commercial and dialyzed yBG.

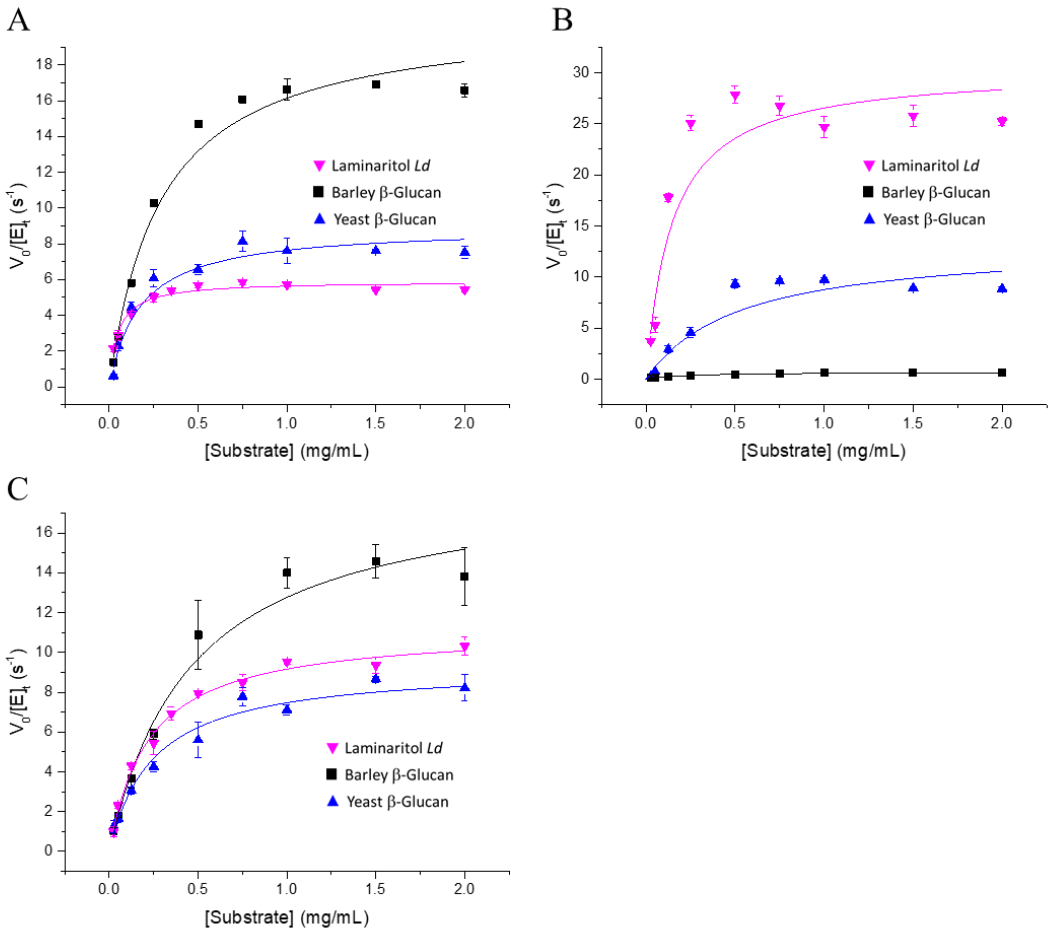

Figure C-14. **Initial-rate kinetic analysis of other 1,3GUL *endo*-glucanases.**
Michaelis-Menten curves for (A) BtGH16, (B) BfGH158, and (C) PlGH16 against laminaritol from *Laminaria digitata* (*Ld*), barley β-glucan, and yeast β-glucan.

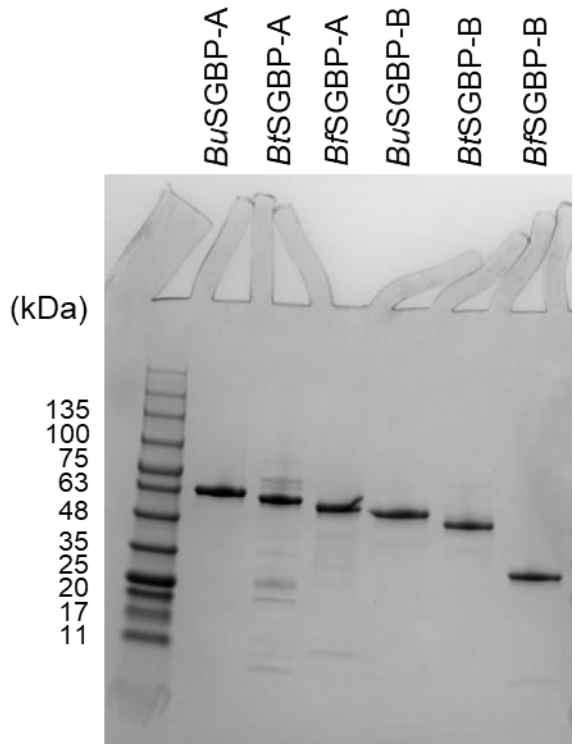

Figure C-15. SDS–PAGE of recombinant Surface Glycan-Binding Proteins (SGBP).
The calculated molecular weights of each protein are: *Bu*SGBP-A, 60615.2 Da; *Bt*SGBP-A, 59314.0 Da; *Bf*SGBP-A, 57639.30 Da; *Bu*SGBP-B, 51352.61 Da; *Bt*SGBP-B, 50274.85 Da; *Bf*SGBP-B, 31227.45 Da.

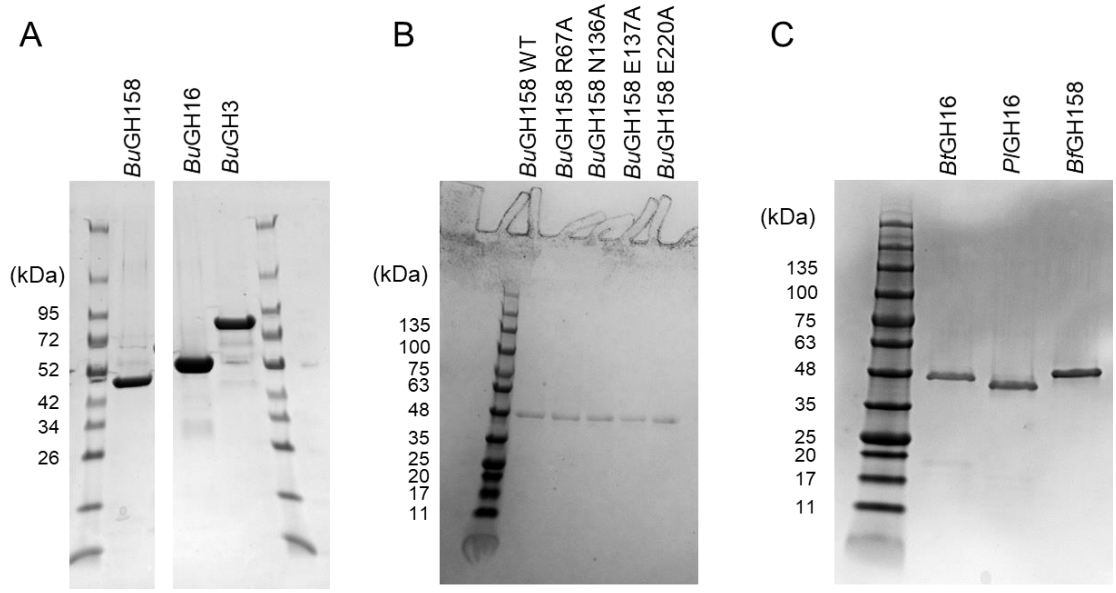

Figure C-16. SDS–PAGE of recombinant Glycoside Hydrolase (GH) proteins.
The calculated molecular weights of each protein are: (A) *Bu*GH158, 48377.5 Da; *Bu*GH16, 50081.62 Da; *Bu*GH3, 81645.3 Da. (B) *Bu*GH158 R67A, 48292.43 Da; *Bu*GH158 N136A, 48334.52 Da; *Bu*GH158 E137A, 48319.51 Da; *Bu*GH158 E220A, 48319.51 Da. (C) *Bt*GH16, 41746.21 Da; *Pl*GH16, 41355.3 Da; *Bf*GH158, 46682.6 Da.

Table C-1. Growth of *B. uniformis* (*Bu*), *B. thetaiotaomicron* (*Bt*), and *B. fluxus* (*Bf*) on β-glucans and glucose

Substrate	*Bu* Rate (OD_{600}/h)	*Bt* Rate (OD_{600}/h)	*Bf* Rate (OD_{600}/h)
***Ld* laminarin**	0.034 ± 0.003	0.061 ± 0.001	0.132 ± 0.016
Barley β-glucan	0.02 ± 0.001	no growth	no growth
Yeast β-glucan	0.02 ± 0.002	0.027 ± 0.003	no growth
Curdlan	no growth	no growth	no growth
Glucose	0.044 ± 0.0003	0.12 ± 0.001	0.069 ± 0.002

Table C-2. Kinetic parameters for the hydrolysis of various substrates by 1,3GUL GH16 and GH158 members.

Bacteria strain	Enzyme	Substrate[a]	k_{cat} (s^{-1})	K_m (mg mL^{-1})	k_{cat}/K_m (s^{-1} mg^{-1} mL)
B. uniformis ATCC 8492	*Bu*GH16	*Ld* laminarin	13.57 ± 0.33	0.05 ± 0.01	270
		Eb laminarin	22.90 ± 1.50	0.14 ± 0.01	170
		Ld laminaritol	17.80 ± 1.36	0.06 ± 0.01	310
		yBG	49.37 ± 0.58	0.18 ± 0.01	270
		bMLG	12.37 ± 0.26	0.16 ± 0.02	78
		Curdlan	1.54 ± 0.27	0.39 ± 0.12	3.9
	*Bu*GH158	*Ld* laminarin	45.00 ± 1.58	0.13 ± 0.02	350
		Eb laminarin	6.18 ± 1.11	0.78 ± 0.2	8.0
		Ld laminaritol	56.13 ± 4.88	0.28 ± 0.04	200
		yBG	12.84 ± 1.67	1.48 ± 0.3	8.7
		bMLG	1.72 ± 0.07	0.69 ± 0.03	2.5
		Curdlan	1.53 ± 0.18	0.60 ± 0.18	2.6
B. thetaiotaomicron NLAE-zl-H207	*Bt*GH16	*Ld* laminaritol	5.90 ± 0.13	0.05 ± 0.01	120
		yBG	8.96 ± 0.69	0.17 ± 0.04	52
		bMLG	20.83 ± 1.06	0.29 ± 0.04	72
B. fluxus YIT 12057	*Bf*GH158	*Ld* laminaritol	30.38 ± 2.79	0.14 ± 0.04	210
		yBG	13.28 ± 2.33	0.52 ± 0.2	26
		bMLG	0.68 ± 0.06	0.22 ± 0.06	3.1
P. loeschii DSM19665	*Pl*GH16	*Ld* laminaritol	11.19 ± 0.34	0.22 ± 0.02	51
		yBG	9.32 ± 0.57	0.25 ± 0.04	38
		bMLG	18.73 ± 1.33	0.47 ± 0.08	40

[a] No activity was observed on tamarind xyloglucan, carob galactomannan, konjac glucomannan, carboxymethylcellulose, hydroxyethylcellulose, and *Xanthomonas campestris* xanthan gum.

Table C-3. Data collection and refinement statistics

	BuGH158-native	BuGH158-SeMet peak	BuGH158-SeMet inflection	BuGH158-SeMet high energy remote
Data Collection				
Beamline	APS 23ID-D	SSRL BL9-2	SSRL BL9-2	SSRL BL9-2
Wavelength (Å)	1.03319	0.97896	0.97922	0.91837
Space Group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions				
a, b, c (Å)	65.7, 106.0, 126.9	66.2, 106.3, 126.5	66.2, 106.4, 126.6	66.3, 106.6, 126.4
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
No. of reflections				
total	1046176 (39102)	697828 (55065)	472562 (49787)	451252 (29850)
unique	80471 (3966)	52896 (4286)	35781 (3686)	35803 (3675)
Resolution (Å)	40.69 – 1.82 (1.85 – 1.82)	39.37 - 2.10 (2.16 - 2.10)	39.42 - 2.40 (2.49 - 2.40)	39.46 - 2.40 (2.49 - 2.40)
R_{merge}	0.113 (1.479)	0.090 (1.373)	0.070 (0.617)	0.107 (2.009)
R_{meas}	0.117 (1.560)	0.098 (1.490)	0.075 (0.665)	0.116 (2.303)
$CC_{1/2}$	0.999 (0.690)	0.999 (0.753)	0.999 (0.921)	0.999 (0.360)
I/σI	14.8 (1.4)	18.0 (2.0)	24.6 (4.4)	16.2 (1.2)
Completeness (%)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)	99.9 (99.7)
Multiplicity	13.0 (9.9)	13.2 (12.8)	13.2 (13.5)	12.6 (8.1)
Anomalous completeness (%)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)	99.9 (99.6)
Anomalous multiplicity	6.8 (5.0)	6.9 (6.6)	7.0 (7.0)	6.7 (4.2)
Refinement				
Resolution (Å)	40.69 – 1.82			
R_{work}	0.168			
R_{free}	0.204			
No. of atoms				
protein	6440			
ligand	23			
water	638			
Avg B-factor (Å^2)				
protein	29.4			
ligand	50.8			
water	37.4			
RMS deviations				
bond length (Å)	0.011			
bond angle (°)	1.61			
Ramachandran statistics				
Favored (%)	96.8			
Outliers (%)	0			

Table C-4. Top 20 Dali search results with BuGH158.

rank	PDB ID	Z score	RMSD	ID (%)	GH family	GH clan	Activity
1	4AWE	22	3.2	17	5	A	endo beta-1,4 mannanase
2	5Z1B	21.3	2.9	17	2	A	beta-glucuronidase
3	4LYP	21.2	3.3	16	5	A	exo beta-1,4 mannosidase
4	6MVH	21.1	3.6	14	2	A	beta-galactosidase
5	6MVG	20.9	4	16	2	A	beta-glucuronidase
6	4D1I	20.8	2.6	14	35	A	beta-galactosidase
7	5DMY	20.7	4.1	14	2	A	beta-galactosidase
8	2W61	20.6	3.3	15	72	A	beta-1,3 transglucosidase
9	6BO6	20.4	3.2	17	2	A	beta-glucuronidase
10	4OOZ	20.3	2.5	14	5	A	beta-1,4 mannanase
11	6MVF	20.2	3.3	18	2	A	beta-galactosidase/ beta-glucuronidase
12	6MPA	19.8	2.8	13	5	A	beta mannosidase
13	1R8L	19.8	3	14	53	A	endo beta-1,4 galactanase
14	5GSM	19.6	4.3	15	42	A	beta-galactosidase
15	5JVK	19.2	3.6	13	39	A	unknown
16	6EON	18.5	7.1	11	35	A	endo galactanase
17	5A8M	18.5	3.2	14	5	A	beta glucanase
18	5XB7	18.3	3.2	14	42	A	alpha-L-arabinopyranosidase
19	1G0C	18.3	3.6	15	1	A	cellulase
20	2ZUN	18.3	3.4	13	1	A	cellulase

Table C-5. Kinetic parameters for the hydrolysis of various substrates by BuGH3[a].

Enzyme	Substrate[b]	k_{cat} (s^{-1})	K_m (mM^{-1})	k_{cat}/K_m (s^{-1} mM^{-1})	Assay
BuGH3	β-Glc-*p*NP	216.47 ± 6.24	2.41 ± 0.45	89.65	*p*NP
	Gentiobiose (G6G)	0.30 ± 0.07	13.31 ± 4.29	0.02	HK/G6PDH[e]
	Laminaribiose (G3G)	21.16 ± 0.92	5.09 ± 0.26	4.15	HK/G6PDH
	Laminaritriose (G3G3G)	26.24 ± 2.51	3.90 ± 0.52	6.73	HK/G6PDH
	Laminaritetraose (G3G3G3G)	24.03 ± 4.84	2.67 ± 0.72	9.01	HK/G6PDH
	Laminaripentaose (G3G3G3G3G)	23.44 ± 2.15	2.22 ± 0.41	10.55	HK/G6PDH
	$MLGO_3$ A (G3G4G)	8.34 ± 0.37	5.76 ± 0.36	1.45	HK/G6PDH
	$MLGO_3$ B (G4G3G)	ND[c]	ND	0.07[d]	HK/G6PDH
	Cellobiose (G4G)	ND	ND	0.01	HK/G6PDH
	Cellotriose (G4G4G)	ND	ND	0.03	HK/G6PDH

[a] See also Supplemental Fig. 9
[b] No activity was observed on α-Glc-*p*NP, β-Gal-*p*NP, β-Man-*p*NP, and β-Xyl-*p*NP
[c] ND, not determined due to insufficient activity
[d] In cases where substrate saturation was not achieved, k_{cat}/K_m values were determined from the slope of linear fits to v vs. [S] plots.
[e] Hexokinase and glucose-6-phosphate dehydrogenase coupled enzyme assay.

Table C-6. Thermodynamic parameters of SGBPs binding to *Ld* laminarin (*Ld*Lam) and barley mixed-linkage β-glucan (bMLG) obtained by isothermal titration calorimetry at 25 °C[a].

	K_a (M^{-1})		ΔG (kcal.mol^{-1})		ΔH (kcal.mol^{-1})		TΔS (kcal.mol^{-1})		*n*	
Proteins	*Ld*Lam	bMLG	*Ld*Lam	bMLG	*Ld*Lam	bMLG	*Ld*Lam	bMLG	*Ld*Lam	bMLG
BuSGBP-A	NB[b]	ND[c]	NB	ND	NB	ND	NB	ND	NB	ND
BuSGBP-B	8.17 (± 0.61) x 10^5	2.99 (± 1.1) x 10^4	-10.0	-6.1	-32.3 ± 0.3	-0.5 ± 0.06	-22.3	5.6	0.893 ± 0.006	1.2 ± 0.01
BtSGBP-A	1.04 (± 0.07) x 10^5	ND	-6.9	ND	-12.3 ± 0.4	ND	-5.4	ND	0.9 ± 0.02	ND
BtSGBP-B	5.15 (± 0.34) x 10^5	ND	-7.8	ND	-21.0 ± 0.2	ND	-13.2	ND	0.918 ± 0.006	ND
BfSGBP-A	2.39 (± 0.85) x 10^4	ND	-6.0	ND	-2.1 ± 0.3	ND	3.9	ND	1	ND
BfSGBP-B	8.63 (± 2.08) x 10^4	8.56 (± 1.96) x 10^4	-6.7	-6.7	-5.2 ± 0.3	-3.4 ± 0.4	1.5	3.3	1	1.01 ± 0.01

[a] Corresponding thermograms are shown in Supplementary Fig. 11. All parameters were allowed to vary independently during data fitting, with the exception of *Bf*SGBP-A and -B binding laminarin, for which *n* was fixed at 1.

[b] No binding observed

[c] Not determined due to a lack of binding observed by affinity PAGE (see Figure 5).

Table C-7. Thermodynamic parameters of BuSGBP-A and BuSGBP-B binding to laminari-oligosaccharides obtained by isothermal titration calorimetry at 25 °C[a].

Carbohydrate	K_a (M^{-1})		ΔG (kcal.mol^{-1})		ΔH (kcal.mol^{-1})		$T\Delta S$ (kcal.mol^{-1})		*n*	
	BuSGBP-A	BuSGBP-B	BuSGBP-A	BuSGBP-B	BuSGBP-A	BuSGBP-B	BuSGBP-A	BuSGBP-B	BuSGBP-A	BuSGBP-B
Gentiobiose	ND[b]	NB[c]	ND	NB	ND	NB	ND	NB	ND	NB
Laminaribiose	NB	NB	NB	NB	NB	NB	NB	NB	NB	NB
Laminaritriose	ND	1.88 (± 0.06) x 10^3	ND	- 4.4	ND	- 55.7 ± 1.3	ND	- 51.3	ND	1
Laminaritetraose	ND	2.45 (± 0.08) x 10^3	ND	- 4.7	ND	- 73.3 ± 1.7	ND	- 68.6	ND	1
Laminaripentaose	NB	6.58 (± 0.19) x 10^3	NB	- 5.2	NB	- 41.3 ± 0.7	NB	- 36.1	NB	1
Laminarihexaose	NB	1.87 (± 0.12) x 10^5	NB	- 7.2	NB	- 25 ± 0.4	NB	- 17.8	NB	0.976 ± 0.012

[a]Corresponding thermograms are shown in Supplementary Fig. 12. All parameters were allowed to vary independently during data fitting, with the exception of *Bu*SGBP-B binding laminaritriose, laminaritetraose, and laminaripentaose, for which *n* was fixed at 1. Laminarihexose was the largest commericially available oligosaccharide.

[b] Not determined due to a lack of binding observed by affinity PAGE (see Figure 5).

[c] No binding observed

Table C-8. 1,3GUL gene conservation in *Bacteroides* species.

Gene	Name	Best homolog (% identity/%similarity)[a]					
***Bacteroides uniformis* ATCC 8492**	**query name**	***Bacteroides thetaiotaomicron* NLAE-zl-H207**	***Bacteroides fluxus* YIT 12057**	***Bacteroides cellulosilyticus* WH2**	***Dysgonomonas mossii* DSM 22836**	***Prevotella loescheii* DSM 19665 = ATCC 15930**	***Bacteroides ovatus* ATCC 8483[b]**
Bacuni_01490	BuHTCS	H207DRAFT_02223 (80/88)	HMPREF9446_00616 (85/91)		HMPREF9456_00137 (33/53)	SusR-like	BACOVA_02740 (23/47)
Bacuni_01489	BuTBDT	H207DRAFT_02224 (60/73)	HMPREF9446_00615 (54/69)	BcellWH2_04351 (87/92)	HMPREF9456_00135 (36/51)	HMPREF1991_02173 (60/74)	BACOVA_02742 (24/39)
Bacuni_01488	BuSGBP-A	H207DRAFT_02225 (52/66)	HMPREF9446_00614 (34/50)	BcellWH2_04352 (82/88)	HMPREF9456_00134 (20/38)	HMPREF1991_02174 (49/65)	BACOVA_02743 (23/40)
Bacuni_01487	BuSGBP-B	H207DRAFT_02226 (51/64)	HMPREF9446_00613 (PKD domain) (36/43)	BcellWH2_04353 (29/40)	HMPREF9456_00133 (28/38)	HMPREF1991_02175 (44/60)	BACOVA_02744 (25/37)
Bacuni_01486	BuGH16	H207DRAFT_02227 (49/61)		BcellWH2_04354 (69/79)	HMPREF9456_00132 (43/56)	HMPREF1991_02176 (46/61)	BACOVA_02741 (32/46)
Bacuni_01485	BuGH158		HMPREF9446_00612 (81/87)	BcellWH2_04355 (metallo-peptidase)	HMPREF9456_00130 (GH5)		
Bacuni_01484	BuGH3	H207DRAFT_02228 (87/93)	HMPREF9446_00611 (80/88)	BcellWH2_04356 (87/93)	HMPREF9456_00131 (57/71)	HMPREF1991_02177 (41/58)	BACOVA_02745 (35/51)

[a] Protein displaying the best homology to *Bacteroides uniformis* ATCC 8492 (% identity, % similarity at amino acid sequence level) with at least 20% amino acid identity over 60 % of the length of both proteins).

[b] Corresponds to Mixed-Linkage Glucan Utilization Locus (MLGUL) (Tamura, K., Hemsworth, G. R., Dejean, G., Rogers, T. E., Pudlo, N. A., Urs, K., Jain, N., Davies, G. J., Martens, E. C., and Brumer, H. 21: 417-430, 2017, DOI) identified in *Bacteroides ovatus* ATCC 8483.

Table C-9. Primers used for qRT-PCR

Gene	Forward primer (5' → 3')	Reverse primer (5' → 3')
BACUNI_01487	GGGAGTAACTTTCAATGTGGC	GCTGTACTTTAACCGTCACGCC
BACUNI_01488	CTGATGTTACTGAGGCCAATCG	CATAGCTTGAGTGACACGTCCC
BACUNI_01489	TGTGAATGGTGGTCAGGCTCC	GGAACGAATAGTCAAACGGTC
H207DRAFT_02224	GGTGACTTCCAACCTCTCTTATG	CGTCGTACACGTTCAGTATGG
H207DRAFT_02225	TGATCTACGGTCTGTGGAAATG	CGATGACCTGAGCTTCGTAATAG
HMPREF9446_00615	GTTACACCCAAGAACAGATTGC	TATCGCTGCCACCACTAATAC
HMPREF9446_00614	AAGGTGTCCAGGAACGTAAG	GCAAAGGCTCCTCGAAATAAG
Bacteroides 16S	GGTAGTCCACACAGTAAACGATGAA	CCCGTCAATTCCTTTGAGTTTC

Table C-10. Cloning primers and *E. coli* expression vectors used in this study

Gene	Vector	Forward primer (5' → 3')[a]	Reverse primer (5' → 3')[a]
BACUNI_01484	pET28	GTCAGTCGACATGCAAGTGCAGGATACAAAA	GGAGGACTCGAGTCAATCTACTACCTTAAA
BACUNI_01485	pMCSG53	TACTTCCAATCCAATGCCATGGGTATTGTAATAGAGAAAGCG	TTATCCACTTCCAATGTTATTAATCTATTACTTGAAATGGAATATTTGCTG
BACUNI_01486	pET28	GTCAGTCGACATGAGTAGCGATGACAAGGAA	GGAGGACTCGAGCTATTTTTTCTGAAACAC
BACUNI_01486_(PFAM13004)$_2$	pET28	GTCAGTCGACATGAGTAGCGATGACAAGGAA	GCAGCACATATGGAAGATCTGGATATCAAT
BACUNI_01487	pMCSG53	TACTTCCAATCCAATGCCATGCCTGAAGATTTTACGGGAGC	TTATCCACTTCCAATGTTATCAATCTCTATGAGTCTGGAA
BACUNI_01488	pMCSG53	TACTTCCAATCCAATGCCATGGAGGTGGAAAATCCGACAGGT	TTATCCACTTCCAATGTTATTATTTATAAGCATTCTGAACCAAAGCC
HMPREF9446_00612	pMCSG53	TACTTCCAATCCAATGCCATGTCACCGAAGAGTGGCCTTGA	TTATCCACTTCCAATGTTATTATTTAATTACTTGAAATGGAA
HMPREF9446_00613	pMCSG53	TACTTCCAATCCAATGCCATGGACGACCATTCATTGGGTGTA	TTATCCACTTCCAATGTTATTATTTTTTAGGAACAAAGCGATACC
HMPREF9446_00614	pMCSG53	TACTTCCAATCCAATGCCATGGATGATTTTTTGACCGCAAA	TTATCCACTTCCAATGTTATTAGTATCCCGGATTCTGTTTCA
H207DRAFT_02225	pMCSG53	TACTTCCAATCCAATGCCATGGAAGTTTCATCTCCTACGGAC	TTATCCACTTCCAATGTTATTAATAATTGTTCTGAACTAAGGTACC
H207DRAFT_02226	pMCSG53	TACTTCCAATCCAATGCCATGTCACCCAACGAAGCCGGCATT	TTATCCACTTCCAATGTTATTAATCTTTATGCTTTTGCAAAATGATGTC
H207DRAFT_02227	pMCSG53	TACTTCCAATCCAATGCCATGGAAACAGAAGTTGCCGTAAT	TTATCCACTTCCAATGTTATTATTTTTTTTGAAATACACGTACAT
F453DRAFT_01541	pMCSG53	TACTTCCAATCCAATGCCATGGAAGAAAATGGTTCTGGGGG	TTATCCACTTCCAATGTTATTATTTTTGATAGATGCGTAGGT

[a]Restriction sites are underlined and pMCSG LIC vector complementary sequences are double underlined.

Table C-11. Primers used for site directed mutagenesis

Gene	Residue change	Forward primer (5' → 3')[a]	Reverse primer (5' → 3')[a]
BACUNI_01485	N136A	CTGGCTTGGGGTATTGGCGCTGAAATTGAGTTGGGAAAT	ATTTCCCAACTCAATTTCAGCGCCAATACCCCAAGCCAG
BACUNI_01485	E137A	GCTTGGGGTATTGGCAATGCTATTGAGTTGGGAAATGCT	AGCATTTCCCAACTCAATAGCATTGCCAATACCCCAAGC
BACUNI_01485	E220A	GGTGCATTTATGATTACCGCTTGGGGACCGACCGGTTGG	CCAACCGGTCGGTCCCCAAGCGGTAATCATAAATGCACC

[a]Bases changed are underline

Appendix D Supporting Information for Chapter 5

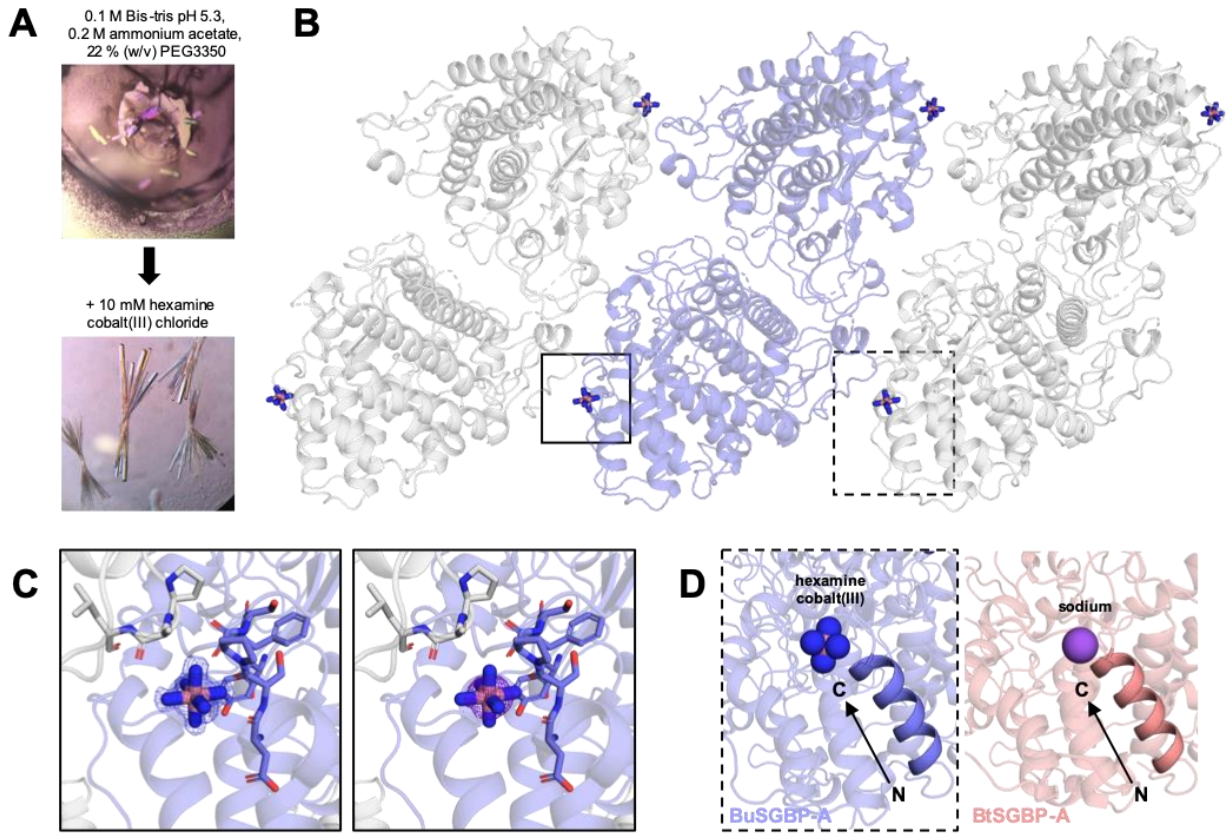

Figure D-1 BuSGBP-A hexamminecobalt (III) chloride crystallography.

A, Photos of BuSGBP-A crystals with and without addition of 10 mM hexamminecobalt(III) chloride. B, Small section of crystal lattice showing one asymmetric unit (containing two molecules) colored slate and neighboring asymmetric unit colored white. Hexamminecobalt(III) chloride is shown as sticks and can be observed bridging molecules in neighboring asymmetric units. C, Close-up of single hexamminecobalt(III) with nearby protein residues shown in sticks. The left panel shows refined $2F_{obs}$-F_{calc} map contoured at $\sigma = 1.0$ (blue mesh) and the right panel an anomalous difference density map contoured at $\sigma = 3.0$ (purple mesh) about hexamminecobalt(III) in thick stick representation. D, Close-up comparison of a hexamminecobalt(III) in BuSGBP-A and a sodium ion in BtSGBP-A both bound to the C-terminal, δ^- end of a TPR3 α-helix.

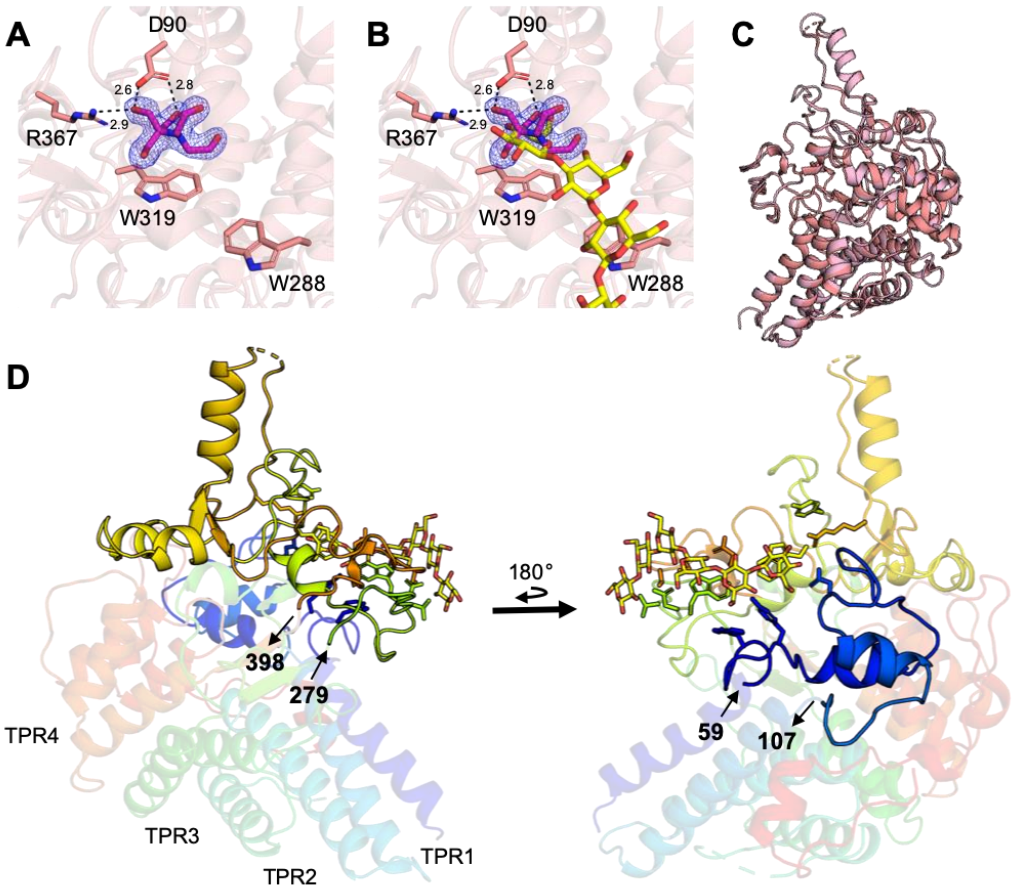

Figure D-2 BtSGBP-B Bis-tris versus laminarihexaose complex.

A, Bis-tris bound to the non-reducing end recognition site of BtSGBP-B in the absence of laminarihexaose. Refined $2F_{obs}$-F_{calc} map contoured at $\sigma = 1.0$ is shown as blue mesh about the Bis-tris molecule (magenta sticks). Dotted lines represent hydrogen bonds whose donor-acceptor distances are labelled in Å. B, Same as panel (A) with additional laminarihexaose overlaid onto bound Bis-tris molecule. C, SSM superposition of BtSGBP-A mainchain with (rose) and without (light rose) laminarihexaose bound. D, Sections of polypeptide harboring binding site residues shown as opaque cartoon and sticks (left: residues 279 – 398 connect TPR3 to TPR4; right: residues 59 – 107 connect the two α-helices comprising TPR1).

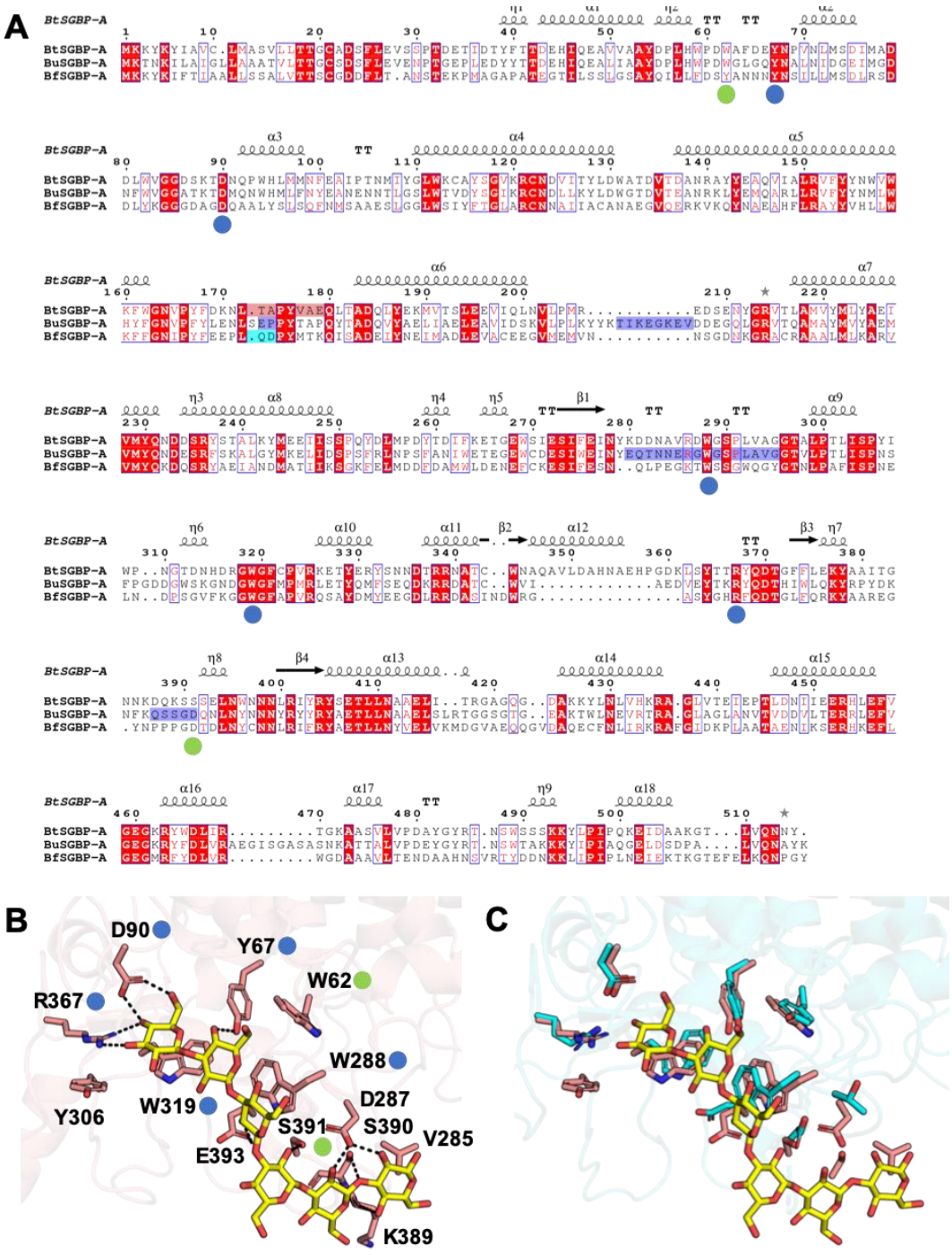

Figure D-3 Conservation of binding site residues in 1,3GUL SGBPs-A.

A, Protein sequence alignment of 1,3GUL SGBPs-A. Unmodeled segments of each SGBP-A are highlighted in their respective color (BuSGBP-A: slate, BtSGBP-A: rose, BfSGBP-A: cyan). Blue and green circles indicate conserved and similar residues involved in binding β(1,3)-glucan, respectively. B, Residue conservation information from panel A mapped onto binding site of BtSGBP-A. C, Conserved binding site residues of BtSGBP-A overlaid with those of BfSGBP-A revealing similar mechanism of laminarihexaose recognition.

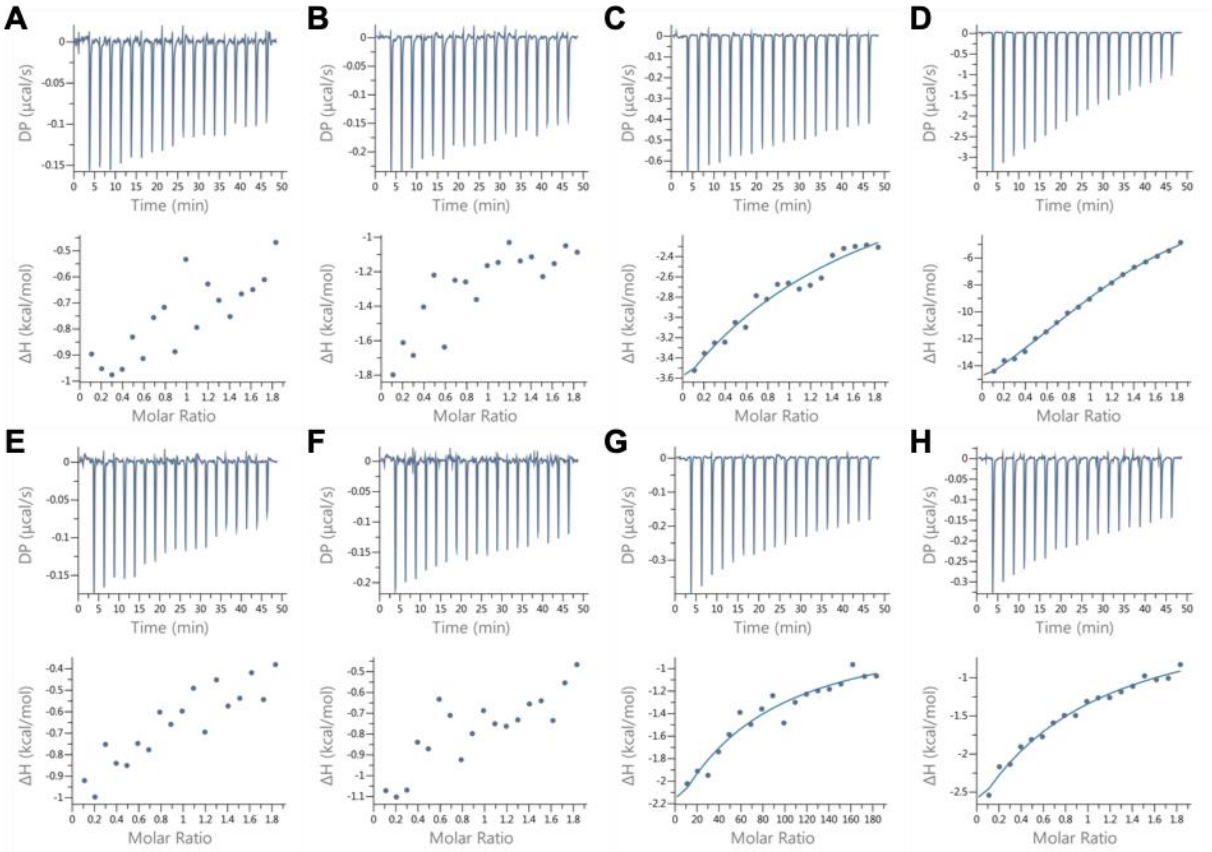

Figure D-4 Representative isothermal titration calorimetry thermograms for BtSGBP-A and BfSGBP-A titrations with laminarioligosaccharides.

All titrations were performed in 20 mM HEPES pH 7.0 at 25 °C. In each case, the upper graph shows the raw injection heat signal and the bottom graph shows the integrated heats. A, 1 mM laminaritriose into 100 µM BtSGBP-A. B, 1 mM laminaritetraose into 100 µM BtSGBP-A. C, 1 mM laminaripentaose into 100 µM BtSGBP-A. D, 1 mM laminarihexaose into 100 µM BtSGBP-A. E, 1 mM laminaritriose into 100 µM BfSGBP-A. F, 1 mM laminaritetraose into 100 µM BfSGBP-A. G, 1 mM laminaripentaose into 100 µM BfSGBP-A. H, 1 mM laminarihexaose into 100 µM BfSGBP-A.

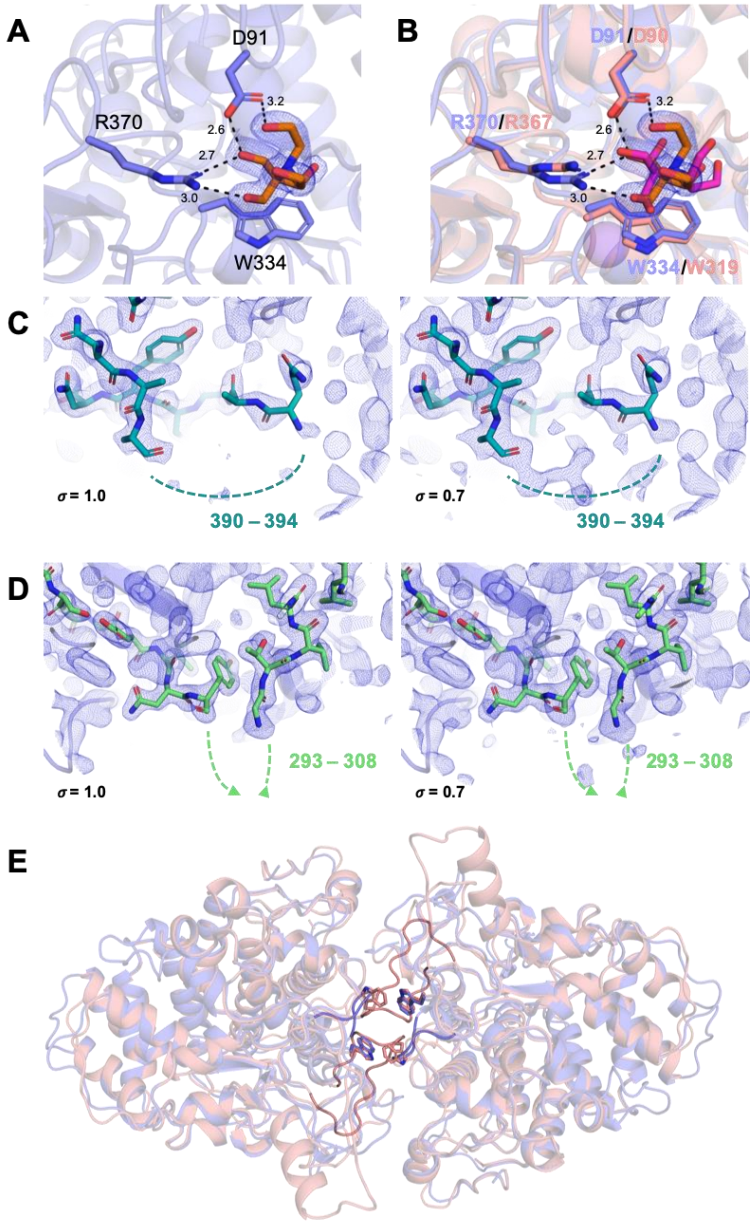

Figure D-5 BuSGBP-A missing key loops.

A, Bis-tris bound to conserved Asp91, Trp334, and Arg370 residues. Refined $2F_{obs}$-F_{calc} map contoured at σ = 1.0 is shown as blue mesh about the Bis-tris molecule (orange sticks). Dotted lines represent hydrogen bonds whose donor-acceptor distances are labelled in Å. B, Same as panel (A) with additional overlaid Bis-tris from BtSGBP-A (magenta sicks) occupying the same space in a different conformation. C, Refined $2F_{obs}$-F_{calc} map around Lys389 and Glu395 showing lack of density corresponding to residues 390 – 394. Map is shown contoured at σ = 1.0 (left panel) and σ = 0.7 (right panel). D, Refined $2F_{obs}$-F_{calc} map around Tyr292 and Gly309 showing lack of density corresponding to residues 293 – 309. Map is shown contoured at σ = 1.0 (left panel) and σ = 0.7 (right panel). E, Asymmetric unit of BuSGBP-A with BtSGBP-A overlaid onto each of two molecules with loop 293 – 308 and binding site tryptophan side chains highlighted in opaque cartoon and stick representations, respectively. A loop from BuSGBP-A which would clash with loop 293 – 308 in this position is also shown in opaque cartoon representation.

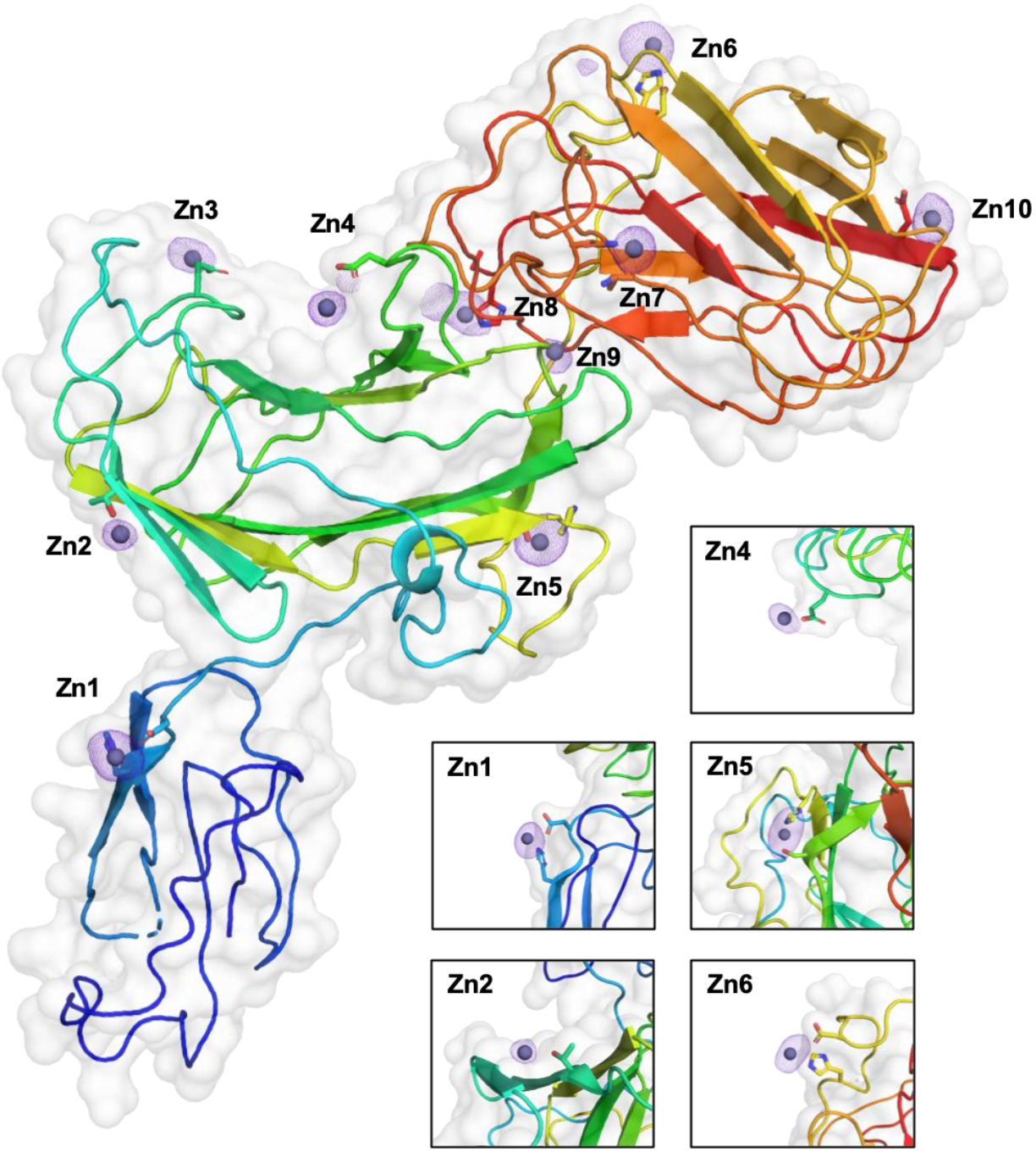

Figure D-6 BtSGBP-B zinc label anomalous signal.

Zinc atoms are shown as grey spheres with anomalous difference density map contoured at σ = 5.0 (purple mesh) about each site. Identical sites were found on both molecules in the asymmetric unit; Chain A is shown as representative. In boxes are close-up views of select zinc site showing the interacting protein sidechains in sticks.

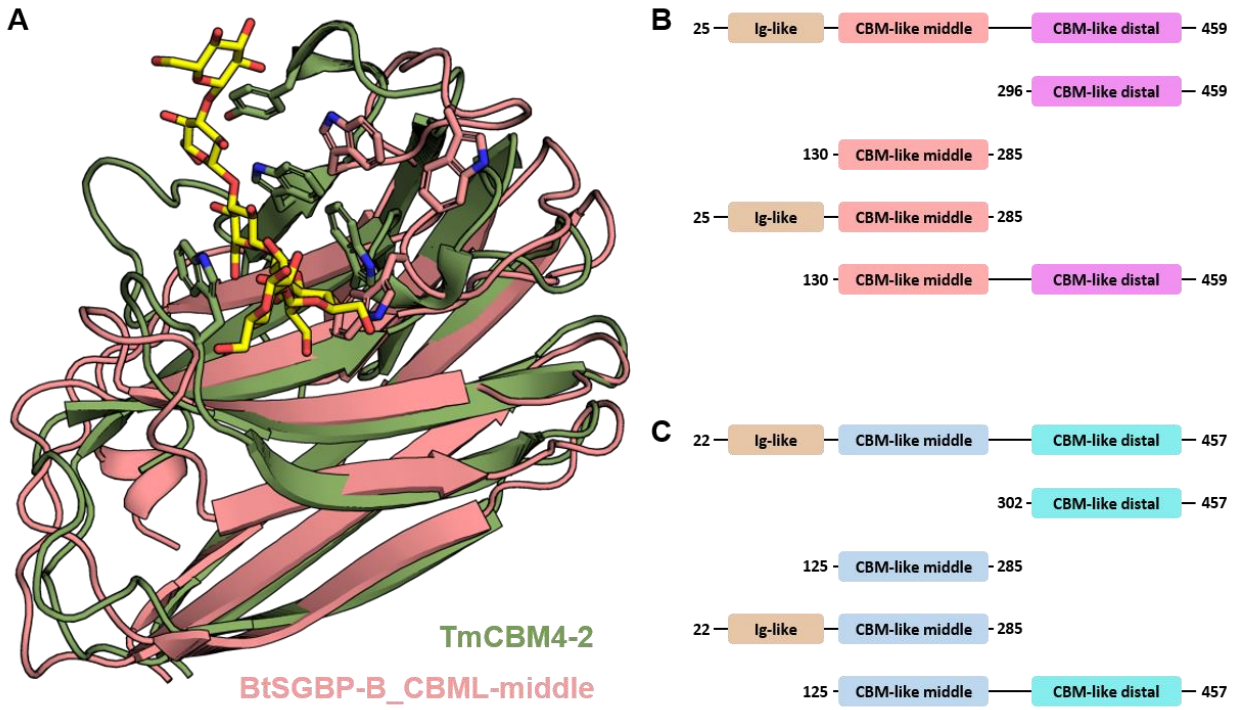

Figure D-7 BtSGBP-B structural comparison and BtSGBP-B/BuSGBP-B domains.
A, SSM superposition of CBML-middle domain of BtSGBP-B (rose) and TmCBM4-2 (green). Laminarihexaose bound to TmCBM4-2 is shown in yellow sticks and key aromatic residues are shown as sticks. B, Domain boundaries and truncation constructs of BtSGBP-B used in this study. C, Domain boundaries and truncation constructs of BuSGBP-B used in this study.

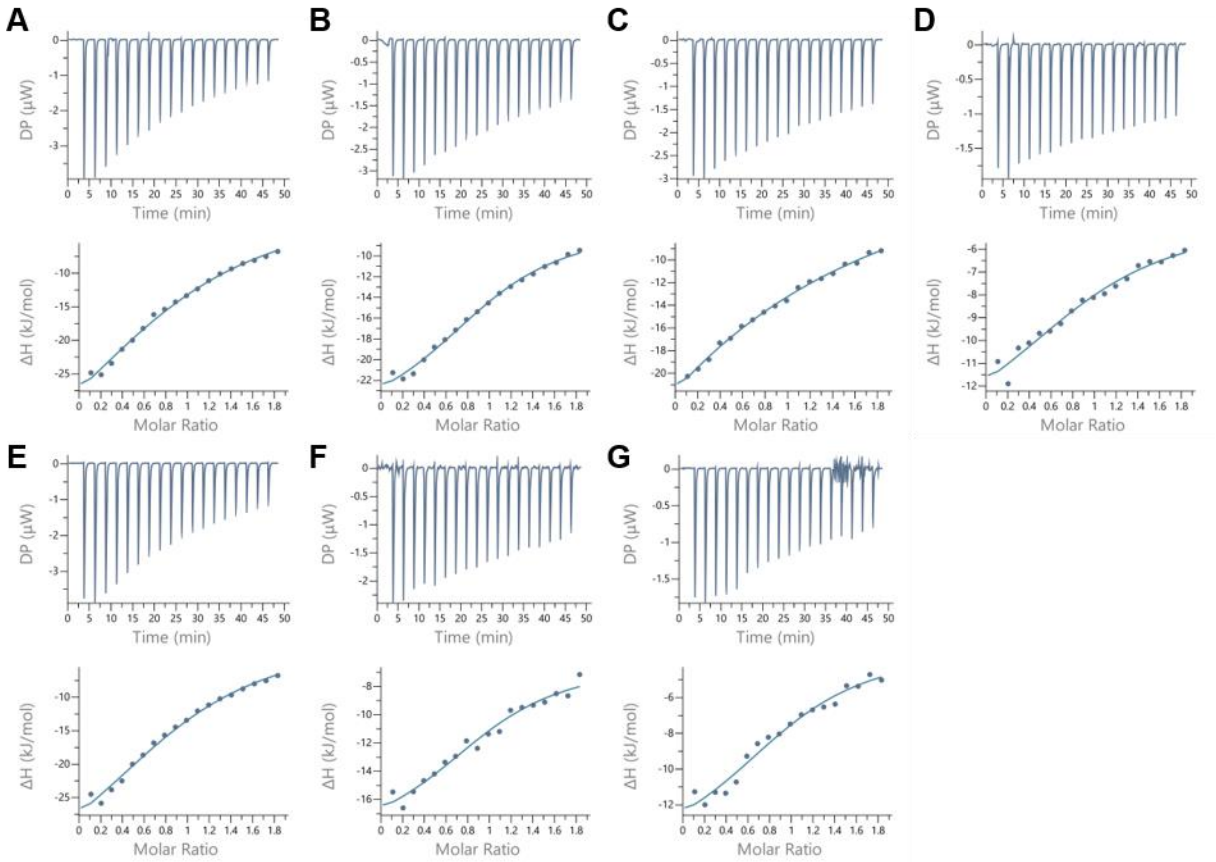

Figure D-8 Representative isothermal titration calorimetry thermograms for BtSGBP-B domains titrations with laminarioligosaccharides.

All titrations were performed in 20 mM HEPES pH 7.0 at 25 °C. In each case, the upper graph shows the raw injection heat signal and the bottom graph shows the integrated heats. A, 1 mM laminarihexaose into 100 µM BtSGBP-B_distal. B, 1 mM laminaripentaose into 100 µM BtSGBP-B_distal. C, 1 mM laminaritetraose into 100 µM BtSGBP-B_distal. D, 1 mM laminaritriose into 100 µM BtSGBP-B_distal. E, 1 mM laminarihexaose into 100 µM BtSGBP-B_middle. F, 1 mM laminaripentaose into 100 µM BtSGBP-B_middle. G, 1 mM laminaritetraose into 100 µM BtSGBP-B_middle.

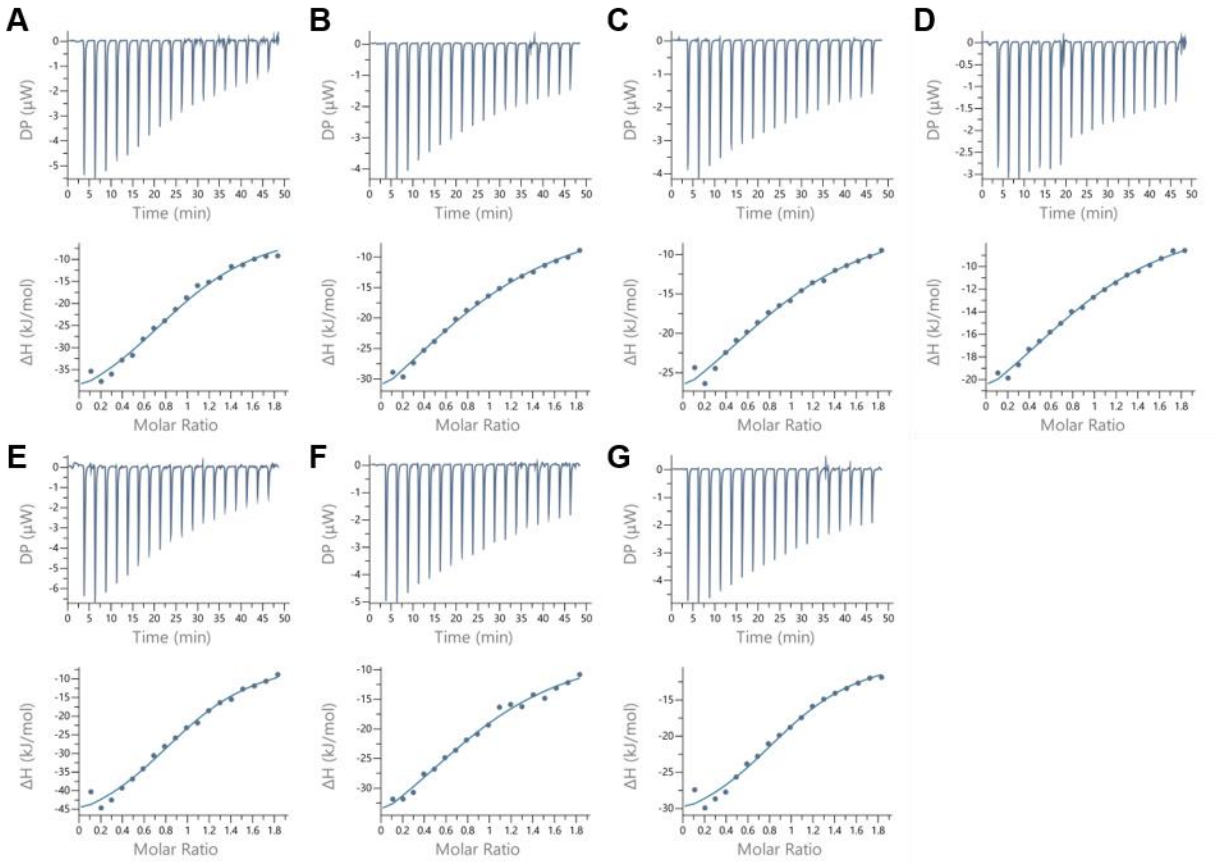

Figure D-9 Representative isothermal titration calorimetry thermograms for BuSGBP-B domains titrations with laminarioligosaccharides.

All titrations were performed in 20 mM HEPES pH 7.0 at 25 °C. In each case, the upper graph shows the raw injection heat signal and the bottom graph shows the integrated heats. A, 1 mM laminarihexaose into 100 µM BuSGBP-B_distal. B, 1 mM laminaripentaose into 100 µM BuSGBP-B_distal. C, 1 mM laminaritetraose into 100 µM BuSGBP-B_distal. D, 1 mM laminaritriose into 100 µM BuSGBP-B_distal. E, 1 mM laminarihexaose into 100 µM BuSGBP-B_middle. F, 1 mM laminaripentaose into 100 µM BuSGBP-B_middle. G, 1 mM laminaritetraose into 100 µM BuSGBP-B_middle.

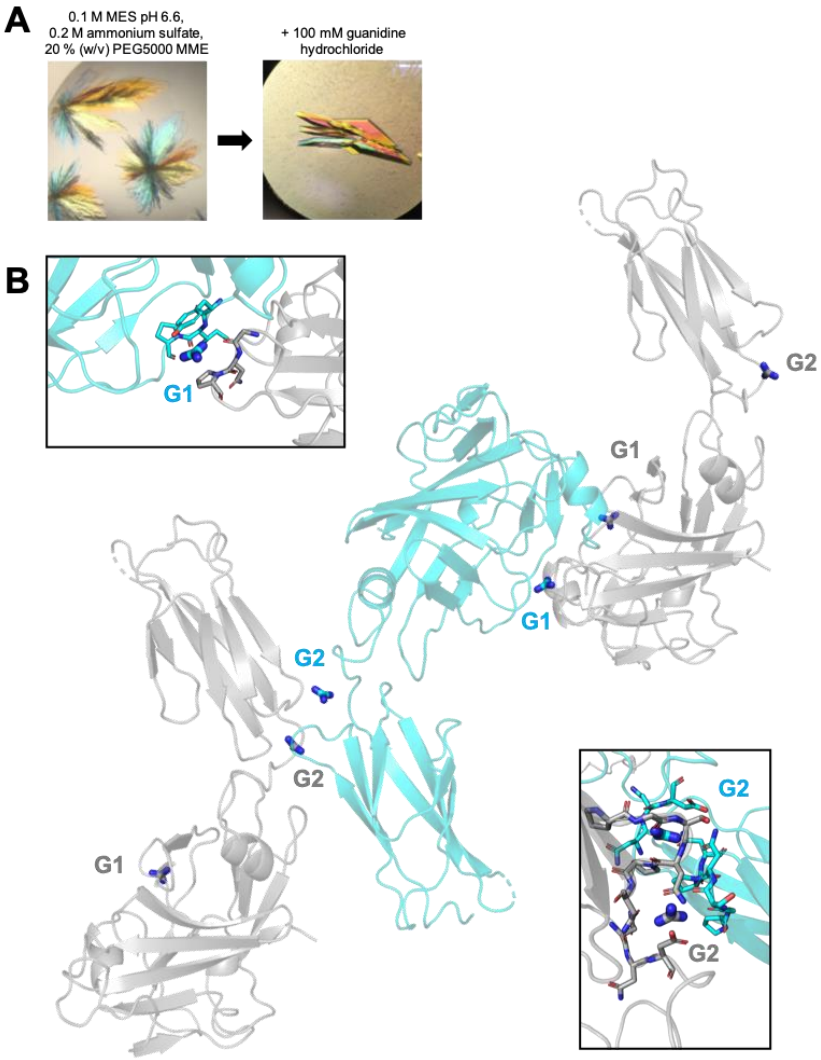

Figure D-10 BfSGBP-B guanidine hydrochloride crystallography.
A, Photos of BfSGBP-B crystals with and without addition of 100 mM guanidine hydrochloride. B, Small section of crystal lattice showing one asymmetric unit colored cyan and neighboring asymmetric unit colored white. Guanidine hydrochloride is shown as sticks and can be observed bridging molecules in neighboring asymmetric units. In boxes are close-up views of guanidine with nearby protein residues shown in sticks.

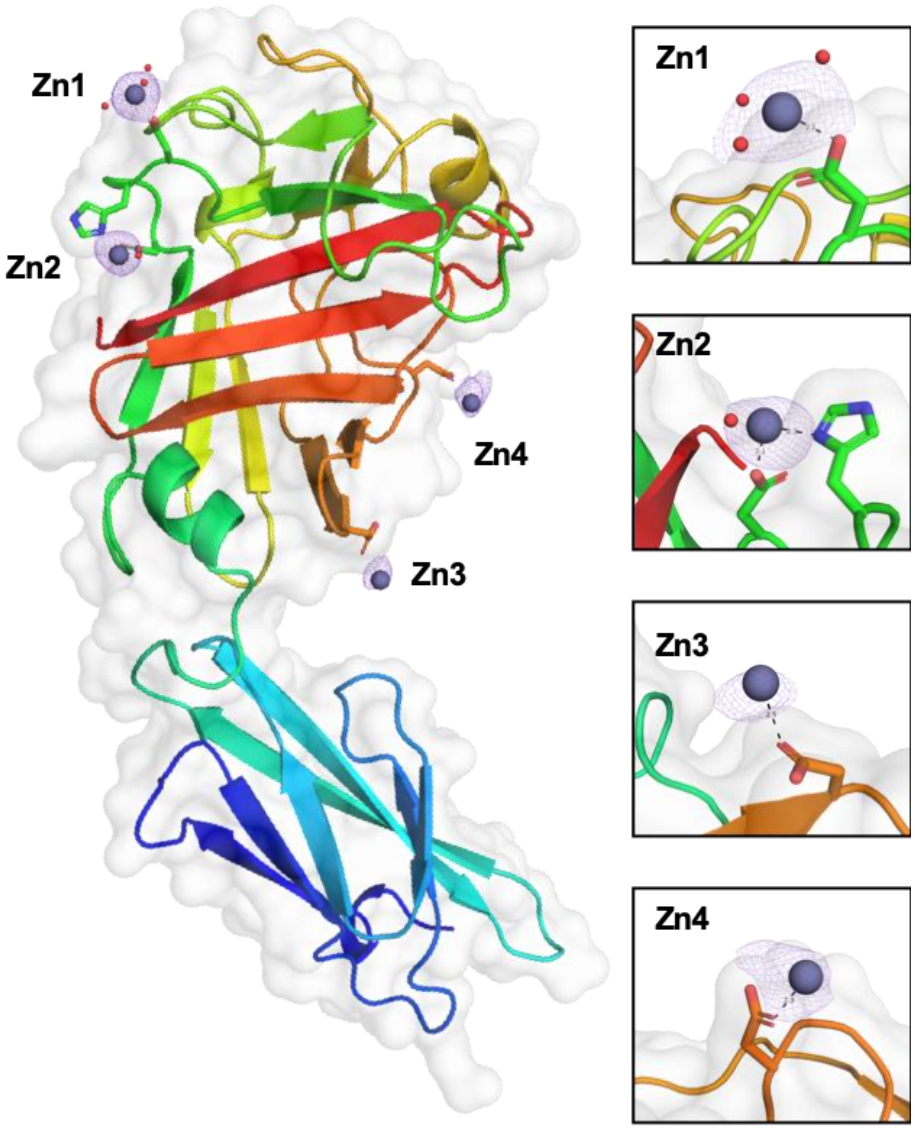

Figure D-11 BfSGBP-B zinc label anomalous signal.

Zinc atoms are shown as grey spheres with anomalous difference density map contoured at σ = 3.0 (purple mesh) about each site. In boxes are close-up views of each zinc site showing the interacting protein sidechains in sticks and water molecules as small red spheres.

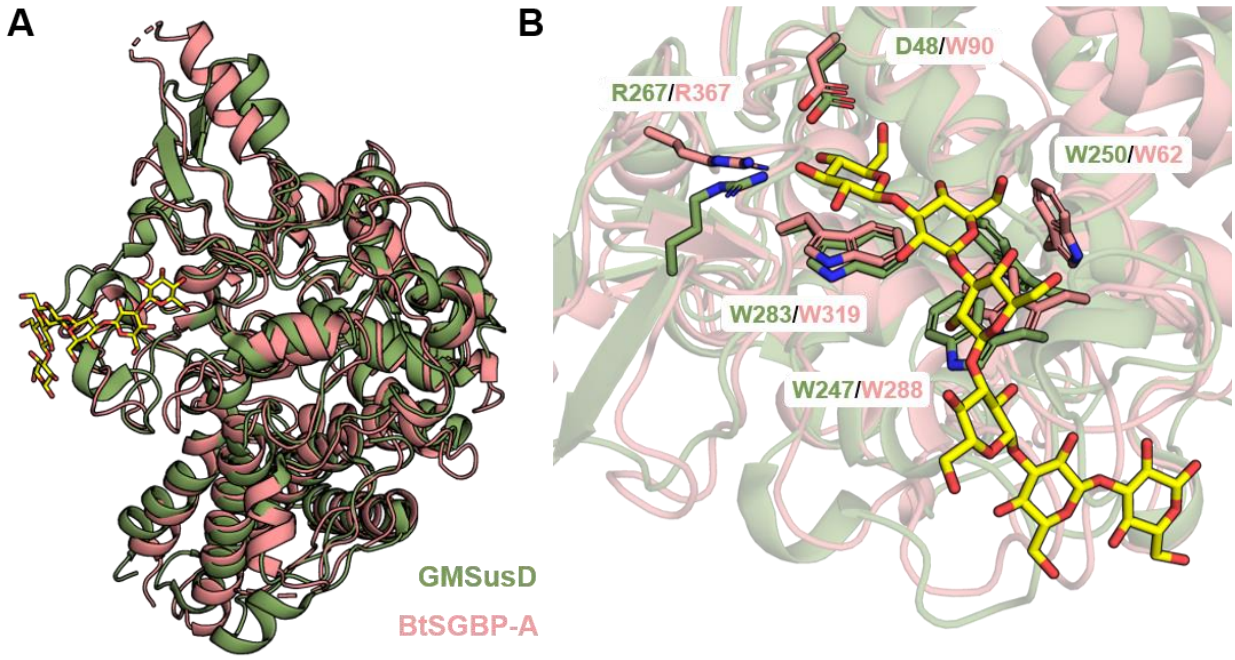

Figure D-12 StGBP-A overlay with GMSusD.
A, SSM superposition of BtSGBP-A (rose) and GMSusD (green). B, Close-up of BtSGBP-A binding site with bound laminarihexaose (yellow sticks) and the superposed GMSusD. BtSGBP-A residues important for binding are shown in rose sticks and those conserved/analogous in GMSusD are shown in green sticks.

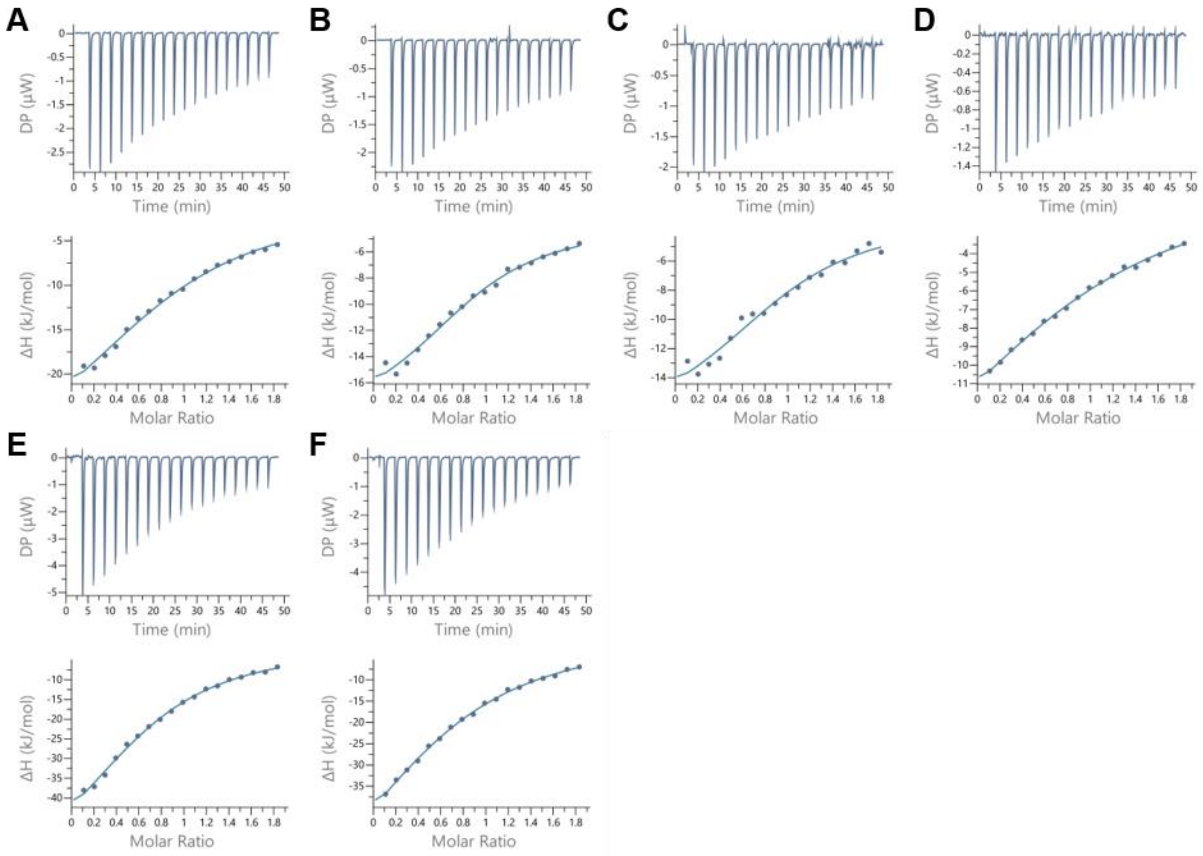

Figure D-13 Representative isothermal titration calorimetry thermograms for BfSGBP-B titrations with laminarioligosaccharides and MLG oligosaccharides.

All titrations were performed in 20 mM HEPES pH 7.0 at 25 °C. In each case, the upper graph shows the raw injection heat signal and the bottom graph shows the integrated heats. A, 1 mM laminarihexaose into 100 µM BtSGBP-B. B, 1 mM laminaripentaose into 100 µM BtSGBP-B. C, 1 mM laminaritetraose into 100 µM BtSGBP-B. D, 1 mM laminaritriose into 100 µM BtSGBP-B. E, 1 mM G4G4G3G into 100 µM BfSGBP-B. F, 1 mM G4G3G into 100 µM BfSGBP-B.

Table D-1 Data collection and refinement statistics for SGBP-A structures

	BuSGBP-A_Co	BuSGBP-A_Co (7KV1)	BtSGBP-A (7KV2)	BtSGBP-A_lam6 (7KV3)	BfSGBP-A (7KV4)
Data Collection					
Beamline	SSRL BL12-2	APS 23ID-D	APS 23ID-D	SSRL BL12-2	SSRL BL12-2
Wavelength (Å)	1.60510	1.03319	1.03319	0.97946	0.97946
Space Group	$P2_1$	$P2_1$	$P2_1$	I2	$C222_1$
Cell dimensions					
a, b, c (Å)	76.5, 53.3, 127.3	76.3, 53.6, 126.9	54.8, 71.7, 75.1	76.0, 72.9, 101.5	63.9, 116.5, 139.9
α, β, γ (°)	90, 93.1, 90	90, 93.1, 90	90, 105.1, 90	90, 105.6, 90	90, 90, 90
Resolution (Å)	38.19 – 2.00 (2.05 – 2.00)[a]	41.81 – 1.86 (1.89 – 1.86)	38.26 – 1.80 (1.83 – 1.80)	39.45 – 2.05 (2.11 – 2.05)	36.40 – 1.84 (1.88 – 1.84)
R_{merge}	0.100 (0.974)	0.127 (1.358)	0.124 (1.067)	0.197 (1.148)	0.224 (1.027)
R_{meas}	0.105 (1.025)	0.138 (1.467)	0.135 (1.206)	0.232 (1.354)	0.241 (1.109)
$CC_{1/2}$	0.999 (0.935)	0.996 (0.578)	0.993 (0.462)	0.992 (0.668)	0.987 (0.793)
$I/\sigma I$	22.5 (3.4)	10.3 (1.6)	10.0 (1.7)	8.0 (2.1)	7.7 (2.3)
Completeness (%)	96.0 (93.7)	99.9 (99.2)	99.1 (89.7)	99.0 (99.4)	99.9 (99.9)
Multiplicity	20.3 (19.9)	6.5 (6.9)	6.3 (4.8)	7.0 (7.0)	13.4 (13.7)
Anomalous completeness (%)	95.5 (92.4)				
Anomalous multiplicity	10.3 (10.1)				
Refinement					
Resolution (Å)		41.81 – 1.86	38.29 – 1.80	39.45 – 2.05	36.40 – 1.84
No. reflections (work / free)		77609 / 3967	51588 / 2482	33221 / 1639	45565 / 2369
R_{work} / R_{free}		0.167 / 0.209	0.164 / 0.194	0.157 / 0.204	0.205 / 0.239
No. of atoms					
protein		7323	3850	3793	3644
ligand		40	30	72	0
water		550	344	295	241
Avg B-factor ($Å^2$)					
protein		29.0	27.6	22.4	43.04
ligand		29.1	39.1	32.2	–
water		33.6	35.4	28.7	43.41
RMS deviations					
bond length (Å)		0.014	0.015	0.014	0.015
bond angle (°)		1.80	1.82	1.78	1.91
Ramachandran statistics					
Favored (%)		98.1	96.4	96.8	95.4
Outliers (%)		0	0	0	0

[a]Values in parentheses represent data in the highest resolution shell.

Table D-2 SGBPs-A % identity and RMSD matrix

(RMSD / seq ID)	**BuSGBP-A**	**BtSGBP-A**	**BfSGBP-A**
BuSGBP-A	0 Å / 100%	0.553 Å / 55.6%	0.885 Å / 35.2%
BtSGBP-A	0.553 Å / 55.6%	0 Å / 100%	0.798 Å / 39.3%
BfSGBP-A	0.885 Å / 35.2%	0.798 Å / 39.3%	0 Å / 100%

Table D-3 Privateer validation results for BtSGBP-A_lam6

Residue	Q[a]	Phi	Theta	Anomer	D/L[b]	Conformation	B-factor	RSCC[c]	Diagnostic[d]
BGC/A 601	0.58	357.4	0.3	beta	D	$^4C_1$	23.52	0.91	Ok
BGC/A 701	0.58	11.5	6.1	beta	D	$^4C_1$	28.41	0.93	Ok
BGC/A 801	0.58	330.6	14.0	beta	D	$^4C_1$	31.79	0.91	Ok
BGC/A 901	0.59	129.9	18.7	beta	D	$^4C_1$	36.79	0.83	Ok
BGC/A 1001	0.57	297.6	4.2	beta	D	$^4C_1$	31.16	0.90	Ok
BGC/A 1101	0.60	93.1	16.0	beta	D	$^4C_1$	36.59	0.89	Ok

[a]Q is the total puckering amplitude measured in Å. [b]D/L is the handedness. [c]RSCC is short for Real Space Correlation Coefficient and measures the agreement between model and positive omit density. [d]Conformation is either acceptable (Ok) or might be mistaken (*).

Table D-4 Thermodynamic parameters of SGBPs-A binding to laminarioligosaccharides obtained by isothermal titration calorimetry[a]

Carbohydrate	K_d (M)		ΔG (kcal mol^{-1})		ΔH (kcal mol^{-1})		-TΔS (kcal mol^{-1})		n	
	BtSGBP-A	BfSGBP-A	BtSGBP-A	BfSGBP-A	BtSGBP-A	BfSGBP-A	BtSGBP-A	BfSGBP-A	BtSGBP-A	BfSGBP-A
Laminarihexaose	2.64 x 10^{-5}	1.49 x 10^{-4}	-26.2	-21.9	-59.3	-23.8	33.1	1.89	1	1
Laminaripentaose	2.08 x 10^{-4}	2.47 x 10^{-4}	-21	-20.6	-28	-335	6.98	314	1	1
Laminaritetraose	NB[b]	NB	NB	NB	NB	NB	NB	NB	NB	NB
Laminaritriose	NB	NB	NB	NB	NB	NB	NB	NB	NB	NB
Laminaribiose	NB	NB	NB	NB	NB	NB	NB	NB	NB	NB

[a]See Fig. D-4 for corresponding thermograms
[b]No binding observed

Table D-5 Data collection and refinement statistics for BtSGBP-B structures

	BtSGBP-B_Zn	BtSGBP-B (7KWB)	BtSGBP-B_28-285 (7KWC)
Data Collection			
Beamline	SSRL BL12-2	SSRL BL12-2	SSRL BL12-2
Wavelength (Å)	1.28212	0.97946	0.97946
Space Group	$C222_1$	$C222_1$	$P4_32_12$
Cell dimensions			
a, b, c (Å)	170.65, 179.83, 93.22	174.10, 182.14, 92.29	80.69, 80.69, 86.03
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	39.77 – 2.75	39.37 – 2.60 (2.69 – 2.60)	19.64 – 2.61 (2.66 – 2.61)
R_{merge}	0.157 (3.868)	0.100 (1.199)	0.061 (0.795)
R_{meas}	0.149 (3.939)	0.104 (1.246)	0.062 (0.810)
$CC_{1/2}$	0.999 (0.754)	0.999 (0.815)	1.000 (0.951)
$I/\sigma I$	27.5 (1.7)	17.0 (2.2)	42.5 (5.3)
Completeness (%)	99.9 (100.0)	99.9 (99.9)	94.5 (98.6)
Multiplicity	53.8 (54.6)	13.5 (13.5)	25.4 (26.8)
Anomalous completeness (%)	100.0 (100.0)		
Anomalous multiplicity	28.0 (27.9)		
Refinement			
Resolution (Å)		39.40 – 2.60	19.65 – 2.61
No. reflections (work / free)		45362 / 2199	8556 / 414
R_{work} / R_{free}		0.238 / 0.282	0.250 / 0.293
No. of atoms			
protein		6121	1861
ligand		0	0
water		87	12
Avg B-factor (Å^2)			
protein		78.46	59.62
ligand		–	–
water		87	41.92
RMS deviations			
bond length (Å)		0.013	0.014
bond angle (°)		1.74	1.66
Ramachandran statistics			
Favored (%)		94.14	95.42
Outliers (%)		0.25	0

[a]Values in parentheses represent data in the highest resolution shell.

Table D-6 Top 15 DALI search results with BtSGBP-B CBML-middle.

DALI rank	PDB ID	Z score	RMSD	% ID	CBM family	Binding target	Notes
1	3K4Z	13.9	3.1	15	4	cellulose	attached to cellulase
2	1GUI	13.7	2.9	18	4	β(1,3)-glucan	attached to GH16 laminarinase
3	4QPW	12.2	2.8	9	1	xylan	attached to GH10 xylanase
4	1K42	11.7	2.7	13	4	xylan	attached to GH10 xylanase
5	2XOM	11.3	2.4	12	61	galactan	attached to GH53 galactanase
6	1CX1	10.9	3.2	19	4	cellulose	attached to cellulase
7	4D0Q	10.9	2.9	12	70	hyaluronan	attached to PL8 hyaluronan lyase
8	5W6H	10.8	2.6	10	non-CBM	–	bacteriophage CBA120 tailspike protein
9	5X7O	10.7	2.9	8	35/61	α-glucan	attached to GT31 glucosyltransferase
10	3SEE	10.1	2.8	10	non-CBM	–	hypothetical protein from *B. thetaiotaomicron*
11	2ZEW	10	2.9	8	16	cellulose/mannan	attached to GH5 mannanase
12	2WYS	10	2.9	9	22	xylan	attached to GH10 xylanase
13	2C26	10	2.9	5	44	cellulose/xyloglucan	attached to GH9 cellulase
14	2W91	9.9	3.1	4	non-CBM (similar to 44)	–	attached to GH85 N-acetylglucoseaminidase
15	4GWM	9.8	3	7	non-CBM	–	human promeprin beta

Table D-7 BtSGBP-B and BuSGBP-B domains % identity matrix

	BtSGBP-B_middle	**BtSGBP-B_distal**	**BuSGBP-B_middle**	**BuSGBP-B_distal**
BtSGBP-B_middle	100.0 %	40.52 %	51.28 %	36.67 %
BtSGBP-B_distal	40.52 %	100.0 %	37.01 %	50.66 %
BuSGBP-B_middle	51.28 %	37.01 %	100.0 %	30.67%
BuSGBP-B_distal	36.67 %	50.66 %	30.67%	100.0 %

Table D-8 Thermodynamic parameters of BtSGBP-B domains binding to laminarioligosaccharides obtained by isothermal titration calorimetry[a]

	K_d (M)		ΔG (kcal mol^{-1})		ΔH (kcal mol^{-1})		-TΔS (kcal mol^{-1})		n	
Carbohydrate	Middle	Distal	Middle	Distal	Middle	Distal	Middle	Distal	Middle	Distal
Laminarihexaose	5.20×10^{-5}	7.37×10^{-5}	-24.5	-23.6	-39.4	-46.9	14.9	23.3	1	1
Laminaripentaose	2.70×10^{-5}	2.61×10^{-5}	-26.1	-26.1	-12.9	-19.4	-13.2	-6.67	1	1
Laminaritetraose	3.12×10^{-5}	4.23×10^{-5}	-25.7	-25.0	-11.8	-20.3	-13.9	-4.69	1	1
Laminaritriose	NB[b]	4.59×10^{-5}	NB	-24.8	NB	-10.1	NB	-14.7	NB	1
Laminaribiose	NB	NB	NB	NB	NB	NB	NB	NB	NB	NB

[a]See Fig. D-8 for corresponding thermograms
[b]No binding observed

Table D-9 Thermodynamic parameters of BuSGBP-B domains binding to laminarioligosaccharides obtained by isothermal titration calorimetry[a]

	K_d (M)		ΔG (kcal mol^{-1})		ΔH (kcal mol^{-1})		-TΔS (kcal mol^{-1})		n	
Carbohydrate	Middle	Distal	Middle	Distal	Middle	Distal	Middle	Distal	Middle	Distal
Laminarihexaose	1.93×10^{-5}	2.50×10^{-5}	-26.9	-26.3	-48.8	-45.5	21.9	19.2	1	1
Laminaripentaose	4.60×10^{-5}	6.25×10^{-5}	-24.8	-24	-41.1	-46.9	16.3	22.9	1	1
Laminaritetraose	1.93×10^{-5}	4.65×10^{-5}	-26.9	-24.7	-25.5	-31.4	-1.4	6.64	1	1
Laminaritriose	NB[b]	4.81×10^{-5}	NB	-24.7	NB	-22.5	NB	-2.16	NB	1
Laminaribiose	NB	NB	NB	NB	NB	NB	NB	NB	NB	NB

[a]See Fig. D-9 for corresponding thermograms
[b]No binding observed

Table D-10 Data collection and refinement statistics for BfSGBP-B structures

	BfSGBP-B_Zn	BfSGBP-B (7KV5)	BfSGBP-B_lam3 (7KV6)	BfSGBP-B_MLG3 (7KV7)
Data Collection				
Beamline	SSRL BL9-2	SSRL BL9-2	SSRL BL12-2	SSRL BL12-2
Wavelength (Å)	1.28215	0.97946	0.97946	0.97946
Space Group	I222	I222	I2	I222
Cell dimensions				
a, b, c (Å)	59.4, 81.0, 144.3	59.5, 77.6, 146.6	74.3, 59.2, 148.1	59.9, 77.2, 148.4
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 102.6, 90	90, 90, 90
Resolution (Å)	39.91 – 2.00 (2.05 – 2.00)[a]	39.71 – 1.82 (1.86 – 1.82)	37.38 – 1.76 (1.79 – 1.76)	74.19 – 1.61 (1.64 – 1.61)
R_{merge}	0.046 (0.335)	0.071 (0.667)	0.049 (0.241)	0.031 (0.649)
R_{meas}	0.050 (0.363)	0.075 (0.722)	0.055 (0.296)	0.032 (0.697)
$CC_{1/2}$	1.000 (0.991)	0.999 (0.934)	0.999 (0.972)	0.999 (0.907)
I/σI	31.0 (6.7)	12.3 (2.4)	21.6 (4.4)	34.4 (2.3)
Completeness (%)	99.3 (99.0)	99.7 (99.7)	93.5 (60.7)	95.5 (68.7)
Multiplicity	13.2 (13.2)	6.6 (6.6)	6.8 (5.8)	11.8 (7.5)
Anomalous completeness (%)	99.4 (99.1)			
Anomalous multiplicity	6.9 (6.8)			
Refinement				
Resolution (Å)		38.84 – 1.82	35.66 – 1.76	74.19 – 1.61
No. reflections (work / free)		30831 / 1532	57885 / 2864	43047 / 2197
R_{work} / R_{free}		0.191 / 0.230	0.159 / 0.201	0.176 / 0.202
No. of atoms				
protein		1859	3867	1919
ligand		30	78	60
water		324	645	337
Avg B-factor (Å^2)				
protein		33.65	24.94	33.57
ligand		45.66	40.72	38.81
water		44.02	38.29	47.49
RMS deviations				
bond length (Å)		0.014	0.015	0.015
bond angle (°)		1.87	1.85	1.89
Ramachandran statistics				
Favored (%)		96.7	96.6	98.1
Outliers (%)		0	0	0

[a]Values in parentheses represent data in the highest resolution shell.

Table D-11 Privateer validation results for BfSGBP-B_lam3

Residue	Q[a]	Phi	Theta	Anomer	D/L[b]	Conformation	B-factor	RSCC[c]	Diagnostic[d]
BGC/A 301	0.64	99.8	4.1	beta	D	$^4C_1$	37.67	0.84	Ok
BGC/A 401	0.56	20.0	8.6	beta	D	$^4C_1$	43.51	0.70	Ok
BGC/B 301	0.61	63.2	4.9	beta	D	$^4C_1$	25.69	0.90	Ok
BGC/B 401	0.57	347.2	7.1	beta	D	$^4C_1$	32.86	0.86	Ok
BGC/B 501	0.60	317.7	7.7	beta	D	$^4C_1$	49.99	0.75	Ok

[a]Q is the total puckering amplitude measured in Å. [b]D/L is the handedness. [c]RSCC is short for Real Space Correlation Coefficient and measures the agreement between model and positive omit density. [d]Conformation is either acceptable (Ok) or might be mistaken (*).

Table D-12 Privateer validation results for BfSGBP-B_MLG3

Residue	Q[a]	Phi	Theta	Anomer	D/L[b]	Conformation	B-factor	RSCC[c]	Diagnostic[d]
BGC/A 301	0.57	15.9	7.9	beta	D	$^4C_1$	31.33	0.93	Ok
BGC/A 401	0.54	18.0	5.3	beta	D	$^4C_1$	35.81	0.92	Ok
BGC/A 501	0.50	120.5	12.1	beta	D	$^4C_1$	43.57	0.80	Ok
GLC/A 601	0.54	234.4	5.2	alpha	D	$^4C_1$	29.70	0.92	*

[a]Q is the total puckering amplitude measured in Å. [b]D/L is the handedness. [c]RSCC is short for Real Space Correlation Coefficient and measures the agreement between model and positive omit density. [d]Conformation is either acceptable (Ok) or might be mistaken (*).

Table D-13 Thermodynamic parameters of BfSGBP-B binding to various oligosaccharides obtained by isothermal titration calorimetry[a]

Carbohydrate	K_d (M)	ΔG (kcal mol^{-1})	ΔH (kcal mol^{-1})	$-T\Delta S$ (kcal mol^{-1})	n
Laminarihexaose	7.61×10^{-5}	-23.5	-36.4	12.8	1
Laminaripentaose	4.38×10^{-5}	-24.9	-18.8	-6.05	1
Laminaritetraose	3.63×10^{-5}	-25.4	-15.2	-10.1	1
Laminaritriose	7.48×10^{-5}	-23.6	-16.8	-6.8	1
Laminaribiose	NB[b]	NB	NB	NB	NB
G4G4G3G	1.16×10^{-5}	-22.5	-106	84	1
G4G3G4G	NB	NB	NB	NB	NB
G3G4G4G	NB	NB	NB	NB	NB
G4G3G	1.23×10^{-5}	-22.3	-102	79.4	1
G3G4G	NB	NB	NB	NB	NB

[a]See Fig. D-13 for corresponding thermograms
[b]No binding observed

Table D-14 Oligonucleotides used for cloning

Construct	Forward primer	Reverse primer
BuSGBP-A_28-529	TACTTCCAATCCAATGCCATGGAGGTGGAAAATCCGACAGGT	TTATCCACTTCCAATGTTATTATTTATAAGCATTCTGAACCAAAGCC
BtSGBP-A_27-515	TACTTCCAATCCAATGCCATGGAAGTTTCATCTCCTACGGAC	TTATCCACTTCCAATGTTATTAATAATTGTTCTGAACTAAGGTACC
BfSGBP-A_24-510	TACTTCCAATCCAATGCCATGGATGATTTTTTGACCGCAAA	TTATCCACTTCCAATGTTATTAGTATCCCGGATTCTGTTTCA
BuSGBP-B_22-457	TACTTCCAATCCAATGCCATGTCGCCTGAAGATTTTACGGGA	TTATCCACTTCCAATGTTATCAATCTCTATGAGTCTGGAA
BuSGBP-B_302-457	TACTTCCAATCCAATGCCATGGAAGACAATCTTTGGAACGGG	TTATCCACTTCCAATGTTATCAATCTCTATGAGTCTGGAA
BuSGBP-B_125-285	TACTTCCAATCCAATGCCATGGTGGAAGATAGTGAATTTAAC	TTATCCACTTCCAATGTTATTCCGTACCGTCATCATTAGC
BuSGBP-B_22-285	TACTTCCAATCCAATGCCATGTCGCCTGAAGATTTTACGGGA	TTATCCACTTCCAATGTTATTCCGTACCGTCATCATTAGC
BuSGBP-B_125-457	TACTTCCAATCCAATGCCATGGTGGAAGATAGTGAATTTAAC	TTATCCACTTCCAATGTTATCAATCTCTATGAGTCTGGAA
BtSGBP-B_31-459	TACTTCCAATCCAATGCCATGTCACCCAACGAAGCCGGCATT	TTATCCACTTCCAATGTTATTAATCTTTATGCTTTTGCAA
BtSGBP-B_296-459	TACTTCCAATCCAATGCCATGCCGGAACCGACATGGAGTGCG	TTATCCACTTCCAATGTTATTAATCTTTATGCTTTTGCAA
BtSGBP-B_130-285	TACTTCCAATCCAATGCCATGGACCCGGAAAGCAATTTCAAT	TTATCCACTTCCAATGTTAGTCATCGTTGGCGTGATCTTT
BtSGBP-B_31-285	TACTTCCAATCCAATGCCATGTCACCCAACGAAGCCGGCATT	TTATCCACTTCCAATGTTAGTCATCGTTGGCGTGATCTTT
BtSGBP-B_130-459	TACTTCCAATCCAATGCCATGGACCCGGAAAGCAATTTCAAT	TTATCCACTTCCAATGTTATTAATCTTTATGCTTTTGCAA
BfSGBP-B_27-290	TACTTCCAATCCAATGCCATGGACGACCATTCATTGGGTGTA	TTATCCACTTCCAATGTTATTATTTTTTAGGAACAAAGCGATACC

Table D-15 Oligonucleotides used for site-directed mutagenesis

Mutant	Forward primer	Reverse primer
BtSGBP-B_W149A	CTCCATCCCGGTGCATACGCGAATGTTGGTTCACTGAT	ATCAGTGAACCAACATTCGCGTATGCACCGGGATGGAG
BtSGBP-B_W154A	GTCCGCAATCTGACTCGCTCCCGGTGCATACCAG	CTGGTATGCACCGGGAGCGAGTCAGATTGCGGAC
BtSGBP-B_W181A	GGCATTTGTGCCTGCGCGGTCTCACTGGTTGC	GCAACCAGTGAGACCGCGCAGGCACAAATGCC
BuSGBP-B_Y144A	CCCCCATCCTGGAGCATAGGCGAAAGATGCCACATTGAAA	TTTCAATGTGGCATCTTTCGCCTATGCTCCAGGATGGGGG
BuSGBP-B_W149A	TGCAATCTGCCCCGCTCCTGGAGCATAGTAGAAAGATGC	GCATCTTTCTACTATGCTCCAGGAGCGGGGCAGATTGCA
BuSGBP-B_W176A	ACATGCATCTGAGCTTGCGCCTGATCGGTAGTAGCTTC	GAAGCTACTACCGATCAGGCGCAAGCTCAGATGCATGT
BfSGBP-B_W164A	GCCGGGCAAGACCACGCACTGGGAGTTCCATT	AATGGAACTCCCAGTGCGTGGTCTTGCCCGGC
BfSGBP-B_W165A	CAGCCGGGCAAGACGCCCAACTGGGAGTTC	GAACTCCCAGTTGGGCGTCTTGCCCGGCTG
BfSGBP-B_K172A	GCACATTCTGCCGCACCCTCAGCCGGGCAAGAC	GTCTTGCCCGGCTGAGGGTGCGGCAGAATGTGC
BfSGBP-B_D221A	CTCCACATCAAAGGCTCCGGCACCGGG	CCCGGTGCCGGAGCCTTTGATGTGGAG
BfSGBP-B_N280A	GCGATACCACCAGCCAGCACCAGACTCTACAGCACTTG	CAAGTGCTGTAGAGTCTGGTGCTGGCTGGTGGTATCGC

Table D-16 Crystallization conditions

structure	protein concentration	crystallization condition – additive/ligand	drop ratio and size	seeding	soak	cryoprotectant
BuSGBP-A	17.2 mg/mL	0.1 M bis-tris pH 5.3, 0.2 M ammonium acetate, 22 % (w/v) PEG3350 – 10 mM hexamminecobalt(III) chloride	1:1, 10 μL	micro		26% ethylene glycol
BtSGBP-A	18.3 mg/mL	0.1 M bis-tris pH 6.5, 0.2 M lithium sulfate, 25% (w/v) PEG3350	1:1, 2 μL	micro		22% ethylene glycol
BtSGBP-A_Lam6	18.3 mg/mL	0.1 M bis-tris pH 6.5, 0.2 M lithium sulfate, 25% (w/v) PEG3350	1:1, 2 μL	micro	10 mM laminarihexaose, 10 minutes	25% ethylene glycol
BfSGBP-A	32.0 mg/mL	0.1 M tris pH 8.1, 0.2 M magnesium chloride, 23% (w/v) PEG8000	1:1, 4 μL	none		20% MPD
BtSGBP-B	23.4 mg/mL	0.1 M HEPES pH 7.4, 1.3 M lithium sulfate	2:1, 1.5 μL	none		2 M lithium acetate
BtSGBP-B_trunc	27.3 mg/mL	0.1 M magnesium chloride, 0.1 M sodium citrate pH 5.0, 15% (w/v) PEG4000	1:1, 1 μL	none		25% ethylene glycol
BfSGBP-B	51.6 mg/mL	0.1 M MES pH 6.6, 0.2 M ammonium sulfate, 20 % (w/v) PEG5000 MME – 100mM guanidine hydrochloride	1:1, 10 μL	none		20% ethylene glycol
BfSGBP-B_Lam3	51.6 mg/mL	0.1 M MES pH 6.6, 0.2 M ammonium sulfate, 20 % (w/v) PEG5000 MME – 100mM guanidine hydrochloride	1:1, 10 μL	none	12.5 mg/mL laminarin partial digest, 1 hour	20% ethylene glycol
BfSGBP-B_MLG3	51.6 mg/mL	0.1 M MES pH 6.6, 0.2 M ammonium sulfate, 20 % (w/v) PEG5000 MME – 100mM guanidine hydrochloride	1:1, 10 μL	none	10 mg/mL MLG partial digest, 1 hour	20% glycerol