



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

Received for publication, January 7, 2021, and in revised form, February 2, 2021. Published, Papers in Press, February 13, 2021.

<https://doi.org/10.1016/j.jbc.2021.100415>

Kazune Tamura^{1,2}, Guillaume Dejean¹ , Filip Van Petegem² , and Harry Brumer^{1,2,3,*} 

From the ¹Michael Smith Laboratories, ²Department of Biochemistry and Molecular Biology, the Life Sciences Institute, ³Department of Chemistry, University of British Columbia, Vancouver, British Columbia, Canada

Edited by Gerald Hart

Complex glycans that evade our digestive system are major nutrients that feed the human gut microbiota (HGM). The prevalence of Bacteroidetes in the HGM of populations worldwide is engendered by the evolution of polysaccharide utilization loci (PULs), which encode concerted protein systems to utilize the myriad complex glycans in our diets. Despite their crucial roles in glycan recognition and transport, cell-surface glycan-binding proteins (SGBPs) remained understudied cogs in the PUL machinery. Here, we report the structural and biochemical characterization of a suite of SGBP-A and SGBP-B structures from three syntenic $\beta(1,3)$ -glucan utilization loci (1,3GULs) from *Bacteroides thetaiotaomicron* (*Bt*), *Bacteroides uniformis* (*Bu*), and *B. fluxus* (*Bf*), which have varying specificities for distinct β -glucans. Ligand complexes provide definitive insight into $\beta(1,3)$ -glucan selectivity in the HGM, including structural features enabling dual $\beta(1,3)$ -glucan/mixed-linkage $\beta(1,3)/\beta(1,4)$ -glucan-binding capability in some orthologs. The tertiary structural conservation of SusD-like SGBPs-A is juxtaposed with the diverse architectures and binding modes of the SGBPs-B. Specifically, the structures of the trimodular *Bt*SGBP-B and *Bu*SGBP-B revealed a tandem repeat of carbohydrate-binding module-like domains connected by long linkers. In contrast, *Bf*SGBP-B comprises a bimodular architecture with a distinct β -barrel domain at the C terminus that bears a shallow binding canyon. The molecular insights obtained here contribute to our fundamental understanding of HGM function, which in turn may inform tailored microbial intervention therapies.

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 (1–4), making HGM manipulation an attractive potential route for therapies (5, 6). HGM composition is easily altered on a short timescale (3, 7) and is driven mainly by our ingestion of complex polysaccharides (indigestible “dietary fiber”) (8–10). 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) (11). 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. 1), individual Bacteroidetes are able to access and grow on broad palettes of complex polysaccharides that nature has to offer (12, 13).

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 (14–25). Not least, these studies have revealed many new glycoside hydrolase (GH) families and specificities (reviewed in (26)). In the PUL paradigm, the noncatalytic 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 (27, 28). Despite their importance to the function of PUL systems, structural studies of the highly diverse types of SGBPs are currently few (29–32). 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 (33–35). 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 $\beta(1,3)/\beta(1,4)$ -glucan (MLG; from cereal crops) (18, 32), $\beta(1,3)$ -glucan (from fungi/yeasts, plants) (25), and $\beta(1,6)$ -glucan (from fungi) (36). In our recent study of homologous $\beta(1,3)$ -glucan utilization loci (1,3GULs) from *Bacteroides uniformis*, *B. thetaiotaomicron*, and *B. fluxus*, we demonstrated that growth on individual branched $\beta(1,3)$ -glucans and/or MLG is dependent on the combinatorial specificities of cognate GHs and SGBPs encoded by each species. (Fig. 1) (25). 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

* For correspondence: Harry Brumer, brumer@msl.ubc.ca.

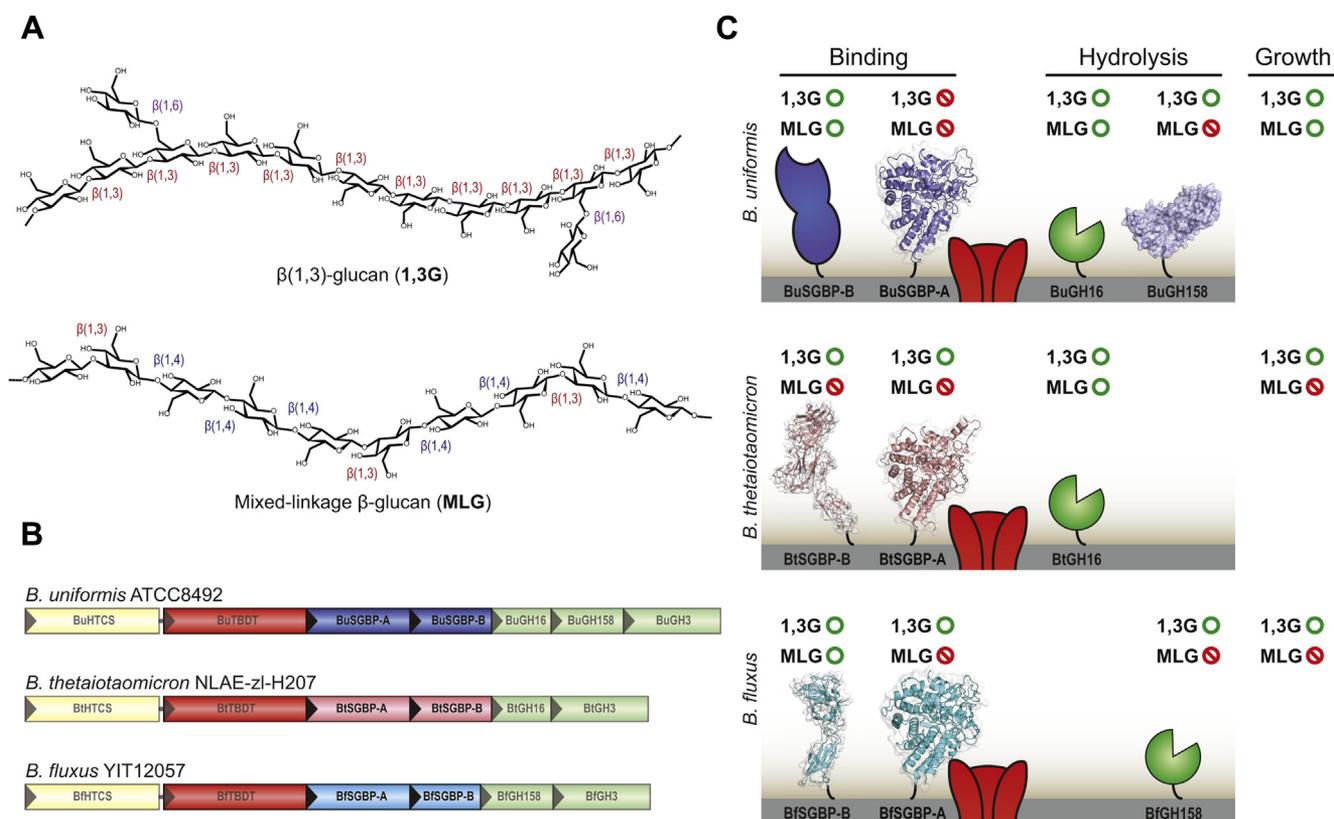


Figure 1. $\beta(1,3)$ -glucan utilization loci (1,3GULs). A, chemical structures of $\beta(1,3)$ -glucan and mixed-linkage β -glucan (MLG). $\beta(1,3)$ -glucans from diverse sources vary in the length and frequency of $\beta(1,6)$ -linked glucose branching. Shown as representative is laminarin from *Laminaria digitata*, which contains single $\beta(1,6)$ -glucose branches at a frequency of around once per every seven $\beta(1,3)$ -linked glucose. MLGs from diverse sources have the same general structure but differ in the ratio of cellotriosyl to cellotetraosyl units separated by $\beta(1,3)$ -linkages. B, genetic organization of the 1,3GULs from three different species of Bacteroidetes. GH, glycoside hydrolase; HTCS, hybrid two-component system sensor/regulator; SGBP, cell surface glycan-binding protein; TBDT, TonB-dependent transporter (SusC homolog). C, specificities of SGBPs and GHs and resulting ability to support growth on $\beta(1,3)$ -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 gray 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.* (25). 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.

unique structural architectures that allow recognition of distinct classes of β -glucan.

Results

PULs classically encode two distinct types 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) (27, 28). Indeed, the tandem TBDT/SGBP-A (*susC/susD* homolog) gene pair is a signature feature used to identify PULs in sequenced Bacteroidetes genomes (12, 37). 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 (30, 32). These are sometimes referred to a “SusE-positioned” proteins due to this genetic organization but lack of sequence similarity with SusE (29). The 1,3GULs from *B. uniformis* ATCC8492, *B. thetaiotaomicron* NLAE-zl-H207, and *B. fluxus* YIT12057 each encode one SGBP-A and one SGBP-B (Fig. 1, B and C),

the tertiary structures of which we determined in free and oligosaccharide-complexed forms.

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 (38, 39)). Hence, the recombinant forms of these proteins were produced with both the native signal peptide and the 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. S1A).

Experimental phases for BuSGBP-A were obtained by single anomalous dispersion at the cobalt absorption edge (Table S1).

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. S1, B and C). Coordinates were refined to a final resolution of 1.85 Å against a higher resolution data set collected at shorter wavelength (Table S1). The completed model of both molecules in the asymmetric unit comprised residues 43 to 529 with unmodeled gaps at residues 176 to 177, 211 to 219, 293 to 308, and 390 to 394, due to lack of corresponding electron density (Fig. 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 S1). 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 to 515 for BtSGBP-A and residues 40 to 510 for BfSGBP-A (Fig. 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

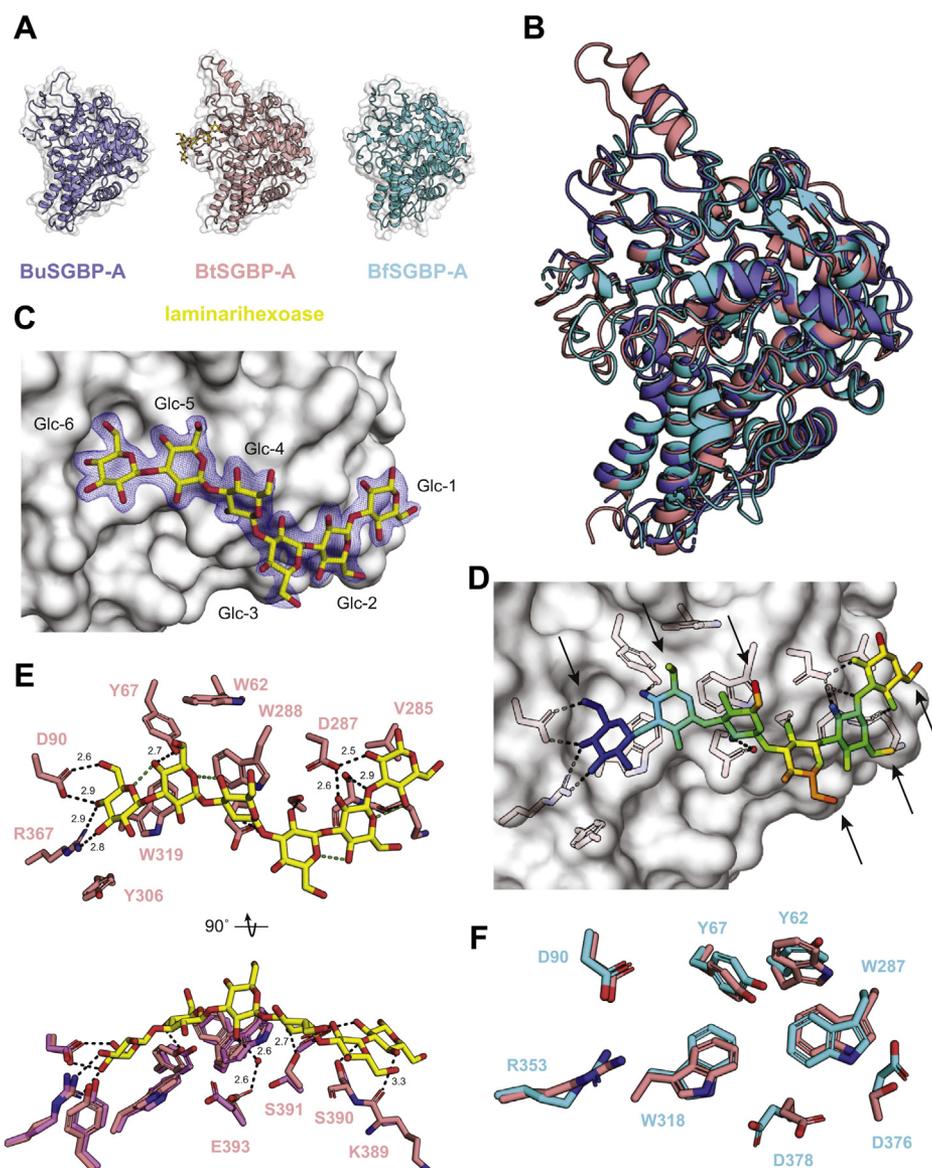


Figure 2. 1,3GUL SGBPs-A and BtSGBP-A laminarihexaose complex. *A*, overall cartoon representation of 1,3GUL SGBPs-A with transparent surface. BuSGBP-A is shown in *slate*, BtSGBP-A in *rose*, BfSGBP-A in *cyan*, and laminarihexaose in *yellow* throughout all figures. *B*, secondary structure matching (SSM) superposition of the three SGBPs-A. *C*, BtSGBP-A in opaque surface representation with bound laminarihexaose. Refined $2F_{\text{obs}} - F_{\text{calc}}$ map contoured at $\sigma = 1.0$ about the modeled 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 $\beta(1,6)$ -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 labeled 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.

BtSGBP-A and BfSGBP-A structures correspond to residues 172 to 179 and residues 172 to 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. 2A) (40). 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. 2B, Table S2). 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.

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

Soaking of the native BtSGBP-A with laminarihexaose yielded a 2.05 Å ligand-complexed structure (Table S1) 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 nonreducing end by multiple hydrogen bonds (Fig. 2). Electron density was observed for all six glucosyl residues, which were convincingly modeled in the favored 4C_1 conformation (Fig. 2C, Table S2). The bound laminarihexaose molecule adopts an extended helical structure with the binding surface complementing a low-energy conformation of the oligosaccharide (41). 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 nonreducing-end glucose (Fig. S2), which displays a very favorable carbohydrate–aromatic interaction with Trp319 (Fig. 2, D–F). In addition, four hydrogen bonds between C3-OH and Arg367-Nⁿ¹, C4-OH and Arg367-Nⁿ², C4-OH and Asp90-O^{δ1}, and C6-OH and Asp90-O^{δ2} firmly anchor the nonreducing-end glucose in place. The strength of these collective interactions is reflected in the comparatively very low B-factor of Glc6 (Fig. 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. 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. 2D). Glc2 and Glc1 both

hydrogen-bond to the same O^δ of Asp287 *via* their respective C2-OH groups (Fig. 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. 2E). Overall, the native and laminarihexaose-complexed structures of BtSGBP-A are virtually identical, superposing with an RMSD value of 0.19 Å over 402 C α pairs (Fig. S2C). 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. 2E).

Despite BtSGBP-A binding the nonreducing end of the glycan chain in a manner reminiscent of a type-C carbohydrate-binding module (CBM) (42), the binding pocket is not prominent and rather constitutes a binding platform that is blocked on one end (Fig. 2, C and D). Indeed, this extended binding site requires a β (1,3)-glucan substrate with a degree of polymerization (DP) ≥ 5 to effect binding, as revealed by isothermal titration calorimetry (ITC). BtSGBP-A bound laminarihexaose (K_D 26 μ M) approximately tenfold tighter than laminaripentaose (K_D 210 μ M), whereas binding to laminaritriose and laminaritetraose was too weak to be quantified (Fig. S4, Table S4). 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 (25). The laminarihexaose complex here reveals the abundance of space around the C6-OH group at every glucose binding subsite, thereby rationalizing this versatility (Fig. 2D).

Despite extensive efforts, including soaking and cocrystallization trials, we were not able to obtain a structure of BfSGBP-A in complex with a laminarioligosaccharide. However, sequence alignment combined with structural superposition reveals that many of the key binding site residues are conserved *vis-à-vis* BtSGBP-A (Fig. 2F, Fig. S3). 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. S3C). This assumption is supported by both β (1,3)-glucan polysaccharide (25) and laminarioligosaccharide-binding data (Fig. S4, Table S4), which indicate that BfSGBP-A requires at least a pentasaccharide for binding and is agnostic to branching frequency and branch length.

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 (25). 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. 3, Fig. S3). Furthermore, a bis-tris molecule was observed in the BuSGBP-A structure at the identical position as the native BtSGBP-A structure, coordinated by a conserved Arg/Asp/Trp triad (Fig. S5, A and B). 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. S5, C and D) correspond to loops that shape the binding site (Fig. 3). Furthermore, a loop comprised of residues 293 to 308 contains the conserved tryptophan residue that in BtSGBP-A (Trp288) forms crucial carbohydrate–aromatic stacking interactions with Glc5 and Glc4 (Fig. 3B). SDS-PAGE analysis of BuSGBP-A crystals after dissolution in 1% (v/v) trifluoroacetic acid (observed molar mass ca. 60 kDa, Fig. 3E) confirmed that unexpected proteolysis was not the cause of the missing density (Fig. S5, C and D).

It is notable that loop 293 to 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. S5E). The two other disordered segments of the polypeptide are thought to be inconsequential to binding. Residues 176 to 177 correspond to a loop distant from the binding site that is also disordered in BtSGBP-A and BfSGBP-A, and residues 211 to 219 correspond to an insertion not found in the other two SGBPs-A (Fig. S3A) but that is also distant from the binding site. On the other hand, another loop that shapes the binding site is loop 60 to 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. 3, F and G). 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 or deformed in BuSGBP-A, which is remarkably coincident with the inability of this protein to bind β (1,3)-glucans (Fig. 3F).

β (1,3)-glucan-specific SGBPs-B comprise unique multidomain 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. 1) (25). 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 for the other two SGBPs-B). To provide insight into the structural features responsible for these different binding properties, we obtained crystal structures of 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-labeled

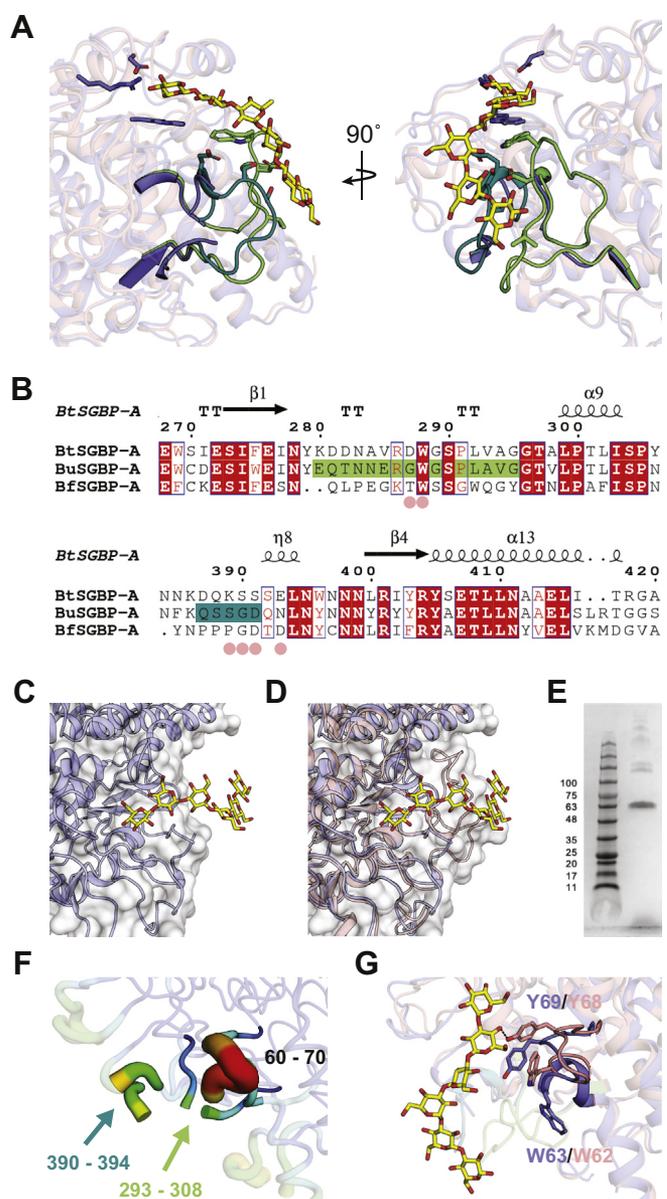


Figure 3. BuSGBP-A disrupted key loops. A, binding site loops in BtSGBP-A corresponding to two of the loops missing in BuSGBP-A; loop 293 to 308 in *lime* and loop 390 to 394 in *teal*. Side chains that participate in ligand binding as well as bound laminarihexaose in BtSGBP-A are shown in sticks. B, 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 to 70 overlaid with those of BtSGBP-A that interact with bound laminarihexaose.

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, namely ten in each of two molecules comprising the asymmetric unit (Fig. S6). The

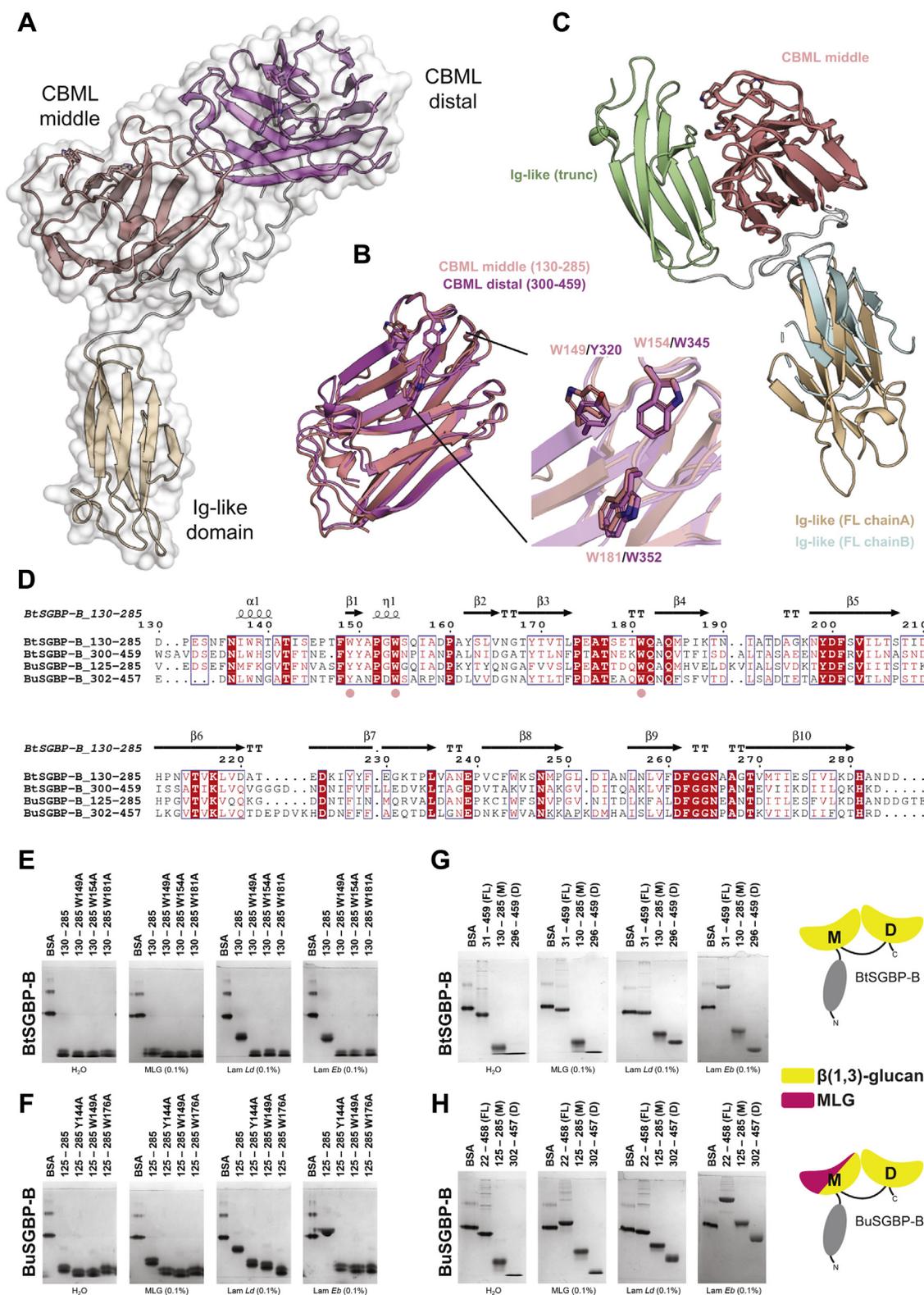


Figure 4. BtSGBP-B structure and biochemistry. A, overall structure of BtSGBP-B in cartoon representation with transparent surface. Individual ordered domains are colored in different tints and interdomain linkers are colored white. B, SSM superposition of CBML-middle (rose) and CBML-distal (magenta). 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 Eb, laminarin from *Eisenia bicyclis*; Lam Ld, laminarin from *Laminaria digitate*. G and H, AGE results of BtSGBP-B and BuSGBP-B domain dissections, respectively. Binding specificity cartoon is shown to the right.

final 2.60 Å native structure of BtSGBP-B was obtained by molecular replacement using the zinc-labeled structure as the search model (Table S5).

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 (29–32) (reviewed in (26)). 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. 4A). The CBML-middle and CBML-distal domains both comprise two β -sheets of five antiparallel β -strands each, which superpose very closely (RMSD = 0.87 Å for 143 C α pairs) (Fig. 4B). DALI analysis (43) using CBML-middle as the search model indicates that the fold of these CBML domains is most similar to that of CBM family 4, including a β (1,3)-glucan-binding module from *Thermotoga maritima* (PDB ID 1GUI, (44)) (Fig. S7A, Table S6). 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 (45), personal communication).

Also notable, the polypeptide linkers connecting the discrete domains are very long. Over ten 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. 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. 4C, Table S5). 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. 4D, Fig. S7, A and B, Table S7).

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. 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. 4, E and F).

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 CBML

domains and assessed their specificities (Fig. 4, G and H). AGE analysis indicates that, while both CBML domains of BuSGBP-B bind β (1,3)-glucan, CBML-middle is exclusively responsible for the binding of MLG (Fig. 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. S8, Table S8). These chain-length dependences and affinity values were precisely recapitulated by BtSGBP-B (Fig. S9, Table S9). The defining difference is that, unlike its counterpart, CBML-middle domain of BtSGBP cannot bind MLG (Fig. 4G).

BfSGBP-B comprises a distinct two-domain architecture

Like BuSGBP-B, BfSGBP-B binds both β (1,3)-glucan and MLG (25). 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. S10; loosely analogous to hexamminecobalt(III) chloride in the case of BuSGBP-A, Fig. S1). 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. S11). The unliganded structure was determined at 1.82 Å resolution by rigid body refinement using coordinates obtained from the phasing model (Table S10).

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. 5A). The PKD domain consists of a β -sandwich with one sheet containing three antiparallel strands and the other containing four. The β -barrel domain consists of eight antiparallel 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. 5A). A β -barrel fold involved in carbohydrate binding is unique among SGBPs-B, which are typified by β -sandwich folds. A DALI search (43) 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).

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

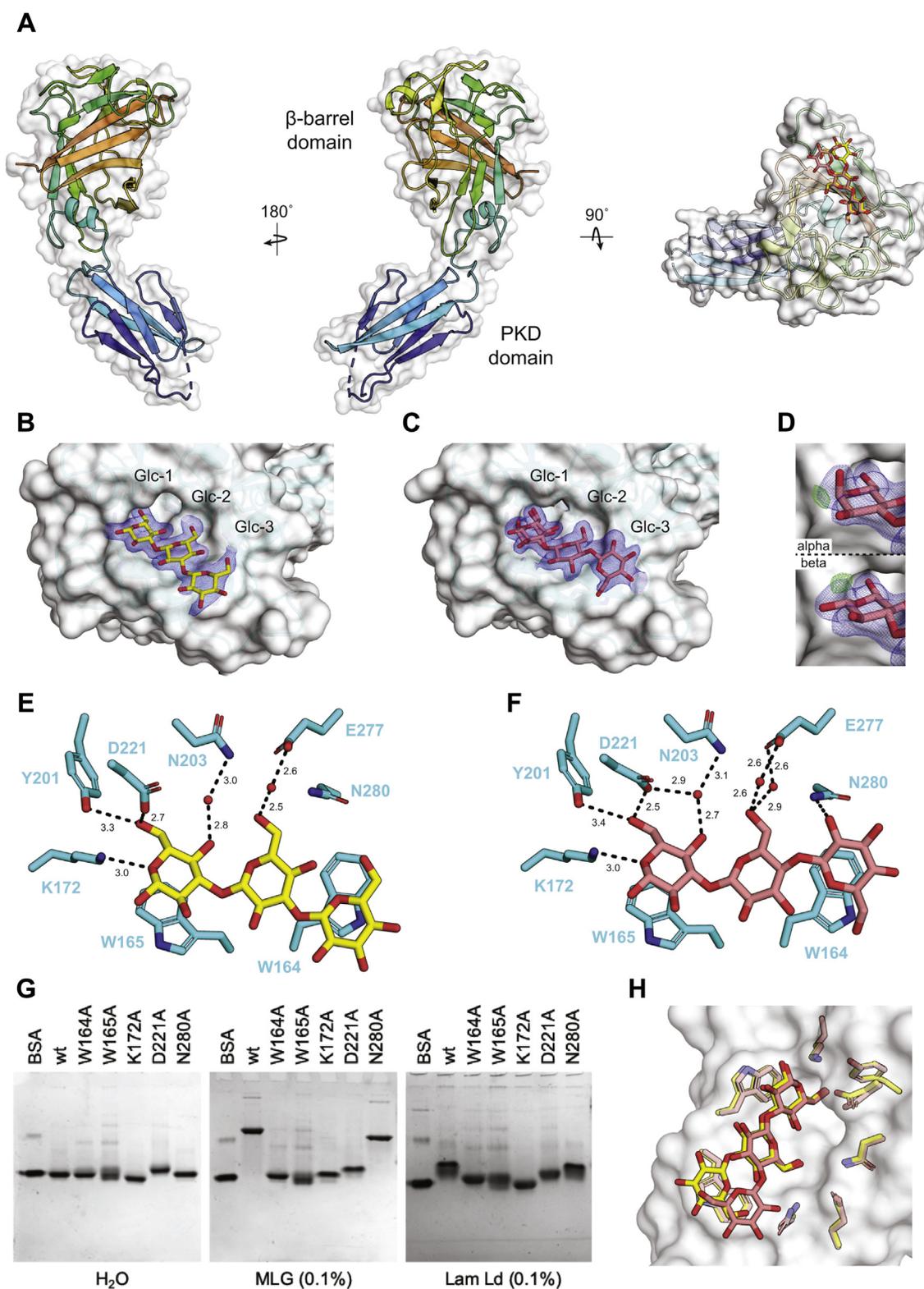


Figure 5. BfSGBP-B laminaritriose and MLG triose (G4G3G) complexes. *A*, overall structure of BfSGBP-B in cartoon representation color ramped from blue (N terminus) to red (C terminus) with transparent surface. Two side views and one top view are presented with the laminaritriose in yellow and G4G3G in salmon shown as sticks in the top view only. *B*, close-up of bound laminaritriose with $F_{\text{obs}}-F_{\text{calc}}$ omit map contoured at $\sigma = 3.0$ about the modeled sugar shown as blue mesh. *C*, close-up of bound G4G3G with $F_{\text{obs}}-F_{\text{calc}}$ omit map contoured at $\sigma = 3.0$ about the modeled sugar shown as blue mesh. *D*, refined $2F_{\text{obs}}-F_{\text{calc}}$ map contoured at $\sigma = 1.0$ and positive $F_{\text{obs}}-F_{\text{calc}}$ map contoured at $\sigma = 3.0$ about the reducing-end glucose modeled 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 labeled 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).

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 S10). In both cases, three glucose residues in the favored 4C_1 conformation were convincingly modeled (Tables S11 and S12) at the same position on the top face of the β -barrel domain (Fig. 5, A–C). 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. 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, E and F). The indispensable nature of not just the carbohydrate–aromatic interaction but also these two hydrogen bonds is highlighted by a total loss of binding when any of these residues is mutated to an alanine (Fig. 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, E and F). In both trisaccharides, the reducing-end Glc1 and Glc2 are linked *via* a $\beta(1,3)$ -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 $\beta(1,3)$ -linked or a $\beta(1,4)$ -linked glucosyl residue (Fig. 5, E, F and H). Like that observed with Trp165, this carbohydrate–aromatic interaction is indispensable (Fig. 5G). A hydrogen bond between Asn280 and the C2-OH is observed only for the $\beta(1,4)$ -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 $F_{\text{obs}}-F_{\text{calc}}$ 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 S13). Further corroborating the crystallography, ITC data also showed that a $\beta(1,3)$ -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. S13, Table S13).

Discussion

Homologous SGBPs-A feed and/or cap the TBDD

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 (30, 32, 46). 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 (40). 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 TBDD (SusC homolog), in which they comprise the extracellular lid of a “pedal-bin” arrangement (28) (Fig. 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 (30, 47). In other cases, both the presence of the SGBP-A at the cell surface and its ability to bind target glycan are necessary (32). In this context, superposition of the three SGBP-A structures with a fructan SGBP-A/TBDD complex (PDB ID 5T3R (28)) highlights that the homologous glycan-binding surfaces are poised over the top of the entrance to the β -barrel (Fig. 6B). In the superposition with BtSGBP-A, the laminarihexose in the binding site is observed to fit within the diameter of the barrel with the nonreducing end positioned more centrally and the reducing end extending toward the edge (Fig. 6C).

In this structural context, it is notable that the syntenic orthologs BtSGBP-A and BfSGBP-A both bind $\beta(1,3)$ -glucans by specifically recognizing the nonreducing 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) (32). 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 (48). Inspection of the modeled BtSGBP-A:laminarihexose/TBDD 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 TBDD by movement of the plug domain.

Intriguingly, the free BtSGBP-A can bind longer $\beta(1,3)$ -glucan polysaccharides, yet these superposed *models* suggest that binding laminarioligosaccharides with DP >6 would create a steric clash with the side of the TBDD in the fully closed conformation. Yet, in the absence of an *experimental* structure of BtSGBP-A in complex with its cognate TBDD, the molecular details of the transport cycle of $\beta(1,3)$ -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 $\beta(1,3)$ -glucan

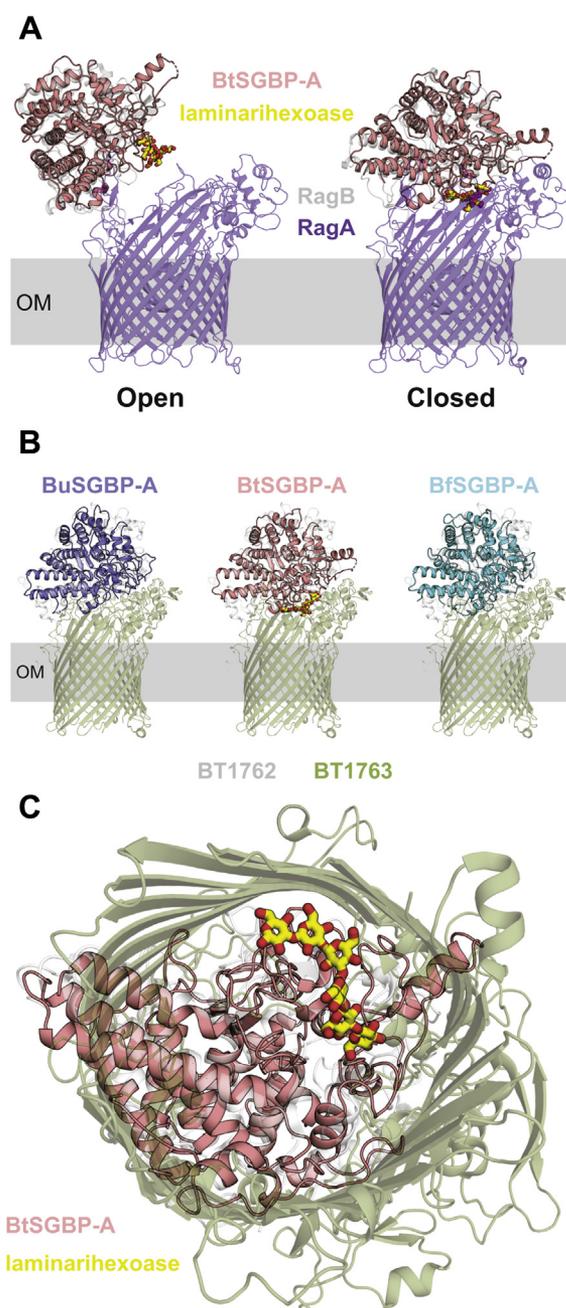


Figure 6. 1,3GUL SGBPs-A overlays with known SusCD-homolog complexes. **A**, SSM superposition of BtSGBP-A (rose) with RagB (transparent white) in complex with RagA (transparent purple) in the open (PDB ID: 6SML) and closed (PDB ID: 6SM3) conformations. Laminarihexaose bound to BtSGBP-A is shown as thick 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.

molecules found in nature from microbial sources (41). Plant MLGs, in contrast, are typically much longer polymers (49). Overall, the binding sites of BtSGBP-A and BfSGBP-A are considerably more round when compared with the flat binding platform of the MLG-specific *B. ovatus* SGBP_{MLG-A} (32). 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. 1) (41, 49).

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 (25). 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. 1) (25).

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 (50). However, three tryptophan side chains determined to be critical to binding in GMSusD are positioned homologously in the BtSGBP-A:laminarihexaose complex (all-C α RMSD 1.26 Å, sequence identity 27.9%) (Fig. S12). 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 (50) 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 (51, 52).

Structurally diverse SGBPs-B provide complementary 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. 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 multimodular architectures (29–32) (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. 4 and 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 intragene 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 that 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 (29). In contrast, those targeting xyloglucan (30), heparin/heparan sulfate (31), and MLG (32) 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 $\beta(1,3)$ -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. 1) (25). 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. 1) (25). Unfortunately, the structural basis by which the single domain of BuSGBP-B recognizes both $\beta(1,3)$ -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 such as $\beta(1,3)$ -glucan and MLG are unusual. Indeed, a survey of the current CBM families (53, 54) reveals that four families contain $\beta(1,3)$ -glucan binders (CBM39, 43, 52, and 56) and six contain MLG binders (CBM11, 72, 78, 79, 80, and 81), but only two families contain dual-function $\beta(1,3)$ -glucan/MLG binders (CBM4 and 6), e.g., *T. maritima* CBM4-2 attached to a GH16 laminarinase (44, 55) and *Cellvibrio mixtus* CBM6-2 attached to a GH5 endo-glucanase (56, 57). 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 (56, 57). Thus, binding sites optimized for $\beta(1,3)$ -glucans generally possess features that are incompatible with binding MLG and vice versa (55, 58). 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 $\beta(1,3)$ -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) (14, 25). Notably, our individual complexes of BfSGBP-B with laminaritrise and G4G3G represent, to our knowledge, the first structures of a single binding site that accommodates $\beta(1,3)$ -glucans and MLGs with equal affinity. In this case, binding MLGs by BfSGBP-B would appear to be a nondeleterious, 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. 1) (25).

Binding chain ends as a strategy for $\beta(1,3)$ -glucan recognition

CBMs have been delineated into three classes based on 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) (42). In this context, the distinction of SGBPs, CBMs, and lectins, which are united by their functions as glycan-binding proteins, is largely semantic (59). The majority of SGBPs characterized to date bind their substrates in a type-A or type-B mode (29–32, 40). Thus, the binding of β -glucan at the chain termini in a type-C or *exo*-mode at the nonreducing 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 $\beta(1,3)$ -glucans, which has precedent outside of the PUL paradigm. That is, diverse CBMs, lectins, receptors, and pattern-recognition proteins also specifically bind $\beta(1,3)$ -glucan chain ends (60). The structure of BhCBM6 from a *Bacillus halodurans* laminarinase was determined with a laminarihexaose molecule bound *via* the nonreducing end (61). Similar to BtSGBP-A and BfSGBP-A, BhCBM6 bound to laminarioligosaccharides with DP \geq 5, suggesting the presence of a comparable number of subsites. In contrast, however, the nonreducing-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* Type-C CBM binding “pocket” (61). Similarly, in CmCBM6-2 (*vide supra*) the higher-affinity laminarin binding site resembles a pocket lined by a tyrosine and tryptophan on opposing sides (57). CsCBM6-3 from a putative *Clostridium stercoarium* xylanase likewise employs a phenylalanine and tyrosine to sandwich the nonreducing end of laminaribiose, despite $\beta(1,3)$ -glucan not being the main target of this CBM (62). In contrast, the binding sites of BtSGBP-A, BfSGBP-A, and BfSGBP-B are better described as shallow binding clefts or platforms that are blocked at one end; this topology is reminiscent of some *exo*-acting $\beta(1,4)$ -glucanases (63, 64). Altogether, these structural data point to convergent evolution of chain-end recognition as an effective strategy for binding $\beta(1,3)$ -glucans in diverse niches.

Conclusion

Within the Bacteroidetes PUL paradigm, noncatalytic, 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 $\beta(1,3)$ -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 (65, 66) with the TBDT/SGBP-A complex (28, 67) will be key to fully illuminating the glycan catch-and-sequester scheme of PULs.

Experimental procedures

Substrates

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

Cloning and site-directed mutagenesis

All full-length and truncated SGBP constructs were cloned as previously described (25). All site-directed mutants were produced as previously described (32). Primers for cloning and site-directed mutagenesis are listed in Tables S14 and S15, respectively.

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 μ g/ml ampicillin. Cells were grown at 37 °C with shaking until culture OD₆₀₀ reached \sim 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 4500g for 30 min.

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 min 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 to 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). His₆-tags were subsequently cleaved from SGBPs by TEV protease (prepared in-house as described in (68)): 1.5 mg TEV protease per 25 to 100 mg SGBP, in 25 to 35 ml reaction volume such that protein concentration does not exceed 5 mg/ml, incubated overnight with gentle rocking at 4 °C. Cleaved His₆-tag, uncleaved SGBP, and His₆-tagged TEV protease were removed by running over freshly charged HisTrap column and collecting the flow-through. 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.

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 S16; 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 μ l into 5 μ l of protein followed by 4 μ l 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. S1D). A molecule in a neighboring asymmetric unit also makes contact with hexamminecobalt(III) near its C-terminal end, similarly *via* δ^- carbonyl oxygens (Fig. S1C). Overall, these hexamminecobalt(III) ions can be seen to aid in the formation of crystal contacts to improve crystal packing (Fig. S1B), 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. S8A). 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. S8B). 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. S8B).

Soaking trials 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 labeling was achieved with a 10-min 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 s) soaked in crystallization solution supplemented with cryoprotectant listed in Table S14 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 (Tables S1, S5, and S10); 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 (69), which utilizes XDS (70), Pointless (71), Aimless (72), and CCP4 (73). Data for BuSGBP-A_Co, BfSGBP-A, BtSGBP-A_lam6, BfSGBP-B_Zn, BfSGBP-B, BfSGBP-B_lam3 were indexed and integrated using XDS (70) and scaled and merged using Aimless (72). Data for BtSGBP-B and BtSGBP-B_Zn were processed and scaled using xia2/DIALS (74, 75).

Experimental phasing by SAD for BuSGBP-A_Co and BtSGBP-B was performed in AutoSol (76), subsequent density modification was performed using RESOLVE (77), and initial models were built using Phenix.autobuild (78). Experimental phasing by SAD for BfSGBP-B was performed in autoSHARP (79), subsequent density modification was performed using SOLOMON (80), and an initial model was built using ARP/wARP (81). All phasing by molecular replacement was performed using Phaser (82) in the Phenix suite (83) or the CCP4i2 suite (84), and initial models were built using Phenix.autobuild (78), Buccaneer (85), or ARP/wARP (81). All subsequent manual model building and refinement were conducted with Coot (86) and Refmac5 (87), respectively, in the CCP4i2 suite (84). The quality of modeled proteins was monitored using Molprobity (88), and that of modeled sugars was validated using Privateer (89).

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 (90). SGBPs (6.0 μ g) and bovine serum albumin (5.0 μ g) were loaded onto gels and subjected to electrophoresis under nondenaturing conditions at 80 V for 3 h at room temperature. Proteins were visualized by staining with Coomassie Brilliant Blue.

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 μ W. The sample cell was loaded with approximately 250 μ l of SGBP at 100 μ M, and the syringe was loaded with approximately 70 μ l of oligosaccharide at 1 mM; all in 20 mM HEPES pH 7.0. An initial injection of 0.2 μ l was followed by 18 subsequent injections of 2 μ l spaced 150 s apart and with an injection duration of 4 s. 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). Based on crystallographically assessed stoichiometries, n was fixed at 1 during data fitting for all isotherms.

Data availability

All atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB) with accession codes as follows: 7KV1 (BuSGBP-A), 7KV2 (BtSGBP-A), 7KV3 (BtSGBP-A_lam6), 7KV4 (BfSGBP-A), 7KWB, (BtSGBP-B), 7KWC (BtSGBP-B_trunc), 7KV5 (BfSGBP-B), 7KV6 (BfSGBP-B_lam3), 7KV7 (BfSGBP-B_MLG3).

Supporting information—This article contains [supporting information](#).

Acknowledgments—We thank the General Medical Sciences and Cancer Institutes Structural Biology Facility at the Advanced Photon Source (GM/CA@APS) for access to beamline 23ID-D. GM/CA@APS has been funded in whole or in part with Federal funds from the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006). This research used resources of the Advanced Photon Source, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357.

We thank the Stanford Synchrotron Radiation Lightsource at the SLAC National Accelerator Laboratory (Menlo Park, CA, USA) for access to beamlines BL9-2 and BL12-2, the use of which is supported by the US DOE, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH.

Author contributions—K. T. conceptualized the project and designed the experiments, generated the gene expression constructs, produced and purified the proteins, performed all

biochemistry and crystallography, analyzed the data, and composed the article. G. D. cloned full-length genes and contributed to project conceptualization. F. V. P. acquired funding and supervised structural biology. H. B. acquired funding and directed overall research.

Funding and additional information—Work in the Brumer Lab was supported by operating grants from the Canadian Institutes of Health Research (MOP-137134 and MOP-142472) and infrastructure support from the Canadian Foundation for Innovation (Project #30663) and the British Columbia Knowledge Development Fund. Work in the Van Petegem Lab was supported by the Canadian Institutes of Health Research (MOP-119404). K. T. was partially supported by a 4-year doctoral fellowship from the University of British Columbia.

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: CAZyme, carbohydrate-active enzyme; CBM, carbohydrate-binding module; GH, glycoside hydrolase; GUL, glucan utilization locus; HGM, human gut microbiota; ITC, isothermal titration calorimetry; PUL, polysaccharide utilization locus; RMSD, root-mean-square deviation; SGBP, surface glycan-binding protein; TBDT, TonB-dependent transporter; TPR, tetratricopeptide repeat.

References

- 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., *et al.* (2013) Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* **341**, 1079
- 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
- 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., *et al.* (2016) A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. *Cell* **167**, 1339–1353
- 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., *et al.* (2018) Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes. *Science* **359**, 1151–1156
- Zmora, N., Soffer, E., and Elinav, E. (2019) Transforming medicine with the microbiome. *Sci. Transl. Med.* **11**, 3
- Kolodziejczyk, A. A., Zheng, D. P., and Elinav, E. (2019) Diet-microbiota interactions and personalized nutrition. *Nat. Rev. Microbiol.* **17**, 742–753
- 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
- Koropatkin, N. M., Cameron, E. A., and Martens, E. C. (2012) How glycan metabolism shapes the human gut microbiota. *Nat. Rev. Microbiol.* **10**, 323–335
- 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
- Zmora, N., Suez, J., and Elinav, E. (2019) You are what you eat: Diet, health and the gut microbiota. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 35–56
- Grondin, J. M., Tamura, K., Dejean, G., Abbott, D. W., and Brumer, H. (2017) Polysaccharide utilization loci: Fueling microbial communities. *J. Bacteriol.* **199**, 1–15
- 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 Biol.* **9**, 1–16
- 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. *Nat. Rev. Microbiol.* **11**, 497–504
- Larsbrink, J., Rogers, T. E., Hemsworth, G. R., McKee, L. S., Tauzin, A. S., Spadiut, O., Kliner, S., Pudlo, N. A., Urs, K., Koropatkin, N. M., Creagh, A. L., Haynes, C. A., Kelly, A. G., Cederholm, S. N., Davies, G. J., *et al.* (2014) A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature* **506**, 498–502
- 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., *et al.* (2015) Glycan complexity dictates microbial resource allocation in the large intestine. *Nat. Commun.* **6**, 15
- Hemsworth, G. R., Thompson, A. J., Stepper, J., Sobala, L., 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**, 160142
- 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., *et al.* (2017) Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature* **544**, 65
- 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 Rep.* **21**, 417–430
- 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., Matar-Mann, M., *et al.* (2017) Carrageenan catabolism is encoded by a complex regulon in marine heterotrophic bacteria. *Nat. Commun.* **8**, 17
- 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., *et al.* (2018) Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic Bacteroides. *Nat. Microbiol.* **3**, 210–219
- 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. *Nat. Commun.* **9**, 14
- 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., *et al.* (2018) A surface endogalactanase in *Bacteroides thetaiotaomicron* confers keystone status for arabinogalactan degradation. *Nat. Microbiol.* **3**, 1314–1326
- Brilluete, 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. *Nat. Microbiol.* **4**, 1571–1581
- 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 /t-carrageenan metabolism pathway of some marine. *Commun. Biol.* **2**, 474
- 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**, e00095-20

26. Tamura, K., and Brumer, H. (2021) Glycan utilization systems in the human gut microbiota: A gold mine for structural discoveries. *Curr. Opin. Struct. Biol.* **68**, 26–40
27. Bolam, D. N., and Koropatkin, N. M. (2012) Glycan recognition by the Bacteroidetes sus-like systems. *Curr. Opin. Struct. Biol.* **22**, 563–569
28. 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
29. 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. *J. Biol. Chem.* **287**, 34614–34625
30. Tausin, 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
31. 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. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 7037–7042
32. 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
33. 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
34. 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
35. 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
36. 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. *J. Biol. Chem.* **292**, 10639–10650
37. Terrapon, N., Lombard, V., Drula, E., Lapebie, P., Al-Masaudi, S., Gilbert, H. J., and Henrissat, B. (2018) PULDB: The expanded database of polysaccharide utilization loci. *Nucleic Acids Res.* **46**, D677–D683
38. Juncker, A. S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H., and Krogh, A. (2003) Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci.* **12**, 1652–1662
39. Nielsen, H. (2017) Predicting secretory proteins with SignalP. *Methods Mol. Biol.* **1611**, 59–73
40. 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
41. Bacic, A., Fincher, G. B., and Stone, B. A. (2009) *Chemistry, Biochemistry, and Biology of 1-3 Beta Glucans and Related Polysaccharides*, Academic Press, Cambridge, MA
42. 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. *Curr. Opin. Struct. Biol.* **23**, 669–677
43. Holm, L. (2020) DALI and the persistence of protein shape. *Protein Sci.* **29**, 128–140
44. 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
45. Davies, G. J., and Sinnott, M. L. (2008) Sorting the diverse: The sequence based classifications of carbohydrate active enzymes. *Biochemist* **30**, 26–32
46. 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
47. Foley, M. H., Martens, E. C., and Koropatkin, N. M. (2018) SusE facilitates starch uptake independent of starch binding in *B. thetaiotaomicron*. *Mol. Microbiol.* **108**, 551–566
48. 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. *Biochemistry* **48**, 1532–1542
49. Lazaridou, A., Biliaderis, C. G., Micha-Screttas, M., and Steele, B. R. (2004) A comparative study on structure-function relations of mixed-linkage (1 \rightarrow 3), (1 \rightarrow 4) linear beta-D-glucans. *Food Hydrocolloids* **18**, 837–855
50. 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 J.* **285**, 4465–4481
51. 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
52. 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. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 19786–19791
53. 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–D495
54. CAZypedia Consortium (2018) Ten years of CAZypedia: A living encyclopedia of carbohydrate-active enzymes. *Glycobiology* **28**, 3–8
55. 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
56. 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. *J. Biol. Chem.* **279**, 21552–21559
57. 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. *J. Biol. Chem.* **279**, 21560–21568
58. 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 J.* **287**, 2723–2743
59. Taylor, M. E., and Drickamer, K. (2014) Convergent and divergent mechanisms of sugar recognition across kingdoms. *Curr. Opin. Struct. Biol.* **28**, 14–22
60. Legentil, L., Paris, F., Ballet, C., Trouvelot, S., Daire, X., Vetricka, V., and Ferrières, V. (2015) Molecular interactions of β -(1 \rightarrow 3)-glucans with their receptors. *Molecules* **20**, 9745–9766
61. 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. *J. Biol. Chem.* **280**, 530–537
62. 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. *J. Mol. Biol.* **327**, 659–669

EDITORS' PICK: *Bacteroides beta-glucan-binding proteins*

63. Sakon, J., Irwin, D., Wilson, D. B., and Karplus, P. A. (1997) Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. *Nat. Struct. Biol.* **4**, 810–818
64. 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
65. 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. *Biophys. J.* **104**, 526
66. Tuson, H. H., Foley, M. H., Koropatkin, N. M., and Biteen, J. S. (2018) The starch utilization system assembles around stationary starch-binding proteins. *Biophys. J.* **115**, 242–250
67. 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., *et al.* (2021) Insights into glycan import by a prominent gut symbiont. *Nat. Commun.* **12**, 44
68. 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
69. 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 Crystallogr. D Struct. Biol.* **67**, 293–302
70. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132
71. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D Struct. Biol.* **62**, 72–82
72. Evans, P. R., and Murshudov, G. N. (2013) How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214
73. 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., *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242
74. 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 Crystallogr. D Struct. Biol.* **74**, 85–97
75. Winter, G. (2010) xia2: An expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190
76. 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 Crystallogr. D Biol. Crystallogr.* **65**, 582–601
77. Terwilliger, T. (2004) SOLVE and RESOLVE: Automated structure solution, density modification, and model building. *J. Synchrotron Radiat.* **11**, 49–52
78. 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 Crystallogr. D Biol. Crystallogr.* **64**, 61–69
79. Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007) Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230
80. Abrahams, J. P., and Leslie, A. G. W. (1996) Methods used in the structure determination of bovine mitochondrial F-1 ATPase. *Acta Crystallogr. D Biol. Crystallogr.* **52**, 30–42
81. 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. *Nat. Protoc.* **3**, 1171–1179
82. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674
83. 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., *et al.* (2010) Phenix: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221
84. 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., *et al.* (2018) CCP4i2: The new graphical user interface to the CCP4 program suite. *Acta Crystallogr. D Struct. Biol.* **74**, 68–84
85. Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011
86. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501
87. 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 Crystallogr. D Biol. Crystallogr.* **67**, 355–367
88. 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., *et al.* (2018) MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315
89. Agirre, J., Iglesias-Fernandez, J., Rovira, C., Davies, G. J., Wilson, K. S., and Cowtan, K. D. (2015) Privateer: Software for the conformational validation of carbohydrate structures. *Nat. Struct. Mol. Biol.* **22**, 833–834
90. Cockburn, D., Wilkens, C., and Svensson, B. (2017) Affinity electrophoresis for analysis of catalytic module-carbohydrate interactions. *Methods Mol. Biol.* **1588**, 119–127



Kazune Tamura is a PhD candidate in the Department of Biochemistry and Molecular Biology, and the Michael Smith Laboratories, at the University of British Columbia. He studies the molecular mechanism of complex carbohydrate metabolism by the human gut microbiota, with a particular emphasis on the utilization of dietary beta-glucans. He has used a wide diversity of molecular techniques in his research and is particularly passionate about using structural biology to tackle the toughest questions in biology. @TamuraKazune @BrumerLab @ubcmsl @UBCBMBGSA