Title

β-1,2-Glucanase superfamily identified by sequential, functional, and structural analyses

Authors

Masahiro Nakajima[1,*], Nobukiyo Tanaka[1], Sei Motouchi[1], Kaito Kobayashi[1,2], Hisaka Shimizu[1], Koichi Abe[1,3], Naoya Hosoyamada[1], Naoya Abara[1], Naoko Morimoto[1], Narumi Hiramoto[1], Ryosuke Nakata[1], Akira Takashima[1], Marie Hosoki[1], Soichiro Suzuki[1], Kako Shikano[1], Takahiro Fujimaru[1], Shiho Imagawa[1], Yukiya Kawadai[1], Ziyu Wang[1], Yoshinao Kitano[1], Takanori Nihira[4,5], Hiroyuki Nakai[4], Hayao Taguchi[1]

Affiliations

[1] Department of Applied Biological Science, Faculty of Science and Technology, Tokyo University of Science, 2641 Yamazaki, Noda Chiba 278-8510, Japan.

[2] Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology (AIST), 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan.

[3] Department of Biotechnology, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan.

[4] Faculty of Agriculture, Niigata University, 8050 Ikarashi 2-no-cho, Nishi-ku, Niigata 950-2181, Japan.

[5] Faculty of Engineering, Niigata Institute of Technology, 1719 Fujihashi, Kashiwazaki, Niigata 945-1195, Japan.

* To whom correspondence should be addressed: Department of Applied Biological Science, Faculty of Science and Technology, Tokyo University of Science, 2641 Yamazaki, Noda, Chiba 278-8510, Japan; Email: m-nakajima@rs.tus.ac.jp; Tel.: +81-471-24-1501; Fax: +81-471-23-9767.

Running title

β-1,2-Glucanase superfamily

Keywords

Enzyme structure, glycoside hydrolase, carbohydrate, enzyme mechanism, glycobiology, β-1,2-glucan, β-1,2-glucanase, glycoside hydrolase family, CAZy

Abstract

β-1,2-Glucans are natural glucose polymers that play important physiological roles, including as symbiotic or pathogenic factors and in osmoregulation. Phylogenetically new glycoside hydrolase (GH) families have recently been identified from β-1,2-glucanase (SGL) sequences from bacteria (GH144 and GH186) and a fungus (GH162). In this study, we identified four phylogenetically new groups (Groups 1–4), and determined that these families, together with GH144, GH162, and GH189, a family of transglycosylase domains in cyclic β-1,2-glucan synthases, form a superfamily. Biochemical analysis of six proteins in these groups revealed that the proteins in Groups 1–3 showed hydrolytic activity specific to β-1,2-glucan. The kinetic parameters of the enzymes of Groups 1–3 were similar to GH144 and GH162 SGLs, indicating that these enzymes were SGLs. Optical rotation analysis revealed that the SGLs followed an anomer-inverting mechanism. Structural analysis and prediction of the proteins in Groups 1–4, GH144, GH162, and GH189 suggested that Groups 1–3 and GH144 had the same reaction mechanism. Nevertheless, Groups 1–3 were dispersed irregularly in the superfamily. Overall, we determined that Groups 1–3 were new GH families, GHxxx, GHyyy, and GHzzz, respectively, and proposed that this superfamily be called an SGL superfamily because of the phylogenetical, functional, and structural relationships within the superfamily.

Introduction

Carbohydrates are used not only as energy sources, but also for the storage of polysaccharides and in cell skeletons, including in the cell walls of plants and fungi (1–3). Indigestible carbohydrates known as prebiotics affect the microbiome in the large intestine by enhancing the growth of beneficial bacteria, which is important for immunostimulation (4–6). The wide variety of the physiological functions of carbohydrates in organisms derives from the complexity of the carbohydrate structures. It is estimated that over 1 billion different structures are possible for a hexasaccharide theoretically (7). In response to such complexity, the enzymes associated with carbohydrate synthesis and degradation have evolved to have diverse amino acid sequences and tertiary structures. Enzymes in the Carbohydrate-Active enZYmes (CAZy) database are classified into families based on the amino acid sequences (8–11), and are called "cazymes." These families are categorized into glycoside hydrolase (GH) families, glycosyltransferase (GT) families, carbohydrate esterase (CE) families, polysaccharide lyase (PL) families, and auxiliary activity (AA) families (AA families contain redox-active enzymes) (9) mainly based on the types of reactions that the enzymes catalyze. Enzymes in the GH families are the most abundant in the CAZy database, and the number of identified GH families continues to increase and had reached 180 as of Jan 2024, suggesting the importance of the GH family of enzymes. However, considering the complexity of carbohydrate structures, it is presumed that many "cazymes" remain as yet unidentified.

Glucose polymers are widely distributed in nature. Cellulose is the most abundant material in biomass, composing the cell walls of plants, and β-glucans from monocotyledonous plants are one of the major hemicelluloses (12, 13). β-1,3-Glucans are one of the major cell wall components in filamentous fungi and act as storage polysaccharides for some algal species (14, 15). Pustulan is a β-1,6-glucan that is produced by lichen and is also a cell wall component in some yeast species (16–18). Starch and glycogen are major α-linked glucose polymers found in nature (19). The functions and structures of the enzymes that act on these polysaccharides have been extensively studied.

β-1,2-Glucans are polysaccharides composed of glucose, which are found in nature mainly in cyclic forms. Although β-1,2-glucans are thought to be rare compared with the glucans described above, β-1,2-glucans are involved in the bacterial infection of animal and plant cells and in hypo-osmotic adaptation (20–23). A β-1,2-glucan was first found as an extracellular polysaccharide from *Agrobacterium tuberculosis*, a plant pathogen forming crown galls on plant roots in 1940 (24, 25). This polysaccharide was later found to be a β-1,2-glucan polymer with a cyclic form (26, 27) and the enzymes for its synthesis were independently identified as cyclic β-1,2-glucan synthases (CGSs) from *Brucella abortus*, *Sinorhizobium meliloti*, and *Agrobacterium tumefacies* (28–32). The overall reaction mechanism of CGSs has been studied biochemically using CGS from *B. abortus* (33, 34). CGSs are composed of three domains: GT84; transglycosylation; and GH94 domains. First, the GT84 domain elongates β-1,2-glucan chains and then the GH94 domain adjusts the chain lengths by phosphorolysis.

Finally, the transglycosylation domain cyclizes the glucans by a transglycosylation reaction. The detailed reaction mechanism of the transglycosylation domain was elucidated using the transglycosylation domain of CGS from *Thermoanaerobacter italicus* (TiCGS), a thermophilic bacterium (35), leading to identification of GH189 (36).

In contrast to the enzymes for the synthesis of β-1,2-glucans, there had been no sequentially identified β-1,2-glucan-degrading enzymes until 1,2-β-oligoglucan phosphorylase (SOGP) was found in *Listeria innocua* in 2014 (37). A method for the large-scale production of β-1,2-glucan has been established (38–40), which has enabled the investigation and identification of genes encoding other β-1,2-glucan-degrading enzymes using the produced glucans. *endo*-β-1,2-Glucanases (SGLs) that release β-1,2-glucooligosaccharides [Sop_ns, where n is the degree of polymerization (DP)] from β-1,2-glucans have been sequentially identified from a bacterium and a fungus (41–43). These enzymes were classified as the new GH families, GH144 and GH162 from the bacterium and fungus, respectively (44, 45). The identification of these families suggested that β-1,2-glucan-associated enzymes and proteins may be widely distributed in various kinds of microorganisms. GH3 β-glucosidases (BGLs) associated with β-1,2-glucans (46, 47); a GH94 glycoside phosphorylase (48); enzymes given new EC (Enzyme Commission) numbers: a sophorose-releasing enzyme (49) and a transglycosylase that transfers a β-1,2-linked glucose (Glc) unit (50); and a solute-binding protein in an ABC transporter specific to Sop_ns (SO-BP) (51) have been characterized functionally and structurally.

In addition, OpgG and OpgD proteins from *Escherichia coli* were found to be SGLs that do not belong to the GH144 or GH162 families (52). Osmoregulated periplasmic glucans (OPG), which are polysaccharides with β-1,2-glucan backbones, are produced by Gram-negative bacteria, such as *E. coli* (32, 53–57). Although OpgG and OpgD from *E. coli* were thought to be involved in OPG biosynthesis (56, 58–60), the exact biochemical functions of these proteins had been previously unknown. The OpgD group has been classified into a new GH family (GH186) and this classification also indicates the wide distribution of β-1,2-glucan-associated enzymes.

GH144 SGL from *Chitinophaga pinensis* (CpSGL), a GH144-establishing enzyme, and GH162 SGL from *Talaromyces funiculosus* (TfSGL), a GH162-establishing enzyme, have only approximately 10% amino acid sequence identity (44, 45). However, CpSGL and TfSGL share a similar overall domain fold, suggesting that there is a relationship between the molecular evolution of these families. The general acid catalyst, one of the catalytic residues in TfSGL, is conserved spatially in CpSGL, and in the complexes of both TfSGL and CpSGL, the conformations and binding modes of the substrates are similar. A structural comparison with other GH enzymes resulted in GH144 and GH162 being categorized into a new GH clan, clan GH-S (36). Although the transglycosylation domain of TiCGS ($TiCGS_{Tg}$) (GH189) shows structural similarity with GH144 and GH162 enzymes, GH189 was not classified as a clan GH-S-related enzyme because of the

differences in the reaction mechanisms of GH189 and GH144 and GH162 (36). Homology searching using CpSGL and TfSGL revealed that several groups were found around the GH144, GH162, and GH189 families like steppingstones, and these groups, along with GH144, GH162, and GH189, were classified as a superfamily. In the present study, we characterized homologs in the newly found groups in the superfamily, biochemically and structurally, and proposed that these groups be classified as new GH families.

Results and Discussion

Superfamily of SGLs

To understand the evolutionary relationship between GH144, GH162, and GH189, a PSI-BLAST search was performed using CpSGL (GH144), TfSGL (GH162), and $TiCGS_{Tg}$ (GH189) as queries. Homologs with very low amino acid sequence identity with CpSGL, TfSGL, and $TiCGS_{TG}$ were collected. Because these homologs do not belong to the GH144, GH162, or GH189 families, the PSI-BLAST search was repeated and the related homologs were collected. Then, a phylogenetic tree was constructed to investigate the phylogenetic relationships between the collected homologs. Homologs were extracted so as not to show more than a certain amino acid sequence identity with the others. The identity cut-off values were set so the number of homologs in each group did not exceed 250. Consequently, four groups were visualized around GH144, GH162, and GH189 as a phylogenetic tree and these groups were named Groups 1–4 (Fig. S1). Considering the scale in Fig. S1, these groups were clearly separated from each other. These results suggested that GH144, GH162, GH189, and Groups 1–4 form a superfamily.

GH189 is the largest group in the superfamily, and GH144 and GH162 are much larger families than were registered in CAZy database as of Jan 2024. While the GH162 proteins were distributed almost exclusively in eukaryotes, the homologs in the superfamily, except for GH162, were distributed widely in bacteria, including uncultured bacteria. Although the species were not clearly separated by the phylogenetic groups, Group 2 proteins, including a homolog from *Sanguibacter keddieii* (KEGG locus tag, Sked_30460; SkSGL), were mainly distributed in Gram-positive bacteria (61). Groups 1 and 3 contained proteins from many marine bacteria and bacteria isolated from marine, sediment, and mudflat environments (62–64). Group 1 was a smaller group than Groups 2 and 3, and Group 4 was the smallest in the proposed superfamily. According to classification by InterPro (https://www.ebi.ac.uk/interpro/) (65, 66), the superfamily proteins, except those in Group 4, were classified as glycoamylase by Pfam (67), while Group 4 proteins could not be classified into any category even by a conserved domain search in NCBI (68) or Pfam (67).

To evaluate the reliability of the branching of the groups in the phylogenetic tree, several proteins were manually selected in each group to perform a bootstrap test using 1000 replications. Groups 1–4 were clearly separated from each other (Fig. 1), which was consistent with the fact that the amino acid sequence identities between the homologs from two different groups were very low (Fig. S2A). Most of the amino acid sequence identity values were less than 20%. The bootstrap probability of the branching points for Group 2, GH162, and at the root of Group 2 were low (6%, 45%, and 22%, respectively), and these low values implies that the order of branching among these groups was not clear.

Targets for analysis in the superfamily

Gene clusters for potential β-1,2-glucan-associated enzymes, such as SOGPs and BGLs that prefer Sop_ns as substrates, were found in the genes encoding the proteins in the superfamily. Intriguingly, *Photobacterium gaetbulicola*, a bacterium isolated from mudflats (69), possesses homologs from multiple groups in the superfamily: PgSGL1 and PgSGL2 from Group 1; PgSGL3 from Group 3; and PgSGL4 from Group 4 (KEGG locus tags; H744_1c0224, H744_2c1936, H744_1c0222, and H744_1c0194, respectively) (Fig. 2).

The PgSGL4 gene forms a gene cluster with genes encoding a GH94 homolog (H744_1c0196, hereafter, H744_ is omitted), an ABC transporter homolog (1c0198–0200), GH3 homologs (1c0192, 1c0202, and 1c0206), a GH43 (subfamily 28) homolog (1c0207), and a Lac I family protein homolog (1c0197). Although there have been no reports on the characteristics of GH94 homologs in the 1c0196 subgroup, the structures and functions of SOGPs from *L. innocua* and *L. phytofermentans* have been reported (37, 48). SOGPs are monomeric enzymes that consist of three major domains, while 1c0196 lacks an *N*-terminal domain and is assumed to form a dimer in a similar manner to other GH94 enzymes, such as cellobiose phosphorylases and chitobiose phosphorylases (70, 71).

1c0192 and 1c0202 are close homologs of GH3 BGLs from *L. innocua* and *B. thetaiotaomicron* that prefer Sop_ns as substrates (46, 47), which suggested that these GH3 homologs from *P. gaetbulicola* were Sop_ns-preferring BGLs. The structure of 1c0205 was the closest to the GH3 BGL from *Kluyveromyces marxianus* that has a wide substrate preference, of all the structurally available homologs (72).

GH43 is a large family mainly containing β-xylosidases and α-L-arabinofuranosidases, which is divided into 39 subfamilies (73–75). A homolog in subfamily 28 from a metagenome has been reported to show hydrolytic activity toward arabinoxylooligosaccharides (76). There have been no reports on the biochemical properties of close homologs of the solute-binding subunit (1c0198) in the ABC transporter. 1c0193 and 1c0203, which are sulfatase homologs, belong to the S1 group (Table S1), suggesting that these proteins are involved in the metabolism of sulfated sugars.

The PgSGL1 and PgSGL3 genes are located near the PgSGL4 gene in the *P. gaetbulicola* genome. These genes formed a gene cluster with several putative sulfatase genes (Fig. 2), although PgSGL2 gene is located independently in the genome. According to SulfAtlas (https://sulfatlas.sb-roscoff.fr/sulfatlas/index.html) (77, 78), a database of sulfatases, 1c0221, 1c0223, 1c0227, and 1c0228 belong to the S1 group (Table S1). Enzymes in the S1 group are believed to be involved in the metabolism of sulfated carbohydrates to acidic sugars by acting on the sulfate groups (79, 80). 1c0214, 1c0216, and 1c0217 also belong to the S1 group, implying that the gene cluster 1c0206–1c0219 is also involved in the metabolism of sulfated carbohydrates. However, there was no gene encoding a sulfoglycosidase homolog (GH4, GH31, GH109, or GH188) in this region (81–84). Overall, PgSGLs 1–4 are expected to be enzymes involved in β-1,2-glucan metabolism. Considering the large number of sulfatase homologs in the region (1c0192–1c0228), sulfated β-1,2-glucans might exist in nature,

although to the best of our knowledge, there have been no reported examples.

Endozoicomonas elysicola, a bacterium isolated from the gastrointestinal tract of a mollusk, the sea slug *Elysia ornate*, possesses a PgSGL1 homolog (NCBI accession number, WP_026258326.1; EeSGL1) and the structure of EeSGL1 has been analyzed (85, 86). *E. elysicola* possesses the same gene cluster, including PgSGL3 and PgSGL4 homologs, as that of *P. gaetbulicola* (Fig. 2).

In Group 2, the function of SGR_2427 from *Streptomyces griseus* can be speculated based on the components of the gene cluster of the *SGR_2427* gene. This gene cluster includes the gene encoding SGR_2426, a putative BGL belonging to GH1, and a solute-binding protein in an ABC transporter homologous with the SO-BP from *L. innocua*. Such gene configurations are found in various genomes, including those of *Microbacterium testaceum* StLB037 and *Nakamurella multipartita* (DSM 44233). These facts suggested that SGR_2427 is likely to be a β-1,2-glucan-associated enzyme. Unfortunately, SGR_2427 could not be obtained as a soluble protein (data not shown). An SkSGL encoding gene is located independently from the putative β-1,2-glucan-associated genes. SkSGL possesses additional domains with unknown functions at the *N*- and *C*-termini. A GH144 homolog from *Parabacteroides distasonis*, a common bacterium in the large intestine (87), with unknown-function domains at the *N*-terminus, was found to be a novel exo-type Sop_2-releasing enzyme. The domains played a critical role in the exolytic degradation activity of the enzyme(49). Therefore, SkSGL was used as a target for Group 2.

A Michaelis complex structure has been reported for GH162 but not for GH144. To compare the functions, structures, and reaction mechanisms of the homologs characterized in the present study, a GH144 enzyme from *Xanthomonas campestris* pv. *campestris*, a phytopathogen (88, 89) (KEGG locus tag, XCC2207; XcSGL), was used to obtain a Michaelis complex with β-1,2-glucans.

General properties of the target homologs in the new groups

Because hydrolytic activity toward β-glucans was detected for PgSGL1, PgSGL2, PgSGL3, EeSGL1, SkSGLc (see the Experimental procedures section for details), and XcSGL, the general properties of these proteins were investigated using β-1,2-glucans ($NaBH_4$) (see the Experimental procedures section for details) as substrates. While many of the enzymes showed optimal pH values around neutral pH, the optimal pH values of PgSGL3 and XcSGL were 8.5 and 5.0, respectively, which suggested that the optimal pH value varied among the superfamily (Figs. S3–8). These enzymes were stable at pH ranges including neutral pH, which was consistent with the results for the optimal pH values. The optimal temperatures were 30–50 °C and the enzymes were stable at least up to 30 °C. The assay for PgSGL2 was performed in the presence of NaCl because PgSGL2 lost its activity without NaCl. The activity of PgSGL2 increased with increasing concentrations of NaCl in the reaction solution up to 500 mM (Fig. 3A). To maintain the maximum activity of PgSGL2, at least 250 mM NaCl was needed in the PgSGL2 solution (Fig. 3B). PgSGL4 did not show activity toward β-1,2-glucans nor Sop_ns (data

not shown). Size-exclusion chromatography was performed using PgSGL1, PgSGL3, and SkSGLc. PgSGL1 and PgSGL3, single domain enzymes, were eluted at the time corresponding 42 kDa, while the elution time of SkSGLc corresponded to 140 kDa (Fig. S9). These results indicated that PgSGL1 and PgSGL3 were monomeric enzymes and SkSGL was a dimeric enzyme. The *N*- and *C*- terminal domains in SkSGL may be involved in the dimer formation.

Characteristics of the SGLs in the new groups

The substrate specificities of the six enzymes toward various polysaccharides were investigated. All the enzymes were specific for β-1,2-glucans (Table 1). The reaction patterns of these enzymes were then investigated using β-1,2-glucans as substrates. EeSGL1 produced $Sop_{3–7}$ endolytically (Fig. 4A), as did PgSGL1 and PgSGL2 (data not shown). SkSGLc released Sop_4 mainly in the initial phase of the reaction and produced $Sop_{4–8}$ as the final products (Fig. 4B). Interestingly, PgSGL3 produced Sop_8 predominantly in the initial stage of the reaction, and then Sop_8 was further hydrolyzed (Fig. 4C left). When Sop_8 was used as a substrate, Sop_8 was hydrolyzed to $Sop_{3–5}$ (Fig. 4C right). All the enzymes were found to be endolytic enzymes, as is CpSGL. However, CpSGL releases $Sop_{2–4}$ as the final products (45). Overall, in the superfamily, the substrate specificity was the same, but the degradation patterns were different. Kinetic analysis of PgSGL1, PgSGL2, PgSGL3, EeSGL1, and SkSGLc was performed (Fig. 5). All examined enzymes showed sufficiently large V_{max} values and sufficiently small K_m values as GH enzymes (Table 2). PgSGL1, PgSGL2, EeSGL1, PgSGL3, and SkSGL were clearly identified as SGLs. XcSGL can be regarded as an SGL because of its narrow specificity for β-1,2-glucans and similar level of specific activity toward β-1,2-glucans as the other SGLs. Finally, the superfamily was named as a β-1,2-glucanase superfamily (SGL superfamily).

To determine the reaction mechanisms of the enzymes in Groups 1–3, i.e., anomer retaining or anomer inverting, the time course of the optical rotation during the hydrolysis of β-1,2-glucans by PgSGL1, SkSGLn (see the Experimental procedures section), and PgSGL3 was analyzed. This analysis is important to understand the fundamental characteristics that are shared in the new phylogenetic groups. The three enzymes (PgSGL1, PgSGL3, and SkSGLn) showed the same patterns as TfSGL and CpSGL that follow an anomer-inverting mechanism (44, 45), indicating that these three enzymes follow an anomer-inverting mechanism (Fig. 6).

Among the PgSGL1–4 enzymes, PgSGL3 and PgSGL4 possess *N*-terminal signal peptides, suggesting that both these proteins are located in the periplasm or extracellular space. Considering that PgSGL3 releases Sop_8 at the beginning of the reaction, *P. gaetbulicola* may have a system to incorporate Sop_8 effectively. The NaCl dependence of PgSGL2 may be related to the large range of environmental conditions experienced in the mudflats where *P. gaetbulicola* is found.

Structural comparison of the SGLs in the superfamily

Ligand-free structures of EeSGL1 and PgSGL3 were obtained with 2.4 and 1.2 Å resolution, respectively (Table S2). A structure of the complex with Sop_7 was obtained with 2.5 Å resolution using XcSGL (Table S2). Overall, the structures of these enzymes are composed of a single $(\alpha/\alpha)_6$-barrel domain, similar to CpSGL and TfSGL (44, 45) (Fig. S10). SkSGL and the catalytic domain of PgSGL4 were predicted using AlphaFold2 (90). The root mean square deviation (RMSD) values between the proteins of the groups in the superfamily are less than 3.5 Å according to a homology search using the DALI server (http://ekhidna2.biocenter.helsinki.fi/dali/) (91, 92) (Fig. S2B), suggesting that the overall structures are similar for the proteins in the superfamily.

The shapes of the substrate pockets in EeSGL1 and PgSGL3 are similar to those of TfSGL and XcSGL in that the Sop_7 molecules in TfSGL and XcSGL are accommodated in the pockets without obvious steric hindrance. Unfortunately, the substrate-binding structures of EeSGL1, SkSGL, and PgSGL3 could not be determined experimentally. Therefore, we generated substrate-binding structures computationally using molecular dynamics (MD) simulations. In the final structures of the MD simulations with Sop_8 as a substrate, the substrate is well-fitted to the pockets of EeSGL1 and PgSGL3 without appreciable structural changes (Fig. S10). The same is true for SkSGL predicted by AlphaFold2. In the absence of a chloride ion in PgSGL3, the position of the 3-OH group of the Glc moiety at subsite +2 is deviated far from that for a putative general acid, which is inappropriate for the catalytic reaction (Fig. S11).

The conservation of residues in each group was investigated using ConSurf (93, 94). The residues forming the substrate pockets are highly conserved in each group, except Group 4, suggesting that the substrate specificity and reaction mechanisms will be the same in each group (Fig. 7). However, the residues involved in substrate recognition that are conserved in each group are different across the groups in the superfamily (Fig. S12). According to the structure of PgSGL4 predicted by AlphaFold2, the substrate pocket of PgSGL4 is smaller than the pockets of the other enzymes investigated in the present study and is too small to accommodate Sop_7, unlike in the complexes of Sop_7 with XcSGL and TfSGL (Figs. 7 and 8). This observation is consistent with the absence of any activity of PgSGL4 toward β-1,2-glucans.

Conserved residues throughout the superfamily

A combination of multiple sequence alignments and superimposition of tertiary structures enabled the important residues shared throughout the superfamily to be identified (Figs. 8 and 9). These residues are E239, Y367, and F286 in XcSGL. E239 corresponds to E262 in TfSGL, which is an acid catalyst. Y367 forms a stacking interaction with a Glc moiety at subsite −3. F286 is a residue located close to a potential nucleophilic water. Although F286 is substituted for a histidine residue that is highly conserved in GH162 (H316 in TfSGL) (44), GH162 is the only family found in eukaryotes. Thus, F286 is conserved in SGL-superfamily homologs from prokaryotes. Overall, the residues

corresponding to E239, Y367, and F286 in XcSGL can be used to define the SGL-superfamily proteins.

The reaction mechanisms of the SGL superfamily

The similarities and differences in the reaction mechanisms between Groups 1–3 and previously reported families (GH144, GH162, GH189) found in the SGL superfamily were investigated. E262 in TfSGL (GH162), a general acid catalyst, is conserved sequentially and spatially in the other groups in the SGL superfamily as described above (Figs. 8 and 9). This residue provides a proton to the oxygen atom in the glycosidic scissile bond through the 3-hydroxy group of the Glc moiety at subsite +2 (44) (Figs. 8A and 10C). In EeSGL1 and SkSGL, the hydrogen atom of the 3-hydroxy group of the Glc moiety at subsite +2 faces the glycosidic bond oxygen atom, according to the MD simulations, which is appropriate for the catalytic reaction, while the position of the corresponding hydrogen atom with PgSGL3 is not suitable (Fig. S13). Both the E214Q mutant (PgSGL3) and E221Q mutant (EeSGL1) had considerably decreased hydrolytic activity toward β-1,2-glucan [less than 0.1% of the specific activity of the wild-type (WT)]. Substitution of the corresponding residues in CpSGL, TfSGL, and $TiCGS_{Tg}$ (E211Q, E262Q, and E1356Q, respectively) also drastically reduced the catalytic activity (36, 44, 45). Thus, E221 (EeSGL1) and E214 (PgSGL3) are considered to be general acid catalysts, and E246 (SkSGL) is expected to be a general acid catalyst as well.

In contrast, D446 acts a general base in TfSGL (GH162) and this residue corresponds to a hydrophobic residue in CpSGL (GH144), indicating that the positions of the general base catalysts are different between TfSGL and CpSGL (Fig. 8G) (44, 95). However, there are several acidic amino acid residues as candidates for the general base catalyst (45). Thus, the general base catalyst in GH144 has not been determined yet. The mutation of D135, D139, and E142 drastically decreased the hydrolytic activity, while the mutation of D400 retained sufficient hydrolytic activity (45), although D400 is conserved in many of the groups in the SGL superfamily (Figs. 8, 9, and S12). D135 is located far from subsite −1, as determined from superimposition of the structure with the XcSGL-Sop_7 complex, and is not conserved in the SGL superfamily (Figs. 8, 9). D139 is also substituted for non-dissociative residues in the other groups in the superfamily. In contrast, E142 is conserved as an acidic residue in all the enzymes, except for PgSGL4 (Group 4) and TfSGL (GH162). Both the D149N mutant (EeSGL1) and D148N mutant (PgSGL3) showed less than 0.1% of the hydrolytic activity of the respective WT enzymes. These results suggested that the residues corresponding to E142 of CpSGL are the most plausible candidates for the general base in GH144 and Groups 1–3. These residues should attract a proton from a 4-hydroxy group to activate a nucleophilic water through a chain of several other water molecules in a particular order. Such a long reaction pathway is rare among GH families, and water molecules for proton transfer were not observed in the structure of the XcSGL complex (Fig. 8B). However, because the α-face of the Glc moiety at subsite −1 is sequestered from the solvent by the substrate, this is the only route possible for proton transfer, which occurs via the

Grotthuss mechanism as has been shown for GH6, GH101, GH136, GH162, GH186, and GH189 (44, 52, 96–100) (+ref Tanaka). In addition, these spaces sequestered by the substrate are somewhat hydrophobic according to the hydrophobic surface representation of Eisenberg et al. (101) (https://web.expasy.org/protscale/pscale/Hphob.Eisenberg.html) (Fig. S14). Recently, OpgD from *E. coli* was found to be a GH186 SGL that showed no amino acid sequence or structural similarity with the SGL superfamily enzymes (52). In this enzyme, the proton transfer route from a nucleophilic water to a general base is via two water molecules and the 4-hydroxy group of the Glc moiety at subsite −1 (Fig. 10A). Furthermore, the three water molecules, including the nucleophilic water, is sequestered from the solvent (52). In the GH186 enzyme from *X. campestris*, the substrate participates in sequestering water molecules for proton transfer (102). Taking the reaction mechanism of GH186 enzymes into account, the reaction route for the Groups 1–3 SGLs shown in Fig. 10A is plausible.

For TiCGS, there is clear evidence that only the transglycosylation domain ($TiCGS_{Tg}$) has transglycosylase activity. Structural analysis of $TiCGS_{Tg}$ suggests that E1356 is an acid/base catalyst and that E1442 is a nucleophile catalyst (36) (Fig. 10B). While E1356 shares its position with the other enzymes in the SGL superfamily as a glutamate residue, the position of the nucleophile is unique in the SGL superfamily (Figs. 8 and 9). $TiCGS_{Tg}$ follows an anomer-retaining mechanism, unlike the homologs in the other SGL superfamily groups, except for Group 4. D1127 can be well superimposed with E142 in CpSGL, implying that there is an evolutional relationship among members of the SGL superfamily. However, L399 in CpSGL, a residue that creates a hydrophobic environment for proton transfer, is replaced with a histidine residue in $TiCGS_{Tg}$ (H1536) and this histidine residue is highly conserved in the GH189 family. The substitution of this residue may be related to the differences in the reaction mechanisms of SGL and CGS.

PgSGL4 has no acidic residue corresponding to E142 in CpSGL according to multiple sequence alignment (Fig. 9) and E398 is located near E142 in CpSGL (Fig. 8). These observations imply that E398 is a candidate for a catalyst, although further investigation is needed.

Conclusion

In this study, we found four groups with unknown functions that were related to GH144, GH162, and GH189, and identified that together these seven groups formed a superfamily. The biochemical functions were determined in three of the four newly identified groups. The three groups with identified functions are believed to share a reaction mechanism with GH144 but not with GH162 or GH189. Multiple sequence alignment revealed that only three residues were mostly conserved in the SGL superfamily. The amino acid sequence identities in the SGL superfamily were mostly less than 20% and Groups 1–4 were placed irregularly in the phylogenetic tree (Figs. 1 and S1).

An example of several GH families forming a superfamily is found with the GH94, GH149, and

GH161 families of glycoside phosphorylases (103–105). In this superfamily, the catalytic residues, inorganic phosphate recognition residues, and subsite −1 recognition residues are conserved (103), suggesting that the reaction mechanism is the same for all the members of this glycoside phosphorylase superfamily. In addition, the amino acid sequence identities in these three families are 16%–20%. Considering the criteria for classification, Groups 1–3 with functions unveiled in this study defines new GH families, GHxxx, GHyyy, and GHzzz, respectively. In addition, the overall structural similarity between the enzymes in the superfamily and the common reaction mechanism between the new families and GH144 suggest that the new GH families belongs to the clan GH-S.

The extensive expansion of the distribution of SGLs found in this study suggests the importance of β-1,2-glucans and related enzymes in nature. Groups with the same reaction mechanism in the SGL superfamily are dispersed phylogenetically, implying that complicated molecular evolution pathways have occurred in this superfamily. The identification of this superfamily is an example of the enormous diversity of carbohydrate-active enzymes and will encourage the further exploration of the large number of unknown carbohydrate-active enzymes.

Experimental procedures

Materials

Glucomannan, polygalacturonic acid, carboxymethyl curdlan, lichenan, arabinogalactan, barley β-glucan, tamarind xyloglucan, arabinan, and beechwood xylan were purchased from Neogen (MI, USA). Laminarin and carboxymethyl cellulose were purchased from Merck (NJ, USA). Soluble starch and pustulan were purchased from FUJIFILM Wako Pure Chemical Corporation (Osaka, Japan) and CarbioChem (CA, USA), respectively.

β-1,2-Glucans with an estimated average DP25 calculated by NMR (37) and average DP121 and DP17.7 calculated from the number average molecular weight (39, 40, 52) [β-1,2-Glucans (DP25), (DP121), and (DP17.7)] were prepared as reported previously. β-1,2-Glucans (DP121) used for colorimetric assay were treated with $NaBH_4$ [β-1,2-glucans ($NaBH_4$)] to reduce the colorization derived from the anomeric hydroxy groups at the reducing end of the polysaccharide (106).

Sequence analysis

The amino acid sequences of TfSGL (GH162) and CpSGL (GH144) were used as queries for a PSI-BLAST search to collect 1000 and 5000 homologs, respectively. PgSGL1, PgSGL3, SkSGL, and CGSs were included in this collection. Then, a BLASTP search was performed using PgSGL1, PgSGL3, SkSGL (residues 72–541 a.a.), CGS from *B. abortus* (residues 1009–1515 a.a.; Genbank accession number, ACD71661.1), and a homolog from *Phycisphaerae bacterium* (Genbank accession number, MBN1513595.1) with the settings to collect up to 5000 homologs for the SkSGL and CGS groups and up to 1000 homologs for the other groups. Because the CGS group contained too many homologs in the database for the search, several organisms with many species were eliminated as targets for the search. The homolog from *P. bacterium* was used as a query because this homolog is located at the far side from CpSGL according to a phylogenetic tree analysis of GH144 (45). Homologs were collected until no new group was obtained. During this collection, the PgSGL4 group, a small group, was found. The number of homolog hits found initially by BLASTP using a query in each group was approximately 5000 for GH144 CpSGL, 4100 for GH144 (MBN1513595.1), 430 for GH162 TfSGL, 1000 for SkSGL, 330 for the SkSGL group (MBN1384282.1), 480 for PgSGL3, 140 for PgSGL1, and 4200 for CGS. Because the GH144 and CGS groups were large, the GH144 proteins and GT84 domains of CGSs in the CAZy database (8, 9) were also collected (1604 and 1930 homologs, respectively). Then, the duplicated sequences were removed from all the collected sequences. Multiple sequence alignment was performed against each group using Clustal Omega (107). The transglycosylation domain and catalytic domain were extracted for GH189 and Group 4, respectively. Then, homologs showing less than 65%−95% amino acid sequence identity with the other homologs in their groups were extracted. The cut-off values were 95%, 95%, 66%, 65%, 55%, 95%, 68% for the groups of PgSGL1, PgSGL3, SkSGL, SkSGL group (MBN1384282.1), CGS, GH162, GH144, and

the homolog from *P. bacterium*, respectively. The latter two groups were combined into the GH144 group. Because the newly found PgSGL4 group was small, all the homologs in this group were used for phylogenetic analysis. Eleven homologs in the PgSGL3 group were transferred to the PgSGL4 group, and one homolog in the SkSGL group was transferred to the PgSGL1 group based on the phylogenetic positions. Finally, 20–220 homologs were obtained from each group (PgSGL1, 42; PgSGL3, 148; SkSGL, 186; PgSGL4, 20; CGS, 144; GH144, 212; and GH162, 168). All the extracted homologs containing the samples used as queries were aligned by multiple alignment using MUSCLE (108). Then, a phylogenetic tree was constructed using the maximum-likelihood method, which was visualized by MEGA11 (109, 110). To simplify the tree, approximately five homologs, including the samples used in the present study, were taken and the tree was constructed based on the maximum-likelihood method using 1000 bootstraps. To investigate conservation of the substrate recognition and catalytic residues in the SGL superfamily, structure-based multiple sequence alignment was performed using the SALIGN server (https://modbase.compbio.ucsf.edu/salign/) (111). This alignment was visualized using the ESPript 3.0 server (http://espript.ibcp.fr/ESPript/ESPript/) (112).

Cloning, expression, and purification

The genomic DNA of *P. gaetbulicola* (DSM 26887), *E. elysicola* (DSM 22380), and *X. campestris* pv. *campestris* (DSM 3586) was purchased from the Leibniz Institute DSMZ (Leibniz, Germany) and *S. keddieii* ST-74 (ATCC 51767) was purchased from the American Type Culture Collection (VT, USA). Genes coding PgSGL1 (KEGG locus tag, H744_1c0224), PgSGL2 (KEGG locus tag, H744_2c01936), PgSGL3 (KEGG locus tag, H744_1c0222), PgSGL4 (KEGG locus tag, H744_1c0194), EeSGL1 (NCBI accession number, WP_026258326.1), SkSGL (KEGG locus tag, Sked_30460), and XcSGL (KEGG locus tag, XCC2207) were amplified by PCR chain reaction using KOD plus (TOYOBO, Osaka, Japan) or PrimeSTAR Max (Takara Bio, Shiga, Japan) for the DNA polymerase and the primer pairs. All the primers used in this study were designed not to include *N*-terminal signal peptides as shown in Table S3. The *N*-terminal signal peptides were predicted using SignlaP5.0 or SignalP4.1 (113, 114). The amplified PCR products were digested by the restriction enzymes shown in Table S3. Vectors (pCold I and pET30a) were also digested by the corresponding restriction enzymes suitable for ligation. The digested PCR products of PgSGL1, PgSGL2, and EeSGL1 were inserted into pCold I (Takara, Shiga, Japan) to fuse a His_6-tag at the *N*-terminus. The digested PCR products of PgSGL3 and XcSGL, and SkSGLc were inserted into pET30a and pET24a (Merck), respectively, to fuse a His_6-tag at the *C*-terminus. Ligation high ver. 2 (TOYOBO, Osaka, Japan) was used for the insertion.

Mutations used in this study were introduced using PrimeSTAR Max according to the manufacturer's instructions using the primer pairs shown in Table S3. The region coding the putative *N*-terminal signal peptide of SkSGL was removed using PrimeSTAR Max and the primer pair (SkSGL Δsignal Fw and Rv). To move the *C*-terminal His_6-tag of SkSGLc to the *N*-terminus, the gene coding

SkSGLc and the pCold I vector were amplified using KOD One (TOYOBO) and the primer pairs (SkSGL pCold Fw and Rv, and pCold I Fw and Rv, respectively). The former amplified PCR product was inserted into pCold I (the latter amplified PCR product) by recombination using the SLiCE method (115). The amplified PCR product of the gene coding PgSGL4 was inserted into pColdI by the same method as for SkSGLn. The resulting products were transferred into JM109 or XL1-blue and then the plasmids were extracted, and the target DNA sequences were checked as described above. This SkSGL fused with a His_6-tag at the *N*-terminus was called SkSGLn.

The constructed plasmids were transferred into *E. coli* BL21(DE3). The recombinant BL21(DE3) were cultured in Luria-Bertani medium to produce target proteins. The strains with the target gene inserted into pET30a were cultured in medium containing kanamycin (30 μg/ml) at 37 °C with vigorous agitation (120 rpm) until the OD_{600} value exceed approximately 0.8 and then isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 0.1 mM. The cells were further cultured at 20 °C overnight. The strains with the target gene inserted into pCold I were cultured in medium containing ampicillin (100 μg/ml) at 37 °C with vigorous agitation (120 rpm) until the OD_{600} value exceeded approximately 0.6. Then, the cells were left at 10–15 °C for 30 min. IPTG was added to a final concentration of 0.1 mM and the cells were agitated again at 10–15 °C overnight. The cells were collected by centrifugation at 6000 × *g* for > 5 min and the cells were suspended in an appropriate volume of 30–50 mM MOPS (pH 7.0–7.5) or Tris-HCl (pH 7.5). The suspended cells were disrupted by sonication and the supernatants were collected by centrifugation at 33000 × *g* for > 10 min. The supernatants, except for SkSGLc, were purified by the HisTrap™ crude method (Cytiva). Each sample was loaded onto a column equilibrated with an equilibration buffer composed of 20−50 mM MOPS (pH 7.5) or Tris-HCl (pH 7.5) and 200–500 mM NaCl. The column was washed with the equilibration buffer containing 10–15 mM imidazole until almost all the unbound compounds were eluted. The samples were eluted with 50 ml of a linear gradient of up to 500 mM imidazole. For SkSGLc, other chromatography methods were used because the protein did not bind to a HisTrap™ crude column (5 ml; Cytiva, MA, USA). The cells were suspended with 50 mM Tris-HCl (pH 7.5). The supernatant was loaded onto tandemly connected HiTrap™ DEAE FF (5 ml; Cytiva) columns equilibrated with 20 mM Tris-HCl (pH 7.5). After the unbound compounds were almost all eluted with the buffer, the target protein was eluted with 100 ml of the buffer with linear gradient of 0–1 M NaCl at a flow rate of 3 ml/min. The fractions containing the target protein were collected and mixed with an equivalent volume of 20 mM Tris-HCl (pH 7.5) containing 60% saturation ammonium sulfate. After the solution was centrifugated at 33000 × *g* for 20 min, the supernatant was loaded onto a HiTrap™ Butyl HP column (5 ml, Cytiva) equilibrated with 20 mM Tris-HCl (pH 7.5) containing 30% saturation ammonium sulfate. The samples were eluted with a linear gradient of 30%–0% saturation ammonium sulfate. The fractions containing the target protein were collected. The activity in the fractionated samples was detected by TLC.

For crystallization of the XcSGL E239Q mutant, the enzyme was purified by nickel affinity chromatography and further purified by hydrophobic interaction chromatography. The fractionated eluates from the HisTrapFF column were mixed with an equivalent volume of 50 mM MOPS (pH 7.0) containing 50% saturation ammonium sulfate. The sample was loaded onto a HiTrap™ Butyl column equilibrated with 50 mM MOPS (pH 7.0) containing 25% saturation ammonium sulfate. The sample was eluted with a linear gradient of 25%–0% saturation ammonium sulfate. The fractions containing the target protein was pooled and dialyzed against 5 mM MOPS (pH 7.0). For PgSGL3 and EeSGL1, the extra purification steps were not used.

The homogeneity of all the target proteins was checked using SDS-PAGE (Fig. S15). DynaMarker Protein MultiColorIII (BioDynamics Laboratory Inc., Japan) was used for protein standards. Each target protein was buffered and concentrated in 5 mM MOPS (pH 7.0) by ultrafiltration with AmiconUltra (Millipore, MA, USA) or VivaSpin (Sartorius, Germany) with molecular weight cut-off values of 10000 or 30000 at 4000 × *g*. For PgSGL2, a buffer containing 5 mM MOPS (pH 7.0) and 150 mM NaCl was used. The protein concentrations were calculated using the theoretical molecular masses and the extinction coefficients at an absorbance of 280 nm according to Pace et al. (116) and are summarized in Table S4.

Temperature and pH profiles

To investigate the optimum pH for each enzyme, 0.2% β-1,2-glucan ($NaBH_4$) was incubated with an appropriate concentration of each enzyme in 50 mM buffer (and 500 mM NaCl in the case of PgSGL2) at 30 °C. To investigate the optimal temperature, 0.2% β-1,2-glucan ($NaBH_4$) was incubated with an appropriate concentration of each enzyme in the assay buffer (50 mM bicine pH 8.5 for PgSGL1; 50 mM MES, pH 6.5 containing 500 mM NaCl for PgSGL2; 50 mM MOPS pH 7.0 for EeSGL1; 50 mM cacodylate pH 6.5 for PgSGL3; 50 mM HEPES, pH 8.0 for SkSGLc; and 50 mM sodium acetate, pH 5.0 for XcSGL). To examine the pH stability, PgSGL1 (10.9 μg/ml), PgSGL2 (0.284 mg/ml), EeSGL1 (4 μg/ml), PgSGL3 (8 μg/ml), and SkSGL (5.2 μg/ml) were incubated with 50 mM buffer solutions (5 mM for EeSGL1 and PgSGL3) at 30 °C for 1 h. Then, the enzymatic reaction was performed in the presence of 0.2% β-1,2-glucan ($NaBH_4$), an appropriate concentration of each enzyme, and the corresponding assay buffer. To examine the temperature stability, PgSGL1 (0.0109 mg/ml), PgSGL2 (0.284 mg/ml), EeSGL1 (4 μg/ml), PgSGL3 (8 μg/ml), and SkSGL (5.2 μg/ml) were incubated in the corresponding assay buffer (except that 5 mM MOPS pH 7.0 and 500 mM NaCl was used for PgSGL2) at various temperatures for 1 h. Then, each enzyme diluted to an appropriate concentration was incubated in the substrate solution containing 0.2% β-1,2-glucan ($NaBH_4$) and the corresponding assay buffer at 30 °C. The incubation times for the enzymatic reactions were 10 min for PgSGL2, PgSGL3, and EeSGL1; 20 min for PgSGL1; and 30 min for SkSGL and XcSGL. The reaction products were colorized by the 3-methyl-2-benzo-thiazolinone hydrazine (MBTH) method described below. The

assays were performed in triplicate.

Size-exclusion chromatography

Superdex™ 200 (HiLoad 16/60; Cytiva) was equilibrated with 50 mM Tris-HCl (pH 8.0) containing 0.15 M NaCl, and each sample (0.8 mg of EeSGL1, 500 µl; 0.2 mg of PgSGL3, 500 µl) or a marker mixture (500 µl) was loaded onto the column at a flow rate of 0.5 ml/min. Ovalbumin (43 kDa), conalbumin (75 kDa), aldolase (158 kDa), ferritin (440 kDa), and thyroglobulin (669 kDa) were used as protein markers. Blue dextran 2000 was also added to the protein markers to determine the void volume of the column. The amounts of ferritin and aldolase were 0.1 and 1.5 mg, respectively, and the amounts of the other marker proteins and blue dextran 2000 were 0.5 mg. For SkSGLc, the 50 mM Tris-HCl (pH 8.0) was replaced with 20 mM Tris-HCl (pH 7.5). Elution of the samples was performed at a flow rate of 0.3 mg/ml. The molecular weights of EeSGL1, PgSGL3, and SkSGLc were calculated using equation 1,

$$K_{\mathrm{av}} = (V_e - V_{\mathrm{o}})/(V_t - V_{\mathrm{o}}) \qquad \text{(Equation 1)}$$

where K_{av} is the gel-phase distribution coefficient; V_{e} is the volume required to elute each protein; V_{o} is the volume required to elute bule dextran 2000; and V_{t} is the bed volume of the column.

Substrate specificity and kinetic analysis

Enzymatic reactions were performed in the assay buffer containing each substrate at 30 °C for the times described above. The assays were performed in triplicate using the assay buffers described in the Temperature and pH profiles section. The relative activity when 0.2% β-1,2-glucan ($NaBH_4$) was used as the substrate were determined for each candidate catalytic residue mutant. For kinetic analysis, the median values were plotted and fitted with a Michaelis-Menten equation (equation 2) using R.

$$v/[E]_{\mathrm{T}} = (k_{\mathrm{cat}}[S])/(K_{\mathrm{m}} + [S]) \qquad \text{(Equation 2)}$$

where v is the initial velocity; [E] is the enzyme concentration; k_{cat} is the turnover number; [S] is the β-1,2-glucan concentration; and K_{m} is the Michaelis constant.

MBTH method

The reducing power in the reaction mixtures was determined using the MBTH method (106). The reaction solution (20 µl), a solution containing 1 mg/ml DTT and 3.0 mg/ml MBTH (20 µl) and 0.5 M NaOH (20 µl) were mixed and then the mixture was heated at 80 °C for 30 min. After the mixture was cooled to room temperature, 40 µl of a solution containing 0.5% $FeNH_4(SO_4)_2 \cdot 12H_2O$, 0.5% sulfamic acid and 0.25 M HCl, and 100 µl water was added. The mixtures (175 µl) were poured into a 96-well microplate (EIA/RIA plate, 96-well half area, Corning, NY, USA) and then the absorbance at 620 nm was measured. Sop_2 was used as a standard for assaying the activity toward β-1,2-glucan ($NaBH_4$) and Glc was used instead of the other substrates.

TLC analysis

The components of the reaction mixtures of EeSGL1, PgSGL3, and SkSGLc were 50 mM MOPS (pH 7.0), 0.2% β-1,2-glucan (DP121), and 4 μg/ml of EeSGL1; 20 mM MOPS (pH 7.0), 0.2% β-1,2-glucan (DP121) or Sop_8, and 0.025–1 mg/ml PgSGL3; and 20 mM MOPS (pH 7.5), 0.2% β-1,2-glucan ($NaBH_4$), and 8.6 μg/ml of SkSGLc, respectively. The reactions were performed at 30 °C for 20 min (PgSGL3) or an appropriate time as shown in Fig. 4 (EeSGL1 and SkSGLc). The reaction was stopped by heating the sample at 80 °C for PgSGL3 or 100 °C for the other enzymes, for 5 min. The reaction mixtures and markers were spotted onto TLC Silica Gel 60 F_{254} (Merck) plates. After the plates were developed with 75% acetonitrile twice or more, the plates were soaked in a 5% (w/v) sulfuric acid/methanol solution. Spots were visualized by heating the plates in an oven. The β-1,2-glucooligosaccharide marker used for lane M1 of PgSGL3 was prepared by incubating a 1% mixture of $Sop_{2–5}$, which are the reaction products of CpSGL, with SOGP from *L. innocua* in the presence of 1 mM sodium phosphate (52). The Sop_ns marker used for SkSGLc was prepared by incubating a 0.5% mixture of $Sop_{3–7}$, which are the reaction products of SGL from *Chloroflexus aurantiacus* (Chy400_4164), with SOGP from *Enterococcus italicus* in the presence of 25 mM sodium phosphate (39, 40).

Polarimetric analysis of the reaction products

To determine the reaction mechanisms of PgSGL1, PgSGL3, and SkSGL, the time course of the degree of optical rotation in the reaction mixture was monitored. Approximately 10 ml of a substrate solution containing 2% β-1,2-glucan (DP25) (44, 45) and 50 mM bicine pH 8.5 was first warmed and then poured into a cylindrical glass cell (100 mm × ø 10.5 mm). The sample was placed in a JASCO P-2200 polarimeter (Jasco, Japan) until the monitored value stabilized. Then, the reaction was started by adding PgSGL1 solution (200 μl, 40.4 mg/ml) at room temperature. When the reaction velocity began to slow down, droplets of 35% aqueous ammonia were added to enhance the mutarotation of the anomers. For PgSGL3, 2.5 mM MOPS (pH 6.5) buffer was used for the substrate solution and 300 μl of 9.6 mg/ml PgSGL3 was added. For SkSGL, 1.6 ml of substrate solution containing 1% β-1,2-glucan (DP25) and 21 mM HEPES (pH 8.0) was poured into a cylindrical glass cell (100 mm × ø 3.5 mm), because it was difficult to obtain a large amount of SkSGLn or SkSGLc and the hydrolytic activity of SkSGL was decreased in the presence of a high substrate concentration. SkSGLn solution (400 μl, 0.91 mg/ml) was added to the substrate solution.

Crystallography

The initial screening of the crystallization conditions for the XcSGL E239Q mutant, PgSGL3, and EeSGL1 was performed using JCSG-plus™ HT-96 and PACT premier™ HT-96 (Molecular

Dimensions, UK). After 1 µl of protein solution (10 mg/ml) was mixed with 1 µl of reservoir solution on a sitting plate (96-well CrystalQuick plates, Greiner Bio-One, Germany), the plate was incubated at 20 °C. The crystallization conditions were optimized from the results of the initial screening by the hanging drop vapor diffusion method using VDX plates (Hampton Research, CA, USA). Finally, crystals of the XcSGL E239Q mutant were prepared by incubation of a mixture of 11.9 mg/ml of the mutant (1 µl) and 1 µl of reservoir solution composed of 20 mM $NaNO_3$, 4.5 % (w/v) PEG 3350, and 0.1 M bis-tris propane (pH 5.5) at 20 °C. For EeSGL1, the crystals obtained from the initial screening were used for data collection. The reservoir component was 0.2 M potassium thiocyanate, 0.1 M bis-tris propane (pH 6.5), and 20% (w/v) PEG 3350. For PgSGL3, a microseeding method was adopted because the crystallization was difficult to reproduce. Reservoir solution (40 mM HEPES pH 7.0, 80 mM $MgCl_2$, and 10% PEG 3350) and 11.7 mg/ml of PgSGL3 (both 1 µl) were mixed and then an appropriate amount of seed crystals created by crushing a PgSGL3 crystal were added to the drop. After incubation of this hanging drop at 20 °C for 10 min, crystals were formed of suitable size for data collection.

Each crystal was soaked in the reservoir solution supplemented with 12% (w/v) PEG 200, 8% glycerol, and 4% β-1,2-glucan (DP17.7) for the XcSGL E239Q mutant, 25% PEG 400 for EeSGL1, and 25% (w/v) PEG 400 for PgSGL3 as cryoprotectants and maintained at 100 K in a nitrogen-gas stream during data collection. All X-ray diffraction data were collected on beamlines (BL-5A and NW-12A) at Photon Factory (Tsukuba, Japan). The data reduction of the diffraction data was performed using the XDS program (117). The initial phase information was obtained by molecular replacement using the MOLREP program (118) and the structures of XcSGL and PgSGL1 predicted by AlphaFold2 (90) were used as model structures. The predicted structures were obtained from the UniProt database for the XcSGL mutant and EeSGL1. For PgSGL3, the iodide single-wavelength anomalous diffraction phasing method using the diffraction data for the iodinated PgSGL3 crystal collected at 1.8 Å was adopted (119). The program used was the Crank2 program (shelxc, shelxd, refmac, solomon, multicomb, buccaneer, parrot) in ccp4 (http://www.ccp4.ac.uk/) (120). Model building for EeSGL1 was performed using Buccaneer (121). Refinement of the structures was performed using Refmac5 (122) for automatic refinement and Coot for manual refinement (123).

MD simulations

For the MD simulations, Sop_8 was placed in the substrate binding pockets of EeSGL1, SkSGL, and PgSGL3, referring to the complex structure of TfSGL [Protein Data Bank (PDB) ID, 6IMW]. The crystal structures of EeSGL1 and PgSGL3, and the predicted structure of SkSGL were used. To reduce calculation costs, only chain A of the EeSGL1 dimer and a catalytic domain of SkSGL (64–545 a.a.) were used for MD. The proteins and Sop_8 were protonated and placed in a dodecahedral box. The box size was determined in order that all molecules were placed at least 1.5 nm from the box edges. The

periodic boundary conditions were applied for all directions. The box was filled with water molecules. Sodium and chloride ions were added to each box to neutralize the total charge as well as to set the ion density to 10 mM. The AMBER ff14SB force field (124) and GLYCAM (125) were used to represent proteins and Sop_8, respectively. The TIP3P model (126) was used for water. After energy minimization, the constant-pressure and constant-temperature (NPT) MD simulations for equilibration were performed at 1 bar and 300 K for 200 ps. Position restraints were applied to the Cα atoms of the proteins and all heavy atoms of Sop_8 during equilibration. The production runs were performed for 100 ns at 300 K. The C-rescale method and Parrinello-Rhaman method were used to maintain the pressure during the equilibrations and the production runs, respectively. The covalent bonds of hydrogen atoms were constrained using the LINCS method, and the integration time step was 2.0 fs. MD simulations were performed by using GROMACS 2022.4 (127).

References

1. Zeeman, S. C., Kossmann, J., and Smith, A. M. (2010) Starch: its metabolism, evolution, and biotechnological modification in plants. *Annu Rev Plant Biol*. **61**, 209–34
2. Popper, Z. A., Michel, G., Hervé, C., Domozych, D. S., Willats, W. G. T., Tuohy, M. G., *et al*. (2011) Evolution and diversity of plant cell walls: From algae to flowering plants. *Annu Rev Plant Biol*. **62**, 567–590
3. Ruiz-Herrera, J., and Ortiz-Castellanos, L. (2019) Cell wall glucans of fungi. A review. *The Cell Surface*. **5**, 100022
4. Noss, I., Doekes, G., Thorne, P. S., Heederik, D. J. J., and Wouters, I. M. (2013) Comparison of the potency of a variety of β-glucans to induce cytokine production in human whole blood. *Innate Immun*. **19**, 10–9
5. Lenardon, M. D., Munro, C. A., and Gow, N. A. R. (2010) Chitin synthesis and fungal pathogenesis. *Curr Opin Microbiol*. **13**, 416–23
6. Nishimoto, M., and Kitaoka, M. (2007) Practical preparation of lacto-*N*-biose I, a candidate for the bifidus factor in human milk. *Biosci Biotechnol Biochem*. **71**, 2101–2104
7. Laine, R. A. (1994) A calculation of all possible oligosaccharide isomers both branched and linear yield 1.05 × 10(12) structures for a reducing hexasaccharides: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. *Glycobiology*. **4**, 759–767
8. Drula, E., Garron, M. L., Dogan, S., Lombard, V., Henrissat, B., and Terrapon, N. (2022) The carbohydrate-active enzyme database: Functions and literature. *Nucleic Acids Res*. **50**, D571–D577
9. 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. *Biotechnol Biofuels*. **6**, 1
10. Coutinho, P. M., Deleury, E., Davies, G. J., and Henrissat, B. (2003) An evolving hierarchical family classification for glycosyltransferases. *J Mol Biol*. **328**, 307–317
11. Henrissat, B., Sulzenbacher, G., and Bourne, Y. (2008) Glycosyltransferases, glycoside hydrolases: surprise, surprise! *Curr Opin Struct Biol*. **18**, 527–533
12. Beguin, P., and Béguin, P. (1990) Molecular biology of cellulose degradation. *Annu Rev Microbiol*. **44**, 219–248
13. Bach Knudsen, K. E. (2001) The nutritional significance of "dietary fibre" analysis. *Anim Feed Sci Technol*. **90**, 3–20
14. Nelson, T. E., and Lewis, B. A. (1974) Separation and characterization of the soluble and insoluble components of insoluble laminaran. *Carbohydr Res*. **33**, 63–74

15. Hu, X., Yang, P., Chai, C., Liu, J., Sun, H., Wu, Y., *et al*. (2023) Structural and mechanistic insights into fungal β-1,3-glucan synthase FKS1. *Nature*. **616**, 190–198
16. Yamanaka, D., Takatsu, K., Kimura, M., Swamydas, M., Ohnishi, H., Umeyama, T., *et al*. (2020) Development of a novel β-1,6-glucan-specific detection system using functionally-modified recombinant endo-β-1,6-glucanase. *J. Biol. Chem*. **295**, 5362–5376
17. Olafsdottir, E. S., and Ingólfsdottir, K. (2001) Polysaccharides from lichens: structural characteristics and biological activity. *Planta Med*. **67**, 199–208
18. Sun, J., Wei, X., and Wei, J. (2014) Research progress of the polysaccharide o-acetylated pustulan-type glucans from umbilicariaceae lichens. in *Lecture Notes in Electrical Engineering*, 10.1007/978-3-642-37925-3_178
19. Brust, H., Orzechowski, S., and Fettke, J. (2020) Starch and Glycogen Analyses: Methods and Techniques. *Biomolecules*. **10**, 1020
20. Rigano, L. A., Payette, C., Brouillard, G., Marano, M. R., Abramowicz, L., Torres, P. S., *et al*. (2007) Bacterial cyclic β-(1,2)-glucan acts in systemic suppression of plant immune responses. *Plant Cell*. **19**, 2077–2089
21. Briones, G., Iñón de Iannino, N., Roset, M., Vigliocco, A., Paulo, P. S., and Ugalde, R. A. (2001) *Brucella abortus* cyclic β-1,2-glucan mutants have reduced virulence in mice and are defective in intracellular replication in HeLa cells. *Infect Immun*. **69**, 4528–4535
22. Arellano-Reynoso, B., Lapaque, N., Salcedo, S., Briones, G., Ciocchini, A. E., Ugalde, R., *et al*. (2005) Cyclic β-1,2-glucan is a brucella virulence factor required for intracellular survival. *Nat Immunol*. **6**, 618–625
23. Javvadi, S., Pandey, S. S., Mishra, A., Pradhan, B. B., and Chatterjee, S. (2018) Bacterial cyclic β-(1,2)-glucans sequester iron to protect against iron-induced toxicity. *EMBO Rep*. **19**, 172–186
24. McIntire, F. C., Peterson, W. H., and Riker, A. J. (1942) A polysaccharide produced by the crown-gall organism. *J Biol Chem*. **143**, 491–496
25. McIntire, F. C., Peterson, W. H., and Riker, A. J. (1940) Factors influencing the carbon metabolism of the crown gall organism. *J Agric Res*. **61**, 313–320
26. Putman, E. W., Potter, A. L., Hodgson, R., and Hassid, W. Z. (1950) The structure of crown-gall polysaccharide. I. *J Am Chem Soc*. **72**, 5024–5026
27. Zevenhuizen, L. P. T. M., and Scholten-Koerselman, H. J. (1979) Surface carbohydrates of Rhizobium. *Antonie Van Leeuwenhoek*. **45**, 165–175
28. Castro, O. A., Zorreguieta, A., Ielmini, V., Vega, G., and Ielpi, L. (1996) Cyclic β-(1,2)-glucan synthesis in Rhizobiaceae: roles of the 319-kilodalton protein intermediate. *J Bacteriol*. **178**, 6043–6048
29. Iannino, N., Briones, G., Tolmasky, M., and Ugalde, R. (1998) Molecular cloning and

characterization of *cgs*, the *Brucella abortus* cyclic β(1-2) glucan synthetase gene: genetic complementation of *Rhizobium meliloti ndvB* and *Agrobacterium tumefaciens chvB* mutants. *J Bacteriol*. **180**, 4392–4400

30. Zorreguieta, A., and Ugalde, R. A. (1986) Formation in *Rhizobium* and *Agrobacterium* spp. of a 235-kilodalton protein intermediate in β-D(1-2) glucan synthesis. *J Bacteriol*. **167**, 947–951
31. Zorreguieta, A., Ugalde, R. A., and Leloir, L. F. (1985) An intermediate in cyclic β 1-2 glucan biosynthesis. *Biochem Biophys Res Commun*. **126**, 352–357
32. Nakajima, M. (2023) β-1,2-Glucans and associated enzymes. *Biologia (Bratisl)*. **78**, 1741–1757
33. Ciocchini, A. E., Guidolin, L. S., Casabuono, A. C., Couto, A. S., de Iannino, N. I., and Ugalde, R. A. (2007) A glycosyltransferase with a length-controlling activity as a mechanism to regulate the size of polysaccharides. *Proc Natl Acad Sci U S A*. **104**, 16492–16497
34. Ciocchini, A. E., Roset, M. S., Briones, G., Iñón de Iannino, N., and Ugalde, R. A. (2006) Identification of active site residues of the inverting glycosyltransferase Cgs required for the synthesis of cyclic β-1,2-glucan, a *Brucella abortus* virulence factor. *Glycobiology*. **16**, 679–691
35. Kozianowski, G., Canganella, F., Rainey, F. A., Hippe, H., and Antranikian, G. (1997) Purification and characterization of thermostable pectate-lyases from a newly isolated thermophilic bacterium, *Thermoanaerobacter italicus* sp. nov. *Extremophiles*. **1**, 171–82
36. Tanaka, N., Saito, R., Kobayashi, K., Nakai, H., Kamo, S., Kuramochi, K., *et al* (2024) Functional and structural analysis of a cyclization domain in a cyclic β-1,2-glucan synthase. *Appl. Microbiol. Biotechnol.* 108, Article number 187
37. Nakajima, M., Toyoizumi, H., Abe, K., Nakai, H., Taguchi, H., and Kitaoka, M. (2014) 1,2-β-oligoglucan phosphorylase from *Listeria innocua*. *PLoS One*. **9**, e92353
38. Abe, K., Nakajima, M., Kitaoka, M., Toyoizumi, H., Takahashi, Y., Sugimoto, N., *et al*. (2015) Large-scale preparation of 1,2-β-glucan using 1,2-β-oligoglucan phosphorylase. *J Appl Glycosci (1999)*. **62**, 47–52
39. Nakajima, M., Tanaka, N., Kobayashi, K., Nakai, H., Kimura, S., Iwata, T., *et al*. (2021) Enzymatic control and evaluation of degrees of polymerization of β-(1→2)-glucans. *Anal Biochem*. **632**, 114366
40. Kobayashi, K., Nakajima, M., Aramasa, H., Kimura, S., Iwata, T., Nakai, H., *et al*. (2019) Large-scale preparation of β-1,2-glucan using quite a small amount of sophorose. *Biosci Biotechnol Biochem*. **83**, 1867–1874
41. Reese, E. T., Parrish, F. W., and Mandels, M. (1961) β-D-1,2-Glucanases in fungi. *Can J Microbiol*. **7**, 309–317
42. Mendoza, N. S., and Amemura, A. (1983) (1,2)-β-D-Glucan-hydrolyzing enzymes in

Cytophaga arvensicola: partial purification and some properties of endo-(1,2)-β-glucanase and β-D-glucosidase specific for (1,2)- and (1,3)- linkages. *J. Ferment. Technol.* **61**, 473–481

43. Kitahata, S., and Edagawa, S. (1987) Cyclic (1→2)-β-D-glucan-hydrolyzing enzymes from *Acremonium* sp. 15: Purification and some properties of endo-(1→2)-β-D-glucanase and β-D-glucosidase. *Agric. Biol. Chem.* **51**, 2701–2708
44. Tanaka, N., Nakajima, M., Narukawa-Nara, M., Matsunaga, H., Kamisuki, S., Aramasa, H., *et al*. (2019) Identification, characterization, and structural analyses of a fungal endo-β-1,2-glucanase reveal a new glycoside hydrolase family. *J Biol Chem*. **294**, 7942–7965
45. Abe, K., Nakajima, M., Yamashita, T., Matsunaga, H., Kamisuki, S., Nihira, T., *et al*. (2017) Biochemical and structural analyses of a bacterial endo-β-1,2-glucanase reveal a new glycoside hydrolase family. *J Biol Chem*. **292**, 7487–7506
46. Nakajima, M., Yoshida, R., Miyanaga, A., Abe, K., Takahashi, Y., Sugimoto, N., *et al*. (2016) Functional and structural analysis of a β-glucosidase involved in β-1,2-glucan metabolism in *Listeria innocua*. *PLoS One*. **11**, e0148870
47. Ishiguro, R., Tanaka, N., Abe, K., Nakajima, M., Maeda, T., Miyanaga, A., *et al*. (2017) Function and structure relationships of a β-1,2-glucooligosaccharide-degrading β-glucosidase. *FEBS Lett*. **591**, 3926–3936
48. Nakajima, M., Tanaka, N., Furukawa, N., Nihira, T., Kodutsumi, Y., Takahashi, Y., *et al*. (2017) Mechanistic insight into the substrate specificity of 1,2-β-oligoglucan phosphorylase from *Lachnoclostridium phytofermentans*. *Sci Rep*. **7**, 42671
49. Shimizu, H., Nakajima, M., Miyanaga, A., Takahashi, Y., Tanaka, N., Kobayashi, *et al*. (2018) Characterization and structural analysis of a novel *exo*-type enzyme acting on β-1,2-glucooligosaccharides from Parabacteroides distasonis. *Biochemistry*. **57**, 3849–3860
50. Kobayashi, K., Shimizu, H., Tanaka, N., Kuramochi, K., Nakai, H., Nakajima, M., *et al*. (2022) Characterization and structural analyses of a novel glycosyltransferase acting on the β-1,2-glucosidic linkages. *J Biol Chem*. **298**, 101606
51. Abe, K., Sunagawa, N., Terada, T., Takahashi, Y., Arakawa, T., Igarashi, K., *et al*. (2018) Structural and thermodynamic insights into β-1,2-glucooligosaccharide capture by a solute-binding protein in *Listeria innocua*. *J Biol Chem*. **293**, 8812–8828
52. Motouchi, S., Kobayashi, K., Nakai, H., and Nakajima, M. (2023) Identification of enzymatic functions of osmo-regulated periplasmic glucan biosynthesis proteins from *Escherichia coli* reveals a novel glycoside hydrolase family. *Commun Biol*. **6**, 961
53. Bohin, J. P. (2000) Osmoregulated periplasmic glucans in Proteobacteria. *FEMS Microbiol Lett*. **186**, 11–19
54. Amemura, A., and Cabrera-Crespo, J. (1986) Extracellular oligosaccharides and Low-M, polysaccharides containing (1→2)-β-D-glucosidic linkages from strains of *Xanthornonas*,

Escherichia coli and *Klebsiella pneumoniae*. *J Gen Microbiol*. **132**, 2443–2452

55. Talaga, P., Fournet, B., and Bohin, J. P. (1994) Periplasmic glucans of *Pseudomonas syringae* pv. *syringae*. *J Bacteriol*. **176**, 6538–6544
56. Lequette, Y., Rollet, E., Delangle, A., Greenberg, E. P., and Bohin, J. P. (2007) Linear osmoregulated periplasmic glucans are encoded by the opgGH locus of *Pseudomonas aeruginosa*. *Microbiology (N Y)*. **153**, 3255–3263
57. Schneider, J. E., Reinhold, V., Rumley, M. K., and Kennedy, E. P. (1979) Structural studies of the membrane-derived oligosaccharides of *Escherichia coli*. *J. Biol. Chem*. **254**, 10135–10138
58. Lequette, Y., Ödberg-Ferragut, C., Bohin, J. P., and Lacroix, J. M. (2004) Identification of mdoD, an mdoG paralog which encodes a twin-arginine-dependent periplasmic protein that controls osmoregulated periplasmic glucan backbone structures. *J Bacteriol*. **186**, 3695–3702
59. Cogez, V., Gak, E., Puskas, A., Kaplan, S., and Bohin, J. P. (2002) The opgGIH and opgC genes of *Rhodobacter sphaeroides* form an operon that controls backbone synthesis and succinylation of osmoregulated periplasmic glucans. *Eur J Biochem*. **269**, 2473–2484
60. Hanoulle, X., Rollet, E., Clantin, B., Landrieu, I., Ödberg-Ferragut, C., Lippens, G., *et al*. (2004) Structural analysis of *Escherichia coli* OpgG, a protein required for the biosynthesis of osmoregulated periplasmic glucans. *J Mol Biol*. **342**, 195–205
61. Ivanova, N., Sikorski, J., Sims, D., Brettin, T., Detter, J. C., Han, C., *et al*. (2009) Complete genome sequence of Sanguibacter keddieii type strain (ST-74). *Stand Genomic Sci*. **1**, 110–8
62. Wiegand, S., Jogler, M., Boedeker, C., Heuer, A., Peeters, S. H., Kallscheuer, N., *et al*. (2020) Updates to the recently introduced family Lacipirellulaceae in the phylum Planctomycetes: isolation of strains belonging to the novel genera Aeoliella, Botrimarina, Pirellulimonas and Pseudobythopirellula and the novel species Bythopirellula polymerisocia and Posidoniimonas corsicana. *Antonie Van Leeuwenhoek*. **113**, 1979–1997
63. Vitorino, I. R., and Lage, O. M. (2022) The Planctomycetia: an overview of the currently largest class within the phylum Planctomycetes. *Antonie Van Leeuwenhoek*. **115**, 169–201
64. Storesund, J. E., and Øvreås, L. (2013) Diversity of Planctomycetes in iron-hydroxide deposits from the Arctic Mid Ocean Ridge (AMOR) and description of *Bythopirellula goksoyri* gen. nov., sp. nov., a novel Planctomycete from deep sea iron-hydroxide deposits. *Antonie Van Leeuwenhoek*. **104**, 569–84
65. Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A., *et al*. (2023) InterPro in 2022. *Nucleic Acids Res*. **51**, D418–D427
66. Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., *et al*. (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics*. **30**, 1236–40
67. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., and Bateman, A. (2021)

Pfam: The protein families database in 2021. *Nucleic Acids Res*. **49**, D412–D419

68. Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., *et al*. (2020) CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res*. **48**, D265–D268
69. Kim, Y. O., Kim, K. K., Park, S., Kang, S. J., Lee, J. H., Lee, S. J., *et al*. (2010) *Photobacterium gaetbulicola* sp. nov., a lipolytic bacterium isolated from a tidal flat sediment. *Int J Syst Evol Microbiol*. **60**, 2587–2591
70. Van Hoorebeke, A., Stout, J., Kyndt, J., De Groeve, M., Dix, I., Desmet, T., Soetaert, W., Van Beeumen, J., and Savvides, S. N. (2010) Crystallization and X-ray diffraction studies of cellobiose phosphorylase from *Cellulomonas uda*. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. **66**, 346–351
71. Hidaka, M., Honda, Y., Kitaoka, M., Nirasawa, S., Hayashi, K., Wakagi, T., Shoun, H., and Fushinobu, S. (2004) Chitobiose phosphorylase from Vibrio proteolyticus, a member of glycosyl transferase family 36, has a clan GH-L-like $(\alpha/\alpha)_6$ barrel fold. *Structure*. **12**, 937–947
72. Yoshida, E., Hidaka, M., Fushinobu, S., Koyanagi, T., Minami, H., Tamaki, H., *et al*. (2010) Role of a PA14 domain in determining substrate specificity of a glycoside hydrolase family 3 β-glucosidase from *Kluyveromyces marxianus*. *Biochemical Journal*. **431**, 39–49
73. Shallom, D., Leon, M., Bravman, T., Ben-David, A., Zaide, G., Belakhov, V., *et al*. (2005) Biochemical characterization and identification of the catalytic residues of a family 43 beta-D-xylosidase from Geobacillus stearothermophilus T-6. *Biochemistry*. **44**, 387–97
74. Flipphi, M. J., Visser, J., van der Veen, P., and de Graaff, L. H. (1993) Cloning of the Aspergillus niger gene encoding alpha-L-arabinofuranosidase A. *Appl Microbiol Biotechnol*. **39**, 335–40
75. 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. *Appl Environ Microbiol*. **82**, 1686–1692
76. Maurício da Fonseca, M. J., Armstrong, Z., Withers, S. G., and Briers, Y. (2020) High-Throughput Generation of Product Profiles for Arabinoxylan-Active Enzymes from Metagenomes. *Appl Environ Microbiol*. 10.1128/AEM.01505-20
77. Stam, M., Lelièvre, P., Hoebeke, M., Corre, E., Barbeyron, T., and Michel, G. (2023) SulfAtlas, the sulfatase database: state of the art and new developments. *Nucleic Acids Res*. **51**, D647–D653
78. Barbeyron, T., Brillet-Guéguen, L., Carré, W., Carrière, C., Caron, C., Czjzek, M., Hoebeke, M., and Michel, G. (2016) Matching the Diversity of Sulfated Biomolecules: Creation of a Classification Database for Sulfatases Reflecting Their Substrate Specificity. *PLoS One*. **11**, e0164846
79. Silchenko, A. S., Rasin, A. B., Zueva, A. O., Kusaykin, M. I., Zvyagintseva, T. N., Kalinovsky,

A. I., *et al*. (2018) Fucoidan Sulfatases from Marine Bacterium *Wenyingzhuangia fucanilytica* CZ1127T. *Biomolecules*. 10.3390/biom8040098

80. Bond, C. S., Clements, P. R., Ashby, S. J., Collyer, C. A., Harrop, S. J., Hopwood, J. J., and Guss, J. M. (1997) Structure of a human lysosomal sulfatase. *Structure*. **5**, 277–89

81. Kaur, A., Pickles, I. B., Sharma, M., Madeido Soler, N., Scott, N. E., Pidot, S. J., *et al*. (2023) Widespread family of NAD^+-dependent sulfoquinovosidases at the gateway to sulfoquinovose catabolism. *J Am Chem Soc*. **145**, 28216–28223

82. Liu, J., Wei, Y., Ma, K., An, J., Liu, X., Liu, Y., Ang, E. L., Zhao, H., and Zhang, Y. (2021) Mechanistically diverse pathways for sulfoquinovose degradation in bacteria. *ACS Catal*. **11**, 14740–14750

83. Yip, V. L. Y., Varrot, A., Davies, G. J., Rajan, S. S., Yang, X., Thompson, J., Anderson, W. F., and Withers, S. G. (2004) An unusual mechanism of glycoside hydrolysis involving redox and elimination steps by a family 4 beta-glycosidase from *Thermotoga maritima*. *J Am Chem Soc*. **126**, 8354–5

84. Teze, D., Shuoker, B., Chaberski, E. K., Kunstmann, S., Fredslund, F., Nielsen, T. S., *et al*. (2020) The catalytic acid–base in GH109 resides in a conserved GGHGG loop and allows for comparable α-retaining and β-inverting activity in an *N*-acetylgalactosaminidase from *Akkermansia muciniphila*. *ACS Catal*. **10**, 3809–3819

85. Kurahashi, M., and Yokota, A. (2007) *Endozoicomonas elysicola* gen. nov., sp. nov., a gamma-proteobacterium isolated from the sea slug Elysia ornata. *Syst Appl Microbiol*. **30**, 202–6

86. Neave, M. J., Michell, C. T., Apprill, A., and Voolstra, C. R. (2017) *Endozoicomonas* genomes reveal functional adaptation and plasticity in bacterial strains symbiotically associated with diverse marine hosts. *Sci Rep*. **7**, 40579

87. Ezeji, J. C., Sarikonda, D. K., Hopperton, A., Erkkila, H. L., Cohen, D. E., Martinez, S. P., *et al*. (2021) *Parabacteroides distasonis*: intriguing aerotolerant gut anaerobe with emerging antimicrobial resistance and pathogenic and probiotic roles in human health. *Gut Microbes*. **13**, 1922241

88. Qian, W., Jia, Y., Ren, S. X., He, Y. Q., Feng, J. X., Lu, L. F., *et al*. (2005) Comparative and functional genomic analyses of the pathogenicity of phytopathogen *Xanthomonas* campestris pv. campestris. *Genome Res*. **15**, 757–67

89. Vieira, P. S., Bonfim, I. M., Araujo, E. A., Melo, R. R., Lima, A. R., Fessel, M. R., *et al*. (2021) Xyloglucan processing machinery in Xanthomonas pathogens and its role in the transcriptional activation of virulence factors. *Nat Commun*. **12**, 1–15

90. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., *et al*. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*. **596**, 583–589

91. Holm, L. (2020) DALI and the persistence of protein shape. *Protein Science*. **29**, 128–140

92. Holm, L. (2022) Dali server: structural unification of protein families. *Nucleic Acids Res*. **50**, W210–W215

93. Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res*. 10.1093/NAR/GKW408

94. Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T., *et al*. (2005) ConSurf 2005: The projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res*. 10.1093/nar/gki370

95. Tanaka, N., Nakajima, M., Aramasa, H., Nakai, H., Taguchi, H., Tsuzuki, W., *et al*. (2018) Synthesis of three deoxy-sophorose derivatives for evaluating the requirement of hydroxy groups at position 3 and/or 3' of sophorose by 1,2-β-oligoglucan phosphorylases. *Carbohydr Res*. **468**, 13–22

96. Cukierman, S. (2006) Et tu, Grotthuss! and other unfinished stories. *Biochim Biophys Acta*. **1757**, 876–85

97. de Grotthuss, C. J. T. (1806) Sur la décomposition de l'eau et des corps qu'elle tient en dissolution à l'aide de l'électricité galvanique. *Ann. Chim.* **58**, 54–73

98. Yamada, C., Gotoh, A., Sakanaka, M., Hattie, M., Stubbs, K. A., Katayama-Ikegami, A., *et al*. (2017) Molecular Insight into Evolution of Symbiosis between Breast-Fed Infants and a Member of the Human Gut Microbiome *Bifidobacterium longum*. *Cell Chem Biol*. 10.1016/j.chembiol.2017.03.012

99. Kawahara, R., Saburi, W., Odaka, R., Taguchi, H., Ito, S., Mori, H., *et al*. (2012) Metabolic mechanism of mannan in a ruminal bacterium, *Ruminococcus albus*, involving two mannoside phosphorylases and cellobiose 2-epimerase: discovery of a new carbohydrate phosphorylase, β-1,4-mannooligosaccharide phosphorylase. *J Biol Chem*. **287**, 42389–42399

100. Gregg, K. J., Suits, M. D. L., Deng, L., Vocadlo, D. J., and Boraston, A. B. (2015) Structural analysis of a family 101 glycoside hydrolase in complex with carbohydrates reveals insights into its mechanism. *J. Biol. Chem*. **290**, 25657–25669

101. Eisenberg, D., Schwarz, E., Komaromy, M., and Wall, R. (1984) Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol*. **179**, 125–142

102. [preprint] Motouchi, S., Komba, S., Nakai, H., and Nakajima, M. (2023) Phytopathogenic cyclic glucohexadecaose from an inverting transglycosylase. *bioRxiv*. 10.1101/2023.11.26.568766

103. Kuhaudomlarp, S., Pergolizzi, G., Patron, N. J., Henrissat, B., and Field, R. A. (2019) Unraveling the subtleties of β-(1→3)-glucan phosphorylase specificity in the GH94, GH149, and GH161 glycoside hydrolase families. *J. Biol. Chem*. **294**, 6483–6493

104. Kuhaudomlarp, S., Stevenson, C. E. M., Lawson, D. M., and Field, R. A. (2019) The structure

of a GH149 β-(1 → 3) glucan phosphorylase reveals a new surface oligosaccharide binding site and additional domains that are absent in the disaccharide-specific GH94 glucose-β-(1 → 3)-glucose (laminaribiose) phosphorylase. *Proteins: Structure, Function, and Bioinformatics*. **87**, 885–892

105. Kuhaudomlarp, S., Patron, N. J., Henrissat, B., Rejzek, M., Saalbach, G., and Field, R. A. (2018) Identification of *Euglena gracilis* β-1,3-glucan phosphorylase and establishment of a new glycoside hydrolase (GH) family GH149. *J. Biol. Chem*. **293**, 2865–2876
106. Kobayashi, K., Aramasa, H., Nakai, H., Nakajima, M., and Taguchi, H. (2018) Colorimetric determination of β-1,2-glucooligosaccharides for an enzymatic assay using 3-methyl-2-benzothiazolinonehydrazone. *Anal Biochem*. **560**, 1–6
107. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., Mcgettigan, P. A., McWilliam, H., *et al*. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*. **23**, 2947–2948
108. Edgar, R. C. (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. **32**, 1792–1797
109. Tamura, K., Stecher, G., and Kumar, S. (2021) MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol*. 10.1093/molbev/msab120
110. Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992) The rapid generation of mutation data matrices from protein sequences. *Bioinformatics*. **8**, 275–282
111. Braberg, H., Webb, B. M., Tjioe, E., Pieper, U., Sali, A., and Madhusudhan, M. S. (2012) SALIGN: a web server for alignment of multiple protein sequences and structures. *Bioinformatics*. **28**, 2072–3
112. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*. **42**, 320–324
113. Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*. **8**, 785–786
114. Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol*. **37**, 420–423
115. Motohashi, K. (2015) A simple and efficient seamless DNA cloning method using SLiCE from *Escherichia coli* laboratory strains and its application to SLiP site-directed mutagenesis. *BMC Biotechnol*. 10.1186/s12896-015-0162-8
116. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci*. **4**, 2411–2423
117. Kabsch, W. (2010) XDS. *Acta Crystallogr D Biol Crystallogr*. **D66**, 125–132
118. Vagin, A., and Teplyakov, A. (2010) Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr*. **66**, 22–25

119. Miyatake, H., Hasegawa, T., and Yamano, A. (2006) New methods to prepare iodinated derivatives by vaporizing iodine labelling (VIL) and hydrogen peroxide VIL (HYPER-VIL). *Acta Crystallogr D Biol Crystallogr*. **62**, 280–9

120. Agirre, J., Atanasova, M., Bagdonas, H., Ballard, C. B., Baslé, A., Beilsten-Edmands, J., *et al*. (2023) The CCP4 suite: integrative software for macromolecular crystallography. *Acta Crystallogr D Struct Biol*. **79**, 449–461

121. Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr*. **62**, 1002–1011

122. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*. **53**, 240–255

123. Emsley, P., and Cowtan, K. (2004) Coot : model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*. **60**, 2126–2132

124. Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput*. **11**, 3696–3713

125. Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., and Woods, R. J. (2008) GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *J Comput Chem*. **29**, 622–655

126. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. **79**, 926–935

127. Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., *et al*. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. **1–2**, 19–25

128. 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. *Appl Environ Microbiol*. **78**, 5682–5689

Competing interests

The authors declare no competing interests.

Contributions

All authors performed experiments and analyzed data. M.N, H.T, T.N, and H.N prepared the substrates for assay and designed the experiments. M.N and H.N conceptualized the study. M.N wrote the draft manuscript, and all authors revised the manuscript.

Data availability

The atomic coordinates and structure factors (codes 8XUJ, 8XUK, and 8XUL) have been deposited in the PDB.

Funding and additional information

This work was supported in part by JSPS KAKENHI (Grant number 23K05041) and JST SPRING (Grant Number JPMJSP2151).

Conflict of interest

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

Acknowledgments

We appreciate the help of all the staff at the Photon Factory for X-ray data collection (proposal nos. 2018G506, 2020G527, and 2022G523). We thank Edanz (https://jp.edanz.com/ac) for editing a draft of this manuscript. Yuta Takahashi, Dr. Naohisa Sugimoto, and Dr. Shinya Fushinobu supported the experiments.

Supporting information

This article contains supporting information (77, 92, 101).

Abbreviations

The abbreviations used are: CAZy, Carbohydrate-Active enZYmes; GH, glycoside hydrolase; GT, glycosyltransferase; CGS, cyclic β-1,2-glucan synthase; TiCGS, CGS from *T. italicus*; SOGP, 1,2-β-oligoglucan phosphorylase; SGL, β-1,2-glucanase; DP, degree of polymerization; Sop_n, β-1,2-glucooligosaccharide (n is DP); BGL, β-glucosidase; Glc, glucose; SO-BP, a solute-binding protein in an ABC transporter specific to Sop_ns; OPG, osmoregulated periplasmic glucans; CpSGL, SGL from *C. pinensis*; TfSGL, SGL from *T. funiculosus*; $TiCGS_{Tg}$, transglycosylation domain of TiCGS; SkSGL, a protein from *S. keddieii* (KEGG locus tag, Sked_30460); PgSGL1, a protein from *P. gaetbulicola*

(KEGG locus tag, H744_1c0224); PgSGL2, a protein from *P. gaetbulicola* (KEGG locus tag, H744_2c1936); PgSGL3, a protein from *P. gaetbulicola* (KEGG locus tag, H744_1c0222); PgSGL4, a protein from *P. gaetbulicola* (KEGG locus tag, H744_1c0194); EeSGL1, a protein from *E. elysicola* (NCBI accession number, WP_026258326.1); XcSGL, a GH144 enzyme from *X. campestris* pv. *campestris* (KEGG locus tag, XCC2207); SkSGLc, (see the Experimental procedures section for details); SkSGLn, (see the Experimental procedures section for details); RMSD, root mean square deviation; MD, molecular dynamics; WT, wild-type; IPTG, isopropyl β-D-1-thiogalactopyranoside; MBTH, 3-methyl-2-benzo-thiazolinone hydrazine.

Table 1. Substrate specificity of SGL-superfamily enzymes.

Substrate	Final concentration (%)	Substrate specificity (U/mg)					
		Group1			Group2	Group3	GH144
		PgSGL1	PgSGL2	EeSGL1	SkSGLc	PgSGL3	XcSGL
β-1,2-Glucan	0.2	8.44	3.13	25.4	12.3	8.96	3.85
($NaBH_4$)		(0.92)[a]	(0.41)	(3.0)	(1.0)	(0.85)	(0.14)
CM-Cellulose	0.1	N.D.[c]	N.D.	N.D.	N.D.	N.D.	N.D.
Barley β-glucan	0.2	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Glucomannan	0.05 (0.1)[b]	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Tamarind xyloglucan	0.2	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Lichenan	0.0125 (0.01)	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Laminarin	0.0125 (0.01)	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
CM-Curdlan	0.05	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Pustulan	0.05	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Arabinogalactan	0.006 (0.01)	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Arabinan	0.1 (0.2)	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
PGA	0.0025 (0.01)	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Xylan (Beechwood)	0.1	N.E.[d]	N.D.	N.E.	N.E.	N.E.	N.E.
Soluble starch	0.02	N.D.	< 0.3%	N.D.	N.D.	N.D.	N.D.

[a]The values in the parentheses after the specific activity represent the large difference between the median value and the other data in triplicate experiments as errors.

[b]The values in the parentheses after the final concentrations represent the substrate concentrations for the assay of EeSGL1.

[c]N.D. indicates that the relative activity was lower than 0.2% of the activity toward β-1,2-glucan ($NaBH_4$).

[d]N.E.: not examined.

Table 2. Kinetic parameters of the homologs in the SGL superfamily.

	V_{max} (U/mg protein)	K_m (mg substrate/ml)	V_{max}/K_m (U/mg protein ml/mg substrate)
This study			
PgSGL1	11.3 ± 0.4	0.484 ± 0.068	23.3 ± 2.6
PgSGL2	8.58 ± 0.71	4.48 ± 0.61	1.92 ± 0.11
EeSGL1	34.7 ± 2.5	0.842 ± 0.182	41.3 ± 6.4
SkSGLc	8.59 ± 0.34	0.455 ± 0.058	18.9 ± 1.8
PgSGL3	21.8 ± 2.5	1.05 ± 0.26	20.8 ± 2.9
References[a]			
CpSGL (GH144)[b,c]	57 ± 4.7	0.83 ± 0.18	81 ± 12
TfSGL (GH162)[b]	33 ± 1.1	0.18 ± 0.02	180 ± 11
OpgD from *E. coli* (GH186)	28.7 ± 2.0	8.6 ± 1.1	3.37 ± 0.198

[a] The kinetic parameters of CpSGL, TfSGL, and OpgD from *E. coli* are cited from Abe et al. (45), Tanaka et al. (44) and Motouchi et al. (52), respectively.

[b] Kinetic parameters were evaluated using β-1,2-glucan (DP25).

[c] Assay was performed using the 4-hydroxybenzhydrazide method (128).

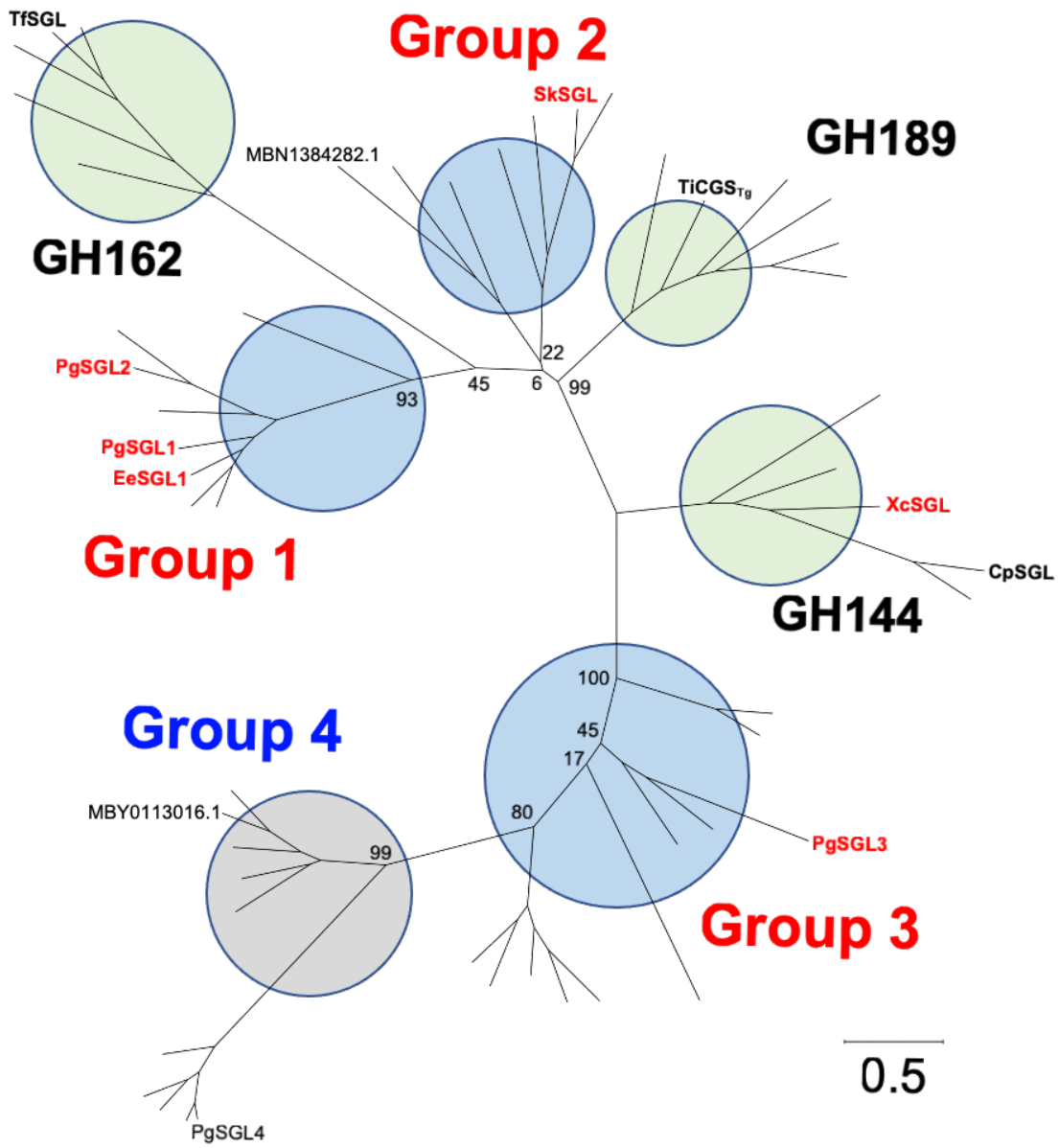

Figure 1. Phylogenetic tree of the superfamily of SGL-related proteins with bootstrap confidence levels

The proteins used in this study are labelled in the tree. TfSGL, CpSGL, and TiCGS$_{Tg}$ with known biochemical functions are shown in bold black letters. The enzymes with biochemical functions that were elucidated in the present study are shown in bold red letters. Known GH families (GH144, GH162, and GH189) and newly established GH families are shown as light green and light blue circles, respectively. A group with an unknown function is indicated by the light gray circle. Bootstrap confidence levels are shown at the root branches for each group as percentages. A scale bar indicating the phylogenetic distance is shown at the bottom right. Multiple sequence alignment was performed by MUSCLE (108) and the phylogenetic tree was created using MEGA11 (109).

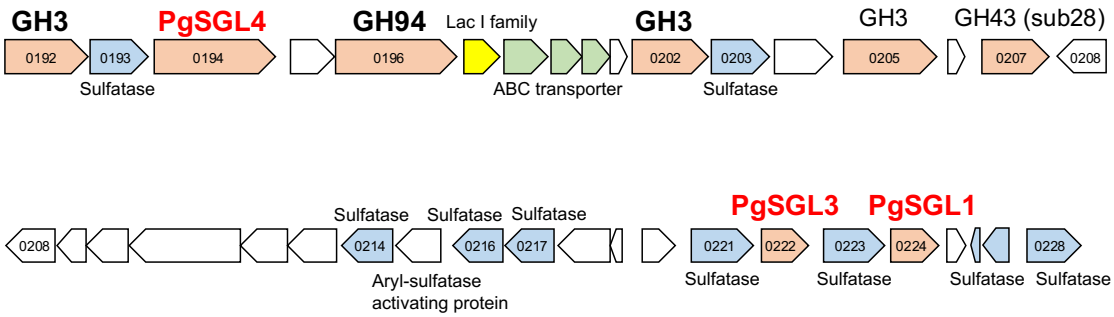

Figure 2. A gene cluster of genes encoding PgSGL1, PgSGL3, and PgSGL4 from *P. gaetbulicola*
The directions of genes are represented as arrows. The arrow boxes for genes registered in CAZy and SulfAtlas databases (8, 9, 77, 78) are colored light red and light blue, respectively. The arrow boxes for putative ABC transporter genes and a putative LacI family transcriptional regulator gene are colored light green and yellow, respectively. KEGG locus tags with the "H744_1c", common letters in the loci omitted are shown in the arrow boxes for genes registered in CAZy or SulfAtlas databases. H744_1c0208 is duplicated to show that the genes in the first and the second lines are shown in the same scale. The target proteins in this study are labelled in bold red letters. Proteins with biochemical functions confidently presumed are labelled in bold black letters.

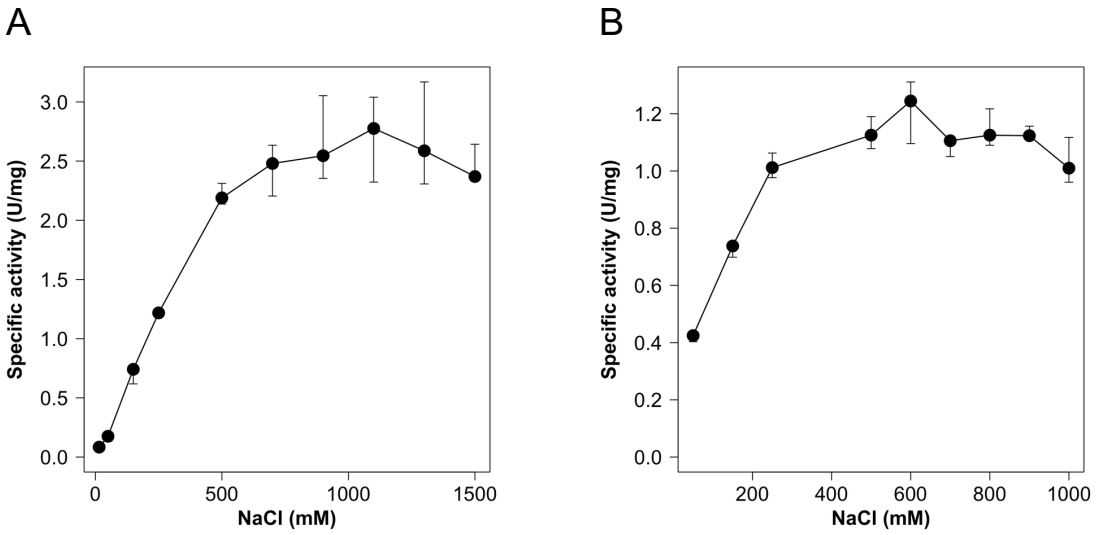

Figure 3. Effect of NaCl on the hydrolytic activity of PgSGL2

Hydrolytic activity (A) and stability (B) of PgSGL2 in the presence of various concentrations of NaCl. Medians in the triplicate experiments are plotted as closed circles and the other data are shown as error bars.

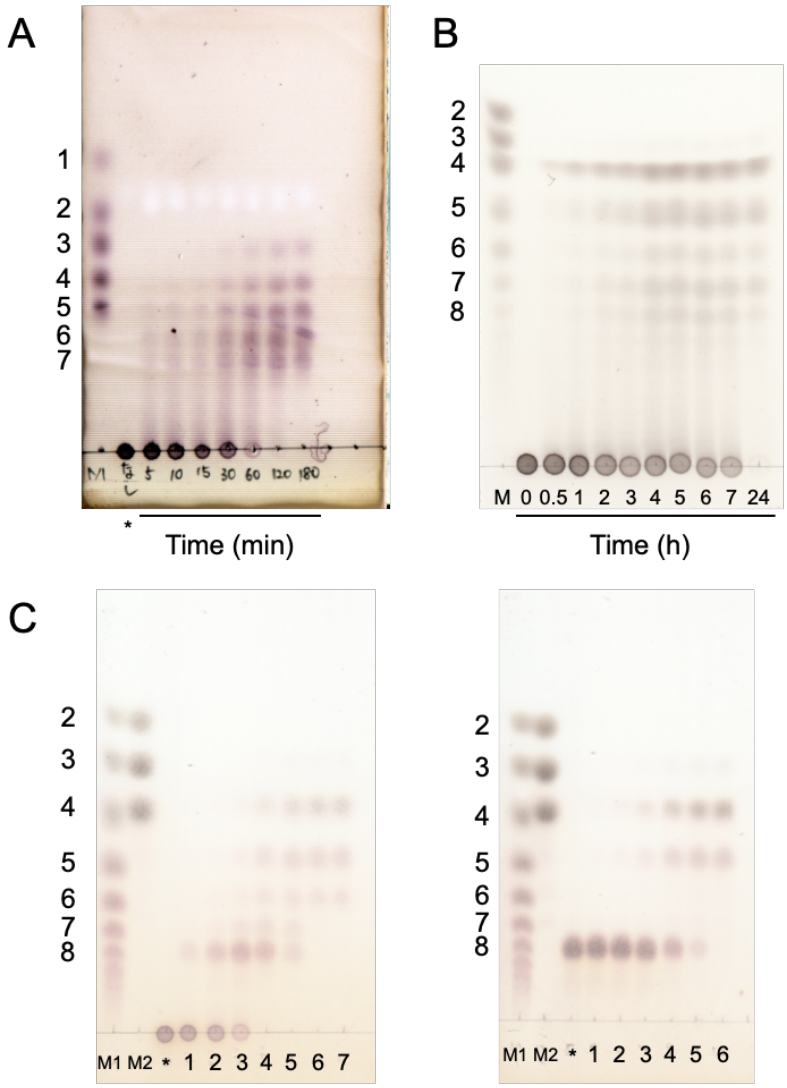

Figure 4. Action pattern analysis using TLC

A, EeSGL1 (Group 1); B, SkSGLc (Group 2); C, PgSGL3 (Group 3). Asterisks indicate the reaction mixture without enzymes. Lanes M, M1, and M2 show the markers. (A) The reaction times are written directly on the TLC plate. Lane M, 0.2 µl of solution containing 0.5% each Glc, Sop_2, Sop_3, Sop_4, and Sop_5. The reaction solutions (0.5 µl) were spotted on the plate. (B) Lane M, 0.5% Sop_ns mixture prepared as described in the Materials and Methods section. A 1-µl aliquot of a marker or sample was spotted on the plate. (C) Lane M1, 1% Sop_ns mixture prepared as described in the Materials and Methods section. Lane M2, a mixture of Sop_2 (10 mM), Sop_3 (7.5 mM), and Sop_4 (5 mM) was spotted. A 0.7-µl aliquot of a marker or sample was spotted on the plate. Lanes 1–5, the reactions were performed with 0.025, 0.05, 0.1, 0.25, and 0.5 mg/ml PgSGL3, respectively, at 30 °C for 20 min. Lanes 6–7 (left), the reactions were performed with 1 mg/ml PgSGL3 at 30 °C for 20 and 220 min, respectively. Lane 6 (right), the reaction was performed with 1 mg/ml PgSGL3 at 30 °C for 30 min.

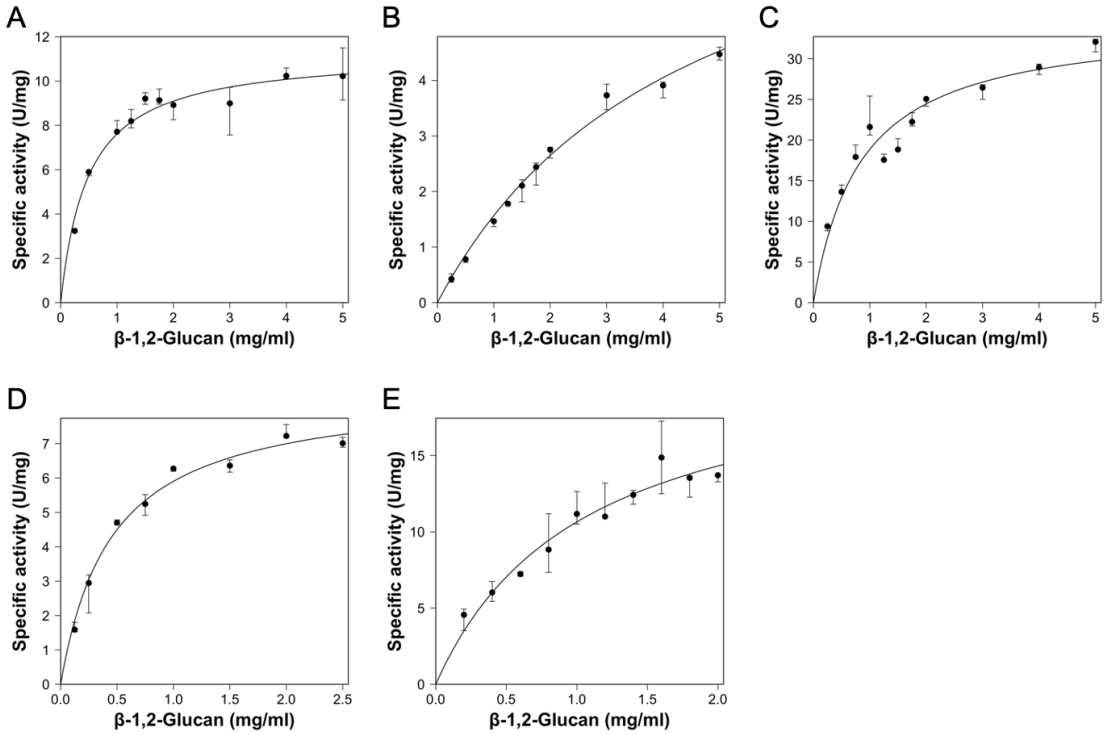

Figure 5. Kinetic analysis of the SGL-superfamily enzymes

Group 1, (A) PgSGL1, (B) PgSGL2, (C) EeSGL1; Group 2, (D) SkSGL; and Group 3, (E) PgSGL3. Medians are plotted as closed circles. The other data in the triplicate experiments are shown as bars. In the case of EeSGL1, two data points (1.25 and 1.5 mg/ml substrate) were eliminated from the graph because they were obviously outliers.

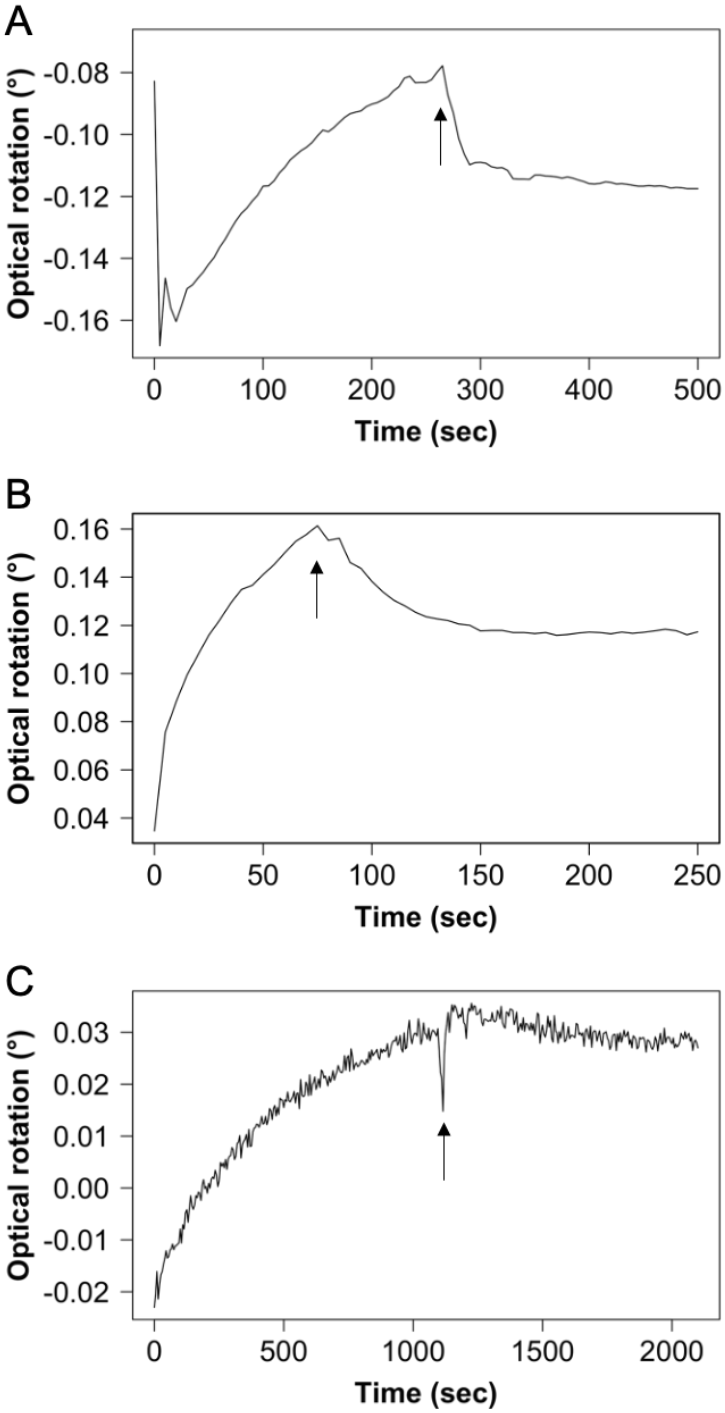

Figure 6. Time-course of optical rotations during the hydrolysis of β-1,2-glucans

(A) PgSGL1 (Group 1), (B) SkSGL (Group2), and (C) PgSGL3 (Group 3). The arrows indicate the time that aqueous ammonia was added to the reaction mixtures.

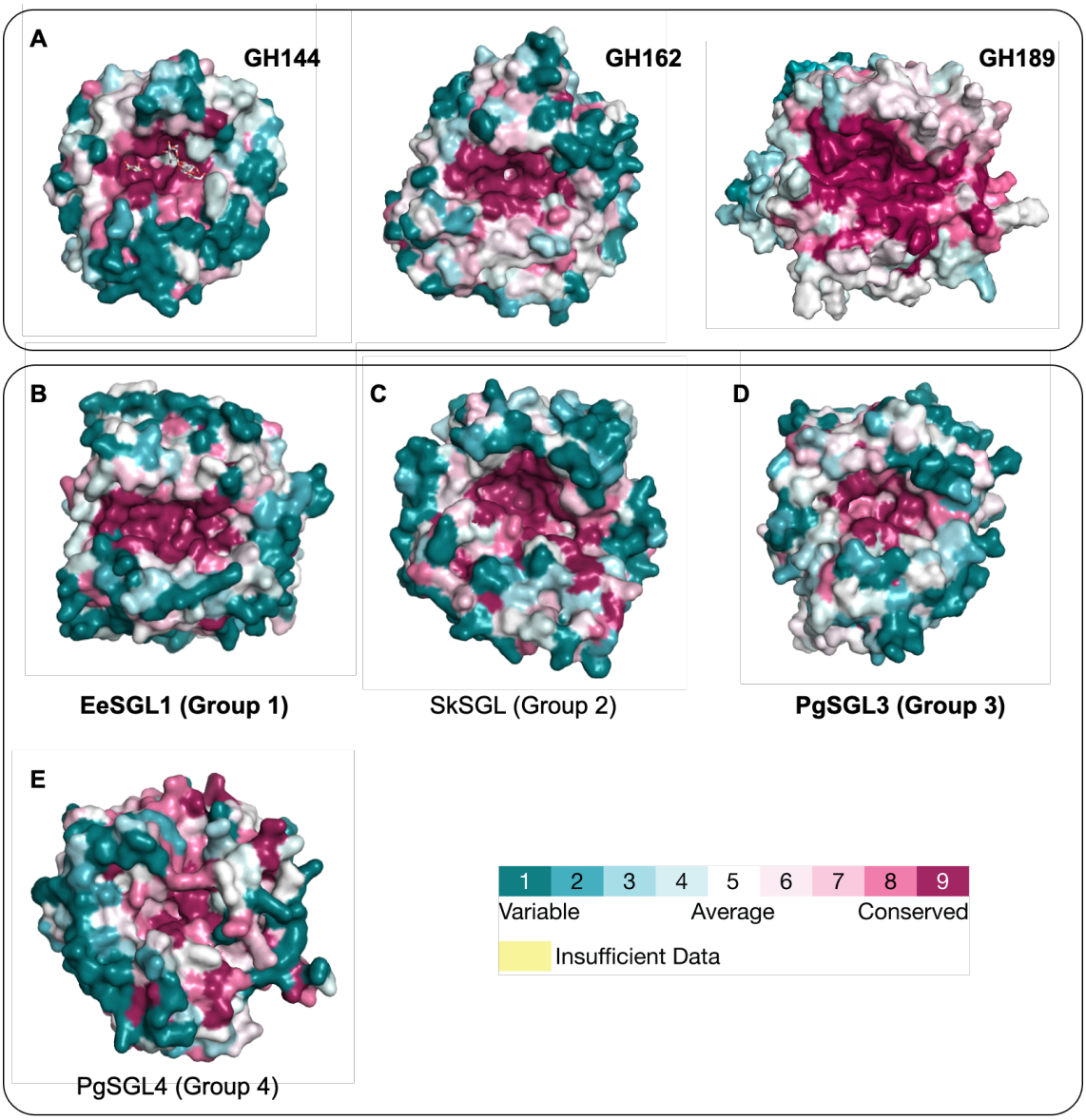

Figure 7. Surface models of the SGL-superfamily proteins

The SGL-superfamily proteins are shown with the surfaces colored based on the conservation scores. The conservation scores were calculated by ConSurf using the structures shown in the figure. The conservation scores are colored according to the color bar provided in the ConSurf server (93, 94). GH families or groups with experimentally solved structures and those with structures only predicted by AlphaFold2 are shown in bold and standard type, respectively. (A) GH families reported previously; (left) CpSGL (GH144, 5GZK), (middle) TfSGL (GH162; PDB ID, 6IMU), and (right) $TiCGS_{Tg}$ (GH189, prediction). The RSMD between the crystal structure of $TiCGS_{Tg}$ (PDB ID, 8WY1) and the predicted structure was 0.54 Å. (B–E) Groups identified in this study. (B) EeSGL1 (Group 1, PDB ID, 8XUJ). (C) SkSGL (Group 2, prediction). (D) PgSGL3 (Group 3, PDB ID, 8XUK). (E) PgSGL4 (Group 4, prediction). "prediction" indicates query structures were predicted by AlphaFold2 (90).

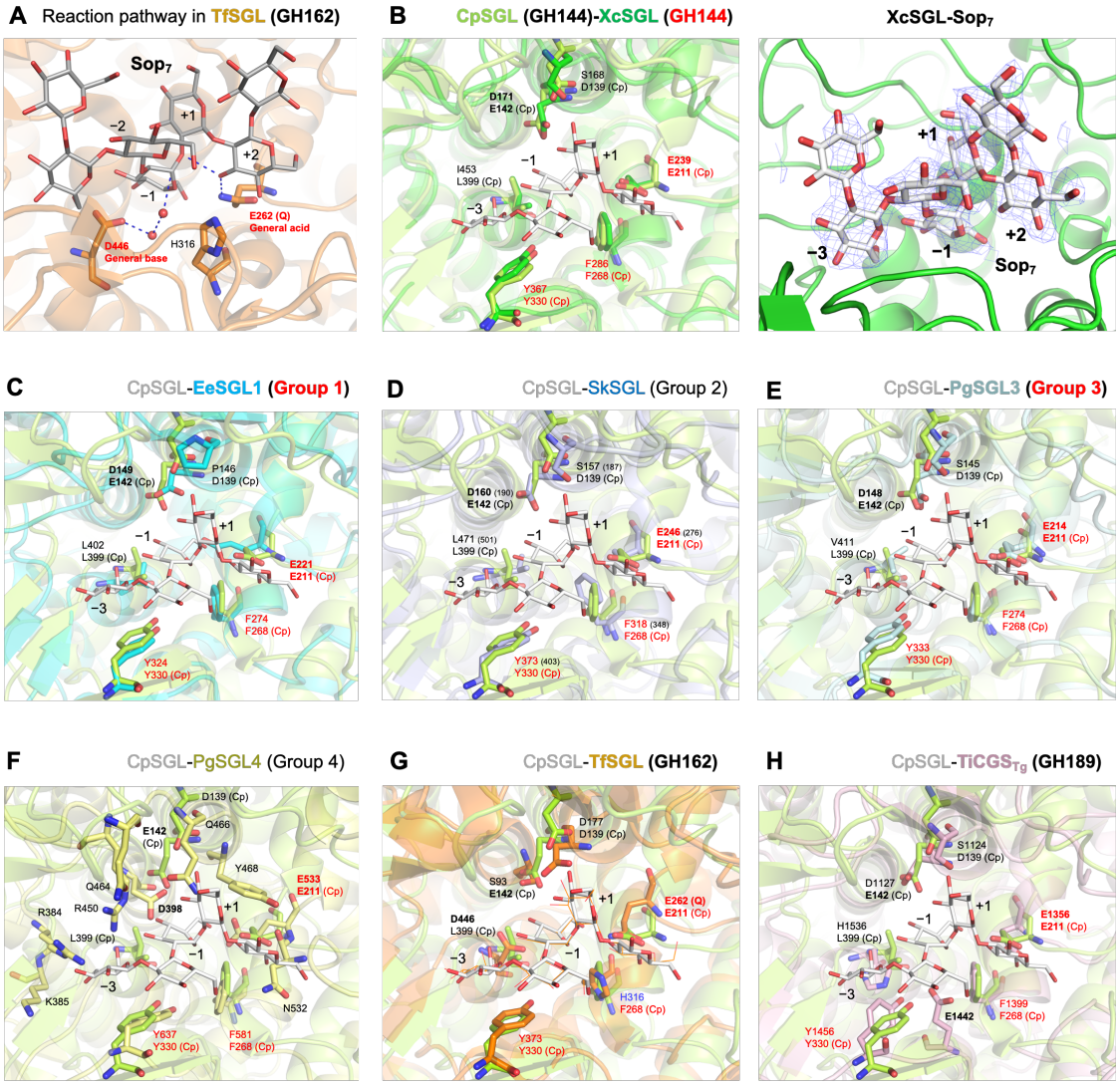

Figure 8. Comparison of the substrate pockets of SGLs in the SGL superfamily

Both Glc moieties at the non-reducing and reducing ends are omitted for clarity. The residues that define the SGL superfamily (E211, F268, and Y330 in the case of CpSGL) are colored red. The H316 in TfSGL is colored blue. Catalytic residues and candidates for catalytic residues are shown in bold letters. (Cp) represents the residues in CpSGL. (A) Reaction pathway in TfSGL (PDB ID, 6IMU). Sop_7 is shown as a gray stick. Water molecules are shown as red spheres. The reaction pathway is shown as blue dashed lines. Q262 in the complex structure was originally E262 in the WT enzyme. The catalytic residues are shown as bold red letters. (B–H) CpSGL (PDB ID, 5GZK) is superimposed with XcSGL (PDB ID, 8XUL) (B, left), EeSGL1 (PDB ID, 8XUJ) (C), SkSGL (prediction) (D), PgSGL3 (PDB ID, 8XUK) (E), PgSGL4 (prediction) (F), TfSGL (PDB ID, 6IMU) (G), and $TiCGS_{Tg}$ (PDB ID, 8WY1) (H). Sop_7 in the XcSGL-Sop_7 complex is superimposed in each panel. The superimposed Sop_7 molecules are represented as white sticks. Experimentally solved structures are

labelled in bold, except for CpSGL in (C–H), which is shown in light gray lettering. Structures solved experimentally in this study are labelled with bold red GH or group names. The colors for each protein are as follows; CpSGL, light green; XcSGL, green; EeSGL1, cyan; SkSGL, light purple; PgSGL3, pale cyan; PgSGL4, pale yellow; TfSGL, orange; and $TiCGS_{Tg}$, light pink. (B) (right) The F_o−F_c omit map for Sop_7 is shown at the 2.0σ contour level and is represented as a blue mesh. (D) The residue numbers of SkSGL were based on the starting methionine residue defined in this study. The residue numbers in parentheses are based on the sequence in the database. (F) PgSGL4 was aligned with CpSGL by pair fitting using E211, F268, and Y330 in CpSGL. (G) TfSGL was aligned with CpSGL using ligands. Sop_7 in TfSGL is shown as a thin orange stick.

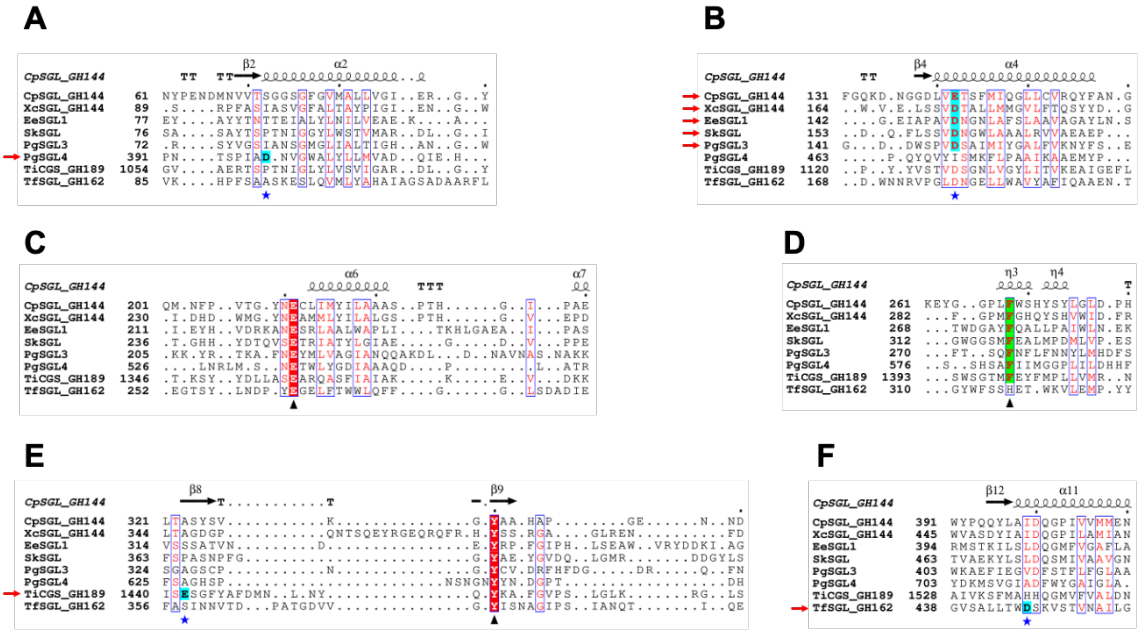

Figure 9. Multiple sequence alignment of the SGL-superfamily proteins

Black triangles represent highly conserved residues in the SGL superfamily, except that the phenylalanine is replaced with a histidine in GH162. The phenylalanine residues indicated by the triangle in (D) are also highlighted in green. The black triangle in (C) represents the position of a nucleophile in $TiCGS_{Tg}$ and (candidate) base catalysts for the other proteins. Residues where blue stars and red arrows intersect are candidates for acid catalysts. These residues are also highlighted in cyan. Multiple alignment was performed using the SALIGN server (111). The ESPript 3.0 server (112) was used for visualization of the alignment.

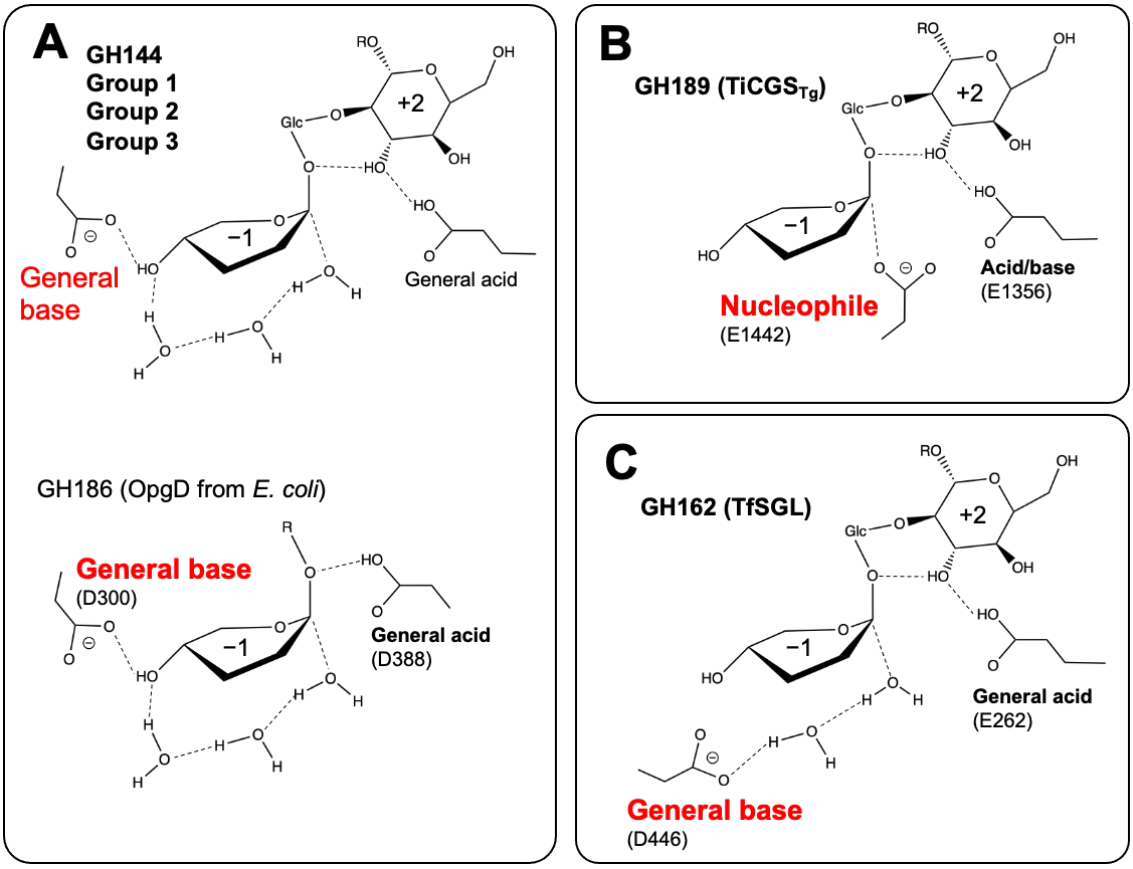

Figure 10. Reaction mechanisms of the SGL superfamily

Reaction mechanisms are shown schematically. (A–C) show different reaction pathways to attack the anomeric carbon at subsite −1. A nucleophile and (potential) general bases are shown in red letters to indicate the diversity in the SGL superfamily. The determined catalytic residues are shown in bold letters. Glc moieties at subsite +1 in the SGL superfamily are indicated by Glc. R represents a β-1,2-glucan moiety. Glc moieties at subsite −2 and beyond are omitted. Some substituents of the Glc moiety at subsite −1 are also omitted. (A) (top) Schematic of the reaction mechanisms of GH144 and Groups 1–3. Three water molecules are depicted in the proposed proton transfer pathways based on OpgD from *E. coli*, although the number of the water molecules was not determined for Groups 1–3. (bottom) The reaction mechanism of OpgD from *E. coli*. (B, C) The reaction mechanisms of $TiCGS_{Tg}$ (B) and TfSGL (C).