

The First Crystal Structure of a Family 129 Glycoside Hydrolase from a Probiotic Bacterium Reveals  
Critical Residues and Metal Co-factors

Mayo Sato<sup>‡</sup>, Dorothee Liebschner<sup>§1</sup>, Yusuke Yamada<sup>§</sup>, Naohiro Matsugaki<sup>§</sup>, Takatoshi Arakawa<sup>‡</sup>, Siobhán S. Wills<sup>¶</sup>, Mitchell Hattie<sup>¶</sup>, Keith A. Stubbs<sup>¶</sup>, Tasuku Ito<sup>‡2</sup>, Toshiya Senda<sup>§</sup>, Hisashi Ashida<sup>||</sup>, and Shinya Fushinobu<sup>‡3</sup>

<sup>‡</sup>From the Department of Biotechnology, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan, <sup>§</sup>Structural Biology Research Center, Photon Factory, Institute of Materials Structure Science, High Energy Accelerator Research Organization, 1-1 Oho, Tsukuba 305-0801, Japan, <sup>¶</sup>School of Molecular Sciences, The University of Western Australia, 35 Stirling Highway, Crawley, WA, Australia, and <sup>||</sup>Faculty of Biology-Oriented Science and Technology, Kindai University, 930 Nishimitani, Kinokawa-shi, Wakayama 649-6493, Japan.

Running title: Structure of GH129  $\alpha$ -N-acetylgalactosaminidase

<sup>1</sup>Present address: Molecular Biophysics & Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

<sup>2</sup>Present address: Yamada Apiculture Center Inc., 194 Ichiba Kagamino-cho, Tomata-gun, Okayama 708-0393, Japan

<sup>3</sup>To whom correspondence should be addressed: Prof. Shinya Fushinobu, Department of Biotechnology, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan, Telephone +81 3 5841 5151; Fax +81 3 5841 5151; E-mail. [asfushi@mail.ecc.u-tokyo.ac.jp](mailto:asfushi@mail.ecc.u-tokyo.ac.jp)

**Keywords:** Glycoside hydrolase family 129, bifidobacteria,  $\alpha$ -N-acetylgalactosaminidase, mucin-type O-glycan, sulfur-SAD phasing

## ABSTRACT

The  $\alpha$ -N-acetylgalactosaminidase from the probiotic bacterium *Bifidobacterium bifidum* (NagBb) belongs to the glycoside hydrolase (GH) family 129 and hydrolyzes the glycosidic bond of Tn antigen (GalNAc $\alpha$ 1-Ser/Thr). NagBb is involved in assimilation of O-glycans on mucin glycoproteins by *B. bifidum* in the human gastrointestinal tract, but its catalytic mechanism has remained elusive because of a lack of sequence homology around putative catalytic residues and of other structural

information. Here we report the X-ray crystal structure of NagBb, representing the first GH129 family structure, solved by the single-wavelength anomalous dispersion method based on sulfur atoms of the native protein. We determined ligand-free, GalNAc and inhibitor complex forms of NagBb and found that Asp-435 and Glu-478 are located in the catalytic domain at appropriate positions for direct nucleophilic attack at the anomeric carbon and proton donation for the glycosidic-bond oxygen, respectively. A highly conserved Asp-330 forms a hydrogen

bond with the O4 hydroxyl of GalNAc in the –1 subsite, and Trp-398 provides a stacking platform for the GalNAc pyranose ring. Interestingly, a metal ion, presumably  $\text{Ca}^{2+}$ , is involved in the recognition of the GalNAc *N*-acetyl group. Mutations at Asp-435, Glu-478, Asp-330, Trp-398, and residues involved in metal coordination (including an all-Ala quadruple mutant) significantly reduced the activity, indicating that these residues and the metal ion play important roles in substrate recognition and catalysis. Interestingly, NagBb exhibited some structural similarities to the GH101 endo- $\alpha$ -N-acetylgalactosaminidases, but several critical differences in substrate recognition and reaction mechanism account for the different activities of these two enzymes.

-----  
*Bifidobacterium* is a well-known representative species of probiotics in human gut microbiota (1). Various health-promoting effects of bifidobacteria have been reported, including prevention of infections by pathogens (2) and alleviation of allergy responses (3). These bacteria mainly reside in the lower intestine of healthy humans, especially during the early life stages of breast-fed infants (4). As digestible carbohydrates such as starch are scarce in the lower intestine, bifidobacteria possess various glycosidases, transporters, and metabolizing enzymes for utilizing indigestible oligosaccharides and glycoconjugates. A well-studied example of this is the system that utilizes human milk oligosaccharides (5,6). Interestingly, it has also been revealed that bifidobacteria utilize mucin glycoproteins which exist on human epithelial cell layers of the

digestive tract (7). It has been recently shown that mammalian gut microbiota degrades mucin glycoproteins as a nutrient source under dietary fiber-deprived conditions (8). The carbohydrates of mucin glycoproteins are highly complex and branched *O*-glycans (9). Eight prevalent core structures of mucin *O*-glycans are defined, all of which are covalently linked via an *N*-acetyl-D-galactosamine (GalNAc) residue to the hydroxyl group of Ser or Thr through an  $\alpha$ -glycosidic bond (10). Previous studies have revealed that *Bifidobacterium bifidum* JCM 1254 possesses various glycosidases such as  $\beta$ -galactosidase (11),  $\beta$ -*N*-acetylhexosaminidase (11), exo- $\alpha$ -sialidase (12), 1,2- $\alpha$ -L-fucosidase (13), and 1,3/1,4- $\alpha$ -L-fucosidase (14) to sequentially liberate sugars from the non-reducing ends of various glycan structures.

Specifically, utilization of GalNAc $\alpha$ 1-Ser/Thr, also referred to as the Tn-antigen, was implicated by the finding of an intracellular  $\alpha$ -N-acetylgalactosaminidase (NagBb, EC 3.2.1.49) from *B. bifidum* JCM 1254 (15). Peptides containing Tn-antigen are thought to be cleaved from the mucin core protein by extracellular proteases and then imported into the cell by an unknown transporter. NagBb then hydrolyzes the  $\alpha$ -linkage between GalNAc and peptide for further metabolism. Of note is that NagBb was found from the genome sequence of *Bifidobacterium bifidum* JCM 1254 by virtue of a very slight sequence similarity (~15%) to an extracellular endo- $\alpha$ -N-acetylgalactosaminidase (EC 3.2.1.97) from *B. longum* JCM 1217 (EngBF) (16), which belongs to glycoside hydrolase (GH) family 101 in the Carbohydrate-Active enZyme (CAZy) database (17). While EngBF is an endo-type enzyme that releases  $\alpha$ -linked Gal $\beta$ 1-3GalNAc

(galacto-*N*-biose) disaccharides from Ser or Thr residues (core 1 or T antigen) in mucin-type glycoproteins, NagBb exhibits an exo-type activity to release GalNAc. As NagBb represented no sequential similarity with any previously identified glycosidases, this enzyme and its homologs formed the basis of a new GH family, 129 (15). Sequence comparison with EngBF suggested that the catalytic nucleophile and a key substrate-recognizing residue (termed as "fixer" or "anchor") of NagBb were Asp-435 and Asp-330, respectively. However, identity of the catalytic acid/base residue has remained elusive due to lack of sequence homology around these positions and X-ray crystallographic information.

Here we report the crystal structure of NagBb, which is the first three-dimensional structure of GH129. A complex structure with GalNAc reveals not only its catalytic residues but also a unique substrate recognition motif involving a metal ion. A possible molecular evolutionary route of this bifidobacterial enzyme is also discussed.

## RESULTS

*Structure determination*—C-terminally His<sub>6</sub>-tagged protein was heterologously expressed in *Escherichia coli* and purified. The molecular masses of NagBb as deduced from the amino acid sequence, estimated by SDS-PAGE, and calibrated gel filtration chromatography were 71.3, 70, and 71.4 kDa, respectively, indicating that it is monomeric in solution. Before gel filtration, NagBb was treated under reductive lysine methylation reaction methodology since untreated protein did not crystallize under any conditions tested. The lysine-methylated protein sample exhibited 95%

catalytic activity compared with the native protein (data not shown). The crystals of NagBb belong to space group  $P2_12_12_1$  containing two molecules in the asymmetric unit, and the crystal structure was solved by the single-wavelength anomalous dispersion (SAD) method using sulfur atoms contained within the native protein (Table 1). Diffraction data for phasing were collected at beamline BL-1A of the Photon Factory, which is designed for long-wavelength experiments (18-20). Fig. 1A shows anomalous difference Fourier map peaks for some sulfur atoms using the phasing data. A ligand-free structure and a product-complex structure with GalNAc were determined at 2.65 Å and 2.10 Å resolution, respectively (Table 2). The  $2mFo-DFc$  electron density maps for the protein contoured at 1  $\sigma$  showed continuous density for all main chain atoms, except for the His<sub>6</sub>-tag and the following residues; 105-108, 565-571 and 633-634 (GalNAc complex, chain A), 104-106, 566-573 and 631-634 (GalNAc complex, chain B), 43-44, 75-76, 103-109, 558-571 and 629-634 (ligand-free, chain A), 43, 102-108, 558-573 and 631-634 (ligand-free, chain B). Apart from these disordered regions, no significant differences between the ligand-free and GalNAc complex structures were observed and both structures superimpose well, as exemplified by the low root mean square deviation (RMSD) of 0.304 Å for all 7,909 protein atoms.

*Overall structure*—NagBb is composed of three domains; an N-domain (residues 1-227), a catalytic barrel-domain (228-591) and a C-domain (592-634) (Fig. 1B). The N-domain adopts a  $\beta$ -sandwich fold with two antiparallel  $\beta$ -sheets containing 18  $\beta$ -strands. The barrel domain adopts a partly broken  $(\beta/\alpha)_8$ -barrel fold,

whose secondary structures of  $\beta 6$ - $\alpha 8$  are not completely formed. A long insertion between  $\beta 3$  and  $\alpha 3$  (loop-2, 376-407) is present in this domain. This insertion corresponds to the common "domain B" of GH13  $\alpha$ -amylases and related enzymes (21). Another insertion is present between  $\beta 1$  and  $\alpha 1$  (loop-1, 274-296) and is a unique structural feature of NagBb. The C-domain consists of an antiparallel  $\beta$ -sheet with 4  $\beta$ -strands. Out of six lysine residues in this protein, electron density of the side chains of Lys-236, Lys-254, Lys-275, Lys-350 were visible. Although the protein was treated using the lysine-methylation protocol, any density peaks for methyl moieties on these lysine residues were not observed.

*Active site architecture*—In the complex structure, clear electron density was observed for the GalNAc ligand, as its  $\alpha$ -anomer, at the center of the barrel-domain (Fig. 2A). Two carboxylate residues are closely located to the anomeric C1 atom of GalNAc; Asp-435 and Glu-478 reside between  $\beta 4$  and  $\alpha 4$ , and  $\beta 5$  and  $\alpha 5$ , respectively. The O $\delta$ 2 atom of Asp-435, which has been previously designated as the nucleophile residue (15), is located 3.1 Å from the C1 atom. The distance and position of Asp-435 are suitable for a nucleophilic attack to the anomeric carbon. Glu-478 forms a hydrogen bond with the C1 hydroxyl group at a distance of 2.7 Å, indicating that this is most likely the catalytic acid/base residue. An amino acid sequence alignment of GH129 members shows that these catalytic residues are completely conserved in this family (Fig. 3B). In addition, the following residues also form hydrogen bonds with GalNAc; Tyr-329 with O3 and O4, Asp-330 (anchor) with O4, Asp-371 with O5, Tyr-433 with the carbonyl

oxygen of the *N*-acetyl group, Asp-435 (nucleophile) with O4, and Asp-561 with O3 and amide nitrogen of *N*-acetyl group (Fig. 3A). In addition, Trp-398 forms a stacking interaction with the pyranose ring of GalNAc.

*Metal-binding sites*—Interestingly, a metal-binding site was observed near the *N*-acetyl group of GalNAc (M1-site, Fig. 2B). The M1-site metal is octahedrally hexacoordinated by the side chains of His-271, His-320, His-366, Asp-322 and two water molecules. Electron density peaks for the metal and two coordinating waters were clearly observed. The water molecules form hydrogen bonds with the carbonyl oxygen of the *N*-acetyl group of GalNAc, suggesting that NagBb utilizes the metal ion for the recognition of the *N*-acetyl group of GalNAc. Apart from the catalytic site, another metal-binding site (M2-site) was observed on the surface of the barrel-domain (Fig. 1B). The M2-site metal ion is tetrahedrally coordinated by the side chains of Cys-407 and Cys-445, and the side chain nitrogen and main chain carbonyl group of His-450 (Fig. 2F). These metal sites were also present in the ligand-free structure. To identify the metal peaks, a crystallographic anomalous scattering analysis was carried out (Table 1). The Bijvoet difference density map of diffraction data measured at 1.283 Å exhibits a strong peak at the M2-site (red mesh in Fig. 2F), whereas the difference map of the data collected at 1.290 Å, which is above the absorption edge of Zn (1.2837 Å), showed no significant peaks ( $<3.0\sigma$ ). At the M1-site, an anomalous difference Fourier peak was always observed at shorter wavelengths than of the absorption edge of Ca (3.0704 Å; see 2.7 Å wavelength data in Fig. 2D), but it

significantly reduced in the difference map of the data collected at 3.15 Å (Table 1 and red mesh in Fig. 2D). At the sulfur atoms contained within the protein (e.g., Met-439), significant peaks were observed in the anomalous difference Fourier maps of the diffraction data collected at long-wavelength (1.9–3.15 Å, Table 1). In addition, no X-ray fluorescence was observed from crystals of NagBb when they were excited by X-rays with the scan energy for Cu, Ni, Co, Fe, and Mn atoms (data not shown). Therefore, based on these measurements it is thought that a  $\text{Ca}^{2+}$  and  $\text{Zn}^{2+}$  ion were present at M1 and M2-sites, respectively. B-factor values of the fully occupied Ca atoms were refined to 22.0 Å<sup>2</sup> and 32.4 Å<sup>2</sup> in chains A and B, respectively, after crystallographic refinement (Table 2) with these values consistent with the average B-factor of the nearby GalNAc ligand (36.0 Å<sup>2</sup> and 44.6 Å<sup>2</sup>). In terms of metal coordination distances of the GalNAc complex structure, the M1–O distances with Asp-322 and two waters (1.93–2.36 Å) are comparable or slightly shorter than the mean Ca–O distances for monodentate carboxylate (2.33–2.38 Å) and H<sub>2</sub>O (2.39–2.40 Å) in PDB and Cambridge Structural Database (22) (Table 3). The M1–N distances with three His residues are 2.16–2.36 Å but from the ionic radii difference between oxygen and nitrogen (~0.1 Å), the expected Ca–N distance is ~2.45 Å. There are only a few examples of Ca–N coordination in the database. Although octahedrally hexacoordinated site is frequently observed for  $\text{Ca}^{2+}$ , most of their coordinating ligand atoms are oxygens, and thus the present type of calcium binding site ( $\text{O}_3\text{N}_3$ ) is unprecedented.

*Activity measurements*—In a previous study it

was reported that the optimal pH for the activity and stable pH range of NagBb were 5.0 and 3.0–11.0, respectively (15). However, protein samples used in this study aggregated at acidic pH range (4.5–6.0) in all buffers tested. Thus, the activity measurements were conducted at pH 6.5 or 7.0. Kinetic parameters of NagBb toward *p*-nitrophenyl- $\alpha$ -D-galactosaminide (*p*NP- $\alpha$ -GalNAc) at pH 6.5 and 37°C were;  $K_m = 2.06 \pm 0.23$  mM,  $k_{\text{cat}} = 11.0 \pm 0.4$  s<sup>-1</sup>,  $k_{\text{cat}}/K_m = 5.34$  s<sup>-1</sup>mM<sup>-1</sup>. The  $k_{\text{cat}}/K_m$  value was comparable to that of the previous report (2.35 s<sup>-1</sup>mM<sup>-1</sup>) measured at pH 5.0 and 37°C (15).

We next evaluated the potency of three synthetic compounds that are inhibitors for  $\alpha$ - and  $\beta$ -N-acetylhexosaminidases (Fig. 4). The 1-deoxynojirimycin-type compound (**1**, GalNAc-DNJ) was a potent inhibitor with a  $K_i$  value of 51 nM. The crystal structure of NagBb complexed with **1** was also determined at 2.79 Å resolution (Table 2), and the inhibitor was bound at the active site in a similar manner with GalNAc (Fig. 2C). In contrast, the PUGNAc-like compound (**2**, Gal-PUGNAc) only slightly inhibited the activity of the enzyme, and this could be attributed to the aglycone binding site of NagBb not being wide enough to accommodate the large phenyl carbamate moiety. This was consistent with what has been observed for **2** with GH27  $\alpha$ -N-acetylgalactosaminidases (23). The thiazoline (**3**, Gal-NAG-thiazoline) is a potent inhibitor of  $\beta$ -N-acetylgalactosaminidases and  $\beta$ -N-acetylhexosaminidases that process GalNAc residues with a substrate-assisted mechanism (24–26). As expected, **3** did not substantially inhibit NagBb, which has been demonstrated to use a retaining mechanism that does not utilize the *N*-acetyl group of the substrate in catalysis.



Specific activities of the wild-type and active site mutants were also measured (Table 4). All of the mutations at the catalytic residues (Asp-435, Glu-478) and substrate-recognizing residues (Asp-330 and Trp-398) exhibited drastic reductions in enzymatic activity. Mutations at the nucleophile (Asp-435) almost completely abolished activity (>1000-fold reduction) while those at the acid/base (Glu-478) retained slight activity (~100-fold reduction). These results are consistent with observed reductions for mutants of retaining GHs when using a synthetic substrate with an activated leaving group (*e.g.* *p*NP) (27).

Mutants at the M1-site were also constructed (Table 4). As proteins of the two single mutants (H271A and H320A) were not stable at pH 6.5, their activities were measured at pH 7.0 and compared with the activity of wild-type at the same pH. Mutations at the M1-site also greatly reduced the activity, with H271A and H320A exhibited only ~3% activity relative to wild-type. The quadruple Ala mutant (H271A/H320A/D322A/H366A, HHDH-A) was stable at pH 6.5, and its activity decreased by 1000-fold. The crystal structure of the HHDH-A mutant was also determined at 1.9 Å resolution (Table 2). The quadruple mutant was confirmed from the clear electron density map of the active site (Fig. 2E), with there being no significant structural difference compared to the wild-type enzyme (RMSD for C $\alpha$  atoms = 0.22 Å with the GalNAc complex for all 8,096 protein atoms). No density peak was observed for the M1-site metal, and a glycerol molecule was bound at a position corresponding to the C4-C5-C6 atoms of GalNAc.

To investigate effects of the metal ions, the activity of EDTA-treated enzyme was measured

in the absence or presence of various metal ions (Fig. 5). The relative activity of the enzyme sample treated with 10 mM EDTA for 10 min at 4°C was 50% to that of untreated native enzyme. Further experiments were conducted using this EDTA-treated enzyme by adding metals or EDTA during the measurements. The activity further decreased in the presence of 1 mM EDTA. Addition of Fe<sup>2+</sup>, Mn<sup>2+</sup>, Co<sup>2+</sup> and Ni<sup>2+</sup> at low concentrations slightly recovered the activity (~30%) but the presence of these metals at high concentrations resulted in inhibition or aggregation of the protein. Other metals (Cu<sup>2+</sup>, Zn<sup>2+</sup>, Ca<sup>2+</sup> and Mg<sup>2+</sup>) showed inhibition regardless of concentrations. Considering that NagBb possesses at least two metal-binding sites with different species, the effects of metals on the activity may be complicated. The increase of the activity in the presence of Fe<sup>2+</sup>, Mn<sup>2+</sup>, Co<sup>2+</sup> and Ni<sup>2+</sup> may indicate that these metal ions are also suitable for use in substrate recognition at M1-site.

Overall the catalytic mechanism and active site architecture can be shown schematically, with the proposed reaction mechanism of NagBb, proceeding via the general mechanism of retaining GHs (28) (Fig. 6).

*Structural comparison with other GHs*—A structural similarity search of NagBb using the DALI Lite v.3 server (29) revealed that it has structural homology to GH13  $\alpha$ -amylase I from *Thermoactinomyces vulgaris* (2D0F; Z-score = 17.7; RMSD for 637 C $\alpha$  atoms = 5.6 Å) (30), GH36  $\alpha$ -galactosidase from *Thermotoga maritima* (5M0X; Z-score = 16.9; RMSD for 445 C $\alpha$  atoms = 4.6 Å) (31), GH101 endo- $\alpha$ -N-acetylgalactosaminidase from *Streptococcus pneumoniae* (SpGH101) (5A56;

Z-score = 16.6; RMSD for 505 C $\alpha$  atoms = 7.9 Å) (32), GH101 EngBF (2ZXQ; Z-score = 15.7; RMSD for 511 C $\alpha$  atoms = 8.3 Å) (33), and several other enzymes in GH13, GH31 and GH36. These structural homologues of NagBb are all retaining GHs acting on  $\alpha$ -galactosides or  $\alpha$ -glucosides. All of the GH13, GH31, GH36 and GH129 members share the following domain architecture; N-terminal  $\beta$ -sandwich domain, core catalytic ( $\alpha/\beta$ )<sub>8</sub>-barrel domain, and C-terminal  $\beta$ -sheet domain. Molecular sizes of GH101 enzymes (SpGH101 and EngBF) are more than twice of NagBb, and they carry peripheral domains in their N- and C-termini (domains 1 and 5–7), in addition to the central three domains (domains 2–4) similar to those of NagBb (34). The catalytic site architecture of these families can be divided into two groups. GH31 and GH36 have two Asp residues as the catalytic residues, whereas GH13, GH101 and GH129 (NagBb) employ Asp (nucleophile) and Glu (acid/base catalyst). In all of the five GH families, the acid/base residue is located in an *anti*-position to the ring oxygen of the carbohydrate in the -1 subsite (*anti*-protonator, Fig. 2A and Nerinckx, W. "syn/anti lateral protonation" in CAZypedia, available at URL <http://www.cazypedia.org/>, accessed 18 January 2017) (35). In addition, when the active site of NagBb was superimposed with these structural homologs, only the GH101 enzymes had their catalytic residues at similar positions (discussed below).

Although the ligand free crystal structures of two GH101 enzymes (SpGH101 and EngBF) have been reported (33,34), detailed substrate recognition and catalytic mechanism of this family only became clear when the structure of SpGH101 complexed with galacto-*N*-biose was

determined (32). Therefore, we compared the NagBb-GalNAc complex with the SpGH101-GNB complex (PDB ID 5A59). Superimposition of these structures revealed that their corresponding domains; N-domain (1-227) of NagBb and domain 2 (307-596) of SpGH101, and core barrel domain (228-591) of NagBb and domain 3 (597-863) of SpGH101, respectively, overlap well (Fig. 7A; RMSD = 2.9 Å for 202 C $\alpha$  atoms for the N-terminal  $\beta$ -sandwich domains, and RMSD = 3.2 Å for 284 C $\alpha$  atoms for the core domains). The N-terminal half of domain 4 of SpGH101 also adopts a similar 4-stranded  $\beta$ -sheet architecture to the C-terminal domain of NagBb (592-630). The active site (nucleophile, acid/base and GalNAc in the -1 subsite) of NagBb and SpGH101 also overlap (Fig. 7B). For SpGH101, a mechanism involving a Grotthuss proton shuttle has been proposed (32). This was due to the distance between the O1 atom of GalNAc (in galacto-*N*-biose) and the acid/base residue (Glu-796) being relatively long (4.3 Å), with a well-ordered water molecule bridging the two moieties. In the case of NagBb, the acid/base and O1 of the GalNAc is directly hydrogen-bonded (2.7 Å) without intervention of a water molecule, indicating that NagBb probably utilizes a common direct proton donation mechanism. Although SpGH101 uses a Grotthuss mechanism, mutations of the catalytic residues gave a similar result with NagBb; Mutations at the nucleophile (D764A) and the acid/base (E796A and E796Q) of SpGH101 decreased the  $k_{\text{cat}}$  value by 700-fold and 30-fold, respectively (32).

The highly conserved anchor residue (Asp-330), which holds the carbohydrate in the -1 subsite, and other residues recognizing the GalNAc moiety in NagBb are also similarly

positioned in SpGH101 (Fig. 7C). Residues involved in the recognition of the galactose moiety (-2 subsite) in SpGH101 (Gln-868, Lys-1156, Glu-1253 and Asp-1254) are not conserved in NagBb which highlights the different substrate specificities of the two enzymes. The M1-site is also not present in SpGH101, and only one histidine residue (His-694), which corresponds to His-366 in NagBb, is conserved.

Comparison of the substrate-binding pockets of NagBb and SpGH101 clearly illustrate their different substrate preferences as the -2 subsite of NagBb is blocked by protein residues (Fig. 7D). In GH101, two conserved Trp residues (Trp-724 and Trp-726 in SpGH101 and Trp-748 and Trp-750 in EngBF) play an important role in substrate recognition by providing aromatic platforms for the -1 and -2 subsites, respectively (33), and they jointly close on substrate binding (32). In NagBb, Trp-398 is solely involved in the recognition in the -1 subsite. The residue corresponding to the second Trp is Gly-400 in NagBb, and there is no substitute residue found in the -2 subsite. Of note is that no conformational change was observed upon substrate binding of NagBb. The activity of W398A mutant was less than 0.6 mU/mg (Table 4). This decrease was as prominent to the most destructive nucleophile mutant (D435A) and thus emphasized the importance of Trp-398 for the catalytic function of NagBb.

## DISCUSSION

*Involvement of metal site in GHs*—Even though the involvement of the M1-site in the substrate recognition of NagBb has been demonstrated in this study, interestingly this site is not conserved in distant members of GH129

(Fig. 3B). Structurally related enzymes in GH13, GH31, GH36 and GH101 also do not possess a metal site for substrate recognition. There are many examples of metal-dependent stabilization of GHs, including the case of  $\text{Ca}^{2+}$  for GH13  $\alpha$ -amylases (36). However, only a limited number of GH enzymes directly involve a metal for substrate recognition and catalysis. Inverting  $\alpha$ -mannosidases in GH38, GH47 and GH92 are prominent examples of the requirement of a divalent cation for enzymatic function (37-39). In these families, a  $\text{Ca}^{2+}$  or  $\text{Zn}^{2+}$  ion is involved in substrate recognition by bridging the O2 and O3 hydroxyls of Man in the -1 subsite. The ion also aids the catalysis by stabilizing the distorted sugar conformation in these  $\alpha$ -mannosidases (39,40). In addition, direct involvement of a divalent cation ( $\text{Ca}^{2+}$ ,  $\text{Zn}^{2+}$ , or  $\text{Mn}^{2+}$ ) in substrate recognition and catalysis has been demonstrated for GH4 (41), GH43 (42), GH62 (43), GH97 (44), GH106 (45), and GH127 (46). In the substrate binding sites of GH43 endo-1,5- $\alpha$ -L-arabinanase and GH62  $\alpha$ -L-arabinofuranosidases, a metal ion, which is hexacoordinated or heptacoordinated by side chain atoms of a His, a Gln and water molecules, was assigned as a  $\text{Ca}^{2+}$  (42,43). Treatment of these enzymes by EDTA had no apparent effect on the activities but mutation of the His residue significantly reduced the activity. To the best of our knowledge, NagBb is the first example of a GH enzyme that utilizes a metal ion for the recognition of the *N*-acetyl group of the sugar substrate.

*Possible molecular evolution of GH129 enzymes*—The structural similarity between GH129 and GH101 strongly suggests that they share a common protein ancestor. However, the



two families currently retain only a slight sequence homology within a limited region (15), and it is also interesting in how they acquired distinct substrate specificities and proton donation mechanisms.

The GH129 family currently consists of ~60 protein sequences exclusively from bacteria, and NagBb is the sole characterized member. Close homologs of NagBb (sequence identity > 60%) are only present in the genomes of *Bifidobacteria* species, and they currently account for more than half of the family member sequences (~30 sequences). It is noteworthy that many infant-associated bifidobacterial species, including *B. bifidum*, *B. breve*, *B. longum* and *B. longum* subsp. *infantis*, possess a NagBb homolog gene. Therefore, these bifidobacterial enzymes probably have the same activity (exo- $\alpha$ -N-acetylgalactosaminidase) and are involved in the utilization of human mucin glycoproteins, along with the endo-type GH101 enzymes. Distant NagBb homologs of GH129 members are mostly found in the genomes of soil bacteria. Interestingly, two human-related bacteria (*Bacillus cereus* and *Zoebellia galactanivorans*) possess a GH129 gene. *B. cereus* is an opportunistic pathogen which sometimes causes food intoxication and survives in the human gastrointestinal tract (47). For the marine bacterium *Z. galactanivorans*, its GH genes, which are active on algal polysaccharides, were thought to be horizontally transferred to the genomes of gut microbes (Bacteroidetes) in Japanese individuals, in conjunction with degradation of seaweeds in daily diet (48). As the M1-site residues are not conserved in the distant homologs of GH129 (Fig. 3B), they may indeed possess different substrate specificities. This may indicate possible horizontal gene

transfer events from food-associated microbes to the bifidobacterial species, and establishment of the metal binding site through molecular evolution.

## EXPERIMENTAL PROCEDURES

**Protein production and purification**—A vector encoding C-terminally His<sub>6</sub>-tagged NagBb (pET23b(+)-*nagbb*, residues 1–634) (15) was used for protein expression. The plasmid was introduced into *E. coli* C43 (DE3)-RIL (Stratagene, La Jolla, CA) for protein expression. The transformants were precultured overnight in Luria-Bertani medium containing 100 mg/L ampicillin and 35 mg/L chloramphenicol at 30°C. A 5 mL portion of the culture was inoculated in 1.5 L Luria-Bertani medium containing 100 mg/L ampicillin and 35 mg/L chloramphenicol at 37°C. When the optical density at 600 nm reached 0.6, isopropyl 1-thio- $\beta$ -D-galactopyranoside was added to a final concentration of 0.5 mM to induce protein expression. Following an additional incubation at 25°C for 10 h, the cells were harvested by centrifugation and suspended in 50 mM Tris-HCl (pH 7.8). Cell extracts were obtained by sonication followed by centrifugation to remove cell debris. The supernatant was applied to a nickel-nitrilotriacetic acid superflow column (Qiagen, Hilden, Germany) pre-equilibrated with 50 mM Tris-HCl (pH 7.8), and the column was washed with 30 mM imidazole and then eluted with 400 mM imidazole at a flow rate of 4 mL/min (on ice). For crystallization, the peak fraction was treated with a lysine methylation reaction (49) while this step was omitted in case of protein sample preparation for activity measurements. The proteins were then subjected

to gel filtration on a Superdex 200 pg 16/60 column (GE Healthcare, Fairfield, CT) pre-equilibrated with 20 mM Tris-HCl (pH 7.8) containing 150 mM NaCl at a flow rate of 1 mL/min (4°C). Protein concentrations were determined using the BCA protein assay kit (Thermo Fisher Scientific, Waltham, MA) with bovine serum albumin as a standard.

*Construction of mutant enzymes*—D435A, D435N, E478A, E478Q, D330A, D330N, H271A, H320A, D322A, H366A, H271A/H320A/D322A/H366A, and W398A mutants were constructed using the QuikChange site-directed mutagenesis method (Stratagene) with the pET23b(+)-*nagBb* plasmid as the template. The following primers and complementary strands were used:

5'-tatctggccgtgttcacctgcaatgaa-3' (D435A),  
 5'-tatctgaacgtgttcacctgcaatgaa-3' (D435N),  
 5'-tcgtcccgagggtgtccgactggg-3' (E478A),  
 5'-tcgtcccgagggtgtccgactggg-3' (E478Q),  
 5'-ggctacgccaacggccatcccgactat-3' (D330A),  
 5'-ggctacaacaacggccatcccgactat-3' (D330N),  
 5'-tgggtggctgtcggtatcaagacgaac-3' (H271A),  
 5'-tacctggctctggacggctggg-3' (H320A),  
 5'-catctggccggtggg-3' (D322A),  
 5'-ggcacggctgaccaataccgcgactat-3' (H366A),  
 5'-gcgatggcgccggcgccgcaaacg-3' (W398A).  
 For the H271A/H320A/D322A/H366A mutant, 5'-tacctggctctggccggtggg-3' (H320A/D322A) primer and complementary strands were used, and H271A and H366A mutations were introduced sequentially. The entire ORF sequence was checked to ensure that no base change other than the designed changes had occurred. The mutant enzymes were expressed and purified using a similar procedure as described above.

*Crystallography*—Crystals of NagBb complexed with GalNAc were obtained at 20°C using the sitting drop vapor diffusion method. Specifically, a 0.5  $\mu$ L protein solution containing 15 mg/mL NagBb and 50 mM GalNAc was mixed with an equal volume of a reservoir solution containing 0.02 M magnesium chloride, 0.1 M HEPES-NaOH (pH 6.5), and 7.5% PEG 3,350. Ligand free crystals were obtained in a similar manner with the above but GalNAc was not added. The quadruple HHDH-A mutant was crystallized with 20 mM GalNAc but the carbohydrate electron density was not observed in the resultant crystal structure. The GalNAc-DNJ complex was crystallized with 50 mM GalNAc-DNJ. Crystals were cryoprotected in the reservoir solutions supplemented with 20% (w/v) glycerol and were flash-cooled at 100 K in a stream of nitrogen gas.

*Data collection, structure determination by sulfur-SAD, and crystallographic refinement*—Diffraction data were collected using a photon counting pixel array detector Eiger X4M (Dectris) and charge-coupled device detector Quantum 270 (Area Detector System Corporation) on beamlines at the Photon Factory of the High Energy Accelerator Research Organization (KEK, Tsukuba, Japan). X-ray fluorescence analysis was performed using a multichannel analyzer installed on the beamlines. Data sets collected at low energy range (at wavelengths of 1.9 Å or longer) were processed with XDS (50), and statistics were calculated with AIMLESS (51). Initial phases were calculated from sulfur-SAD data sets from native NagBb crystals. Thirteen datasets, from two crystals, collected at 1.9 Å were merged, and used for substructure search with SHELXD (52).

A promising substructure was found with a resolution cut-off at 3.7 Å and 70 sites as input for the parameter FIND. This number for the FIND parameter was chosen due to the two molecules of NagBb have 76 sulfur-containing amino acids. Of the resulting heavy atom positions, only higher occupancy sites (~30) were selected for subsequent steps. Autobuilding was performed with PHENIX AutoSol (53) using the following inputs: substructure found by SHELXD, NagBb sequence, and merged reflection file from 12 datasets collected at 2.7 Å. Data sets collected at high energy range (at wavelengths of 1.29 Å or shorter) were processed using HKL2000 (54). Manual model rebuilding and refinement were achieved using COOT (55) and REFMAC5 (56), respectively. The statistics for data collection and refinement are listed in Tables 1 and 2. Molecular graphic images were prepared using PyMOL (DeLano Scientific, Palo Alto, CA).

**Activity measurement**—The hydrolysis activities of the enzymes were determined by measuring the amount *p*NP liberated from *p*NP- $\alpha$ -GalNAc (*p*-nitrophenyl- $\alpha$ -D-galactosaminide; Santa Cruz Biotechnology, TX). The standard assay mixture (50  $\mu$ L) contained 50 mM MES-NaOH (pH 6.5 or 7.0), 0.25 mM *p*NP- $\alpha$ -GalNAc, and NagBb protein (0.25  $\mu$ g for wild-type and 5.0  $\mu$ g for mutants). The reaction was started by mixing the NagBb protein solution (10  $\mu$ L) and remaining assay mixture (40  $\mu$ L), both of which had been preincubated at 37°C for 10 min. The reaction was monitored using a Benchmark plus microplate reader (Bio-Rad, Hercules, CA) with

a 96-well flat-bottomed transparent microplate #9018 (Corning, Corning, NY) at 37°C. The amount of 4-nitrophenolate produced was measured by the absorbance increase at 405 nm at 10 s intervals. The reaction was monitored for up to 20 min, and the enzyme concentration was set to observe a linear absorbance increase (initial rate) within this time frame. For the determination of kinetic constants, 0.125 to 8.0 mM *p*NP- $\alpha$ -GalNAc was used. Nonlinear regression curve fitting was calculated using the program Kaleidagraph (Synergy Software, Reading, PA).

**Inhibition**—The synthesis of the inhibitors **1-3** was conducted using previously described procedures (23,24,57). The  $K_i$  value for compound **1** was determined by linear regression of data tested at six concentrations of **1** (10 – 500 nM) and two concentrations of *p*NP- $\alpha$ -GalNAc (0.25 and 0.50 mM) at pH 6.5 using a Dixon plot (58).

**Metal dependency**—To remove metal ions, a NagBb protein solution was incubated with 10 mM Na-EDTA (pH 8.0) in 20 mM Tris-HCl (pH 7.8) at 4°C for 10 min. After removing excess EDTA by buffer exchange, its activity was measured as an “EDTA-treated” sample. For further analysis of the effect of metal ions ( $\text{Fe}^{2+}$ ,  $\text{Mn}^{2+}$ ,  $\text{Co}^{2+}$ ,  $\text{Ni}^{2+}$ ,  $\text{Cu}^{2+}$ ,  $\text{Zn}^{2+}$ ,  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$  in the form of chloride salts) or Na-EDTA (pH 8.0), reagents were added to the EDTA-treated enzyme to give final concentrations of 1.0  $\mu$ M – 1.0 mM. Activity measurements were conducted as described above.

**Acknowledgments:** We thank the staff of the Photon Factory for the X-ray data collection. This research is partially supported by the Platform Project for Supporting in Drug Discovery and Life Science Research (Platform for Drug Discovery, Informatics, and Structural Life Science) from Japan Agency for Medical Research and Development (AMED). We also thank the Centre for Microscopy, Characterisation and Analysis at The University of Western Australia, which is supported by University, State and Federal Government funding.

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

**Author contributions:** SF, TA and HA conceived the project and built the experimental design. MS conducted most of the experiments. MS and TI performed protein preparation and crystallization screening. MS, TA, YY, NM, TS and SF designed the normal crystallographic experiments and collected the X-ray data. DL, YY, NM, and TS designed and performed the sulfur-SAD experiments. DL and YY calculated the initial phase. MS, TA and SF refined the crystal structures. SSW, MH and KAS synthesized the inhibitors. MS, DL, YY, NM, TA and SF discussed the data. MS, SF, DL and YY wrote the paper. All authors commented on the manuscript.

## REFERENCES

1. Picard, C., Fioramonti, J., Francois, A., Robinson, T., Neant, F., and Matuchansky, C. (2005) Review article: bifidobacteria as probiotic agents - physiological effects and clinical benefits. *Aliment Pharmacol. Ther.* **22**, 495-512
2. Fukuda, S., Toh, H., Hase, K., Oshima, K., Nakanishi, Y., Yoshimura, K., Tobe, T., Clarke, J. M., Topping, D. L., Suzuki, T., Taylor, T. D., Itoh, K., Kikuchi, J., Morita, H., Hattori, M., and Ohno, H. (2011) Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature* **469**, 543-547
3. Kirjavainen, P. V., Arvola, T., Salminen, S. J., and Isolauri, E. (2002) Aberrant composition of gut microbiota of allergic infants: a target of bifidobacterial therapy at weaning? *Gut* **51**, 51-55
4. Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R. N., Anokhin, A. P., Heath, A. C., Warner, B., Reeder, J., Kuczynski, J., Caporaso, J. G., Lozupone, C. A., Lauber, C., Clemente, J. C., Knights, D., Knight, R., and Gordon, J. I. (2012) Human gut microbiome viewed across age and geography. *Nature* **486**, 222-227
5. Katayama, T. (2016) Host-derived glycans serve as selected nutrients for the gut microbe: human milk oligosaccharides and bifidobacteria. *Biosci. Biotechnol. Biochem.* **80**, 621-632
6. Matsuki, T., Yahagi, K., Mori, H., Matsumoto, H., Hara, T., Tajima, S., Ogawa, E., Kodama, H., Yamamoto, K., Yamada, T., Matsumoto, S., and Kurokawa, K. (2016) A key genetic factor for fucosyllactose utilization affects infant gut microbiota development. *Nat. Commun.* **7**, 11939
7. Ruas-Madiedo, P., Gueimonde, M., Fernandez-Garcia, M., de los Reyes-Gavilan, C. G., and Margolles, A. (2008) Mucin degradation by *Bifidobacterium* strains isolated from the human intestinal microbiota. *Appl. Environ. Microbiol.* **74**, 1936-1940
8. Desai, M. S., Seekatz, A. M., Koropatkin, N. M., Kamada, N., Hickey, C. A., Wolter, M., Pudlo, N. A., Kitamoto, S., Terrapon, N., Muller, A., Young, V. B., Henrissat, B., Wilmes, P., Stappenbeck, T. S., Nunez, G., and Martens, E. C. (2016) A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell* **167**, 1339-1353
9. Larsson, J. M., Karlsson, H., Sjoval, H., and Hansson, G. C. (2009) A complex, but uniform O-glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MSn. *Glycobiology* **19**, 756-766
10. Brockhausen, I., Schachter, H., and Stanley, P. (2009) O-GalNAc Glycans. in *Essentials of Glycobiology* (Variki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., and Etzler, M. E. eds.), 2nd Ed., Cold Spring Harbor Laboratory Press, New York. pp
11. Miwa, M., Horimoto, T., Kiyohara, M., Katayama, T., Kitaoka, M., Ashida, H., and



- Yamamoto, K. (2010) Cooperation of beta-galactosidase and  $\beta$ -N-acetylhexosaminidase from bifidobacteria in assimilation of human milk oligosaccharides with type 2 structure. *Glycobiology* **20**, 1402-1409
12. Kiyohara, M., Tanigawa, K., Chaiwangsri, T., Katayama, T., Ashida, H., and Yamamoto, K. (2011) An *exo*- $\alpha$ -sialidase from bifidobacteria involved in the degradation of sialyloligosaccharides in human milk and intestinal glycoconjugates. *Glycobiology* **21**, 437-447
13. Katayama, T., Sakuma, A., Kimura, T., Makimura, Y., Hiratake, J., Sakata, K., Yamanoi, T., Kumagai, H., and Yamamoto, K. (2004) Molecular cloning and characterization of *Bifidobacterium bifidum* 1,2- $\alpha$ -L-fucosidase (AfcA), a novel inverting glycosidase (glycoside hydrolase family 95). *J. Bacteriol.* **186**, 4885-4893
14. Ashida, H., Miyake, A., Kiyohara, M., Wada, J., Yoshida, E., Kumagai, H., Katayama, T., and Yamamoto, K. (2009) Two distinct  $\alpha$ -L-fucosidases from *Bifidobacterium bifidum* are essential for the utilization of fucosylated milk oligosaccharides and glycoconjugates. *Glycobiology* **19**, 1010-1017
15. Kiyohara, M., Nakatomi, T., Kurihara, S., Fushinobu, S., Suzuki, H., Tanaka, T., Shoda, S., Kitaoka, M., Katayama, T., Yamamoto, K., and Ashida, H. (2012)  $\alpha$ -N-acetylgalactosaminidase from infant-associated bifidobacteria belonging to novel glycoside hydrolase family 129 is implicated in alternative mucin degradation pathway. *J. Biol. Chem.* **287**, 693-700
16. Fujita, K., Oura, F., Nagamine, N., Katayama, T., Hiratake, J., Sakata, K., Kumagai, H., and Yamamoto, K. (2005) Identification and molecular cloning of a novel glycoside hydrolase family of core 1 type O-glycan-specific *endo*- $\alpha$ -N-acetylgalactosaminidase from *Bifidobacterium longum*. *J. Biol. Chem.* **280**, 37415-37422
17. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490-495
18. Yamada, Y., Matsugaki, N., Chavas, L. M., Hiraki, M., Igarashi, N., and Wakatsuki, S. (2013) Data Management System at the Photon Factory Macromolecular Crystallography Beamline. *J. Phys.: Conf. Ser.* **425**, 012017
19. Hiraki, M., Matsugaki, N., Yamada, Y., and Senda, T. (2016) Development of sample exchange robot PAM-HC for beamline BL-1A at the photon factory. *AIP Conf. Proc.* **1741**, 030029
20. Liebschner, D., Yamada, Y., Matsugaki, N., Senda, M., and Senda, T. (2016) On the influence of crystal size and wavelength on native SAD phasing. *Acta Crystallogr. D Struct. Biol.* **72**, 728-741
21. Matsuura, Y., Kusunoki, M., Harada, W., and Kakudo, M. (1984) Structure and possible catalytic residues of Taka-amylase A. *J. Biochem.* **95**, 697-702
22. Harding, M. M. (2006) Small revisions to predicted distances around metal sites in proteins. *Acta Crystallogr. D Struct. Biol.* **62**, 678-682

23. Stubbs, K. A., Macauley, M. S., and Vocadlo, D. J. (2009) A selective inhibitor Gal-PUGNAc of human lysosomal  $\beta$ -hexosaminidases modulates levels of the ganglioside GM2 in neuroblastoma cells. *Angew. Chem. Int. Ed. Engl.* **48**, 1300-1303
24. Amorelli, B., Yang, C., Rempel, B., Withers, S. G., and Knapp, S. (2008) N-Acetylhexosaminidase inhibitory properties of C-1 homologated GlcNAc- and GalNAc-thiazolines. *Bioorg. Med. Chem. Lett.* **18**, 2944-2947
25. Sumida, T., Fujimoto, K., and Ito, M. (2011) Molecular cloning and catalytic mechanism of a novel glycosphingolipid-degrading  $\beta$ -N-acetylgalactosaminidase from *Paenibacillus* sp. TS12. *J. Biol. Chem.* **286**, 14065-14072
26. Sumida, T., Stubbs, K. A., Ito, M., and Yokoyama, S. (2012) Gaining insight into the inhibition of glycoside hydrolase family 20 *exo*- $\beta$ -N-acetylhexosaminidases using a structural approach. *Org. Biomol. Chem.* **10**, 2607-2612
27. Miyanaga, A., Koseki, T., Matsuzawa, H., Wakagi, T., Shoun, H., and Fushinobu, S. (2004) Crystal structure of a family 54  $\alpha$ -L-arabinofuranosidase reveals a novel carbohydrate-binding module that can bind arabinose. *J. Biol. Chem.* **279**, 44907-44914
28. Rye, C. S., and Withers, S. G. (2000) Glycosidase mechanisms. *Curr Opin Chem Biol* **4**, 573-580
29. Holm, L., and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545-549
30. Abe, A., Yoshida, H., Tono-zuka, T., Sakano, Y., and Kamitori, S. (2005) Complexes of *Thermoactinomyces vulgaris* R-47  $\alpha$ -amylase 1 and pullulan model oligosaccharides provide new insight into the mechanism for recognizing substrates with  $\alpha$ -(1,6) glycosidic linkages. *FEBS J.* **272**, 6145-6153
31. Adamson, C., Pengelly, R. J., Shamsi Kazem Abadi, S., Chakladar, S., Draper, J., Britton, R., Gloster, T. M., and Bennet, A. J. (2016) Structural Snapshots for Mechanism-Based Inactivation of a Glycoside Hydrolase by Cyclopropyl Carbasugars. *Angew. Chem. Int. Ed. Engl.* **55**, 14978-14982
32. Gregg, K. J., Suits, M. D., 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
33. Suzuki, R., Katayama, T., Kitaoka, M., Kumagai, H., Wakagi, T., Shoun, H., Ashida, H., Yamamoto, K., and Fushinobu, S. (2009) Crystallographic and mutational analyses of substrate recognition of *endo*- $\alpha$ -N-acetylgalactosaminidase from *Bifidobacterium longum*. *J. Biochem.* **146**, 389-398
34. Caines, M. E., Zhu, H., Vuckovic, M., Willis, L. M., Withers, S. G., Wakarchuk, W. W., and Strynadka, N. C. (2008) The structural basis for T-antigen hydrolysis by *Streptococcus pneumoniae*: a target for structure-based vaccine design. *J. Biol. Chem.* **283**, 31279-31283
35. Nerinckx, W., Desmet, T., Piens, K., and Claeysens, M. (2005) An elaboration on the *syn-anti* proton donor concept of glycoside hydrolases: electrostatic stabilisation of the

- transition state as a general strategy. *FEBS Lett.* **579**, 302-312
36. Fitter, J. (2005) Structural and dynamical features contributing to thermostability in  $\alpha$ -amylases. *Cell Mol. Life Sci.* **62**, 1925-1937
  37. Vallee, F., Lipari, F., Yip, P., Sleno, B., Herscovics, A., and Howell, P. L. (2000) Crystal structure of a class I  $\alpha$ 1,2-mannosidase involved in *N*-glycan processing and endoplasmic reticulum quality control. *EMBO J.* **19**, 581-588
  38. van den Elsen, J. M., Kuntz, D. A., and Rose, D. R. (2001) Structure of Golgi  $\alpha$ -mannosidase II: a target for inhibition of growth and metastasis of cancer cells. *EMBO J.* **20**, 3008-3017
  39. Zhu, Y., Suits, M. D., Thompson, A. J., Chavan, S., Dinev, Z., Dumon, C., Smith, N., Moremen, K. W., Xiang, Y., Siriwardena, A., Williams, S. J., Gilbert, H. J., and Davies, G. J. (2010) Mechanistic insights into a  $\text{Ca}^{2+}$ -dependent family of alpha-mannosidases in a human gut symbiont. *Nat. Chem. Biol.* **6**, 125-132
  40. Mulakala, C., and Reilly, P. J. (2002) Understanding protein structure-function relationships in Family 47  $\alpha$ -1,2-mannosidases through computational docking of ligands. *Proteins* **49**, 125-134
  41. Varrot, A., Yip, V. L., Li, Y., Rajan, S. S., Yang, X., Anderson, W. F., Thompson, J., Withers, S. G., and Davies, G. J. (2005)  $\text{NAD}^+$  and metal-ion dependent hydrolysis by family 4 glycosidases: structural insight into specificity for phospho- $\beta$ -D-glucosides. *J. Mol. Biol.* **346**, 423-435
  42. de Sanctis, D., Inacio, J. M., Lindley, P. F., de Sa-Nogueira, I., and Bento, I. (2010) New evidence for the role of calcium in the glycosidase reaction of GH43 arabinanases. *FEBS J.* **277**, 4562-4574
  43. Siguier, B., Haon, M., Nahoum, V., Marcellin, M., Burlet-Schiltz, O., Coutinho, P. M., Henrissat, B., Mourey, L., O'Donohue, M. J., Berrin, J. G., Tranier, S., and Dumon, C. (2014) First structural insights into alpha-L-arabinofuranosidases from the two GH62 glycoside hydrolase subfamilies. *J. Biol. Chem.* **289**, 5261-5273
  44. Okuyama, M., Yoshida, T., Hondoh, H., Mori, H., Yao, M., and Kimura, A. (2014) Catalytic role of the calcium ion in GH97 inverting glycoside hydrolase. *FEBS Lett.* **588**, 3213-3217
  45. Ndeh, D., Rogowski, A., Cartmell, A., Luis, A. S., Basle, A., Gray, J., Venditto, I., Briggs, J., Zhang, X., Labourel, A., Terrapon, N., Buffetto, F., Nepogodiev, S., Xiao, Y., Field, R. A., Zhu, Y., O'Neill, M. A., Urbanowicz, B. R., York, W. S., Davies, G. J., Abbott, D. W., Ralet, M. C., Martens, E. C., Henrissat, B., and Gilbert, H. J. (2017) Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature* **544**, 65-70
  46. Ito, T., Saikawa, K., Kim, S., Fujita, K., Ishiwata, A., Kaeothip, S., Arakawa, T., Wakagi, T., Beckham, G. T., Ito, Y., and Fushinobu, S. (2014) Crystal structure of glycoside hydrolase family 127  $\beta$ -L-arabinofuranosidase from *Bifidobacterium longum*. *Biochem. Biophys. Res. Commun.* **447**, 32-37
  47. Berthold-Pluta, A., Pluta, A., and Garbowska, M. (2015) The effect of selected factors on the survival of *Bacillus cereus* in the human gastrointestinal tract. *Microb. Pathog.* **82**, 7-14

48. Hehemann, J. H., Correc, G., Barbeyron, T., Helbert, W., Czjzek, M., and Michel, G. (2010) Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464**, 908-912
49. Rayment, I. (1997) Reductive alkylation of lysine residues to alter crystallization properties of proteins. *Methods Enzymol.* **276**, 171-179
50. Kabsch, W. (2010) Xds. *Acta Crystallogr. D Struct. Biol.* **66**, 125-132
51. Evans, P. R., and Murshudov, G. N. (2013) How good are my data and what is the resolution? *Acta Crystallogr. D Struct. Biol.* **69**, 1204-1214
52. Sheldrick, G. M. (2008) A short history of SHELX. *Acta Crystallogr. A* **64**, 112-122
53. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Struct. Biol.* **66**, 213-221
54. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326
55. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Struct. Biol.* **66**, 486-501
56. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Struct. Biol.* **53**, 240-255
57. Best, D., Chairatana, P., Glawar, A. F. G., Crabtree, E., Butters, T. D., Wilson, F. X., Yu, C., Wang, W., Jia, Y., Adachi, I., Kato, A., and Fleet, G. W. J. (2010) Synthesis of 2-acetamido-1,2-dideoxy-D-galacto-nojirimycin [DGJNAc] from D-glucuronolactone: the first sub-micromolar inhibitor of  $\alpha$ -N-acetylgalactosaminidases. *Tetrahedron Lett.* **51**, 2222-2224
58. Dixon, M. (1953) The determination of enzyme inhibitor constants. *Biochem. J.* **55**, 170-171

## FOOTNOTES

This work was supported by JSPS-KAKENHI (15H02443 and 26660083 to SF and 15K07448 to HA), by Grant-in-Aid for JSPS Research Fellow (16J03683 to MS), and by the Science and Technology Research Promotion Program for Agriculture, Forestry, Fisheries and Food Industry (25010A). This work was also supported by Australian Postgraduate Awards from the University of Western Australia (to SSW and MH), a Bruce and Betty Green Scholarship (to SSW), a Jean Rogerson Postgraduate Scholarship (to MH) and the Australian Research Council (to KAS).

The abbreviations used are: NagBb,  $\alpha$ -N-acetylgalactosaminidase from *B. bifidum* JCM 1254; EngBF, endo- $\alpha$ -N-acetylgalactosaminidase from *B. longum* JCM 1217; GH, glycoside hydrolase; CAZy,

Carbohydrate-Active enZyme; SAD, single-wavelength anomalous dispersion; RMSD, root mean square deviation; *p*NP, *p*-nitrophenyl; SpGH101, endo- $\alpha$ -N-acetylgalactosaminidase from *Streptococcus pneumoniae*.



TABLE 1

Data collection statistics and peak height of metal sites of datasets using anomalous scattering

Dataset	S-SAD 1.9 Å <sup>a</sup>	S-SAD 2.7 Å <sup>a</sup>	Zn peak	Zn low energy	Ca high remote <sup>a</sup>	Ca low energy <sup>a</sup>
Data collection statistics						
Beamline	BL1A	BL1A	AR-NW12A	AR-NW12A	BL1A	BL1A
Wavelength (Å)	1.900	2.700	1.283	1.290	2.700	3.150
Space group	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Unit cell (Å)	<i>a</i> = 62.75 <i>b</i> = 127.23 <i>c</i> = 176.17	<i>a</i> = 63.06 <i>b</i> = 127.97 <i>c</i> = 176.92	<i>a</i> = 65.810 <i>b</i> = 128.21 <i>c</i> = 176.86	<i>a</i> = 65.790 <i>b</i> = 128.18 <i>c</i> = 176.85	<i>a</i> = 64.03 <i>b</i> = 126.70 <i>c</i> = 175.50	<i>a</i> = 64.09 <i>b</i> = 126.84 <i>c</i> = 175.61
Resolution (Å) <sup>b</sup>	47.43–2.26 (2.31–2.26)	47.66–2.4 (2.48–2.40)	50.0–3.00 (3.05–3.00)	50.00–3.20 (3.26–3.20)	50.00–2.61 (2.75–2.61)	50.00–3.20 (3.38–3.20)
Total reflections	9,659,079	6,820,294	204,414	167,605	441,314	246,371
Unique reflections	66,294	49,581	30,385	25,229	38,234	42,346
Completeness (%) <sup>b</sup>	99.2 (92.1)	87.3 (42.4)	100.0 (100.0)	99.9 (100.0)		
Anom. completeness (%) <sup>b</sup>	99.1 (91.5)	86.9 (41.6)	100.0 (100.0)	99.8 (99.9)	83.3 (25.6)	90.4 (50.4)
Redundancy <sup>b</sup>	145.7 (48.1)	137.6 (70.5)	6.7 (6.4)	6.6 (6.5)	13.6 (8.5)	13.0 (8.4)
Anom. redundancy <sup>b</sup>	74.9 (23.4)	70.8 (33.9)	3.6 (3.4)	3.6 (3.4)	11.5 (4.1)	5.8 (2.7)
Mean <i>I</i> / $\sigma$ ( <i>I</i> ) <sup>b</sup>	42.4 (2.5)	50.9 (2.4)	6.4 (0.7)	8.7 (1.4)	13.6 (1.6)	12.0 (4.5)
<i>R</i> <sub>merge</sub> (%) <sup>b</sup>	16.4 (194.9)	11.3 (245.9)	29.1 (228)	23.2 (147)	11.9 (70.9)	11.5 (20.1)
CC <sub>1/2</sub> <sup>b</sup>	(0.732)	(0.913)	(0.510)	(0.615)	(0.909)	(0.975)
Number of molecules per asymmetric unit	2	2	2	2	2	2
Peak height (σ) <sup>c</sup>						
M1 (chain A/chain B)	3.9/2.3	4.3/3.3	5.1/6.2	6.5/5.4	5.0/3.8	2.2/3.0
M2 (A/B)	7.6/6.3	10.2/10.9	12.3/9.6	—/—	12.8/10.3 <sup>d</sup>	10.3/9.0 <sup>d</sup>
Cys439-S (A/B)	4.7/4.1	7.3/4.8	—/—	—/—	7.5/4.9	5.8/4.2

<sup>a</sup> Data processed by XDS.<sup>b</sup> Values for the highest resolution shell are shown in parenthesis.<sup>c</sup> Peak heights of the anomalous difference Fourier map. "—", no significant peak observed.<sup>d</sup> These high values are due to the long tail of the L edges of Zn and the anomalous peaks of the two sulfur atoms of nearby cysteines.

TABLE 2

Data collection and refinement statistics for crystallographic structure determination

Dataset	GalNAc complex	Ligand free	HHDH-A mutant	Gal-NHAc- DNJ complex
Data collection statistics				
Beamline	BL1A	AR-NW12A	BL1A	BL1A
Wavelength (Å)	1.000	1.000	1.100	1.100
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell (Å)	$a = 65.640$	$a = 63.640$	$a = 62.890$	$a = 64.630$
	$b = 127.62$	$b = 127.76$	$b = 127.84$	$b = 128.48$
	$c = 176.82$	$c = 176.76$	$c = 176.36$	$c = 176.89$
Resolution (Å) <sup>a</sup>	50.0–2.10	103.55–2.65	103.51–1.90	103.95–2.79
	(2.14–2.10)	(2.70–2.65)	(1.93–1.90)	(2.85–2.80)
Total reflections	525,023	297,103	200,484	235,268
Unique reflections	87,389	42,830	77,231	37,110
Completeness (%) <sup>a</sup>	99.9 (99.7)	99.6 (99.2)	68.7 (85.4)	99.7 (100.0)
Redundancy <sup>a</sup>	6.0 (5.7)	6.9 (6.6)	2.6 (3.3)	6.3 (6.7)
Mean $I/\sigma(I)$ <sup>a</sup>	24.1 (2.5)	18.3 (1.9)	13.7 (2.4)	19.6 (2.4)
$R_{\text{merge}}$ (%) <sup>a</sup>	6.6 (37.2)	8.6 (62.7)	5.7 (43.9)	5.7 (51.8)
$CC_{1/2}$	(0.943)	(0.849)	(0.921)	(0.979)
Refinement statistics				
Resolution range (Å)	44.29–2.10	103.55–2.64	103.51–1.90	103.95–2.79
No. of reflections	82,971	40,656	73,352	35,202
$R$ -factor/ $R_{\text{free}}$ (%)	19.4/24.6	23.1/31.1	17.2/22.5	22.9/31.3
RMSD from ideal values				
Bond lengths (Å)	0.024	0.037	0.017	0.009
Bond angles (°)	1.777	1.713	1.765	1.461
Coordinate error (Å) <sup>b</sup>	0.186	0.409	0.187	0.459
Average $B$ -factor (Å <sup>2</sup> )				
Protein (chain A/chain B)	39.5/46.3	60.8/70.3	18.5/22.1	75.4/86.0
Active site ligand (A/B)	35.7/44.1	—	22.5/23.0	50.4/65.0
	(GalNAc)		(Glycerol)	(Gal-NHAc-DNJ)
Ca <sup>2+</sup> (M1-site, A/B)	21.7/31.9	55.8/36.3	—	37.3/50.0
Zn <sup>2+</sup> (M2-site, A/B)	27.3/33.0	51.4/57.4	11.7/13.8	50.9/56.5
Water	42.3	47.2	26.9	56.1
Ramachandran plot (%) <sup>c</sup>				
Favored	95.1	92.6	95.9	88.1

*Structure of GH129  $\alpha$ -N-acetylgalactosaminidase*

Allowed	4.5	6.2	3.5	9.5
Outlier	0.4	1.2	0.6	2.4
Number of molecules per asymmetric unit	2	2	2	2
PDB code	5WZN	5WZP	5WZQ	5WZR

<sup>a</sup> Values for the highest resolution shell are shown in parenthesis.

<sup>b</sup> Estimated based on  $R_{\text{free}}$  value.

<sup>c</sup> Determined by RAMPAGE server.

TABLE 3

Metal coordination distances in the GalNAc complex structure

Coordination type	Coordination atom	Distance (Å, A/B)
M1–N	His-271 N $\epsilon$ 2	2.28/2.33
M1–N	His-320 N $\delta$ 1	2.17/2.16
M1–N	His-366 N $\delta$ 1	2.17/2.36
M1–O	Asp-322 O $\delta$ 1	1.93/2.19
M1–O	water 1 <sup>c</sup>	2.36/2.31
M1–O	water 2 <sup>c</sup>	2.00/2.14
M2–N	His-450 N $\delta$ 1	1.99/2.08
M2–O	His-450 carbonyl O	1.96/2.05
M2–S	Cys-407 S $\gamma$	2.14/1.78
M2–S	Cys-445 S $\gamma$	1.86/1.91

<sup>c</sup> Water 1 is located at the opposite side of Asp-322, whereas water 2 is located in the same plane with the three His residues.

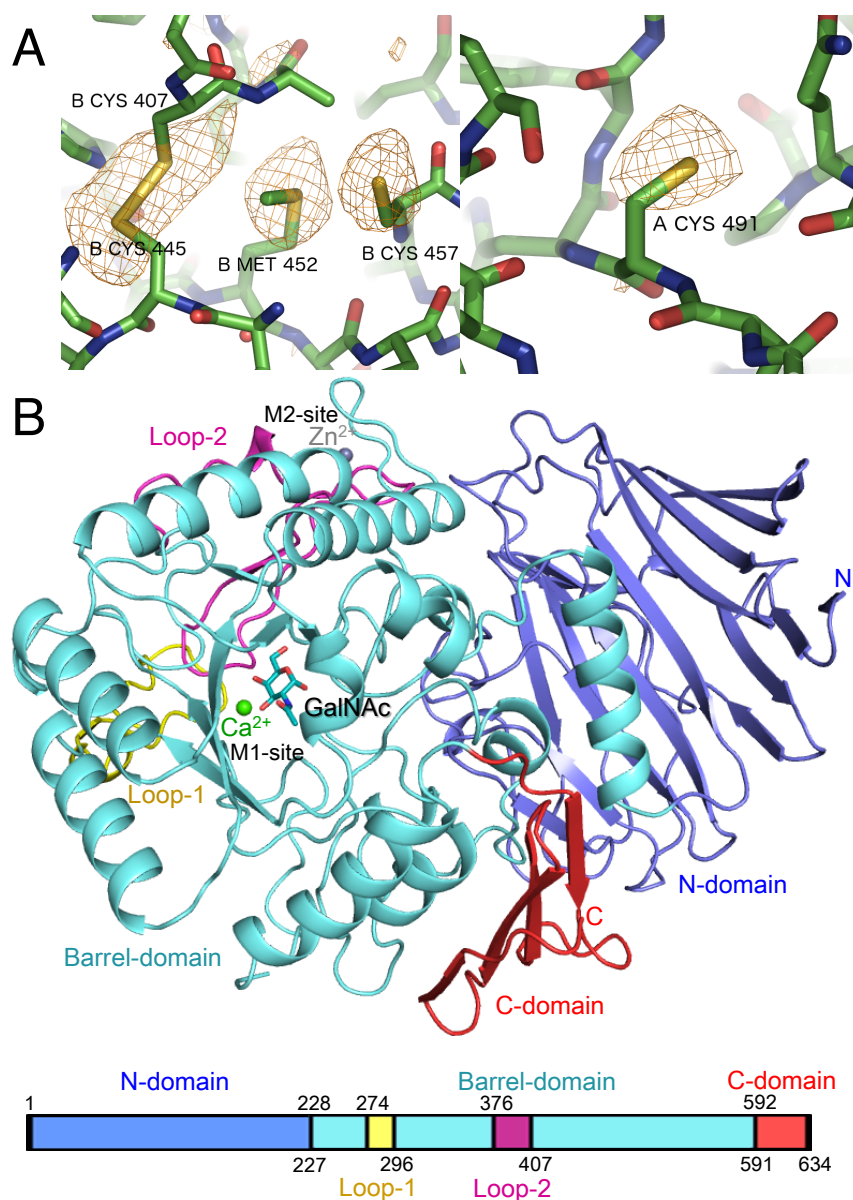
TABLE 4  
Activities of NagBb mutants

Enzyme	pH	Relative activity (mU/mg) <sup>a</sup>	Ratio (%)	Note
Wild type	6.5	628 ± 60	100	
Wild type	7.0	274 ± 26	43.7	
D435A	6.5	< 0.6	—	Nucleophile
D435N	6.5	0.7 ± 0.1	0.11	Nucleophile
E478A	6.5	8.6 ± 0.6	1.4	Acid/base
E478Q	6.5	8.8 ± 1.5	1.4	Acid/base
D330A	6.5	1.3 ± 0.0	0.21	Anchor
D330N	6.5	2.2 ± 0.6	0.35	Anchor
H271A	7.0	19 ± 1	3.01	Metal binding
H320A	7.0	19 ± 1	2.97	Metal binding
D322A	6.5	3.1 ± 0.3	0.50	Metal binding
H366A	6.5	2.6 ± 0.2	0.41	Metal binding
HHdHA <sup>b</sup>	6.5	0.6 ± 0.1	0.09	Metal binding
W398A	6.5	< 0.6	—	Subsite -1

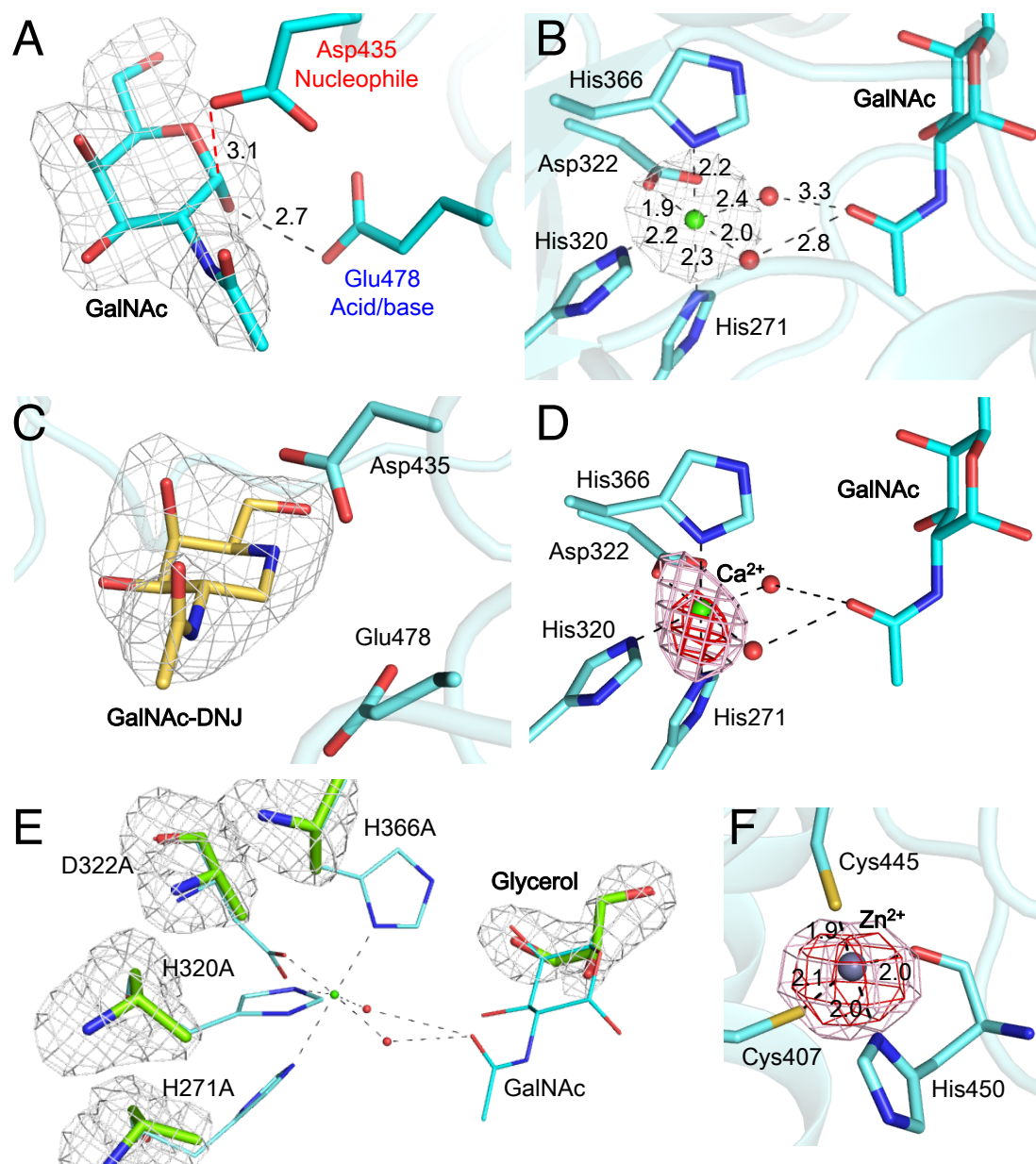
<sup>a</sup> Activities were measured at 0.25 mM *p*NP- $\alpha$ -GalNAc in 50 mM MES-NaOH (pH 6.5 or 7.0) at 37°C.

<sup>b</sup> Quadruple mutant (H271A/H320A/D322A/H366A).

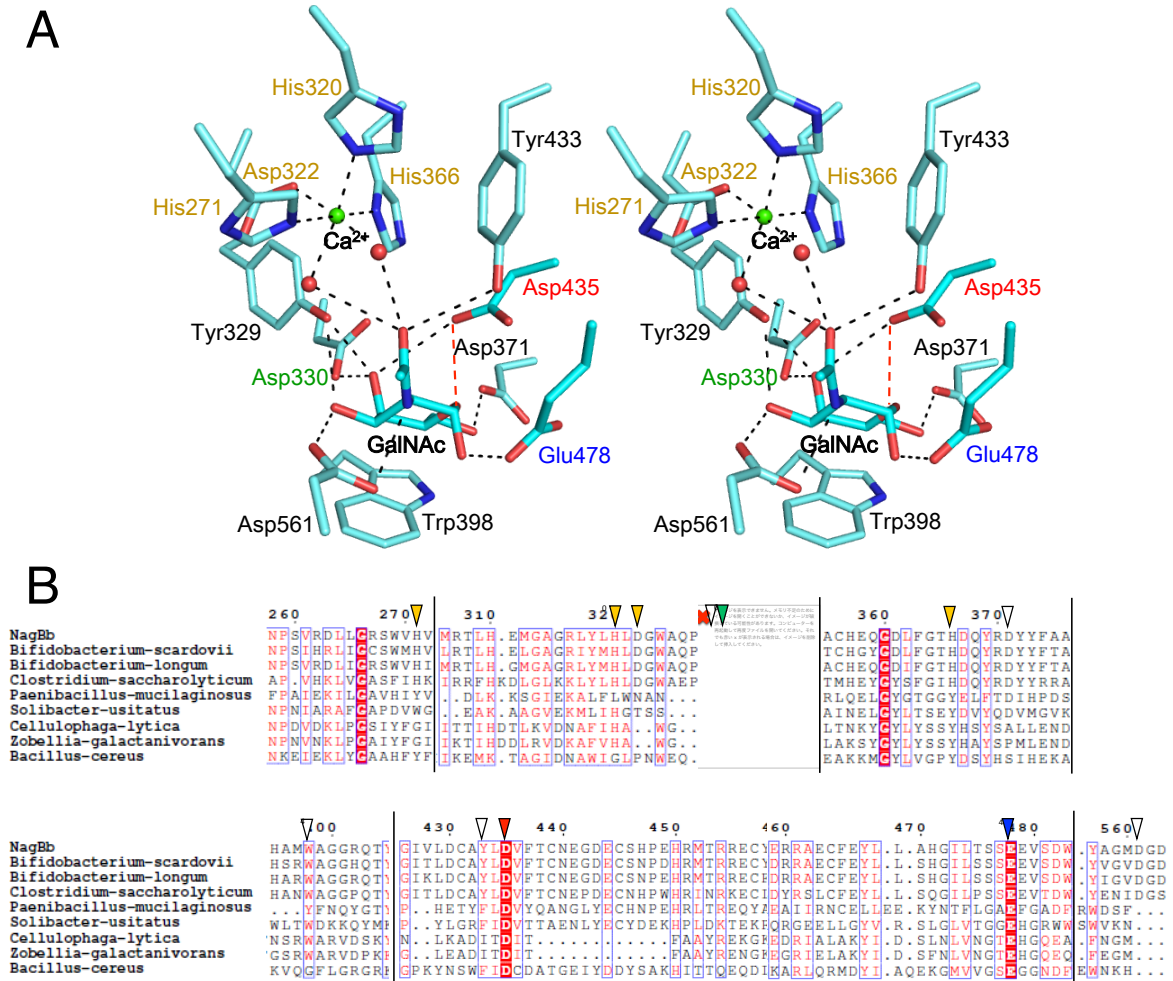




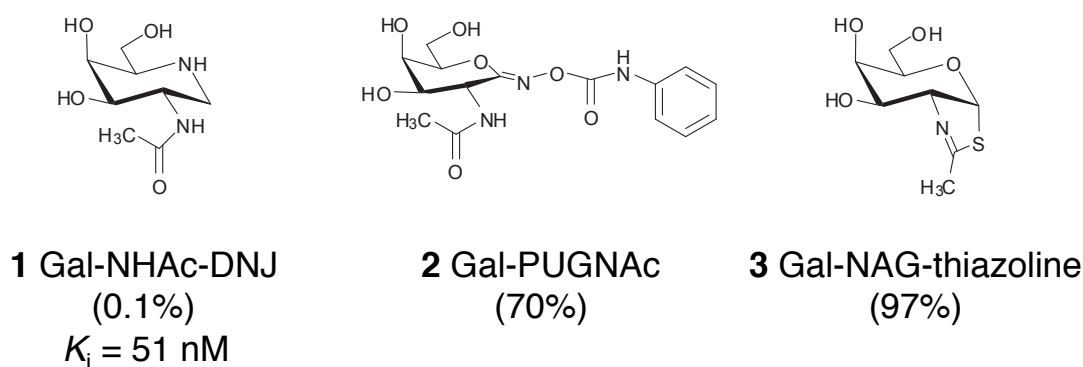
**FIGURE 1.** Sulfur SAD phasing and overall structure of NagBb. (A) Sulfur SAD peaks in the anomalous difference Fourier map (yellow mesh) of the phasing data (2.7 Å). Cys and Met residues in chain B (left, 3 $\sigma$ ) and Cys-491 in chain A (right, 4 $\sigma$ ) are shown. (B) Overall structure of NagBb complexed with GalNAc (cyan sticks). Bound Ca<sup>2+</sup> (green) and Zn<sup>2+</sup> (grey) atoms are shown as spheres. The protein is colored according to domain division as shown.



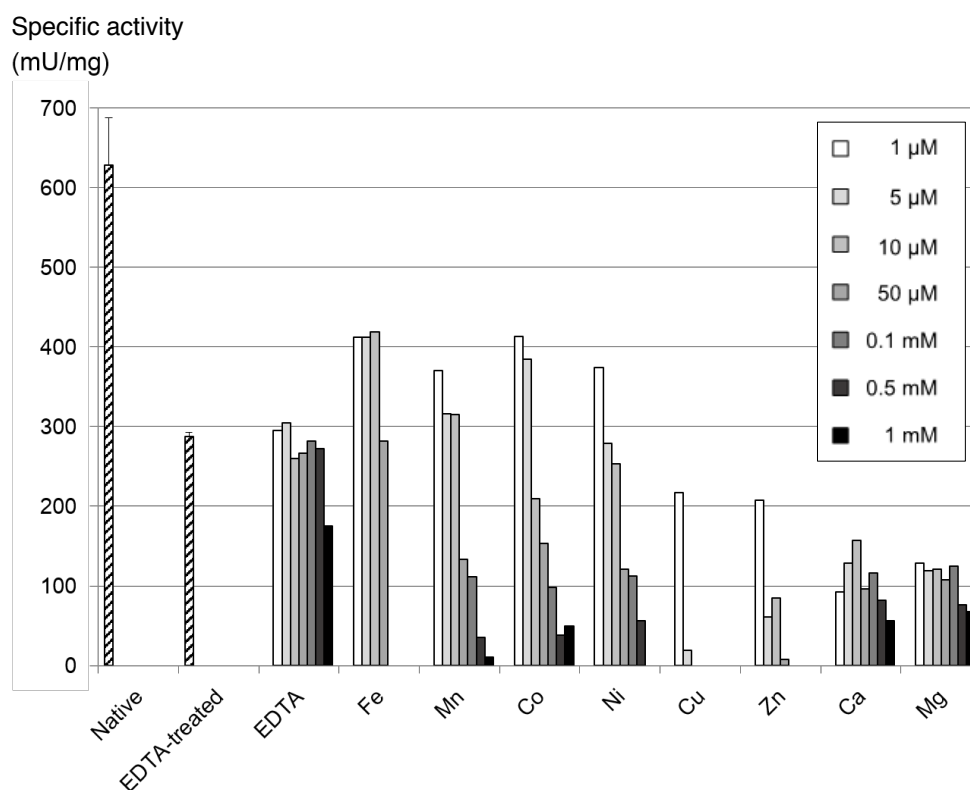
**FIGURE 2.** Electron density maps. (A) *mFo*-*DFc* omit electron density map ( $4.0\sigma$ , grey mesh) of GalNAc in the complex structure. (B) *mFo*-*DFc* omit electron density map ( $6.0\sigma$ , grey mesh) of M1-site metal. Two metal-coordinating waters are shown as red spheres. (C) *mFo*-*DFc* omit electron density map ( $3.5\sigma$ , grey mesh) of Gal-NHAc-DNJ in the complex structure. (D) M1-site and anomalous difference Fourier map (light pink mesh at  $3.5\sigma$  and red mesh at  $4.5\sigma$ ) of the data collected at a wavelength of  $2.700 \text{ \AA}$  (Ca high remote). (E) *mFo*-*DFc* omit electron density map ( $4.0\sigma$ , grey mesh) of quadruple HHDH-A mutant (green) for the mutated residues and glycerol. The GalNAc complex structure (cyan lines) is superimposed. (F) M2-site and anomalous difference Fourier map (light pink mesh at  $6.0\sigma$  and red mesh at  $8.0\sigma$ ) of the data collected at a wavelength of  $1.2826 \text{ \AA}$ .



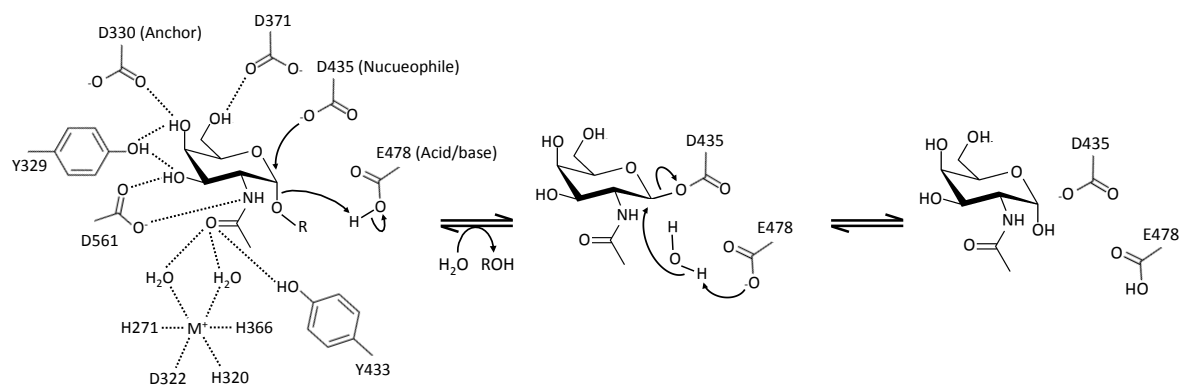
**FIGURE 3.** Active site residues. (A) Stereoview of the active site. Nucleophile (Asp-435), acid/base catalyst (Glu-478), highly conserved anchor residue (Asp-330), and residues coordinating M1-site are indicated by red, blue, green, and brown characters. (B) Partial amino acid sequence alignment of GH129 proteins showing the active site residues. The active site residues are indicated by colored triangles.



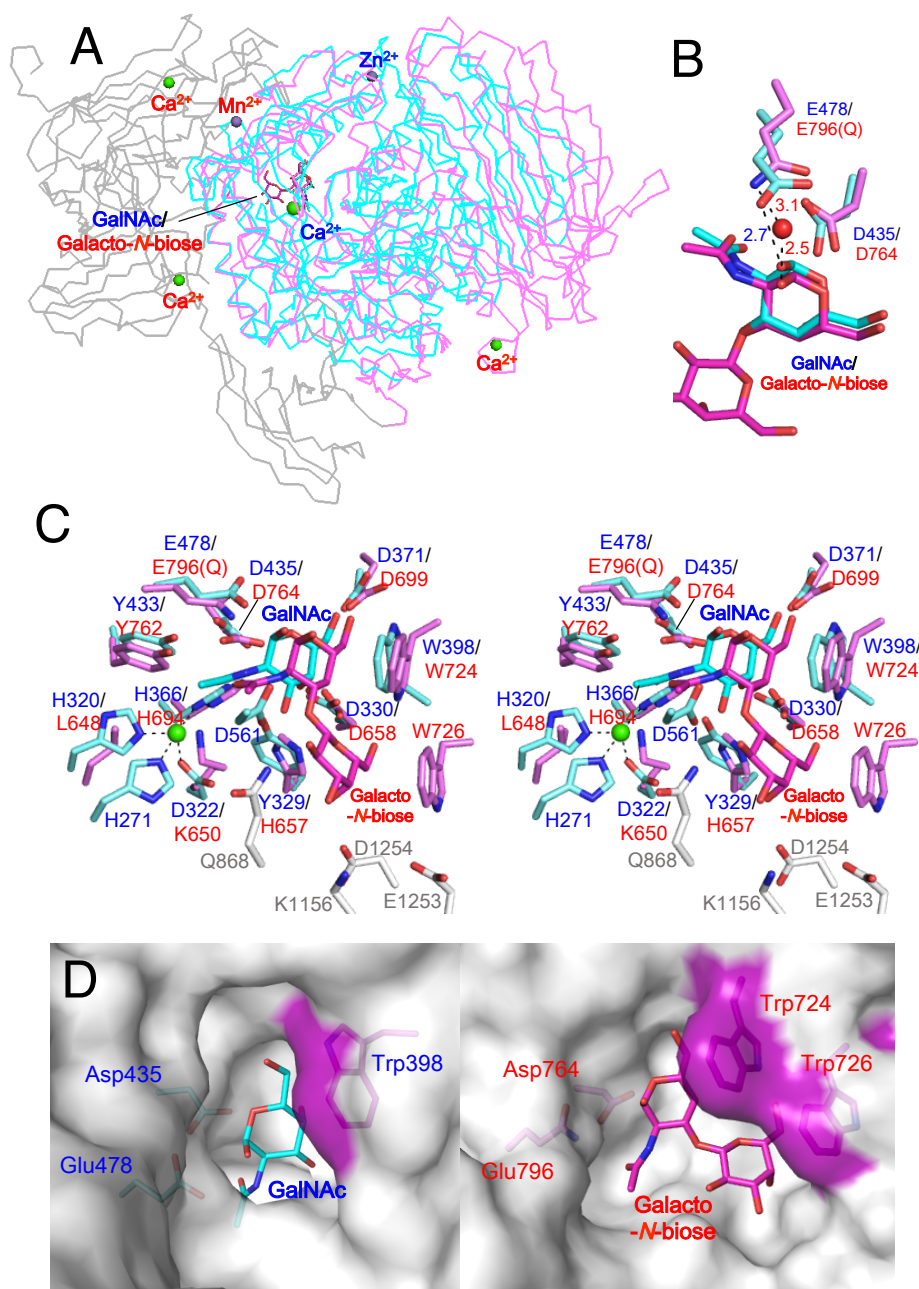
**FIGURE 4.** Structures of inhibitors used in the study. Relative activity of NagBb in the presence of 100  $\mu$ M of inhibitor using 0.25 mM *p*NP- $\alpha$ -GalNAc at pH 6.5, compared with that in the absence of the inhibitor (in parentheses). For compound **1**, the competitive  $K_i$  value is also shown.



**FIGURE 5.** Effect of EDTA and metal ions on the enzyme activity of NagBb. EDTA-treated enzyme was prepared by incubation with 10 mM EDTA at 4°C for 10 min and dialysis. The EDTA-treated enzyme was used for the assay in the presence of EDTA or a metal ion at indicated concentrations. Presence of high concentrations of  $\text{Fe}^{2+}$  ( $\geq 0.1$  mM),  $\text{Cu}^{2+}$  ( $\geq 10$   $\mu$ M), and  $\text{Zn}^{2+}$  ( $\geq 0.1$  mM) caused precipitation of NagBb protein.



**FIGURE 6.** Active site architecture and proposed reaction mechanism of NagBb.



**FIGURE 7.** Structural comparison with SpGH101. (A) Superimposition of  $\text{Ca}^{2+}$  backbone models of NagBb (cyan) and SpGH101 (domains 2 and 3 in pink and domains 4-7 in grey). Bound ligands and metal ions are indicated by blue and red characters for NagBb and SpGH101, respectively. (B) Carbohydrate ligand (GalNAc in NagBb and galacto-*N*-biose in SpGH101), catalytic residues, and a bridging water molecule for proton donation (SpGH101). (C) Stereographic superimpositions of the active sites of NagBb (blue) and SpGH101 (pink). (D) Molecular surfaces of the active site pocket of NagBb (left) and SpGH101 (right).

**The First Crystal Structure of a Family 129 Glycoside Hydrolase from a Probiotic Bacterium Reveals Critical Residues and Metal Co-factors**

Mayo Sato, Dorothee Liebschner, Yusuke Yamada, Naohiro Matsugaki, Takatoshi Arakawa, Siobhan S. Wills, Mitchell Hattie, Keith A. Stubbs, Tasuku Ito, Toshiya Senda, Hisashi Ashida and Shinya Fushinobu

*J. Biol. Chem.* published online May 25, 2017

---

Access the most updated version of this article at doi: [10.1074/jbc.M117.777391](https://doi.org/10.1074/jbc.M117.777391)

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at  
<http://www.jbc.org/content/early/2017/05/25/jbc.M117.777391.full.html#ref-list-1>