J. Biochem. 2009;146(3)389–398
doi:10.1093/jb/mvp086

Crystallographic and Mutational Analyses of Substrate Recognition of Endo-α-*N*-acetylgalactosaminidase from *Bifidobacterium longum*

Ryuichiro Suzuki[1], Takane Katayama[2], Motomitsu Kitaoka[3], Hidehiko Kumagai[2], Takayoshi Wakagi[1], Hirofumi Shoun[1], Hisashi Ashida[4], Kenji Yamamoto[4] and Shinya Fushinobu[1,*]

[1]Department of Biotechnology, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan; [2]Research Institute for Bioresources and Biotechnology, Ishikawa Prefectural University, Nonoichi-machi, Ishikawa 921-8836, Japan; [3]Enzyme Laboratory, National Food Research Institute, National Agriculture and Food Research Organization, 2-1-12 Kannondai, Tsukuba, Ibaraki 305-8642, Japan; and [4]Graduate School of Biostudies, Kyoto University, Kitashirakawa, Sakyo-ku, Kyoto 606-8502, Japan

Received March 24, 2009; accepted May 25, 2009; published online June 5, 2009

Endo-α-*N*-acetylgalactosaminidase (endo-α-GalNAc-ase), a member of the glycoside hydrolase (GH) family 101, hydrolyses the *O*-glycosidic bonds in mucin-type *O*-glycan between α-GalNAc and Ser/Thr. Endo-α-GalNAc-ase from *Bifidobacterium longum* JCM1217 (EngBF) is highly specific for the core 1-type *O*-glycan to release the disaccharide Galβ1-3GalNAc (GNB), whereas endo-α-GalNAc-ase from *Clostridium perfringens* (EngCP) exhibits broader substrate specificity. We determined the crystal structure of EngBF at 2.0 Å resolution and performed automated docking analysis to investigate possible binding modes of GNB. Mutational analysis revealed important residues for substrate binding, and two Trp residues (Trp748 and Trp750) appeared to form stacking interactions with the β-faces of sugar rings of GNB by substrate-induced fit. The difference in substrate specificities between EngBF and EngCP is attributed to the variations in amino acid sequences in the regions forming the substrate-binding pocket. Our results provide a structural basis for substrate recognition by GH101 endo-α-GalNAc-ases and will help structure-based engineering of these enzymes to produce various kinds of neo-glycoconjugates.

Key words: bifidobacteria, endo-α-*N*-acetylgalactosaminidase, galacto-*N*-biose, glycoside hydrolase family 101, mucin-type *O*-glycan.

Abbreviations: CBM, carbohydrate-binding module; core 1-*p*NP, Galβ1-3GalNAcα1-*p*NP; core 2-*p*NP, Galβ1-3(GlcNAcβ1-6)GalNAcα1-*p*NP; core 3-*p*NP, GlcNAcβ1-3GalNAcα1-*p*NP; endo-α-GalNAc-ase, endo-α-*N*-acetylgalactosaminidase; EngBF, endo-α-GalNAc-ase from *Bifidobacterium longum*; EngCP, endo-α-GalNAc-ase from *Clostridium perfringens*; EngEF, endo-α-GalNAc-ase from *Enterococcus faecalis*; EngPA endo-α-GalNAc-ase from *Propionibacterium acnes*; EngSP, endo-α-GalNAc-ase from *Streptococcus pneumoniae* R6; GH, glycoside hydrolase; GNB, galacto-*N*-biose (Galβ1-3GalNAc); LGA, Lamarckian genetic algorithm; LNB, lacto-*N*-biose I (Galβ1-3GlcNAc); MAD, multiple-wavelength anomalous dispersion; mGNB, 1-α-methyl-GNB; MPD, 2-methyl-2,4-pentanediol; RMSD, root mean square deviations.

Mucin-type *O*-glycans, in which an α-GalNAc binds to the hydroxyl group of Ser or Thr side chains, are involved in various important biological events including cell-to-cell communication in higher eukaryotes (*1*), bacterial adhesion to host cells (*1*, *2*), processing of hormones (*3*), endocytosis (*4*) and protein sorting (*5*). To date, at least eight different core structures of mucin-type *O*-glycans have been identified (*1*). Abnormal glycosylations in carcinoma patients (*6*) result in expression of the carcinoma markers, Tn-antigen (GalNAcα1-Ser/Thr) and T-antigen or core 1 (Galβ1-3GalNAcα1-Ser/Thr) (*7*, *8*). These markers are involved in the metastasis and invasion of cancer by acting as cell adhesion molecules (*8–10*). Thus, T-antigen analogues have the potential to serve as vaccines for cancer (*11*, *12*). In this context, many attempts have been made to produce the mucin-type glycan analogues containing T-antigen by applying various methods, e.g. transglycosylation activity (*13–18*).

Endo-α-*N*-acetylgalactosaminidase (endo-α-GalNAc-ase; EC 3.2.1.97) catalyses the hydrolysis of *O*-glycosidic bonds in mucin-type *O*-glycan between α-GalNAc and Ser/Thr to release oligosaccharides. The enzyme activity has been reported in several bacteria, i.e. *Streptococcus pneumoniae* (*19*), *Alcaligenes* sp. (*20*), *Bacillus* sp. (*13*), *Streptomyces* sp. (*21*) and so on. We have previously found that several bifidobacterial strains exhibit endo-α-GalNAc-ase activity, and the gene *engBF* encoding endo-α-GalNAc-ase was cloned from *Bifidobacterium longum* JCM1217 (*22*). A new glycoside hydrolase (GH) family 101 has been established in the Carbohydrate-Active enZyme (CAZy) database (http://www.cazy.org) (*23*) for this enzyme and its homologues. To date, GH101 endo-α-GalNAc-ase genes have been cloned from

*To whom correspondence should be addressed.
Tel/Fax: +81+3-5841-5151,
E-mail: asfushi@mail.ecc.u-tokyo.ac.jp

several bacteria and their gene products characterized: EngBF from *B. longum* JCM1217 (*24*), EngCP (CPE0693) from *Clostridium perfringens* strain 13 (*24*), EngCP (CEF0685) from *C. perfringens* ATCC13124 (*25*), EngPA (PPA1569) from *Propionibacterium acnes* ATCC25746 (*25*), EngEF from *Enterococcus faecalis* ATCC700802 (*25*) and EngEF from *E. faecalis* NBRC3971 (*26*). EngBF is highly specific for the core 1-type *O*-glycan to release the disaccharide Galβ1-3GalNAc (GNB, galacto-*N*-biose), but EngCP, EngEF and EngPA have broader substrate specificities. The broader substrate specificity of EngCP may contribute to the pathogenicity of *C. perfringens* (*24*).

Bifidobacteria constitute a major part of the microflora in the gastrointestinal tract, and they have received special attention due to their health-promoting effects in humans (*27*). EngBF is an extracellular membrane-bound enzyme that plays a critical role in the degradation of intestinal mucin (*28*). EngBF is possibly linked to the recently found metabolic pathway of Bifidobacteria specific for GNB and lacto-*N*-biose I (LNB, Galβ1-3GlcNAc) (*29*). LNB-containing oligosaccharides are abundantly present in human milk (*30*), and the GNB/LNB pathway is thought to be involved in the intestinal colonization of bifidobacteria (*29*, *31*). Extracellular enzymes of bifidobacteria, e.g. 1,2-α-L-fucosidase (*32*) and lacto-*N*-biosidase (*33*), liberate LNB from human milk oligosaccharides. The GNB released from *O*-glycans can be transported into the bifidobacterial cells via an ABC-type transporter specific for GNB and LNB (*34*), after which it is further metabolized by intracellular enzymes (*31*).

Very recently, the crystal structure of endo-α-GalNAc-ase from *S. pneumoniae* R6 (EngSP, PDB code 3ECQ) was determined at 2.9 Å for a selenomethionine-labelled protein as the first three-dimensional structure of GH101 enzyme (*35*). However, details of the structure–function relationship of GH101 endo-α-GalNAc-ase are still unclear. Here we report the crystal structure of native EngBF protein at 2.0 Å resolution. On the basis of structure, automated docking and mutational analyses, we have elucidated the critical residues for substrate recognition of GH101 endo-α-GalNAc-ases.

MATERIALS AND METHODS

Protein Expression and Purification—The overexpression vectors for N-terminally $(His)_6$-tagged EngBF (340–1,528) and EngBF (340–1,694) were constructed by inserting the PCR-amplified fragments of the *engBF* gene (*22*) into the *Nde*I and *Bam*HI sites of pET28b plasmid (Novagen, Madison, WI). The primers used were 5′-GCA TAT GGT CGC CTC CGA AAC GCT GAA GAC-3′ (forward) and 5′-CGG ATC CCG GCT CTC GTT GCC GGG AGC GGT C-3′ or 5′-GCG GAT CCT CGC GGT CCG TGG TCA AAC GC-3′ (reverse). Site-directed mutagenesis was performed with a QuikChange site-directed mutagenesis kit (Stratagene, La Jolla, CA). The following primers and their complementary primers were used: D682A, 5′-TGA AGG CCA CGC TAG CGG CCA CCC GGA-3′; N720A, 5′-TCG CTT CGG TGT TCA CGT CGC CGC CTC CGA AAT-3′; W748A, 5′-TGA GCT ACG GCG CCA ACT GGC TAG ATC AGG GTG T-3′; W748F, 5′-TGA GCT ACG GCT TTA ACT GGC TAG ATC AGG GTG T-3′; W748Y, 5′-TGA GCT ACG GCT ATA ACT GGC TAG ATC AGG GTG T-3′; W750A, 5′-TGA GCT ACG GCT GGA ACG CCC TAG ATC AGG GTG T-3′; W750F, 5′-TGA GCT ACG GCT GGA ACT TTC TAG ATC AGG GTG T-3′; W750Y, 5′- TGA GCT ACG GCT GGA ACT ATC TAG ATC AGG GTG T-3′; Y787F, 5′-ATG GAC TTC ATC TTC CTC GAC GTC TGG GGC AAC CT-3′; D789A, 5′-TAC CTC GCC GTG TGG GGC AAC CTG ACG TCT TCC GGT-3′; E822A, 5′-TAT GAC CAC CGC CTG GGG ATC CGG CAA CGA GT-3′; Q894A, 5′-AAG GAC TTC GAG GGC TGG GCG GGC CGC AAC GA-3′; K1199A, 5′-TCG CCA AGA ACT ACA TCG CAG CGT ACG GCC ACA ACA CGA A-3′; and D1295A, 5′-AAG GCG TTG AGG CCA ACC GGA TCC ACC TCT-3′. The expression plasmid was introduced into *Escherichia coli* BL21 CodonPlus(DE3)-RIL (Stratagene), and the transformants were grown in 3 l of Luria-Bertani broth containing 50 μg/ml kanamycin at 37°C to an A_{600} of 0.8. The protein expression was induced by adding 1 mM of isopropyl-1-thio-β-D-galactopyranoside, and then cultivated additionally for 16 h at 25°C. The harvested cells were resuspended in 20 mM sodium phosphate buffer (pH 7.2) and sonicated. The supernatant was filtrated using a 0.45-μm filter and purified with Ni-NTA agarose gel (Qiagen, Valencia, CA), MonoQ HR 5/5 and Hiload 16/60 Superdex 200 pg (GE Healthcare, Piscateway, NJ) column chromatographies. The protein concentration was determined with a BCA protein assay kit (Pierce, Rockford, IL) using bovine serum albumin as the standard. The selenomethionine-labelled protein was expressed in methionine auxotroph *E. coli* B834 (DE3) (Novagen, Madison, WI) in Se-Met core medium (Wako, Osaka, Japan) supplemented with 10 g/l D-glucose, 250 mg/l $MgSO_4 \cdot 7H_2O$, 4.2 mg/l $FeSO_4 \cdot 7H_2O$, 8.3 μl/l H_2SO_4, 10 ml/l Kao and Michayluk Vitamin Solution (Sigma-Aldrich, St Louis, MO), 50 μg/ml kanamycin, and 25 mg/l seleno-L-methionine (Wako). The purification procedures were similar to that of the native enzyme.

Enzyme Assay—Core 1-*p*NP was prepared from α-galactose 1-phosphate and GalNAcα1-*p*NP by using the reverse reaction of GNB/LNB phosphorylase (*36*, *37*). The standard assay buffer was composed of 50 mM HEPES–NaOH (pH 7.0) and 0.25 mM core 1-*p*NP. The amount of liberated *p*-nitrophenol was continuously measured at 30°C by monitoring the absorbance at 400 nm using a spectrophotometer DU-7400 (Beckman Coulter, Fullerton, CA). To determine the kinetic parameters, the reaction was continuously measured in the assay buffer composed of 80 μl of McIlvaine buffer (a mixture of 0.1 M citric acid and 0.2 M Na_2HPO_4, pH 7.0), 10 μl of 1% bovine serum albumin, 100 μl of 0.025–1.0 mM core 1-*p*NP and 10 μl of the purified enzyme. The pH dependence of the enzyme activity was determined by the discontinuous assay method using McIlvaine buffer. The reaction was stopped by adding an equal volume of 0.2 M Na_2CO_3, and then liberated *p*-nitrophenol was measured.

Crystallography—The crystals were grown at 20°C using the hanging-drop vapour-diffusion method.

Native EngBF crystals were obtained by mixing 1 μl of a protein solution (25 mg/ml) with 1 μl of a reservoir solution comprising 0.1 M 2-(*N*-morpholino)ethanesulphonic acid (MES) buffer (pH 6.9), 3% PEG 20000, 25% 2-methyl-2, 4-pentanediol (MPD), 0.2 M NaCl and 0.01 M $MnCl_2$. The selenomethionine-labelled EngBF crystals were obtained in the same way as the native ones, except that the pH of MES buffer was 6.5. The X-ray diffraction data were collected using charge-coupled device cameras on the BL6A station at the Photon Factory, and the NW12 station at the Photon Factory AR, High Energy Accelerator Research Organization (KEK), Tsukuba, Japan. The crystals were flash-cooled in a nitrogen stream at 100 K. Diffraction images were processed using the HKL2000 program suite (*38*). Twenty-four of the expected 26 Se sites were located using SnB (*39*), and initial phases were calculated using SOLVE/RESOLVE (*40*, *41*). Initial structural model building was conducted by using the program ARP/wARP (*42*). Manual model rebuilding, introduction of water molecules and refinement were achieved using Coot (*43*) and Refmac5 (*44*). Table 1 shows data collection and refinement statistics. Figures were prepared using PyMol (*45*).

Automated Docking—The AUTODOCK program version 4.0 (*46*) was used for automated docking analysis. The GNB and 1-α-methyl-GNB (mGNB) ligand models were prepared with the PCModel program (Serena Software, Bloomington, IN) and optimized using the MMX force field. Rotatable ligand bonds (12 in GNB and mGNB) were defined using the AutoDockTools interface. All water molecules in the EngBF structure were removed for docking analysis. After adding polar hydrogens, Gasteiger charges were calculated for the ligand and protein. Grid maps were prepared with $60 \times 60 \times 60$ points covering the substrate-binding pocket with a point spacing of 0.375 Å. For the Lamarckian genetic algorithm (LGA) search, the size of the initial random population was 150 individuals, the maximal number of energy evaluations was 2.5×10^6, the maximal number of generations was 27,000, the number of top individuals that survived into the next generation was 1, the rate of mutation was 0.02, the rate of crossover was 0.80 and

Table 1. **Data collection and refinement statistics.**

Data set	Native	MAD		
		Peak	Edge	Remote
Data collection statistics				
Space group	$P6_5$		$P6_5$	
Beam line	BL6A	NW12A	NW12A	NW12A
Wavelength (Å)	0.96787	0.97923	0.97939	0.96416
Unit cell (Å)	$a = b = 192.3$	$a = b = 192.2$	$a = b = 192.3$	$a = b = 192.3$
	$c = 123.0$	$c = 123.0$	$c = 123.0$	$c = 123.0$
Resolution (Å)	50–2.00 (2.07–2.00)	50–2.50 (2.59–2.50)	50–2.50 (2.59–2.50)	50–2.50 (2.59–2.50)
Total reflections	1,239,271	1,003,496	1,005,884	1,004,086
Unique reflections[a]	172,447	176,048	176,300	175,955
Completeness (%)	99.1 (99.8)	100 (100)	100 (100)	100 (100)
R_{merge} (%)	7.5 (46.7)	8.1 (32.5)	8.2 (35.6)	8.3 (38.8)
I/σI	27.2 (3.5)	29.0 (4.8)	28.0 (4.3)	26.1 (3.8)
Redundancy	7.2 (6.5)	5.7 (5.3)	5.7 (5.3)	5.7 (5.3)
Refinement statistics				
Resolution range (Å)	36.49–1.99			
No. of reflections	163,203			
R/R_{free} (%)	17.4/19.6			
RMSD from ideal values				
Bond lengths (Å)	0.014			
Bond angles (°)	1.412			
Average B-factor (Å^2)				
Protein	28.5			
Water	38.4			
Manganese ion	35.3			
MPD	51.4			
Ramachandran plot (%)[b]				
Favoured	88.7			
Allowed	11.2			
Disallowed	0.1			
Estimated overall coordinate error (Å)				
Based on R_{free}	0.101			
Based on maximum likelihood	0.066			
PDB code	2ZXQ			

[a]Bijvoet pairs unmerged for MAD data sets. [b]Calculated by PROCHECK. Values in parentheses are for highest resolution shell.

the average of the worst energy was calculated over a window of 10 generations. After 256 docking runs, all structures generated for a single compound were assigned to clusters based on a tolerance of 2.0 Å for all atom root mean square deviations (RMSDs) from the lowest energy structure. The best docking result of GNB was found in the third ranked cluster. The number of conformations in this cluster, the lowest binding energy and the mean binding energy were 5, −5.79 kcal/mol and −5.14 kcal/mol, respectively. The best docking result of mGNB was found in the second ranked cluster. The number of conformations in this cluster, the lowest binding energy and the mean binding energy were 3, −5.93 and −5.00 kcal/mol, respectively.

RESULTS AND DISCUSSION

Crystallography and Overall Structure—EngBF is a multidomain protein comprising 1,966 residues (Fig. 1). To facilitate crystallization screening, we first constructed various deletion mutants and determined a minimal region that retains catalytic activity. The region of residues 340–1,528 showed similar kinetic parameters measured against Galβ1-3GalNAcα1-*p*NP (core 1-*p*NP) as compared to those reported for the full-length enzyme (*22*) (Table 2). A region including the carbohydrate-binding module (CBM) family 32 domain (residues 340–1,694) also showed high catalytic activity. The optimal temperature and pH of this construct were 50°C and 5.0, respectively. The biochemical features were similar to the reported values of the full-length enzyme (*22*). Therefore, we used this construct (340–1,694) for our subsequent study.

The crystal structure of EngBF was determined by the multiple-wavelength anomalous dispersion (MAD) method using a selenomethionine derivative. Then we determined the crystal structure of the native EngBF protein at 2.0 Å resolution (Table 1 and Fig. 1). Supplementary Fig. S1 displays the electron density map of the substrate-binding pocket. The final model contains a monomer of EngBF (residues 341–1,475 and 1,482–1,524), four Mn^{2+} ions, three MPD molecules, and 1,403 waters. The C-terminal 170 residues, including the CBM32 domain, and residues 1,476–1,481 are not included due to disorder. We checked the length of the protein dissolved from the crystals by SDS–PAGE, and confirmed that it is not truncated by proteolysis during the crystallization (data not shown). There seems to be enough space for the C-terminal 170 residues in the loose crystal packing. The calculated V_M value and solvent content of the crystal were 4.44 $Å^3$/Da and 72.3%, respectively. The overall structure of EngBF is very similar to that of EngSP, and the RMSD for 1,024 Cα atoms is 1.28 Å. Figure 2A shows the superimposition of the overall structures of EngBF (colour coded by the domains) and EngSP (chains A and B shown in dark and light grey, respectively). The crystal structure of EngSP contains two chains in the asymmetric unit, and residues 119–307, 319–1,439 and 1,451–1,476 (chain A) and residues 122–222, 226–307 and 318–1,481 (chain B) are included in the final model. Seven domains were assigned for the EngSP structure. The EngBF structure includes domains 2–7, but lacks domain 1 (Fig. 1). Domain 3 of EngBF has a $(\beta/\alpha)_8$ barrel-like fold, which resembles the GH13 α-amylase family. The catalytic domain of the α-amylase family comprises a complete $(\beta/\alpha)_8$ barrel, but EngBF lacks β6–β8 strands and α6–α8 helices to form a broken barrel. In the structures of the α-amylase enzyme family, a conserved subdomain (domain B) is inserted between the β3 strand and α3 helix in the $(\beta/\alpha)_8$ barrel of domain A (*47*). A similar insertion corresponding to domain B was also found in

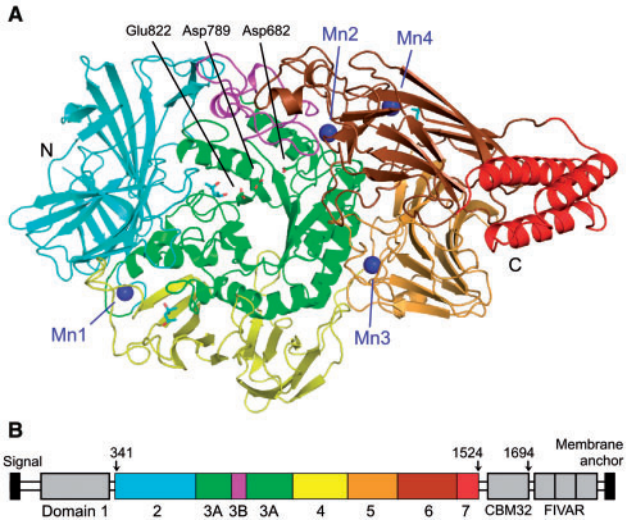

Fig. 1. **(A) The overall structure of EngBF in a ribbon representation and (B) a schematic illustration of its domain organization.** The final refined model contains the domains 2–7 (residues 341–1,475 and 1,482–1,524). Manganese ions are shown as blue spheres. The catalytically important residues (green) and MPD molecules (cyan) are shown as stick models.

Table 2. **Kinetic parameters of wild-type EngBF and its mutant enzymes against core 1-*p*NP.**

Enzyme[a]	Mutant group[b]	k_{cat} (sec^{-1})	K_m (μM)	k_{cat}/K_m ($sec^{-1}mM^{-1}$)
WT (340–1,528)		23.1 ± 0.5	43 ± 1.8	536.1 ± 12.0
WT (340–1,694)		23.8 ± 0.6	32 ± 3	762.7 ± 66.0
D682A	A	NA[c]		
D789A	A	NA[c]		
E822A	A	0.25 ± 0.02	225 ± 18	1.13 ± 0.07
Y787F	B	0.24 ± 0.001	9.8 ± 0.36	24.3 ± 0.97
D1295A	B	42.0 ± 0.72	575 ± 27	73.1 ± 2.2
N720A	C	43.8 ± 2.6	93.0 ± 14	477.3 ± 58.1
Q894A	C	43.8 ± 2.5	67.2 ± 9.6	657.3 ± 65.0
K1199A	C	41.0 ± 2.2	62.0 ± 1.6	660.9 ± 23.1
W748A	D	ND[d]	ND[d]	1.76 ± 0.097
W748F	D	ND[d]	ND[d]	6.1 ± 0.7
W748Y	D	ND[d]	ND[d]	8.3 ± 0.4
W750A	D	ND[d]	ND[d]	12.0 ± 0.14
W750F	D	30.7 ± 4.9	104 ± 24	294 ± 17
W750Y	D	ND[d]	ND[d]	116 ± 7.9

[a]Numbers in parentheses are residues in the construct. All mutant enzymes were made based on the construction comprising residues 340–1,694. [b]See text for detail. [c]NA, no activity was detected. [d]ND, K_m and k_{cat} could not be determined individually due to high K_m value.

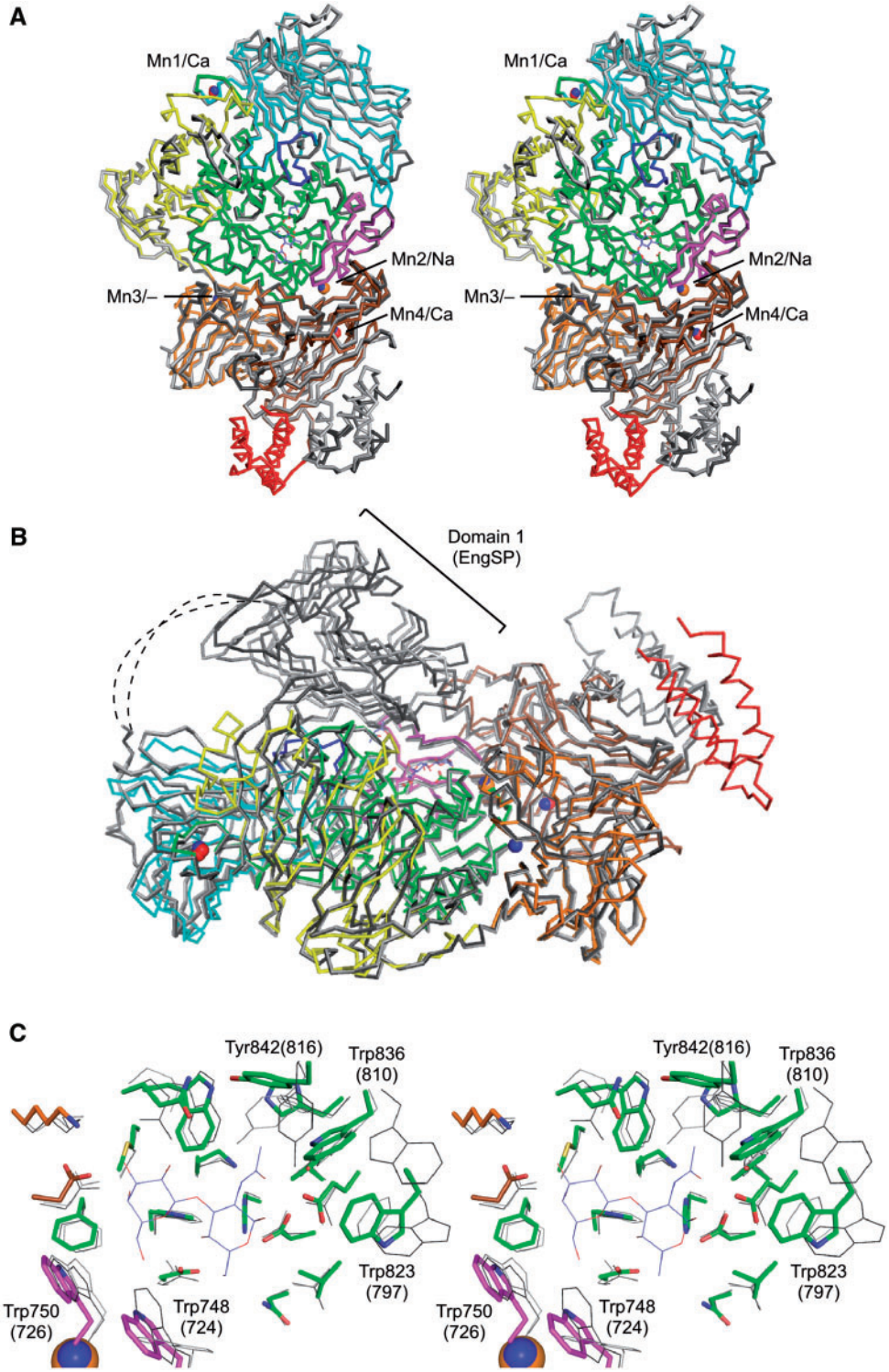

Fig. 2. **Structural comparison between EngBF (coloured by domains) and EngSP.** Both chains of EngSP (chains A and B in dark and light grey, respectively) are shown to illustrate the flexible regions in this molecule. (A) Stereoview of the overall structure. Domain 1 of EngSP is not shown. The 534–545 loop in EngBF (see text) is coloured in blue. MPD and docked GNB molecules are shown as thin blue lines. (B) The overall structure is with domain 1 of EngSP. Orientation of the molecule is rotated about 90° around the horizontal axis from Fig. 1A. The disordered linker between domains 1 and 2 in EngSP is indicated by dotted lines. (C) Stereoview of the active site. Deviated aromatic residues are labelled. The numbers in parentheses are those of EngSP.

Table 3. **Effect of metal ions for the stability of EngBF[a].**

Additive	Remaining activity (%)
(No additive)	29.3
$CaCl_2$	10.1
$MgCl_2$	45.4
$NiCl_2$	33.0
$ZnSO_4$	1.3
$MnCl_2$	53.8
$CdCl_2$	1.4
$CoCl_2$	36.6
Li_2SO_4	22.7
EDTA	14.6

[a]Remaining activity of EngBF after incubation at 60°C for 30 min in 50 mM HEPES–NaOH buffer (pH 7.0) with or without 5 mM additives.

EngBF. Therefore, we divide domain 3 of EngBF into subdomains 3A and 3B. The active site is located at the centre of domain 3, and subdomain 3B forms a wall of the substrate-binding pocket. Domain 3 is surrounded by four β-sandwich domains (domains 2, 4, 5 and 6). Domain 7 has a three-helix bundle fold, and the disordered region (residues 1,476–1,481) is a loop connecting the first and second helices. We found three out of the four manganese ions (Mn1–Mn3) at the interface between domain 3 and the surrounding domains. Mn4 is located in domain 6. See Supplementary Table S1 for details of the metal coordination. The EngSP structure contains two calcium ions and a sodium ion in both the chains. Mn1 and Mn4 in EngBF correspond to the calcium ions in EngSP, and Mn2 corresponds to the sodium ion (Fig. 2A). We examined the effect of metal ions on the stability of EngBF (Table 3). Mn^{2+} and Mg^{2+} showed a significant stabilizing effect, whereas metal extraction by addition of ethylenediaminetetraacetic acid (EDTA) led to destabilization. The stabilization effect of Mn^{2+} and Mg^{2+} was maximal at 6 μM and 0.3 mM, respectively (Supplementary Fig. S2). Considering the environment of the habitat of *Bifidobacteria* (the human gut), Mg^{2+} is most likely to stabilize this enzyme *in vivo*.

EngBF and EngSP share 45.8% sequence identity within the region of domains 2–7, and there are several insertions and deletions. Most of the insertions and deletions are located at the periphery of the molecule, and they do not appear to be related to substrate binding. However, an insertion of six residues in EngBF forms a loop of residues 534–545 as compared with EngSP (the blue in Fig. 2A). The 534–545 loop is located at the aglycon side and appears to be able to interact with the core protein of the mucin substrate. The distance between the Glu822 residue (acid/base catalyst) and this loop is ~10 Å. Domain 7 is located at largely different positions between EngBF and EngSP (Fig. 2A). Since these extracellular enzymes are anchored by the C-terminal transmembrane region, flexibility in the region after domain 7 may facilitate access to the large substrate, mucin. Domain 6 also shows relatively larger deviations compared to other domains (*2–5*) (Fig. 2A). Domain 1 in chain B of EngSP structure is poorly ordered, and the linker between domains 1 and 2 (residues 308–317) is completely disordered (*35*). Domain 1 in EngSP directly covers the substrate-binding pocket in the both chains, and it appears to hinder the accessibility of large substrates (Fig. 2B). As discussed in the report by Caines *et al*., the position of domain 1 in the EngSP structure seems to be an artefact by crystal packing (*35*).

Active Site—The catalytic residues of EngSP were inferred to be Asp764 (nucleophile) and Glu796 (acid/base) because they are located at the same positions as the catalytic residues of α-amylase (*35*). These residues correspond to Asp789 and Glu822 in EngBF, and it has been shown that mutation at these positions significantly impairs catalytic activity (*22*). In EngBF, Asp789 and Glu822 are located at the C-terminal loop of the fourth and fifth β-strand in domain 3A, respectively. There is a pocket filled by a number of water molecules in the vicinity of the catalytic residues (Fig. 3A). An extensive hydrogen bond network connects these water molecules. An MPD molecule is located on the edge of the pocket, and it is bound by a hydrophobic interaction with the Trp823 side chain as well as a hydrogen bond with a water molecule in the pocket.

Caines *et al*. (*35*) modelled a substrate (Galβ1-3GalNAcα1-Thr) in the active site of EngSP based on the glucose moiety at subsite −1 in the structure of α-amylase I from *Thermoactinomyces vulgaris* R-47 (TVA I) complexed with a pullulan model oligosaccharide, P2 (*48*). In this model, the sugar ring plane of Gal at subsite −2 is almost perpendicular to that of GalNAc at subsite −1; detailed modelling method is not described here. To obtain a probable substrate-bound model, we performed an automated docking analysis. The AutoDock 4.0 program was used to dock GNB or mGNB (ligand) to the crystal structure of EngBF from which the water molecules were removed (receptor). Results of 256 LGA runs were clustered with tolerance of the RMSD <2.0 Å. Obviously inadequate results, with positions of GalNAc moiety that were not adequately placed against the two catalytic residues, were rejected. Figure 3B shows the best docking results of GNB and mGNB with a blue stick model and thin orange lines, respectively. The position of docked mGNB is slightly different from that of GNB, but they are basically recognized by the same interactions. The RMSD between the non-hydrogen atoms of docked GNB and mGNB was 1.44 Å. The anomeric C1 atom of docked GNB is located in close proximity with the Oδ2 atom of Asp789 (3.1 Å), and the α-C1 hydroxyl group points in the opposite direction from the nucleophile. The C1 hydroxyl group forms a hydrogen bond with the Oε2 atom of Glu822 (2.7 Å). These results indicate that the reaction of EngBF proceeds through the standard double displacement mechanism of retaining GHs. Asp682 has been identified as the third catalytically important residue since mutation at this residue completely abolishes the activity (*22*). The side chain of Asp682 forms hydrogen bonds with three hydroxyl groups of the docked GNB (O4 and O6 of GalNAc and O6 of Gal), suggesting that this residue plays a critical role in substrate binding. The *N*-acetyl group of GalNAc fits into a hydrophobic pocket formed by the side chains of His835, Trp836 and Tyr842, and the

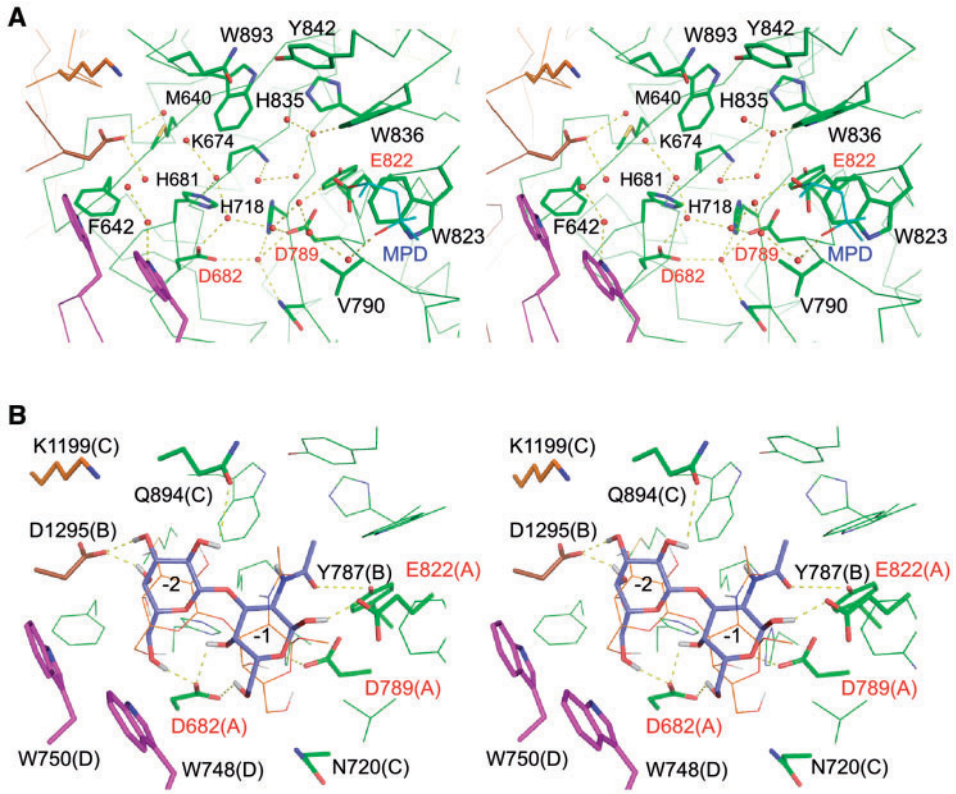

Fig. 3. **Stereoview of the active site of EngBF (A) and that with the docked molecules (B).** The catalytically important residues are labelled by red characters. (A) Active site residues (stick models) and water molecules (red spheres) in the substrate-binding pocket. The MPD molecule located near the active site is shown as thin lines (cyan). (B) Docked GNB (blue sticks) and mGNB (orange thin lines) in the pocket. Potential hydrogen bonds between docked GNB and EngBF protein are shown as yellow dotted lines. Groupings of the residues for mutational analysis are shown in parentheses in the labels (see text).

carbonyl oxygen forms a hydrogen bond with the side chain hydroxyl group of Tyr787. Asp1295 is located far from the catalytic residues, but it forms hydrogen bonds with the C3 and C4 hydroxyls of Gal at the non-reducing end in the docking model. The side chain of Gln894 is located near the C2 hydroxyl of Gal. Asn720 and Lys1199 do not directly interact with the docked GNB, but they form the rim of the pocket.

Stacking interactions between the sugar β-faces and aromatic residues are essential for almost all protein–carbohydrate interactions (*49*, *50*). In our docking model, however, the hydrophobic β-faces of both sugar units of the docked GNB molecule face upward and are exposed to the solvent. Interestingly, two aromatic residues—Trp748 and Trp750—form the wall of the pocket, and they are likely to form the stacking interaction when they lay down on the sugar β-faces by substrate-induced fit. To provide experimental evidence for the computational docking results and the hypothesis of induced fit, we prepared various mutant enzymes with amino acid substitutions in the substrate-binding pocket and examined their catalytic properties. We assigned the mutated residues to four different groups (Fig. 3B): residues essential for catalysis (group A), those involved in direct substrate recognition (group B), those involved in indirect substrate recognition (group C), and aromatic residues for potential stacking interactions (group D). Table 2 lists the kinetic parameters of the mutant enzymes against core 1-*p*NP. The results for the group A mutants were virtually the same as those of our previous report (*22*). We detected no catalytic activity for D682A and D789A, but the acid/base mutant (E822A) retained about 700-fold reduced activity for k_{cat}/K_m. The k_{cat}/K_m values of the group B mutants were significantly decreased. The decreased k_{cat} value of Y787F implies a supporting role of Tyr787 in the catalysis. The distance between the Oη atom of Tyr787 and the Oδ1 atom of Asp789 (nucleophile) is 3.3 Å. In contrast to Y787F, D1295A showed a significantly elevated K_m value as compared to the wild-type enzyme, suggesting that this residue is involved in substrate binding. The group C mutants (N720A, Q894A and K1199A) show similar features in their kinetic parameters; the increased k_{cat} and K_m values led to almost the same k_{cat}/K_m values as those of the wild-type enzyme. Therefore, these three residues are less important for substrate recognition than the group B residues. The group D mutants replaced with Ala (W748A and W750A) exhibited drastically decreased k_{cat}/K_m values, and we could not determine individual K_m and k_{cat} values due to the elevation of the K_m value. Therefore, Trp748 and Trp750 are more critical residues for substrate binding compared to group B and C residues. When the group D residues were replaced with aromatic residues (W748F, W748Y, W750F and W750Y),

they exhibited relatively high k_{cat}/K_m values compared to those of the Ala mutants. This result strongly supported our hypothesis for substrate-induced formation of stacking interactions. Because the k_{cat}/K_m values of the Trp748 mutants were smaller than those of Trp750 mutants (Table 2), the former residue is more important for substrate binding than the latter.

Comparison with EngSP—Figure 2C shows superimposition of EngBF and EngSP at the active site. The residues located around the docked GNB molecule are conserved and well overlapped. However, the side chains of several aromatic residues that are located relatively far from the GNB show significant deviations. Trp724 in EngSP takes different conformations in both chains from the corresponding Trp748 residue in EngBF, suggesting that this residue is flexible. We also performed GNB docking with flexible side chains at group D residues by liberating χ^1 and χ^2 torsion angles of Trp748 and Trp750. However, we could not obtain any plausible docking results (data not shown), suggesting that main chain movements are also necessary to create stacking interactions with the β-faces of GNB. When we soaked the EngBF crystals in buffers containing 1–10 mM GNB, the crystals cracked immediately (data not shown). Therefore, GNB binding likely induced a conformational change.

The right side of Fig. 2B shows a region that is presumed to bind the protein core region of the mucin substrates including the Ser or Thr residue. This region appears to define the acceptor specificity in the transglycosylation reaction. The 534–545 loop, which we described earlier, is located in close vicinity to this region (Fig. 2A) and may interact with larger substrates. There are significant structural differences in this region between EngBF and EngSP; Trp797, Trp810 and Tyr816 in EngSP exhibit large displacements from the corresponding residues in EngBF. The structural difference between the A and B chains is also significant. In the EngBF structure, the side chain of Trp823 forms a hydrophobic platform, and a cryoprotectant-derived MPD molecule is bound to it (Fig. 3A). Therefore, a certain flexibility exists in this region. Such feature seems to be suitable for binding the protein core region of mucin glycoproteins, which basically behave as random coils in the solution (*51*). We also tried to dock Galβ1-3GalNAcα1-Ser and Galβ1-3GalNAcα1-Thr to EngBF, but we could not obtain any plausible docking results (data not shown). The side chain of the aromatic residues in this region, e.g. that of Trp823, might have hindered docking at the aglycone-binding site.

Substrate Specificity for Glycans—We have previously reported relative activities of EngBF and EngCP from *C. perfringens* strain 13 towards 14 *p*NP-α-glycoside substrates in the same condition (*24*). For example, EngCP showed higher activities towards core 2-*p*NP (Galβ1-3(GlcNAcβ1-6)GalNAcα1-*p*NP), core 3-*p*NP (GlcNAcβ1-3GalNAcα1-*p*NP), GalNAcβ1-3GalNAcα1-*p*NP, and Glcβ1-3GalNAcα1-*p*NP than EngBF. Core 2 has a β1–6 branch of a GlcNAc unit at GalNAc (subsite −1) of core 1. The structures of the latter three substrates are similar to core 1, but the Gal unit at subsite −2 is substituted by GlcNAc, GalNAc and Glc, respectively.

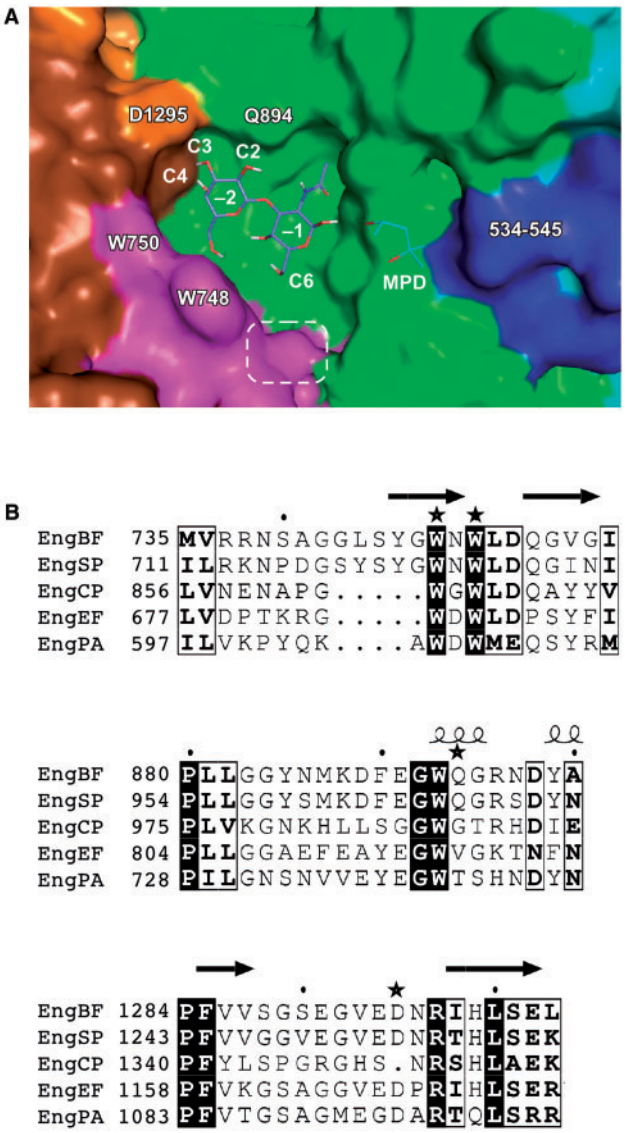

Fig. 4. **Molecular surface of the substrate-binding pocket of EngBF (A) and partial amino acid sequence alignment of GH101 endo-α-GalNAc-ases (B).** (A) The surface is coloured by domains as in Fig. 1 except for the 534–545 loop (blue). MPD (cyan thin lines) and docked GNB (blue stick models) molecules are shown. An area circled by a white dotted line is a putative binding site for the β1–6 GlcNAc branch in the core 2 glycan. (B) Sequence alignment in the region around the two group D Trp residues (upper), Gln894 (middle), and Asp1295 (bottom). Arrows and coils above the sequence alignments are β-strands and helices in the EngBF structure. Stars indicate the positions of Trp748, Trp750, Gln894 and Asp1295 in EngBF. The sequence alignment was prepared using clustalW (*52*) and ESPript (*53*).

Figure 4 shows the molecular surface of the substrate-binding pocket of EngBF and the partial sequence alignments of GH101 endo-α-GalNAc-ases. The extrapolated surface area from the C6 hydroxyl group of GalNAc at subsite −1 (circled by a white dotted line in Fig. 4A) is formed by the N-terminal region adjacent to Trp748. EngCP, EngEF and EngPA have a deletion of four or five residues in this region, suggesting that these

enzymes have space to accommodate the β1–6 GlcNAc branch in this region. The side chain of Gln894 is located at the area recognizing the C2 hydroxyl group of Gal at subsite–2, and the corresponding residues in EngCP, EngEF and EngPA are smaller ones (Gly, Val and Thr). Moreover, EngCP lacks the residue corresponding to Asp1295 in EngBF. The amino acid replacement and deletion in these regions might broaden the substrate specificity of EngCP and make the enzyme more active against core 3-*p*NP, GalNAcβ1-3GalNAcα1-*p*NP, and Glcβ1-3GalNAcα1-*p*NP than EngBF. Extensive hydrogen bond interactions between Asp682 and both sugars of docked GNB suggest that this enzyme strictly recognizes the β1–3 bond between the disaccharide. Actually, all GH101 endo-α-GalNAc-ases studied to date display no activity towards core 5-*p*NP, core 6-*p*NP, and core 7-*p*NP, which have α1–3, β1–6 and α1–6 bond, respectively. However, EngCP showed a slight activity towards core 8-*p*NP (Galα1-3GalNAcα1-*p*NP) (*24*). The relatively wider subsite at −2 may be able to accommodate the Gal moiety even when oriented differently. In summary, the amino acid sequence varieties in the area recognizing the docked GNB molecule can explain the difference in the substrate specificities of GH101 endo-α-GalNAc-ases.

Conclusions—On the basis of the crystal structure, automated docking and mutational analysis of EngBF, we could provide a detailed structural basis for the substrate specificities of GH101 endo-α-GalNAc-ases. Endo-α-GalNAc-ase is currently the sole enzyme that can release intact *O*-gylcan from mucin-type glycoproteins, and thus can be used as an important tool for glycan analyses. Moreover, this enzyme can synthesize mucin-type glycoconjugates due to its transglycosylation activity with wide acceptor specificity. However, the reaction efficiency requires further improvements, and the donor specificity is narrow since sugars other than core 1 disaccharide are not efficient donor substrates. Our results will help structure-based engineering of endo-α-GalNAc-ases to produce various kinds of neo-glycoconjugates, which have the potential to be novel vaccines for cancer.

ACKNOWLEDGEMENTS

We thank the staff of the Photon Factory for the X-ray data collection.

FUNDING

Program for Promotion of Basic Research Activities for Innovative Biosciences (PROBRAIN).

CONFLICT OF INTEREST

None declared.

REFERENCES

1. Variki, A., Cummings, R., Esko, J., Freeze, H., Hart, G., and Marth, J. D. (1999) *Essentials of Glycobiology*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
2. Hanisch, F.G. (2001) *O*-glycosylation of the mucin type. *Biol. Chem.* **382**, 143–149
3. Birch, N.P., Estivariz, F.E., Bennett, H.P., and Loh, Y.P. (1991) Differential glycosylation of *N*-POMC1-77 regulates the production of gamma 3-MSH by purified pro-opiomelanocortin converting enzyme. A possible mechanism for tissue-specific processing. *FEBS Lett.* **290**, 191–194
4. Atiya-Nasagi, Y., Cohen, H., Medalia, O., Fukudan, M., and Sagi-Eisenberg, R. (2005) *O*-glycosylation is essential for intracellular targeting of synaptotagmins I and II in non-neuronal specialized secretory cells. *J. Cell Sci.* **118**(Pt. 7), 1363–1372
5. Tian, E. and Ten Hagen, K.G. (2007) A UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase is required for epithelial tube formation. *J. Biol. Chem.* **282**, 606–614
6. Brockhausen, I. (2006) Mucin-type *O*-glycans in human colon and breast cancer: glycodynamics and functions. *EMBO reports.* **7**, 599–604
7. Anglin, J.H. Jr., Lerner, M.P., and Nordquist, R.E. (1977) Blood group-like activity released by human mammary carcinoma cells in culture. *Nature* **269**, 254–255
8. Springer, G.F. (1984) T and Tn, general carcinoma autoantigens. *Science* **224**, 1198–1206
9. Springer, G.F. (1995) T and Tn pancarcinoma markers: autoantigenic adhesion molecules in pathogenesis, prebiopsy carcinoma-detection, and long-term breast carcinoma immunotherapy. *Crit. Rev. Oncog.* **6**, 57–85
10. Hollingsworth, M.A. and Swanson, B.J. (2004) Mucins in cancer: protection and control of the cell surface. *Nat. Rev. Cancer.* **4**, 45–60
11. Danishefsky, S.J. and Allen, J.R. (2000) From the laboratory to the clinic: a retrospective on fully synthetic carbohydrate-based anticancer vaccines. *Angew. Chem. Int. Ed. Engl.* **39**, 836–863
12. Tarp, M.A. and Clausen, H. (2008) Mucin-type *O*-glycosylation and its potential use in drug and vaccine development. *Biochim. Biophys. Acta.* **1780**, 546–563
13. Ashida, H., Yamamoto, K., Murata, T., Usui, T., and Kumagai, H. (2000) Characterization of endo-alpha-N-acetylgalactosaminidase from *Bacillus* sp. and syntheses of neo-oligosaccharides using its transglycosylation activity. *Arch. Biochem. Biophys.* **373**, 394–400
14. Ashida, H., Yamamoto, K., and Kumagai, H. (2001) Enzymatic syntheses of T antigen-containing glycolipid mimicry using the transglycosylation activity of endo-alpha-*N*-acetylgalactosaminidase. *Carbohydr. Res.* **330**, 487–493
15. Marcaurelle, L.A. and Bertozzi, C.R. (2002) Recent advances in the chemical synthesis of mucin-like glycoproteins. *Glycobiology.* **12**, 69R–77R
16. Hanisch, F.G. (2005) Design of a MUC1-based cancer vaccine. *Biochem. Soc. Trans.* **33**(Pt 4), 705–708
17. DeFrees, S., Wang, Z.G., Xing, R., Scott, A.E., Wang, J., Zopf, D., Gouty, D.L., Sjoberg, E.R., Panneerselvam, K., Brinkman-Van der Linden, E.C., Bayer, R.J., Tarp, M.A., and Clausen, H. (2006) GlycoPEGylation of recombinant therapeutic proteins produced in *Escherichia coli*. *Glycobiology.* **16**, 833–843
18. Amano, K., Chiba, Y., Kasahara, Y., Kato, Y., Kaneko, M.K., Kuno, A., Ito, H., Kobayashi, K., Hirabayashi, J., Jigami, Y., and Narimatsu, H. (2008) Engineering of mucin-type human glycoproteins in yeast cells. *Proc. Natl. Acad. Sci. USA* **105**, 3232–3237
19. Bhavanandan, V.P., Umemoto, J., and Davidson, E.A. (1976) Characterization of an endo-alpha-*N*-acetyl galactosaminidase from *Diplococcus pneumoniae*. *Biochem. Biophys. Res. Commun.* **70**, 738–745
20. Fan, J.Q., Kadowaki, S., Yamamoto, K., Kumagai, H., and Tochikura, T. (1998) Purification and characterization of

endo-alpha-*N*-acetylgalactosaminidase from *Alcaligenes* sp. *Agric. Biol. Chem.* **52**, 1715–1723
21. Ishii-Karakasa, I., Iwase, H., Hotta, K., Tanaka, Y., and Omura, S. (1992) Partial purification and characterization of an endo-alpha-*N*-acetylgalactosaminidase from the culture medium of *Streptomyces* sp. OH-11242. *Biochem. J.* **288**, 475–482
22. 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
23. Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2008) The Carbohydrate-active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic. Acid. Res.* **37**(Database issue), D233–D238
24. Ashida, H., Maki, R., Ozawa, H., Tani, Y., Kiyohara, M., Fujita, M., Imamura, A., Ishida, H., Kiso, M., and Yamamoto, K. (2008) Characterization of two different endo-alpha-*N*-acetylgalactosaminidases from probiotic and pathogenic enterobacteria, *Bifidobacterium longum* and *Clostridium perfringens*. *Glycobiology.* **18**, 727–734
25. Koutsioulis, D., Landry, D., and Guthrie, E.P. (2008) Novel endo-alpha-*N*-acetylgalactosaminidases with broader substrate specificity. *Glycobiology.* **18**, 799–805
26. Goda, H.M., Ushigusa, K., Ito, H., Okino, N., Narimatsu, H., and Ito, M. (2008) Molecular cloning, expression, and characterization of a novel endo-alpha-*N*-acetylgalactosaminidase from *Enterococcus faecalis*. *Biochem. Biophys. Res. Commun.* **375**, 541–546
27. Salminen, S., von Wright, A., Morelli, L., Marteau, P., Brassart, D., de Vos, W.M., Fonden, R., Saxelin, M., Collins, K., Mogensen, G., Birkeland, S.E., and Mattila-Sandholm, T. (1998) Demonstration of safety of probiotics—a review. *Int. J. Food Microbiol.* **44**, 93–106
28. 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
29. Kitaoka, M., Tian, J., and Nishimoto, M. (2005) Novel putative galactose operon involving lacto-*N*-biose phosphorylase in *Bifidobacterium longum*. *Appl. Environ. Microbiol.* **71**, 3158–3162
30. Sumiyoshi, W., Urashima, T., Nakamura, T., Arai, I., Saito, T., Tsumura, N., Wang, B., Brand-Miller, J., Watanabe, Y., and Kimura, K. (2003) Determination of each neutral oligosaccharide in the milk of Japanese women during the course of lactation. *Br. J. Nutr.* **89**, 61–69
31. Nishimoto, M. and Kitaoka, M. (2007) Identification of *N*-acetylhexosamine 1-kinase in the complete lacto-*N*-biose I/galacto-*N*-biose metabolic pathway in *Bifidobacterium longum*. *Appl. Environ. Microbiol.* **73**, 6444–6449
32. 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
33. Wada, J., Ando, T., Kiyohara, M., Ashida, H., Kitaoka, M., Yamaguchi, M., Kumagai, H., Katayama, T., and Yamamoto, K. (2008) *Bifidobacterium bifidum* lacto-*N*-biosidase, a critical enzyme for the degradation of human milk oligosaccharides with a type 1 structure. *Appl. Environ. Microbiol.* **74**, 3996–4004
34. Suzuki, R., Wada, J., Katayama, T., Fushinobu, S., Wakagi, T., Shoun, H., Sugimoto, H., Tanaka, A., Kumagai, H., Ashida, H., Kitaoka, M., and Yamamoto, K. (2008) Structural and thermodynamic analyses of solute-binding protein from *Bifidobacterium longum* specific for core 1 disaccharide and lacto-*N*-biose I. *J. Biol. Chem.* **283**, 13165–13173
35. 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
36. Farkas, E., Thiem, J., Krzewinski, F., and Bouquelet, S. (2000) Enzymatic synthesis of Galbeta1-3GlcNAc derivatives utilising a phosphorylase from Bifidobacterium bifidum 20082. *Synlett.* **5**, 728–730
37. 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
38. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326
39. Rappleye, J., Innus, M., Weeks, C. M., and Miller, R. (2002) *SnB* version 2.2: an example of crystallographic multiprocessing. *J. Appl. Cryst.* **35**, 374–376
40. Terwilliger, T.C. and Berendzen, J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **55**(Pt 4), 849–861
41. Terwilliger, T.C. (2000) Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.* **56**(Pt 8), 965–972
42. Perrakis, A., Morris, R., and Lamzin, V.S. (1999) Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**, 458–463
43. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**(Pt 12 Pt 1), 2126–2132
44. 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**(Pt 3), 240–255
45. DeLano, W.L. (2002) *The PyMOL Molecular Graphics System.* DeLano Scientific, San Carlos, CA
46. Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., and Olson, A.J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **19**, 1639–1662
47. MacGregor, E.A., Janecek, S., and Svensson, B. (2001) Relationship of sequence and structure to specificity in the alpha-amylase family of enzymes. *Biochim. Biophys. Acta.* **1546**, 1–20
48. Abe, A., Yoshida, H., Tonozuka, 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
49. Vyas, N.K. (1991) Atomic features of protein-carbohydrate interactions. *Curr. Opin. Struct. Biol.* **1**, 732–740
50. Weis, W.I. and Drickamer, K. (1996) Structural basis of lectin-carbohydrate recognition. *Annu. Rev. Biochem.* **65**, 441–473
51. Bansil, R. and Turner, B.S. (2006) Mucin structure, aggregation, physiological functions and biomedical applications. *Curr. Opin. Colloid Interface Sci.* **11**, 164–170
52. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2007) ClustalW and ClustalX version 2. *Bioinformatics.* **23**, 2947–2948
53. Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics.* **15**, 305–308