Crystal Structures of a Glycoside Hydrolase Family 20 Lacto-N-Biosidase from *Bifidobacterium bifidum*

Tasuku Ito¹, Takane Katayama², Mitchell Hattie³, Haruko Sakurama²,⁶, Jun Wada⁴,⁷, Ryuichiro Suzuki¹,⁸, Hisashi Ashida⁵, Takayoshi Wakagi¹, Kenji Yamamoto², Keith A. Stubbs³ and Shinya Fushinobu¹,⁹

¹From Department of Biotechnology, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan
²Research Institute for Bioresources and Biotechnology, Ishikawa Prefectural University, Nonoi, Ishikawa 921-8836, Japan
³School of Chemistry and Biochemistry, University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia
⁴Graduate School of Biostudies, Kyoto University, Kitashirakawa, Sakyoku, Kyoto 606-8502, Japan
⁵Department of Science and Technology on Food Safety, Faculty of Biology-Oriented Science and Technology, Kinki University, Kinokawa, Wakayama 649-6493, Japan

*Running title: Structure of lacto-N-biosidase

Keywords: lacto-N-biosidase, human milk oligosaccharides, *Bifidobacterium bifidum*, glycoside hydrolase family 20, lacto-N-biose I.

**Background:** Infant gut-associated bifidobacteria possess lacto-N-biosidase, which releases lacto-N-biose I (LNB) from human milk oligosaccharides. Human milk oligosaccharides (HMOs) contain a large variety of oligosaccharides, of which lacto-N-biose I (Gal-β1,3-GlcNAc; LNB) predominates as a major core structure. A unique metabolic pathway specific for LNB has recently been identified in the human commensal, bifidobacteria. Several strains of infant gut-associated bifidobacteria possess lacto-N-biosidase, a membrane-anchored extracellular enzyme, that liberates LNB from the non-reducing end of HMOs and plays a key role in the metabolic pathway of these compounds. Lacto-N-biosidase belongs to the glycoside hydrolase (GH) family 20, and its reaction proceeds via a substrate-assisted

**Results:** The crystal structures of lacto-N-biosidase complexed with LNB and LNB-thiazoline were determined.

**Conclusion:** The intermediate analogue complex allows the proposal of a conformational reaction coordinate.

**Significance:** The structures of a key enzyme in the colonization of human commensal bacteria provided its structural basis and insight into the development of inhibitors.
catalytic mechanism. Several crystal structures of GH20 β-N-acetylhexosaminidases, which release monosaccharide GlcNAc from its substrate, have been determined but to date a structure of a lacto-N-biosidase is unknown. Here we have determined the first three-dimensional structures of lacto-N-biosidase from Bifidobacterium bifidum JCM1254 in complex with LNB and LNB-thiazoline (Gal-β1,3-GlcNAc-thiazoline) at 1.8 Å resolution. Lacto-N-biosidase consists of three domains, and the C-terminal domain has a unique β-trefoil-like fold. Compared with other β-N-acetylhexosaminidases, lacto-N-biosidase has a wide substrate-binding pocket with a –2 subsite specific for β-1,3-linked Gal, and the residues responsible for Gal recognition were identified. The bound ligands are recognized by extensive hydrogen bonds at all of their hydroxyls consistent with the enzyme’s strict substrate specificity for the LNB moiety. The GlcNAc sugar ring of LNB is in a distorted conformation near 4E, whereas that of LNB-thiazoline is in a 4C1 conformation. A possible conformational pathway for the lacto-N-biosidase reaction was discussed.

Human milk contains various oligosaccharides collectively termed human milk oligosaccharides (HMOs) (1,2). HMOs function as prebiotics, promoting the growth of bifidobacteria in the gastrointestinal tracts of breastfed infants, which in turn promotes optimal health (3,4). Most HMOs contain a lactose moiety (Gal-β1,4-Glc) at their reducing end, which is elongated by β1,3-linked lacto-N-biose I (Gal-β1,3-GlcNAc; LNB, to give type I HMOs) and/or β1,3/6-linked N-acetyllactosamine (Gal-β1,4-GlcNAc; LacNAc, to give type II HMOs). A unique feature of the composition of HMOs is the predominance of type I over type II oligosaccharides. Indeed such a composition has not been observed in milk oligosaccharides from other mammals, including anthropoids (5). Further elongation of these core structures is made by the addition of fucose and sialic acid residues via α1,2/3/4- and α2,3/6- linkages, respectively. HMOs are composed of more than 130 different oligosaccharide structures that account for 2% of the solid components of dried human milk. Of note is that four oligosaccharides constitute 25-33% of total HMOs (6): 2'-fucosyllactose (Fuc-α1,2-Gal-β1,4-Glc), lacto-N-fucopentaose I (Fuc-α1,2-Gal-β1,3-GlcNAc-β1,3-Gal-β1,4-Glc; LNFP I), lacto-N-difucohexaoae I (Fuc-α1,2-Gal-β1,3-(Fuc-α1,4-)GlcNAc-β1,3-Gal-β1,4-Glc; LNDFH I), and lacto-N-tetraose (Gal-β1,3-GlcNAc-β1,3-Gal-β1,4-Glc; LNT). Three of these oligosaccharides (LNFP I, LNDFH I, and LNT) contain an LNB unit, highlighting the importance of this component in these systems.

In 2005, a novel metabolic pathway specific to LNB and galacto-N-biose (Gal-β1,3-GalNAc; GNB) was uncovered in bifidobacteria (7) and was termed as the GNB/LNB pathway (8). Considering the living environment of bifidobacteria (the gastrointestinal tract of infants), LNB and GNB are hypothesized to originate from HMOs and intestinal mucin glycoproteins, respectively. Proteins related to this pathway have been actively investigated, and crystallographic analyses (9) has uncovered the structures of an extracellular endo-α-N-acetylgalactosaminidase that releases GNB from mucin-type O-glycans (10), a solute-binding protein of an ABC transporter specific to GNB and LNB (11), and an intracellular phosphorylase that cleaves both GNB and LNB (12). In addition, it has been demonstrated that LNB is the “bifidus factor” that selectively promotes growth of infant gut-associated bifidobacteria (13,14) but interestingly a free form of LNB has not been found in HMOs (1).

Lacto-N-biosidase (LNBase, EC 3.2.1.140), which liberates LNB from the non-reducing end of oligosaccharides, was first found in the soil actinomycete Streptomyces sp. 142 (15,16). Since then LNBase activity has also been found
in several strains of *Bifidobacterium bifidum* and *Bifidobacterium longum* (17), which are particularly predominant in the intestines of infants (18). Bifidobacterial LNBase is a membrane-anchored extracellular enzyme, which suggests that it may play a key role in the excision of LNB from HMOs to supply LNB to the GNB/LNB pathway. Recently, the gene for LNBase from *B. bifidum* JCM1254 (*Bb*LNBase) was cloned, and its recombinant protein was characterized in detail. *Bb*LNBase consists of 1,112 amino acids and contains a signal sequence, a glycoside hydrolase (GH) family 20 domain, a carbohydrate-binding module (CBM) family 32 domain, a bacterial Ig-like domain, and a transmembrane region (Fig. 1A). The enzyme’s activity favored LNT as a substrate to produce LNB and lactose, and it was found that it could not act on oligosaccharides where the LNB unit is modified with fucose. In addition, it was found that *Bb*LNBase specifically releases LNB at the non-reducing end of type I oligosaccharides but does not hydrolyze type II oligosaccharides. Therefore, a potential use for this enzyme is in identifying type I structures in glycoconjugates. Intriguingly, most of the cancer-associated oligosaccharide antigens (including sialyl Leα, sialyl Leβ, and their derivatives) have a core structure containing type I or type II chains (19).

LNBase are members of the GH20 family of glycoside hydrolases (20) along with β-N-acetylatedhexosaminidases (β-HexNAcases); however, they exhibit very low sequence homology. For example, *Bb*LNBase exhibits less than 24% amino sequence identity to all the characterized β-HexNAcases. GH20 enzymes cleave the glycosidic linkage at the reducing end of GlcNAc via a retaining substrate-assisted catalytic mechanism in which the 2-acetamido group of the substrate acts as the catalytic nucleophile (21-23). Whereas β-HexNAcases release a monosaccharide (GlcNAc), LNBase releases a disaccharide (LNB), implying that the latter has an extended −2 subsite. The crystal structures of multiple GH20 β-HexNAcases have been reported to date (21,24-32); however, the three-dimensional structure of LNBase is not yet available.

Here we report two crystal structures of *Bb*LNBase complexed with LNB and LNB-thiazoline, a potent inhibitor of *Bb*LNBase (Ki = 125 ± 8 nM at pH 4.5) (33). Using the sugar ring conformations of LNB and LNB-thiazoline molecules found in the active site, the reaction mechanism and possible conformational changes of the substrate are discussed.

**EXPERIMENTAL PROCEDURES**

**Protein production and purification**—The overexpression vector for N-terminally (His)6-tagged *Bb*LNBase (residues 41 to 663) was constructed by inserting the PCR-amplified fragment of the *lnbB* gene (17) into the NdeI and EcoRI sites of the pET28b plasmid (kanr; Novagen, Madison, WI). The primers used were 5'-gggaattcataattggtacagtgccacggcctcc-3’ and 5’-ccggaattctcagtcgctgaccaggtcag -5’ (restriction sites are underlined). The plasmid was introduced into *Escherichia coli* BL21 CodonPlus (DE3)–RIL (Stratagene, La Jolla, CA) for native protein expression. For selenomethionine-labeled protein expression, a NcoI–XhoI fragment of the pET28b-based expression plasmid was inserted into the pET19b plasmid (*amp*), which was subsequently introduced into *E. coli* BL21 CodonPlus (DE3)-RIL-X (*kan*; Stratagene). The transformants were cultured in Luria-Bertani medium (native protein) or LeMaster medium (selenomethionine-labeled protein) containing 100 mg/L kanamycin or ampicillin and 20 mg/L chloramphenicol at 25°C for 20 h. Isopropyl 1-thio-β-D-galactopyranoside was added to a final concentration of 1.0 mM to induce protein expression. Following an additional incubation at 25°C for 20 h, the cells were harvested by centrifugation and suspended in 50 mM HEPES-NaOH (pH 7.5). Cell extracts were obtained by sonication followed by centrifugation to remove cell debris. The protein was purified to homogeneity by sequential column chromatography involving Ni-NTA
Structural determination of BbLNBase—We first constructed various deletion mutants to determine a minimal region that retained activity towards the substrate pNP-LNB (Supplementary Table S1 and Fig. S1). Constructs 37–663, 41–663, and 46–663 (numbers indicate the residues) exhibited virtually the same activity as the full-length construct (31–1064, which has
only the signal and membrane anchor regions deleted). Thus, these three constructs were selected for crystallization screening but only diffraction-quality crystals were obtained with the 41–663 construct in the presence of LNB. The crystal structure of BbLNBase was determined by the single-wavelength anomalous dispersion method using a selenomethionine derivative. Subsequently, we determined the crystal structure of native (non-labeled) protein crystals (complexed with LNB and LNB-thiazoline) both at 1.8 Å resolution (Table 1 and Fig. 1B). The crystal contains two molecules in the asymmetric unit, and the final model contains residues from Ser-30 to Ser-662 of the A chain and from Ser-30 to Val-661 of the B chain. The protein has an N-terminal His₆-tag containing 21 amino acid residues derived from the pET-28b vector (MGSSHHHHHHSSGLVPRGSHM), in which 11 amino acids (SSGLVPRGSHM, residues 30–40) were visible in the electron density map. A region of the artificial tag sequence (Ser-30 to Arg-36) contributes to the packing of the dimer in the asymmetric unit (Supplementary Fig. S2).

The molecular weights of the 41–663 construct of BbLNBase as deduced from the amino acid sequence, estimated by SDS-PAGE and calibrated gel filtration chromatography were 71.8, 72.4, and 64.9 kDa, respectively (data not shown), suggesting that it is monomeric in solution. The root mean square deviations (RMSD) for the Cα atoms of all pairs of the four molecules (two chains in the two crystal structures) are less than 0.5 Å. The two ligand molecules bound to chains A and B are also virtually identical (RMSD = 0.047 and 0.069 Å for LNB and LNB-thiazoline, respectively). Hence, our subsequent descriptions will refer to chain A, unless otherwise noted.

Overall structure—The BbLNBase monomer consists of three domains; an N-terminal domain (N-domain, residues 41-178), a catalytic (β/α)₈ barrel domain (barrel domain, residues 179-496) and a C-terminal domain (C-domain, residues 497-662) (Fig. 1B). The N-domain has an α/β topology with a seven-stranded β-sheet exposed to the surface, and two α helices buried in the interface with the subsequent barrel domain. The N-domain and the barrel domain correspond to the two conserved domains of typical GH20 β-HexNAcases (21,24,25,27-32). The N-domain and the barrel domain of BbLNBase are structurally similar to the domains I and II, respectively, of β-HexNAcase from Streptomyces plicatus (SpHex) (Fig. 1C). The N-domain is conserved in most GH20 enzymes, although its function remains unknown.

The C-domain on the other hand is not common to GH20 enzymes. Two examples of GH20 β-HexNAcases that have a distinct domain in the C-terminal region of the catalytic barrel domain are the β-HexNAcases from Serratia marcescens (chitobiase) and Streptococcus gordonii (GcnA) which have a small (67 amino acids) immunoglobulin-like β-sandwich domain (domain IV) and a large (227 amino acids) α-helical domain (domain III), respectively (24,28). In the case of the C-domain of BbLNBase, there is no resemblance to these GH20 C-domain structures. The C-domain of BbLNBase is located on the side of the barrel domain that is farthest from the active site. A deletion of only 5 residues at the C-terminus of this domain (construct 37–658) significantly reduced the catalytic activity, and further deletions caused complete inactivation (Supplementary Table S1). These results indicate that the C-domain of BbLNBase is essential for protein stability and catalytic activity. Of interest is that the C-domain possesses a broken β-trefoil fold. According to a structural similarity search using the DALI server (40), the C-domain is slightly similar to an R-type lectin (MOA) from Marasmius oreades (CBM13) (Z-score = 11.0, RMSD = 2.0 Å for 102 Cα atoms), which also has a typical β-trefoil fold (41). The β-trefoil fold of MOA is composed of three subdomains that are similar in structure and are assembled at a 3-fold axis (Supplementary Fig. S3C) and each subdomain contains four-stranded β-hairpin turns (β1–β4) (Supplementary Fig. S3E). The
Structure of lacto-N-biosidase

C-domain of BbLNBase can be divided into an α-subdomain (residues 522-550), β-subdomain (residues 551-615) and γ-subdomain (residues 616-662) (Supplementary Fig. S3A and S3B), with each subdomain having at least one additional α-helix. The β-subdomain retains the typical topology of the β-trefoil fold with a complete set of four β-strands (Supplementary Fig. S3D). However, the α- and γ-subdomains lack one or two β-strand(s), and the C-domain is largely broken at the interface between these subdomains. In the β-subdomain, a disulfide bond is formed between Cys-564 (within β2) and Cys-589 (loop after β3). Interestingly, these structural features (a disulfide bond in the β-subdomain and an additional α-helix after β3) are also present in a CBM13-related arabinose-binding domain (CBM42) of the fungal GH54 α-L-arabinofuranosidase (42). The CBM42 domain also partially lacks the three-fold symmetry, and its α-subdomain lacks the arabinose binding site. However, the C-domain of BbLNBase appears to be more broken than CBM42.

Active site of BbLNBase—We observed a clear electron density for ligands at the center of the barrel domain (Fig. 2A and 2B). The pyranose ring of the GlcNAc residue, in LNB, is in the 4E conformation and that of GlcNAc-thiazoline, in LNB-thiazoline, is in the 4C1 conformation (Fig. 2C and 2D, discussed below). For LNB, GlcNAc and Gal are bound to the –1 and –2 subsites, respectively with all the hydroxyl groups of LNB forming hydrogen bonds with surrounding amino acids. Such extensive recognitions give evidence for the strict substrate specificity observed for this enzyme (17) as demonstrated for pNP-β-GNB, the 4-epimer of pNP-β-LNB which shows highly reduced activity. In the BbLNBase structure, the C4-hydroxyl group of GlcNAc of LNB forms a hydrogen bond with the side chain of Asp-467 (Fig. 2) and this amino acid would block GalNAc-containing substrates from binding such as GNB. In accordance with other biochemical studies, BbLNBase does not hydrolyze fucosylated substrates such as pyridylamino (PA)-LNFP I and PA-LNFP II (17) which can be rationalised as the active site of BbLNBase, has no space for the fucosyl group at the C2-hydroxyl of Gal (blocked by Asn-259 and Asp-320).

The substrate binding site of BbLNBase was compared with that of SpHex, which is one of the most well-studied GH20 β-HexNAcases (Fig. 3) (21-23). Amino acids surrounding the GlcNAc at –1 subsite (Asp-320, Glu-321, Tyr-419, and Asp-467 in BbLNBase) are highly conserved with SpHex and among GH20 enzymes (Supplementary Fig. S4). The two catalytic residues of GH20, Asp-320 (polarizing residue) and Glu-321 (acid/base catalytic residue), form hydrogen bonds with the amide nitrogen of the 2-acetamido group and the O1-hydroxyl, respectively. Tyr-419 is a highly conserved residue in GH20 enzymes, and its side-chain hydroxyl group forms a hydrogen bond with the carbonyl oxygen atom of the 2-acetamido group. Asp-467 forms bifurcated hydrogen bonds with the O4- and O6-hydroxyl groups of the GlcNAc residue. The corresponding residue is often substituted with Glu in β-HexNAcases (Glu-444 in SpHex), which forms a hydrogen bond with the O4-hydroxyl group alone. The hydrogen bond with Asp-467 fixes the O6-hydroxyl group in a gauche-gauche orientation in BbLNBase (Fig. 3A), whereas the O6-hydroxyl group of GlcNAc in SpHex is in the gauche-trans orientation due to hydrogen bonds with Asp-395 and Trp-408 (Fig. 3B). In addition, His-263, Trp-373, Trp-394 and Trp-465 form hydrophobic interactions with the substrate in the active site (Fig. 2) and are highly conserved in GH20 (Supplementary Fig. S4).

To investigate the importance of these residues in the –1 subsite, we constructed site-directed mutants D320A, D320N, Y419F, and H263F (Table 2). Mutations at the catalytically important Asp-320 residue (D320A and D320N) did not affect the $K_m$ values, but they exhibited significantly reduced $k_{cat}$ values, respectively which is consistent with those of the corresponding mutants of SpHex (D313A and
D313N) (22), human HexB (D354N) (43), and Paenibacillus sp. TS12 Hex1 (D321N) (29). In the case of Y419F, the \( K_m \) and \( k_{cat} \) were both significantly reduced compared to the wild-type enzyme. The mutation at His-263 also showed a reduced \( k_{cat} \), which is consistent with this residue being critical in the hydrophobic interaction with GlcNAc.

As expected from the disaccharide-releasing characteristics of \( BbLNBase \), a –2 subsite specific for \( \beta \)-1,3-linked Gal is clearly defined in the crystal structure (Fig. 3A). The side chains of Gln-190 and Glu-216 form hydrogen bonds with the O4- and O3-hydroxyl groups of Gal, respectively. Gln-190 and Glu-216 in \( BbLNBase \) are replaced by Arg-162 and His-188, respectively, in \( SpHex \) (Fig. 3C) and jointly block the space for a potential –2 subsite in this enzyme (Fig. 3B). Furthermore, in most \( \beta \)-HexNAcases, a 6-amino acid loop insertion is present just after the His-188 residue (Fig. 3C), and a relatively conserved Asp residue (Asp-191 in \( SpHex \)) located in the loop occupies a position corresponding to the −2 subsite (Gal binding site) of \( BbLNBase \) (Fig. 3B). Comparison of the molecular surfaces of \( BbLNBase \) and \( SpHex \) (Fig. 4) illustrates pockets that are suitable in size for binding disaccharide and monosaccharide units, respectively. However, we were unable to determine the positive subsites in the \( BbLNBase \) structure. Attempts at co-crystallization and soaking experiments with lactose, Gal, or Glc were undertaken but no electron density for these sugars were found (data not shown). \( BbLNBase \) has been previously shown to efficiently hydrolyze pyridylamino-LNT as well as \( pNP-\beta-LNB \) (17). However, the enzymatic activity against LNT, the natural form of a major HMO component, has not been fully studied. Thus, we measured the kinetic parameters for hydrolysis of LNT by \( BbLNBase \). The \( K_m \), \( k_{cat} \), and \( k_{cat}/K_m \) values were determined to be 626 ± 23 \( \mu \)M, 42.1 ± 0.8 s \(^{-1} \), and 67.2 ± 1.9 s \(^{-1} \)mM \(^{-1} \), respectively. The \( K_m \) and \( k_{cat} \) values were higher compared with those for \( pNP-LNB \) (Table 2), demonstrating that this enzyme is an exo-acting enzyme and does not have positive subsites specific for sugar moieties. In addition, the entrance of the substrate-binding pocket of \( BbLNBase \) appears to be wider than that of \( SpHex \) (Fig. 4). This is in agreement with the finding that LNBase from \( Streptomyces \) sp. 142 releases LNB from the non-reducing end of various oligosaccharides, including a large triantennary sugar chain (15,16).

**Reaction mechanism and conformational changes of the substrate** – In the widely accepted mechanism of GH20 (Fig. 5), which is one of substrate-assisted catalysis, the 2-acetamido group of a substrate is polarized and oriented by a deprotonated Asp residue that is located at the neighboring position of the acid/base Glu residue (44). The oxygen atom of the carbonyl group acts as the nucleophile, which attacks the anomic carbon, resulting in the formation of an oxazoline or oxazolinium ion intermediate. The protonated acid/base Glu residue facilitates bond cleavage by providing a proton to the glycosidic oxygen. In the second step of the reaction, the acid/base Glu activates a water molecule, thereby facilitating its nucleophilic attack at the anomic carbon. During the course of this reaction, two oxocarbenium ion-like transition states are thought to exist on each side of the oxazoline or oxazolinium ion intermediate.

When investigating the catalytic mechanism of carbohydrate-active enzymes, the dynamic behavior of a substrate sugar molecule during the catalytic route in the active site is of particular interest. In this work, we observed distorted GlcNAc sugars of LNB in the −1 subsite of both chains (A and B) of the \( BbLNBase \) structure (Fig. 2C), while the GlcNAc-thiazoline group of the LNB-thiazoline complex is in a \( ^4C_1 \) conformation (Fig. 2D). Therefore to gain insight into the conformations of the sugar at binding, we analyzed the sugar ring conformations in detail using the Cremer-Pople system, which is widely used to define the puckering conformations of a pyranose (45). The Cremer-Pople parameters of GlcNAc in chains A and B were $\phi = 245.9^\circ$, $\theta = 60.1^\circ$, and $Q = 0.595$, and $\phi = 252.6^\circ$, $\theta = 56.7^\circ$, and $Q = 0.616$,
respectively, indicating that they are in a conformation close to \( ^4E \) (ideally, \( \phi = 240^\circ \) and \( \theta = 54.7^\circ \)). On the other hand, the parameters of GlcNAc-thiazoline of LNB-thiazoline in chains A and B were \( \phi = 237.6^\circ, \theta = 13.2^\circ, \) and \( Q = 0.547, \) and \( \phi = 265.4^\circ, \theta = 17.9^\circ, \) and \( Q = 0.597, \) respectively, indicating that they are in a conformation close to \( ^4C_1. \)

The 2-acetamido group of GlcNAc in LNB is nestled in amongst three aromatic residues (Trp-373, Trp-394, and Trp-465) and fixed by two hydrogen bonds with Asp-320 and Tyr-419 (Fig. 2C). The 2-acetamido group at its fixed position elevates the C1 atom of GlcNAc, and the hydrogen bond between Asp-467 and O4-hydroxyl also fixes the C4 atom of GlcNAc in an elevated position. These interactions potentially distort the conformation of GlcNAc to \( ^4E, \) which is not intrinsically stable as evidenced by crystallographic analysis and molecular simulation (46,47).

Currently, various substrates (e.g., chitobiose), inhibitors (e.g., GlcNAc-thiazoline), and reaction products (e.g., GlcNAc) of \( \beta \)-HexNAcases have been observed in the crystal structures of other GH20 enzymes (21,22,24,25,27-31,48-52) (Supplementary Table S2). In an effort to gain insight into the global binding of these ligands against GH20 enzymes, their Cremer-Pople parameters (\( \phi \) and \( \theta \)) were evaluated and plotted (Fig. 6) together with those of the GlcNAc sugars of LNB and LNB-thiazoline in \( BbLNBase. \)

GlcNAc-thiazoline and GalNAc-thiazoline are specific inhibitors of GHs that utilize the substrate-assisted mechanism (53,54) because they structurally resemble the oxazoline or oxazolinium ion intermediate. A recent study using \textit{ab initio} molecular dynamic simulations predicts that the oxazoline or oxazolinium ion intermediate is distorted to a conformational region of \( ^4H_3E^3H_5 \) (44). However, in GH20 \( \beta \)-HexNAcases, the thiazoline inhibitors always adopt conformations close to \( ^4C_1 (\theta < 23^\circ) \) (21,25,27,52). Our crystallographic result also indicates that the LNB-thiazoline in LNBase also adopt a \( ^4C_1 \) conformation (Fig. 6). These ample crystallographic data indicate that the oxazolinium ion intermediate adopts the \( ^4C_1 \) conformation (21).

Various substrates that carry sugar moieties at positive subsites have also been observed in crystal structures of GH20 \( \beta \)-HexNAcases. Complexes of a chitobiase (\( \beta \)-HexNAcase) from \textit{S. marcescens} with chitobiose (di-N-acetyl-D-glucosamine) were trapped using the mutants D537A and E540A (catalytic residues) (48) or by a simple soaking method (24). The complex structures of the N-terminal and C-terminal modules (GH20A and GH20B) of the large multimodular \( \beta \)-HexNAcase StrH from \textit{Streptococcus pneumoniae} with disaccharide (in GH20A), tetrasaccharide (in GH20B), and bisected glycan heptasaccharide (in GH20B) substrates were reported using their acid/base residue mutants (E361Q of GH20A and E805Q of GH20B) (31). The GlcNAc moieties of these substrates bound to the \(-1\) subsite were all in conformations close to a \( ^{1}B \) conformation (229° < \( \phi < 254^\circ \) and \( \theta > 66^\circ \); ideally \( \phi = 240^\circ \) and \( \theta = 90^\circ \)). Moreover, the complex structure of TMG-chitotriomycin, a linear tetrasaccharide inhibitor (in GH20B) substrates were reported using their acid/base residue mutants (E361Q of GH20A and E805Q of GH20B) (30). The GlcNAc moiety of these substrates bound to the \(-1\) subsite were all in conformations close to a \( ^{1}B \) conformation (229° < \( \phi < 254^\circ \) and \( \theta > 66^\circ \); ideally \( \phi = 240^\circ \) and \( \theta = 90^\circ \)). Therefore, the substrate of the Michaelis complex for GH20 enzymes is thought to adopt a \( ^{1}B \) conformation.

Transition state analogs that have an \( sp^2 \)-hybridized planar anomeric carbon were also used to study \( \beta \)-HexNAcases and related enzymes as these compounds are found to be potent inhibitors of these enzymes. In the complex structure of human lysosomal \( \beta \)-HexNAcase B, 2-acetamido-2-deoxy-D-glucono-1,5-lactone (\( \delta \)-lactone) was bound in conformations close to \( ^{4}E (236^\circ < \phi < 253^\circ \) and \( 48^\circ < \theta < 64^\circ ) \) (50). \( O-(2\text{-Acetamido}-2\text{-deoxy}-D\text{-glucopyranosylidene})\text{-amino}-N\text{-phenylcarbamate (PUGNAc)} \) is a potent inhibitor of a wide range
of β-HexNAcases that utilize the classical double displacement mechanism (e.g., GH3) as well as the substrate-assisted mechanism used by GH20 enzymes. In β-HexNAcases from *O. furnacalis* and *Paenibacillus* sp. TS12 (Hex1), PUGNAc adopts conformations ranging from $4^E$ to $1^B/4^E$ ($240° < \phi < 245°$ and $60° < \theta < 78°$) (29,51). Our results strongly support the hypothesis that the conformation at the transition states of GH20 enzymes is nearly $4^E$, which is in compliance with the requirement of coplanarity at the C2, C1, O5, and C5 pyranose ring atoms to attain the oxocarbenium ion-like transition state. $4^E$-like transition states are also predicted for other GH enzymes that utilize the substrate-assisted mechanism, namely a GH18 chitinase (by a QM/MM modeling study) (56), and a GH84 O-GlcNAcase (57).

According to previous knowledge regarding GH20 β-HexNAcases and the present crystal structures, we propose a conformational itinerary pathway for the *BbLNBase*-catalyzed reaction that obeys the “principle of least nuclear motion” (21,58): $1^B$ (Michaelis complex) – $4^E$ (transition state 1) – $4^C_1$ (oxazolinium ion intermediate) – $4^E$ (transition state 2) – $4^E$ (product complex) (Fig. 5). Of interest is the $4^E$ product complex observed in this study. In GH20 β-HexNAcases, several complexes observed with reaction products (GlcNAc or GalNAc) have been reported. However, the ring conformations and binding modes of these molecules differ depending on the conditions of complex formation. In the structures of β-HexNAcase (Hex1) from *Paenibacillus* sp. TS12, GlcNAc is in a $4^C_1$ conformation, whereas GalNAc is in a $1^B/4^C_1$ conformation (29). In *SpHex*, GlcNAc is bound in a $4^C_1$ conformation with the wild-type enzyme, but it adopts the alternative conformations of $1^B$ and $4^C_1$ in the D313A mutant (22). In the D313N mutant, the bound GlcNAc is in a $4^C_1$ conformation but is tilted and dropped from its normal position in the wild-type complex. In *BbLNBase*, we speculate that there are two major factors that fix LNB with its GlcNAc moiety in a $4^E$ conformation. Firstly, the Gal moiety bound to −2 subsite acts as an anchor to lock in place the LNB disaccharide in the active site. This anchor is lacking in other GH20 β-HexNAcases as they only have confined space in the active site to accommodate a GlcNAc/GalNAc residue. Secondly, the bifurcated hydrogen bond between Asp-467 and the O4/O6-hydroxyls of the GlcNAc moiety strongly fix it in the $4^E$ and gauche-gauche conformations. These factors are unique to LNBases and have to date not been shown in β-HexNAcases.

**C-terminal domains**—In the crystal structure of *BbLNBase*, we observed a novel β-trefoil-like domain (C-domain) that is required for protein stability. One of our most striking findings was that the C-domain has several features similar to those of the arabinose-binding domain CBM42. However, there is currently no evidence indicating that the C-domain in *BbLNBase*, has carbohydrate-binding ability. We attempted co-crystallization and soaking experiments with various sugars such as lactose, Gal, or Glc at high concentrations (up to 1 M), but we could not observe any extra electron density in this domain (data not shown). In a study examining β-HexNAcases from *Paenibacillus* sp. TS12, the C-terminal region of one of the two β-HexNAcases (residues 503–978 of Hex1) was suggested to help the enzyme in its interaction with glycosphingolipid substrates in the absence of detergent (29). However, the three-dimensional structure of this domain remains unknown because only a truncated structure of Hex1 is available (deletion mutant 1–502), and there is no amino acid sequence homology between the C-terminal region of Hex1 and the C-domain of *BbLNBase*. The full length *BbLNBase* also has a CBM32 domain (residues 784–932) in a region adjacent to the C-domain (Fig. 1A). CBM32 domains of GHs from enteric bacteria have been shown to recognize a terminal Gal or GalNAc, or disaccharide motifs such as LacNAc and GlcNAc-$\alpha$1,4-Gal (59). Therefore, the CBM32 domain of *BbLNBase* may also function in binding β-1,3-linked Gal, LNB or other moieties.
of HMOs.

**Biological implications**—In the type I HMO-degradation system of *B. bifidum* JCM1254, *Bb*LNBase plays a central role in releasing LNB from HMOs (9,60). However, the assistance of other enzymes that remove modifying sugars such as fucose or sialic acid is required. Many extracellular glycosidases involved in HMO degradation have been found in *B. bifidum* JCM1254. GH95 1,2-α-L-fucosidase (AfεA) (61,62), GH29 1,3-1,4-α-L-fucosidase (AfεB) (63,64), and GH33 exo-α-sialidases (SiaBb1 and SiaBb2) (65) all share roles in the complete degradation of α1,2- and α1,3/4-fucosylated and sialylated HMOs. In this study, we provide a structural basis for the substrate specificity of *Bb*LNBase. This enzyme cannot accommodate any modified LNB moieties because of a substrate binding pocket that is constrained to allow only LNB binding. This feature is suitable for the specificity of the solute-binding protein of the GNB/LNB transporter (GL-BP) in the GNB/LNB pathway, which specifically binds unmodified LNB disaccharide (11). GL-BP binds LNB and GNB with low *K*ₐ values (< 0.09 µM), whereas LNT, a major HMO tetrasaccharide containing LNB, exhibits a significantly higher *K*ₐ value (11 µM). Following uptake, LNB is subsequently metabolized by GNB/LNB phosphorylase and other enzymes including *N*-acetylhexosamine kinase, UDP-glucose hexose 1-phosphate uridylyltransferase, and UDP-glucose 4-epimerase (8). The positive subsites of *Bb*LNBase are wider and appears to be capable of accommodating various groups, suggesting that this enzyme acts on various type I HMOs after the action of fucosidases and sialidases. Moreover, *B. bifidum* JCM1254 has three extracellular enzymes including one GH2 β-galactosidase (BbgIII) and two GH20 β-HexNAcases (BbhI and BbhII) (66). Two of these enzymes (BbgIII and BbhI) are suggested to play essential roles in degrading type II HMOs because they catalyze the complete hydrolysis of LnNT to monosaccharides. BbgIII hydrolyzes β-1,4- and β-1,6-linked Gal but not β-1,3-linked Gal in LNB and LNT, suggesting that it is also involved in the complete degradation of type I HMOs after the action of LNBase (e.g. cleavage of lactose released from LNT).

In contrast to *B. bifidum* JCM1254, *B. longum* subsp. *infantis* ATCC15697 lacks the LNBase gene. A genomic analysis of *B. longum* subsp. *infantis* identified a unique gene cluster, the HMO cluster, containing various intracellular GHs that lack signal sequences (67,68). A GH29 fucosidase, a GH95 fucosidase, a GH33 sialidase, a GH20 β-HexNAcase, a GH2 β-galactosidase, and at least 4 putative sugar transporters were found in the HMO cluster. The GH2 β-galactosidase is specific for lactose and type II HMOs (69). Furthermore, the intracellular GH42 β-galactosidase Bga42A, which is distant from the HMO cluster, is highly specific for LNT and functions as the sole β-galactosidase acting on type I HMOs. In the HMO degradation system of this strain, it was suggested that low-molecular-weight HMOs (degree of polymerization ranging from 3 to 8) are directly imported into the cells by transporters and subsequently cleaved by intracellular GHs (70). This analysis as well as recent research on commensal bacteria, is unveiling the different HMO consumption modes of bifidobacterial strains. Extracellular bifidobacterial LNBase may also aid in the HMO consumption of other bifidobacterial strains that lack endogenous LNBase. Importantly, in this study, we identified residues that are important for β-1,3-linked Gal recognition at -2 subsite (viz., Gln-190 and Glu-216). Conservation of these residues is a critical marker that helps to predict putative LNBase genes. A BLAST search against protein databases indicates that putative LNBase genes are present in many bacteria, mainly in the Actinobacteridae subclass. The presence (or absence) of LNBase in the genome of microbes in an infant’s gastrointestinal tract will provide important information that will advance our understanding of how these
microbes metabolize HMOs.
REFERENCES


47. Sattelle, B. M., and Almond, A. (2011) Is N-acetyl-D-glucosamine a rigid 4C1 chair?
Glycobiology 21, 1651-1662
catalysis by glycosidases *Biochem. J.* **224**, 817-821


**Acknowledgments**—We thank Drs M. Nishimoto and M. Kitaoka (NFRI-NARO, Japan) for providing LNB and for helpful discussion. We thank the staff of the Photon Factory and SPring-8 for the X-ray data collection. A part of this work was performed under the Priority Program for Disaster-Affected Quantum Beam Facilities (proposal numbers PF 2011G080 and SPring-8 2011A1908).

**FOOTNOTES**

*This work was supported by the Program for the Promotion of Basic Research Activities for Innovative Bioscience (PROBRAIN) in Japan (to SF, KY, and TK), by JSPS KAKENHI Grant Number 24380053 (to SF and HA), and by the Australian Research Council.

The atomic coordinates and structure factors (codes 4H04 and 4JAW) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

†The abbreviations used are: HMOs, human milk oligosaccharides; LNB, lacto-N-biose I (Gal-β1,3-GlcNAc); LacNAc, N-acetyllactosamine (Gal-β1,4-GlcNAc); LNFP I, lacto-N-fucopentaose I (Fuc-α1,2-Gal-β1,3-GlcNAc-β1,3-Gal-β1,4-Glc); LNDFH I, lacto-N-difucohexaose I (Fuc-α1,2-Gal-β1,3-(Fuc-α1,4-)GlcNAc-β1,3-Gal-β1,4-Glc); LNT, lacto-N-tetraose (Gal-β1,3-GlcNAc-β1,3-Gal-β1,4-Glc); GNB, galacto-N-biose (Gal-β1,3-GalNAc); LNBase, lacto-N-biosidase; *Bb* LNBase, LNBase from *B. bifidum* JCM1254; GH, glycoside hydrolase; CBM, carbohydrate-binding module; LNFP II, lacto-N-fucopentaose II (Gal-β1,3-(Fuc-α1,4-)GlcNAc-β1,3-Gal-β1,4-Glc); LNNT, lacto-N-neotetraose (Gal-β1,4-GlcNAc-β1,3-Gal-β1,4-Glc); β-HexNAcase, β-N-acetylhexosaminidase; pNP, p-nitrophenyl; PA, pyridylamino; RMSD, root mean square deviation; *Sp*Hex, β-HexNAcase from *Streptomyces plicatus*.

**FIGURE LEGENDS**
**FIGURE 1.** *A*, Domain structure of *Bb*LNBase. *B*, Overall structure of *Bb*LNBase (residues 41-663) complexed with GlcNAc (PDB ID 1M01). In *B* and *C*, α-helices and β-strands in the catalytic (α/β)_8 barrel domain are shown in red and yellow, respectively, and α-helices and β-strands in the N-terminal domain are shown in cyan and magenta, respectively. Bound ligands are shown as blue sticks.

**FIGURE 2.** Stereoviews of LNB (*A* and *C*, yellow) and LNB-thiazoline (*B* and *D*, cyan) bound to the active site of *Bb*LNBase. Catalytic residues, residues forming hydrogen bonds, and residues forming hydrophobic interactions are shown in magenta, green, and white, respectively. Hydrogen bonds are shown as red dashed lines. *A* and *B*, |F_o|−|F_c| omit electron density map (contoured at 4σ) and interactions with protein atoms. *C* and *D*, GlcNAc and GlcNAc-thiazoline sugar ring and surrounding residues.

**FIGURE 3.** Comparison of the active sites of GH20 enzymes. *A* and *B*, Active sites of *Bb*LNBase (*A*) and *Sp*Hex (*B*), respectively. Inset, loop regions following β1 and β2 that determine the presence or absence of −2 subsite. *C*, Amino acid sequence alignment of the loop regions of GH20 enzymes (2 LNBases and 8 β-HexNAcases). Residues important for the presence or absence of −2 subsite are labeled with red characters. Six-amino acid loop insertion of β-HexNAcases is indicated by a magenta bar. LNBase142, *Streptomyces* sp. 142 LNBase; DspB, *Actinobacillus actinomycetemcomitans* dispersin B; GenA, *Streptococcus gordonii* GenA; HexA, human lysosomal β-HexNAcase A; HexB, human lysosomal β-HexNAcase B; SmChb, *Serratia marcescens* chitobiase; PsHex1, *Paenibacillus* sp. TS12 Hex1; OfHex1, *Ostrinia furnacalis* β-HexNAcase.

**FIGURE 4.** The molecular surfaces of *Bb*LNBase (*A*) and *Sp*Hex (*B*) showing the substrate binding pockets. According to the electrostatic potential, the surface is colored blue for positive and red for negative charges. LNB and GlcNAc bound to the pocket are shown as green sticks.

**FIGURE 5.** Proposed reaction pathway of *Bb*LNBase and possible conformations of the GlcNAc sugar ring. R’ = β1,3-linked Gal.

**FIGURE 6.** Distribution map of sugar ring conformations observed in the −1 subsite of GH20 enzymes as analyzed by Cremer-Pople angle parameters. Cross symbols indicate ideal positions of 1^E (φ = 240°, θ = 90°), 4^E (φ = 240°, θ = 54.7°), and 1^S (φ = 270°, θ = 90°) conformations. Open squares (GlcNAc + R) include chitobiose, GlcNAc-β1,2-Man, GMMG, NGAB2, and TMG-chitotriomycin. Closed downward triangles (PUGNAcs) include PUGNAc and Gal-PUGNAc. Open upward triangles (thiazolines) include GlcNAc-thiazoline and GalNAc-thiazoline. Open downward triangles (isofagomine) indicate GalNAc-isofagomine. Closed squares indicate LNB-thiazoline. Details of the sugar compound names and their parameters are listed in Supplementary Table S2.
Table 1. Data collection and refinement statistics

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Selenomethionine-labeled Peak</th>
<th>LNB</th>
<th>LNB-thiazoline</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDB entry</td>
<td>4H0H</td>
<td>4JAW</td>
<td></td>
</tr>
<tr>
<td>Data collection statistics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.97898</td>
<td>1.00000</td>
<td>1.00000</td>
</tr>
<tr>
<td>Space group</td>
<td>( P2_12_12 )</td>
<td>( P2_12_12 )</td>
<td>( P2_12_12 )</td>
</tr>
<tr>
<td>Unit cell (Å)</td>
<td>( a = 117.3 )</td>
<td>( a = 116.8 )</td>
<td>( a = 116.5 )</td>
</tr>
<tr>
<td></td>
<td>( b = 131.6 )</td>
<td>( b = 131.0 )</td>
<td>( b = 131.6 )</td>
</tr>
<tr>
<td></td>
<td>( c = 104.6 )</td>
<td>( c = 104.4 )</td>
<td>( c = 104.8 )</td>
</tr>
<tr>
<td>Resolution (Å)(^a)</td>
<td>50.0–2.20</td>
<td>50.0–1.80</td>
<td>50.0–1.80</td>
</tr>
<tr>
<td></td>
<td>(2.28–2.20)</td>
<td>(1.86–1.80)</td>
<td>(1.83–1.80)</td>
</tr>
<tr>
<td>Total reflections</td>
<td>607,378</td>
<td>1,107,867</td>
<td>1,106,951</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>158,438</td>
<td>148,717</td>
<td>150,089</td>
</tr>
<tr>
<td>Completeness (%)(^a)</td>
<td>100 (99.9)</td>
<td>100 (100)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>Redundancy (^a)</td>
<td>3.8 (3.8)</td>
<td>7.4 (7.4)</td>
<td>7.4 (7.3)</td>
</tr>
<tr>
<td>Mean ( I/\sigma(I) ) (^a)</td>
<td>14.22 (3.41)</td>
<td>27.64 (4.67)</td>
<td>23.97 (4.85)</td>
</tr>
<tr>
<td>( R_{merge} )(^a)</td>
<td>10.0 (31.2)</td>
<td>7.9 (28.1)</td>
<td>10.2 (48.5)</td>
</tr>
</tbody>
</table>

Refinement statistics

| Resolution range (Å)      | 47.51–1.80                   | 43.61–1.80 |
| No. of reflections        | 141,195                      | 141,665 |
| \( R\)-factor/\( R_{free} \) (%) | 15.5/19.1                | 18.2/21.3 |
| RMSD from ideal values    |                             |           |
| Bond lengths (Å)          | 0.028                        | 0.027 |
| Bond angles (°)           | 2.503                        | 2.365 |
| Average \( B \)-factor (Å\(^2\)) | 16.0/15.7                | 11.6/11.3 |
| Protein (chain A/B)       | 11.5/10.1                    | 11.6/9.12 |
| Ligand (chain A/B)        | 26.6                         | 12.3 |
| Water                     | 32.3                         | 23.8 |
| \( \text{SO}_4^{2-} \)    |                             |           |
| Ramachandran plot (%)\(^b\) | 97.8/97.3               | 98.3/98.4 |
| Favored (chain A/B)       | 1.7/2.4                      | 1.7/1.4 |
| Allowed (chain A/B)       | 0.5/0.3                      | 0.0/0.2 |

\(^a\) Values in parentheses are for the highest resolution shell.
\(^b\) Calculated by RAMPAGE (71).
### Table 2. Kinetic parameters of wild-type and mutants of *BbLNBase*<sup>a</sup>

<table>
<thead>
<tr>
<th>Enzyme</th>
<th><em>K</em>&lt;sub&gt;m&lt;/sub&gt; (×10&lt;sup&gt;-6&lt;/sup&gt; M)</th>
<th><em>k</em>&lt;sub&gt;cat&lt;/sub&gt; (s&lt;sup&gt;-1&lt;/sup&gt;)</th>
<th><em>k</em>&lt;sub&gt;cat&lt;/sub&gt;/&lt;sub&gt;<em>K</em>&lt;sub&gt;m&lt;/sub&gt; (×10&lt;sup&gt;3&lt;/sup&gt; s&lt;sup&gt;-1&lt;/sup&gt;/M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>99 ± 11</td>
<td>15 ± 1</td>
<td>160 ± 10</td>
</tr>
<tr>
<td>H263F</td>
<td>120 ± 20</td>
<td>0.15 ± 0.01</td>
<td>1.3 ± 0.1</td>
</tr>
<tr>
<td>D320A</td>
<td>130 ± 20</td>
<td>0.0068 ± 0.0005</td>
<td>0.052 ± 0.006</td>
</tr>
<tr>
<td>D320N</td>
<td>380 ± 50</td>
<td>0.011 ± 0.001</td>
<td>0.028 ± 0.002</td>
</tr>
<tr>
<td>Y419F</td>
<td>_&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.086&lt;sup&gt;c&lt;/sup&gt;</td>
<td>–</td>
</tr>
</tbody>
</table>

<sup>a</sup> Activities of *pNP-LNB* hydrolysis were measured in 50 mM McIlvaine buffer (pH 4.5) at 30°C.

<sup>b</sup> The *K*<sub>m</sub> value was too low to be determined.

<sup>c</sup> The *v/*[E]<sub>0</sub> value at 100 µM substrate.
Structure of lacto-N-biosidase

Fig. 1, Ito et al.
Fig. 2, Ito et al.
Fig. 2, Ito et al. (continued)
Fig. 3, Ito et al.
Fig. 4, Ito et al.
Structure of lacto-N-biosidase

Fig. 5, Ito et al.
Fig. 6, Ito et al.
Crystal structures of a glycoside hydrolase family 20 lacto-N-biosidase from
Bifidobacterium bifidum

Tasuku Ito, Takane Katayama, Mitchell Hattie, Haruko Sakurama, Jun Wada, Ryuichiro
Suzuki, Hisashi Ashida, Takayoshi Wakagi, Kenji Yamamoto, Keith A. Stubbs and
Shinya Fushinobu

J. Biol. Chem. published online March 11, 2013

Access the most updated version of this article at doi: 10.1074/jbc.M112.420109

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

Supplemental material:
http://www.jbc.org/content/suppl/2013/03/11/M112.420109.DC1

This article cites 0 references, 0 of which can be accessed free at
http://www.jbc.org/content/early/2013/03/11/jbc.M112.420109.full.html#ref-list-1