

Original Article

Open Access



# Substrate recognition mode of a glycoside hydrolase family 42 $\beta$ -galactosidase from *Bifidobacterium longum* subspecies *infantis* (BiBga42A) revealed by crystallographic and mutational analyses

Aina Gotoh<sup>1,2</sup>, Masafumi Hidaka<sup>3</sup>, Haruko Sakurama<sup>2</sup>, Mamoru Nishimoto<sup>4</sup>, Motomitsu Kitaoka<sup>4,5</sup>, Mikiyasu Sakanaka<sup>1</sup>, Shinya Fushinobu<sup>6</sup>, Takane Katayama<sup>1,2</sup>

<sup>1</sup>Graduate School of Biostudies, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan.

<sup>2</sup>Ishikawa Prefectural University, Nonouchi, Ishikawa 921-8836, Japan.

<sup>3</sup>Graduate School of Agricultural Science, Tohoku University, Sendai, Miyagi 980-8572, Japan.

<sup>4</sup>Institute of Food Research, National Agriculture and Food Research Organization, Tsukuba, Ibaraki 305-8642, Japan.

<sup>5</sup>Faculty of Agriculture, Niigata University, Niigata 950-2102, Japan.

<sup>6</sup>Department of Biotechnology, The University of Tokyo, Bunkyo-ku, Tokyo 113-8657, Japan.

**Correspondence to:** Takane Katayama, Graduate School of Biostudies, Kyoto University, Kitashirakawa Sakyo-ku, Kyoto 606-8502, Japan. E-mail: takane@lif.kyoto-u.ac.jp; Shinya Fushinobu, Department of Biotechnology, The University of Tokyo, Hongo Bunkyo-ku, Tokyo 113-8657, Japan. E-mail: asfushi@mail.ecc.u-tokyo.ac.jp

**How to cite this article:** Gotoh A, Hidaka M, Sakurama H, Nishimoto M, Kitaoka M, Sakanaka M, Fushinobu S, Katayama T. Substrate recognition mode of a glycoside hydrolase family 42  $\beta$ -galactosidase from *Bifidobacterium longum* subspecies *infantis* (BiBga42A) revealed by crystallographic and mutational analyses. *Microbiome Res Rep* 2023;2:20. <https://dx.doi.org/10.20517/mrr.2023.14>

**Received:** 21 Feb 2023 **First Decision:** April 12 2023 **Revised:** May 2 2023 **Accepted:** May 9 2023 **Published:** 26 May 2023

**Academic Editor:** Christian Milani **Copy Editor:** Dong-Li Li **Production Editor:** Dong-Li Li

## Abstract

**Aim:** *Bifidobacterium longum* subsp. *infantis* uses a glycoside hydrolase (GH) family 42  $\beta$ -galactosidase (BiBga42A) for hydrolyzing lacto-*N*-tetraose (LNT), which is the most abundant core structure of human milk oligosaccharides (HMOs). As such, BiBga42A represents one of the pivotal enzymes underpinning the symbiosis between bifidobacteria and breastfed infants. Despite its importance, the structural basis underlying LNT hydrolysis by BiBga42A is not understood. Moreover, no substrate-complexed structures are available to date for GH42 family members.



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



**Methods:** X-ray crystallography was used to determine the structures of *BiBga42A* in the apo- and liganded forms. The roles of the amino acid residues that were presumed to be involved in catalysis and substrate recognition were examined by a mutational study, in which kinetic parameters of each mutant were determined using 4-nitrophenyl- $\beta$ -D-galactoside, lacto-*N*-biose I, LNT, and lacto-*N*-neotetraose (LNnT) as substrates. Conservation of those amino acid residues was examined among structure-determined GH42  $\beta$ -galactosidases.

**Results:** Crystal structures of the wild-type enzyme complexed with glycerol, the E160A/E318A double mutant complexed with galactose (Gal), and the E318S mutant complexed with LNT were determined at 1.7, 1.9, and 2.2 Å resolutions, respectively. The LNT molecule (excluding the Gal moiety at subsite +2) bound to the E318S mutant is recognized by an extensive hydrogen bond network and several hydrophobic interactions. The non-reducing end Gal moiety of LNT adopts a slightly distorted conformation and does not overlap well with the Gal molecule bound to the E160A/E318A mutant. Twelve of the sixteen amino acid residues responsible for LNT recognition and catalysis in *BiBga42A* are conserved among all homologs including  $\beta$ -1,6-1,3-galactosidase (*BiGal42A*) from *Bifidobacterium animalis* subsp. *lactis*.

**Conclusion:** *BiGal42A* is active on 3- $\beta$ -galactobiose similarly to *BiBga42A* but is inactive on LNT. Interestingly, we found that the entrance of the catalytic pocket of *BiGal42A* is narrower than that of *BiBga42A* and seems not easily accessible from the solvent side due to the presence of two bulky amino acid side chains. The specificity difference may reflect the structural difference between the two enzymes.

**Keywords:** lacto-*N*-tetraose, glycoside hydrolase family 42,  $\beta$ -galactosidase, bifidobacteria, human milk oligosaccharides, crystal structure

## INTRODUCTION

Human milk not only fulfills the nutritional requirements of newborns but also guides microbiota development in the infant gut<sup>[1]</sup>. Human milk oligosaccharides (HMOs), a collective term for sugars with a degree of polymerization of  $\geq 3$ , are bioactive compounds shown to be instrumental in the formation of bifidobacteria-rich gut microbiota<sup>[2-6]</sup>. HMOs are present at a concentration of 10~20 g/L in breast milk and are the third most abundant solid material after lactose (Lac) and lipids. HMOs are assumed to reach the colon intact because they are resistant to pancreatic digestion<sup>[7]</sup>. A unique feature of HMOs that distinguishes them from other mammalian milk oligosaccharides is the richness of type-1 chains (Gal $\beta$ 1-3GlcNAc-O-R)<sup>[8]</sup>. As such, lacto-*N*-tetraose (LNT, Gal $\beta$ 1-3GlcNAc $\beta$ 1-3Gal $\beta$ 1-4Glc) is present as the most abundant core structure of HMOs<sup>[3,8,9]</sup>. The type-1 chain is resistant to most bacterial  $\beta$ -galactosidases<sup>[3]</sup>, and its hydrolysis requires a specific subgroup of glycoside hydrolase family (GH) 42 enzymes that infant gut-associated bifidobacterial species possess. We previously reported that *Bifidobacterium longum* subsp. *infantis* (*B. infantis*) ATCC 15697, which directly internalizes HMOs into its cells using ABC transporters<sup>[3,10]</sup>, encodes three intracellular GH42 homologs in its genome. Among the three paralogs, only a  $\beta$ -galactosidase termed *BiBga42A* (Blon\_2016) hydrolyzes LNT<sup>[11]</sup>. One of the remaining two GH42 enzymes, *BiBga42B* (Blon\_2013), was later shown to be specific for  $\beta$ -1,4-galactooligosaccharides<sup>[12]</sup>. A study by James *et al.* showed that inactivation of *lntA*, a *BiBga42A* homolog gene (Bbr\_0529), in *Bifidobacterium breve* UCC2003 abolishes the ability of the strain to grow on LNT as a sole carbon source, while the mutation neither affects its ability to utilize Lac nor lacto-*N*-neotetraose (LNnT, Gal $\beta$ 1-4GlcNAc $\beta$ 1-3Gal $\beta$ 1-4Glc)<sup>[13]</sup>. The genome of *B. breve* UCC2003 encodes an additional GH42 homolog<sup>[14]</sup>, although its specificity has not been determined. GH42 members have thus acquired diversified functions during evolution. Note that *B. bifidum* and several strains of *B. longum* subsp. *longum* (*B. longum*) employ an alternative strategy to degrade LNT. These bifidobacteria extracellularly degrade LNT into lacto-*N*-biose I (LNB) and lactose (Lac) by lacto-*N*-biosidases belonging to GH20<sup>[15,16]</sup> (for *B. bifidum*) and GH136<sup>[5,17,18]</sup> (several strains of *B. longum*). The released LNB is then imported by a specific ABC transporter (GltABC) for assimilation<sup>[19,20]</sup>.

As our diet contains different sets of  $\beta$ -galactosides, which are represented by HMOs during breastfeeding and by galactan oligosaccharides after weaning, precise characterization of GH42 members is of fundamental importance to obtain a better understanding of how the gut microbiota is shaped and affected by dietary  $\beta$ -galactosides during the early stages of life.

Since our first report of the crystal structure of a GH42  $\beta$ -galactosidase from *Thermus thermophilus* A4<sup>[23]</sup>, the three-dimensional structures of 8 members of GH42 have been reported<sup>[24-31]</sup>. Nonetheless, only a few GH42 members have been subjected to substrate specificity analysis, and crystal structures of GH42 enzymes are available only in the apo form and/or the product Gal-complexed form. Here, we report the crystal structures of *BiBga42A* wild-type (WT) enzyme complexed with glycerol, E160A/E318A mutant complexed with Gal, and E318S mutant complexed with LNT. This study is the first to show a substrate-complexed form of GH42 members.

## METHODS

### Chemicals

4-Nitrophenyl- $\beta$ -D-galactoside (*p*NP-Gal) was purchased from Wako Pure Chemical Industries (Osaka, Japan). LNT and LNnT were provided as gifts from Glycom A/S (Hørsholm, Denmark). LNT was further purified by the method described by Ojima *et al.*<sup>[32]</sup>. LNB was prepared as described previously<sup>[33]</sup>. Other reagents of analytical grade were obtained from commercial sources.

### Site-directed mutagenesis

The QuikChange site-directed mutagenesis method (Agilent Technologies, CA, USA) was used for introducing amino acid substitutions. pET23b (Merck Millipore, MA, USA) carrying the gene for WT *BiBga42A* with a C-terminal hexahistidine-tag [*Bga42A*-(His)<sub>6</sub>] was used as the template<sup>[11]</sup>. The following primers and their complementary strands were used for the mutagenesis (mutation sites are underlined): 5'-gccagcccgggtgccgcacagcactggcgccg-3' (for R121A), 5'-catgtgagcaacgcgtacggctccac-3' (for E160A), 5'-cgacggcaacgcaatgaaccggg-3' (for F221A), 5'-cagaccaccaactcggcgtctccg-3' (for M262G), 5'-tccaacgaccattttctctcgcccggc-3' (for Y287F), 5'-tggttctcatggcgattccacgtcc-3' (for E318A), 5'-tggttctcatgggcatccacgtcc-3' (for E318G), 5'-tggttctcatgcagcattccacgtcc-3' (for E318Q), 5'-tggttctcatgagccattccacgtcc-3' (for E318S), 5'-ccgctcaacgcagcccgaaccaac-3' (for W326A), and 5'-cgtcaactgggcaccgaccaactac-3' (for R327A). After sequence confirmation, respective plasmids were introduced into the *Escherichia coli* BL21 (DE3)  $\Delta$ *lacZ* strain for expression<sup>[11]</sup>.

### Expression and purification of *Bga42A* variants

The WT *Bga42A* enzyme from *B. infantis* ATCC 15697 and its mutant enzymes were expressed and purified by a previously described method<sup>[12]</sup>, which involved Ni-nitrilotriacetic acid (NTA) affinity chromatography (Qiagen, Hilden, Germany), MonoQ 5/50 GL anion exchange chromatography (GE Healthcare Bio-Sciences, Uppsala, Sweden), and Superdex 200 10/300 GL size exclusion chromatography (GE Healthcare Bio-Sciences). When comparing LNT-hydrolyzing activity among E318A, E318G, E318Q, and E318S mutants, the enzymes were purified by Ni-NTA column chromatography only. The purified enzymes were concentrated using Amicon Ultra Centrifugal Filters (30 K) (Merck Millipore), dialyzed against 50 mM sodium phosphate buffer (pH 7.0) containing 0.05% Tween-20, and stored at 4 °C. The enzymes were stable for at least one month under the conditions. The protein was quantified using a theoretical absorption coefficient at 280 nm, calculated based on the sequence (<https://web.expasy.org/protparam/>).

### Enzyme assay

*p*NP-Gal, LNB, LNT, and LNnT were used as the substrates. The release of Gal was colorimetrically quantified as described previously<sup>[12,34]</sup>, in which the released Gal was continuously converted to 6-phospho-

glucono-1,5-lactone with the production of an equimolar of Thio-NADH. The reaction mixture contained varied concentrations of each of the substrates in 100 mM citrate-phosphate buffer (pH 6.5) supplemented with 10 U/mL hexose mutarotase (FUJIFILM Wako Pure Chemical, Osaka, Japan), 5 U/mL galactokinase, 2.5 U/mL UDP-glucose-hexose-1-phosphate uridylyltransferase, 10 U/mL phosphoglucomutase (Sigma-Aldrich, MO, USA), 5 U/mL glucose 6-phosphate dehydrogenase (Sigma-Aldrich), 1 mM UDP-Glc, 1 mM ATP, 0.625 mM Thio-NAD<sup>+</sup> (Oriental Yeast, Tokyo, Japan), 0.0125 mM Glc-1,6-bisphosphate, and 12.5 mM MgCl<sub>2</sub> in a total volume of 40 μL. Galactokinase (GalK, BLLJ\_0339) and UDP-glucose-hexose-1-phosphate uridylyltransferase (GalT, BLLJ\_0398) from *B. longum* JCM 1217 were prepared as described previously<sup>[33,34]</sup>. The mixture was preincubated at 37 °C for 5 min, to which 10 μL of similarly preincubated enzyme solution diluted with 50 mM sodium phosphate buffer (pH 6.5) containing 0.05% Tween-20 was added to initiate the reaction. The reaction was monitored at 37 °C by measuring the absorbance at 400 nm (Thio-NADH) every 2 min for 120 min. A Multiskan GO microplate reader (Thermo Fisher Scientific, MA, USA) was used for spectrophotometry. The kinetic parameters were calculated by curve-fitting the experimental data with the Michaelis-Menten equation, using KaleidaGraph 4.0 (Synergy Software, Tokyo, Japan).

### Crystallography

The WT enzyme, E160A/E318A double mutant, and E318S mutant were purified as described above. The WT and E160A/E318A proteins were dialyzed against 20 mM 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid-KOH buffer (pH 7.0) containing 150 mM NaCl, while E318S protein was dialyzed against 20 mM 3-(*N*-morpholino)propanesulfonic acid (MES)-NaOH buffer (pH 6.0) containing 0.05% Tween-20. The hanging drop vapor diffusion method was used for crystallization. The crystal of WT complexed with glycerol was obtained by mixing 1 μL of a protein solution (20 mg/mL) containing 100 mM Gal with an equal volume of a reservoir solution consisting of 0.1 M KSCN, 30% PEG MME 2000, and 25% glycerol (cryoprotectant). The crystal of E160A/E318A complexed with Gal was obtained by mixing 1 μL of a protein solution (20 mg/mL) containing 60 mM LNT with an equal volume of a reservoir solution consisting of 0.1 M MES-NaOH buffer (pH 6.5), 0.1 M (CH<sub>3</sub>COO)<sub>2</sub>Mg, 10% PEG 10000, and 25% glycerol (cryoprotectant). The crystal of E318S complexed with LNT was obtained by mixing 1 μL of a protein solution (20 mg/mL) containing 100 mM LNT with an equal volume of a reservoir solution consisting of 5 mM MES-NaOH buffer (pH 5.8), 0.1 M KSCN, 30% PEG MME 2000, and 20% ethylene glycol (cryoprotectant). The crystals grew at 20 °C within 2 days in all cases. The crystals were flash-cooled in a nitrogen stream at 100 K. X-ray diffraction data were collected at 100 K at the beamline BL-5A at the Photon Factory of the High Energy Accelerator Research Organization (KEK, Tsukuba, Japan, λ = 1.0 Å). Preliminary diffraction data were collected at other beamlines at Photon Factory and SPring-8 (Hyogo, Japan). The data sets were processed using XDS (Jan 10, 2022)<sup>[35]</sup> and Aimless (0.7.9)<sup>[36]</sup>. Molecular replacement was performed using MOLREP (11.9.02)<sup>[37]</sup>. Model building and refinement were performed using Coot (0.9.8.6)<sup>[38]</sup> and Refmac (5.8.0403)<sup>[39]</sup>. The Dali Server was used for structural comparison<sup>[40]</sup>. Molecular interface analysis was performed using the PDBePISA server<sup>[41]</sup>. Molecular graphic images were prepared using PyMOL (2.5.4) (Schrödinger, NY, USA). Stereographic figures were created using the “ray angle =” command, specifying -3 and +3 degrees for the left and right panels, respectively.

### Phylogenetic tree construction

The structure-determined β-galactosidases<sup>[23-30]</sup> and α-arabinopyranosidase<sup>[31]</sup> listed in GH42 of the Carbohydrate-Active enZYmes database<sup>[42]</sup> and *BiBga42A* were used for the tree construction. A β-galactosidase from *Bifidobacterium adolescentis* was omitted from the analysis because the structure does not contain the catalytic domain (Midwest Center for Structural Genomics). The maximum likelihood tree was constructed using MegaX, based on the sequences aligned using the ClustalW algorithm with default settings<sup>[43]</sup>.

### Search for *BiBga42A* homologs among several bifidobacterial species

The occurrence of *BiBga42A* homologs among several *Bifidobacterium* species was examined using the Tblastn program (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Genome sequence-completed, type strains of *B. adolescentis* ATCC 15703, *Bifidobacterium angulatum* JCM 7096, *Bifidobacterium animalis* subsp. *animalis* ATCC 25527, *B. animalis* subsp. *lactis* (*B. lactis*) DSM 10140, *B. bifidum* JCM 1255, *B. breve* JCM 1192, *Bifidobacterium catenulatum* subsp. *catenulatum* JCM 1194, *B. catenulatum* subsp. *kashiwanohense* JCM 15439, *Bifidobacterium dentium* JCM 1195, *Bifidobacterium eulemuris* DSM 100216, *Bifidobacterium lemurum* DSM 28807, *B. longum* JCM 1217, *Bifidobacterium pseudocatenulatum* JCM 1200, and *Bifidobacterium pseudolongum* subsp. *globosum* DSM 20092 were used for the homolog search. Alignment of the retrieved homologs was carried out using the ClustalW program with default settings<sup>[44]</sup>.

## RESULTS AND DISCUSSION

### Crystal structures of *BiBga42A*

First, we determined the structure of the WT enzyme complexed with glycerol (WT-GOL) at 1.7 Å resolution [Supplementary Table 1]. The electron density of a glycerol molecule, which was used as a cryoprotectant, was observed at the active site, even though the crystals were grown in the presence of Gal [Supplementary Figure 1A]. *BiBga42A* has a three-dimensional structure similar to other GH42 enzymes that have three domains consisting of a ( $\beta/\alpha$ )<sub>8</sub> barrel (domain A), an  $\alpha/\beta$  fold (domain B), and an anti-parallel  $\beta$ -sandwich (domain C) [Figure 1A]. A Dali structural search<sup>[45]</sup> revealed that *BiBga42A* is most similar to GH42  $\beta$ -1,6-galactosidase from *B. bifidum* S17 (BbgII, PDB ID: 4UCF)<sup>[24]</sup> with the root-mean-square deviation (RMSD) = 0.6 Å for 684 Ca atoms (Z score = 56.6 and sequence identity = 76%). The second hit was GH42  $\beta$ -1,6-1,3-galactosidase from *B. lactis* Bl-04 (*BiGal42A*, PDB ID: 4UOZ)<sup>[25]</sup> with RMSD = 0.6 Å for 685 Ca atoms (Z score = 55.9 and sequence identity = 62%). The asymmetric unit of the WT-GOL crystal contained one molecule of *BiBga42A*, and it forms a homotrimer with symmetry-related molecules by a crystallographic 3-fold axis [Figure 1B]. A PISA molecular interface analysis calculated that approximately 27% of the entire surface area of the trimer (17,082 Å<sup>2</sup> of 64,278 Å<sup>2</sup>) was buried. Our previous study using calibrated gel filtration chromatography also suggested that *BiBga42A* forms a trimer in solution<sup>[11]</sup>. The trimeric structure of *BiBga42A* resembles a flowerpot, like other GH42 enzymes<sup>[23-25]</sup>.

Glu-160 and Glu-318 in *BiBga42A* were predicted to be the acid/base catalyst and the nucleophile, respectively, based on previous studies of GH42 enzymes and sequence alignment<sup>[46,47]</sup>. The  $k_{\text{cat}}$  values of the respective alanine mutants (E160A and E318A) towards *p*NP-Gal dropped by 170- and 2,900-fold compared to the WT enzyme, respectively, without affecting  $K_m$  values (within 2.5-fold change) [Table 1 and Supplementary Figure 2]. Based on these results, we first attempted co-crystallization of E318A with LNT. However, in a preliminary experiment, we found that the purified mutant can hydrolyze LNT when present at a high concentration (9 mg/mL) [Supplementary Figure 2]. Therefore, we also replaced the acid/base catalyst Glu-160 with alanine and performed co-crystallization with LNT using the preparation (E160A/E318A double mutant). The crystals obtained within 2 days were used for X-ray diffraction data collection. However, LNT was also hydrolyzed during the crystallization under these conditions, because only the electron density of the product Gal in the  $\alpha$ -configuration ( $\alpha$ -Gal) was observed in subsite -1 of the active site [Supplementary Figure 1B]. The Gal-complexed structure was determined at 1.9 Å resolution (E160A/E318A-Gal) [Supplementary Table 1].  $\alpha$ -Gal was also observed in the crystal structures of several other GH42 enzymes<sup>[23-26]</sup>. The asymmetric unit contained six molecules (two trimers) of *BiBga42A*. The main chain structures of the six polypeptides (chains A-F) were virtually the same (C $\alpha$  RMSD < 0.107 Å for all chain pairs), and an  $\alpha$ -Gal molecule was similarly bound in each chain. Therefore, we mainly describe the chain A molecule. The pyranose ring of the  $\alpha$ -Gal adopts an undistorted <sup>4</sup>C<sub>1</sub> conformation, and the C6 hydroxyethyl group takes a *gt* conformation [Figure 2A and Supplementary Figure 1B]. The glycerol molecule in WT-GOL occupies the C2-C4 position of the Gal. The Gal recognition by the protein involves

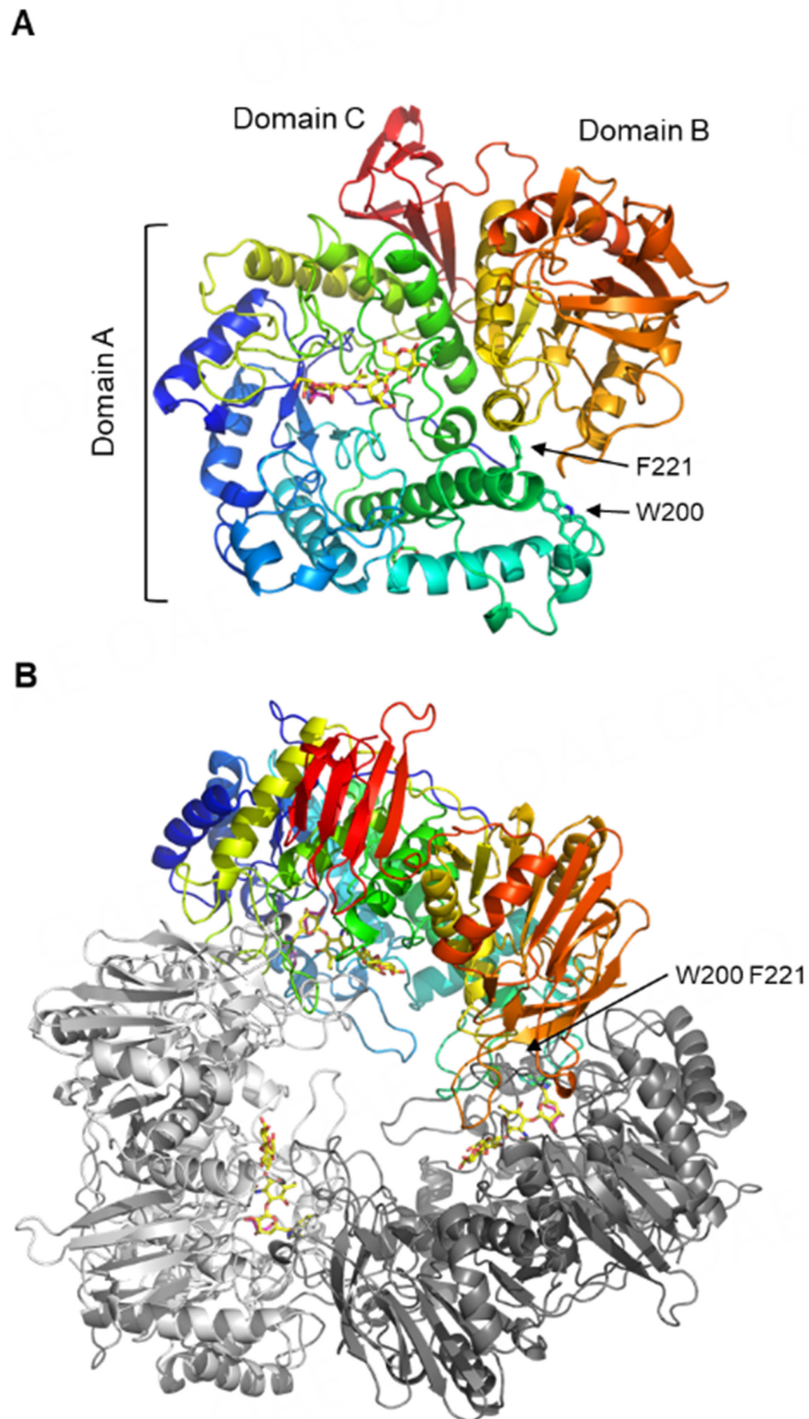
**Table 1. The kinetic parameters of *BiBga42A* variants for pNP-Gal, LNB, LNT, and LNnT**

		pNP-Gal	LNB	LNT	LNnT
<b>WT</b>	$K_m$ (mM)	0.57	25	2.7	15
	$k_{cat}$ ( $s^{-1}$ )	580	31	86	15
<b>R121A</b>	$K_m$ (mM)	1.4	42	16	35
	$k_{cat}$ ( $s^{-1}$ )	32	0.79	3.6	0.53
<b>E160A</b>	$K_m$ (mM)	0.94	nd <sup>b</sup>	nd	nd
	$k_{cat}$ ( $s^{-1}$ )	3.4	nd	nd	nd
<b>F221A<sup>a</sup></b>	$K_m$ (mM)	1.2	27	8.4	24
	$k_{cat}$ ( $s^{-1}$ )	110	1.4	11	6.4
<b>M262G</b>	$K_m$ (mM)	1.1	14	3.0	13
	$k_{cat}$ ( $s^{-1}$ )	12	0.99	2.9	0.13
<b>Y287F</b>	$K_m$ (mM)	1.0	16	2.6	18
	$k_{cat}$ ( $s^{-1}$ )	6.0	0.18	1.1	0.13
<b>E318A</b>	$K_m$ (mM)	1.4	nd	nd	nd
	$k_{cat}$ ( $s^{-1}$ )	0.20	nd	nd	nd
<b>W326A</b>	$K_m$ (mM)	12	33	7.7	64
	$k_{cat}$ ( $s^{-1}$ )	44	0.054	0.14	0.13
<b>R327A</b>	$K_m$ (mM)	0.92	53	11	52
	$k_{cat}$ ( $s^{-1}$ )	380	21	60	11

<sup>a</sup>The amino acid residue from the neighboring subunit in the trimer; <sup>b</sup>not determined. *BiBga42A*: a glycoside hydrolase family 42  $\beta$ -galactosidase; LNB: lacto-*N*-biose I; LNT: lacto-*N*-tetraose; LNnT: lacto-*N*-neotetraose; pNP-Gal: 4-Nitrophenyl- $\beta$ -D-galactoside; WT: wild-type.

many hydrogen bonds and a stacking interaction. The O6 atom of the sugar forms hydrogen bonds with the side chain nitrogen atoms of Trp-326 and His-369, while the axial O4 atom is recognized by Arg-121 and Glu-366. Arg-121 is also involved in the recognition of the O3 atom of the sugar. The O3 atom also forms water-mediated hydrogen bonds with the N $\epsilon$ 1 atom of Trp-200 from the neighboring subunit of the trimer. Trp-200 is located at the tip of a long helix in domain A [Figure 1]. The O2 atom of the Gal makes hydrogen bonds with the side chains of Asn-159 and Asp-285. The O2 atom and Asp-285 also form a hydrogen bond via a water molecule. The water molecule is located at the corresponding position of a side chain oxygen atom of Glu-318 (nucleophile) observed in WT-GOL. The  $\alpha$ -anomeric O1 atom is recognized by the side chains of Asp-285 and Tyr-287 via hydrogen bonds. The side chain of Phe-356 makes a stacking interaction with the hydrophobic C4 region of the pyranose ring. All of these residues are conserved within the members of GH42  $\beta$ -galactosidases with reported structures [Figure 3A]. In the crystal structure of the  $\beta$ -galactosidase from *Niallia circulans* (formerly called *Bacillus circulans*) subsp. *alkalophilus* (Bca- $\beta$ -Gal), the residue corresponding to Trp-326 of *BiBga42A* is shifted away<sup>[26]</sup>.

As mentioned above, even in the E160A/E318A double mutant, the substrate hydrolysis occurred within 2 days of crystal growth. Crystal soaking experiments were unsuccessful. We then replaced Glu-318 with glycine (E318G), glutamine (E318Q), or serine (E318S). Using the His-tag affinity-purified preparations at high concentrations (9 mg/ml), we examined their remaining LNT-hydrolyzing activity. As a result, the lowest activity was detected for the E318S mutant [Supplementary Figure 2]. An LNT-complexed crystal was indeed obtained for the mutant, and the structure was determined at 2.2 Å resolution (E318S-LNT) [Supplementary Table 1]. The asymmetric unit of the E318S-LNT crystal contained two molecules (chains A and B) of *BiBga42A*, and they are virtually the same (C $\alpha$  RMSD = 0.106 Å). Each chain in the asymmetric



**Figure 1.** The overall structure of *BiBga42A*. WT-GOL structure (rainbow color) with bound glycerol (magenta) is shown. LNT in E318S-LNT structure (yellow) is superimposed [Figure 2]. (A) Monomer structure; (B) Trimer structure related by a crystallographic 3-fold axis. Trp-200 and Phe-221, which form the substrate binding site of a neighboring subunit, are indicated by arrows. *BiBga42A*: a glycoside hydrolase family 42  $\beta$ -galactosidase; LNT: lacto-*N*-tetraose; WT-GOL: WT enzyme complexed with glycerol.

unit forms a trimer with crystallographic symmetry-related molecules. The electron density map of LNT was clearly visible at the active site [Supplementary Figure 1C]. We mainly describe chain B because the electron density of LNT was more clearly observed. The pyranose sugar ring of the Gal moiety in subsite -1

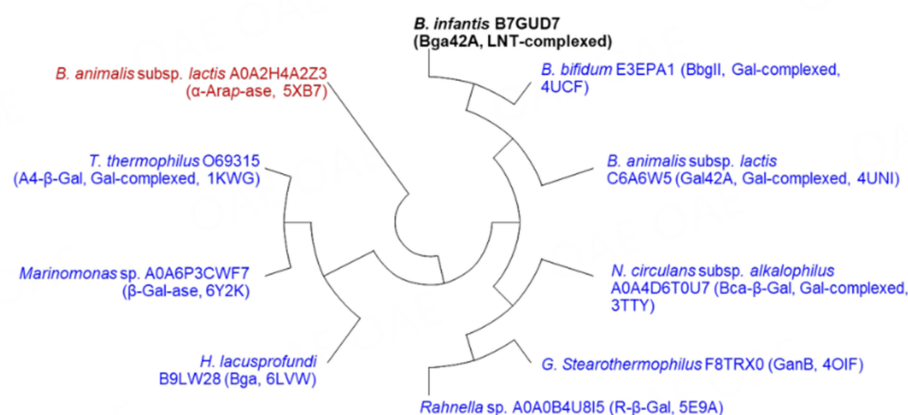




A

Interacting sugar moiety	Gal (-1)	Gal (-1)	Gal (-1)	Gal (-1)	Acid/Base	Gal (-1)	GlcNAc (+1)	GlcNAc (+1)	Gal (-1)	Gal (-1)	Nucleophile	GlcNAc (+1)	Glc (+3)	Gal (-1)	Gal (-1)	Gal (-1)
Residue number	24	121	156	159	160	200*	221*	262	285	287	318	326	327	356	366	369
<i>B. infantis</i> Bga42A	D	R	H	N	E	W	F	M	D	Y	E	W	R	F	E	H
<i>B. bifidum</i> BbgII (4UCF)	D	R	H	N	E	W	F	M	D	Y	E	W	R	F	E	H
<i>B. lactis</i> Gal42A (4UNI)	D	R	H	N	E	W	F	M	D	Y	E	W	R	F	E	H
<i>N. alkalophilus</i> Bca-β-Gal (3TTY)	D	R	H	N	E	W	T	M	D	Y	E	W (shifted)	Q	F	E	H
<i>G. stearothermophilus</i> GanB (4OIF)	D	R	H	N	E	W	H	M	D	Y	E	W	H	F	E	H
<i>Rahnella</i> sp R-β-Gal (5E9A)	D	R	H	N	E	W	N	M	D	Y	E	W	Q	F	E	H
<i>H. lacusprofundi</i> Bga (6LVW)	C	R	Q	N	E	W	E	M	D	Y	E	W	P	F	E	H
<i>Marinomonas</i> sp. β-Gal-ase (6Y2K)	C	R	Q	N	E	W	E	M	D	Y	E	W	A	F	E	H
<i>Thermus</i> sp. A4-β-Gal (1KWG)	C	R	Q	N	E	W	E	M	D	Y	E	W	A	F	E	H

B



**Figure 3.** Sequence comparison within the structure-determined GH42 members. (A) The amino acid residues involved in ligand recognition in *Bi*Bga42A were aligned with eight structure-determined GH42 β-galactosidases using the ClustalW program<sup>[44]</sup>. The residue numbers are based on *Bi*Bga42A. The sugar moieties interacting with the respective amino acid residues are indicated by their subsite positions. Asterisks indicate the residues from the neighboring subunit; (B) Structure-determined enzymes, i.e., eight β-galactosidases (blue) and one α-arabinopyranosidase (brown), and *Bi*Bga42A (bold black) were used for the tree construction. The Uniprot numbers, PDB IDs, and taxonomic names are shown. β-Galactosidase from *B. adolescentis* (PDB ID: 5VYM) was omitted from the analysis because the structure does not contain the catalytic domain. The maximum likelihood tree was constructed using MegaX based on the sequences aligned using ClustalW with default settings<sup>[43]</sup>. The lowest amino acid identity detected among the members is 20% between *B. lactis* A0A2H4A2Z3 and *B. bifidum* E3EPA1, *G. stearothermophilus* F8TRX0, or *Thermus* sp. O69315, while the highest one is 76% between *B. infantis* B7GUD7 (*Bi*Bga42A) and *B. bifidum* E3EPA1 (BbgII). The sequences are: *B. bifidum* E3EPA1<sup>[24]</sup>, *B. animalis* subsp. *lactis* C6A6W5<sup>[25]</sup>, *N. circulans* subsp. *alkalophilus* A0A4D6T0U7<sup>[26]</sup>, *G. stearothermophilus* F8TRX0<sup>[27]</sup>, *Rahnella* sp. A0A0B4U8I5<sup>[28]</sup>, *H. lacusprofundi* B9LW28<sup>[29]</sup>, *Marinomonas* sp. A0A6P3CWF7<sup>[30]</sup>, *T. thermophilus* O69315<sup>[23]</sup>, and *B. animalis* subsp. *lactis* A0A2H4A2Z3<sup>[31]</sup>. *Bi*Bga42A: a glycoside hydrolase family 42 β-galactosidase; GH: glycoside hydrolase.

of Glu-160 is suitably poised for proton donation to the leaving group as it is located 3.3 Å from the glycosidic bond oxygen of the LNB (Galβ1-3GlcNAc) unit in LNT. The stacking interaction between the phenolic ring of Phe-356 and the C4 region of the Gal remains unchanged. As the Gal-recognition modes between E160A/E318A-Gal and E318S-LNT differ, the two Gal do not overlap well with each other even at the C1 atom [Figure 2B], to which the nucleophile Glu-318 residue is thought to attack to form a galactosyl-enzyme intermediate during catalysis<sup>[47]</sup>. The C1 atom of the Gal moiety at subsite -1 in E318S-LNT has moved away from the side chain of Ser-318 with a distance of 6.2 Å. In addition, the C1 atom was located 4.3 Å from the nearest side chain oxygen atom (Oε1) of the superimposed Glu-318 of WT-GOL, and the configuration was not suitable for the in-line nucleophilic attack to the anomeric carbon. These observations might suggest that LNT was bound in a catalytically incompetent mode in the structure. The

side chain of Glu-366 changes its conformation upon the substrate binding and makes a hydrogen bond with the O6 atom of the Gal moiety. It is unclear whether the difference between the Gal molecules in the two crystallographic structures is caused by an artifact during crystallization or reflects dynamics that could occur during catalysis. Nonetheless, it is noteworthy that the Gal moiety in the E318S-LNT structure adopts a slightly distorted  ${}^0H_5$ -like ( $\psi \sim 324^\circ$ ) conformation. In the case of GH2 LacZ  $\beta$ -galactosidase from *Escherichia coli*, the Gal moiety of allolactose bound to subsite -1 took the  ${}^3H_4$  ( $\psi \sim 210^\circ$ ) conformation<sup>[49]</sup>. The  $\beta$ -galactosidic bond of allolactose in LacZ was in a pseudo-axial orientation that is suitable for the nucleophilic attack from the catalytic Glu residue, and a conformation itinerary through  ${}^4C_1$ - ${}^3H_4$ - ${}^4C_1$  was proposed for the catalysis. Since it has been suggested that sugar distortion is generally involved in the catalysis of  $\beta$ -glycosidases<sup>[50]</sup>, our observation of the sugar distortion in BiBga42A might suggest that a similar conformational change occurs during catalysis in GH42.

The leaving group trisaccharide moiety [lacto-*N*-triose II (LNT2), GlcNAc $\beta$ 1-3Gal $\beta$ 1-4Glc] of LNT is recognized by the protein via both stacking and hydrogen bonding interactions. The sugar ring of the GlcNAc in subsite +1 forms a stacking interaction with the side chain of Trp-326 [Figure 2B]. The nitrogen atom of the *N*-acetyl group is recognized by a direct hydrogen bond with the carboxyl group of Glu-160 (the catalytic acid/base). The *N*-acetyl group is fixed by a hydrophobic patch created by the side chain of Met-262 and by an intramolecular hydrogen bond with the O2 atom of the Gal moiety of the Lac unit. The C6 hydroxymethyl group of the GlcNAc moiety, which faces solvent, is constrained in a gg conformation by a packing interaction with Phe-221 from the neighboring molecule of the trimer. Phe-221 is located on the tip of a long helix next to Trp-200 [Figure 1]. As for the Gal moiety of the Lac unit, no specific interaction with the protein was observed [Figure 2C]. In addition, the reducing-end Glc is recognized by two direct and one water-mediated hydrogen bonds. The direct contacts are formed between the O6 atom of the Glc moiety and the side chain of Arg-327 and between the O2 atom of the Glc moiety and the carbonyl oxygen of the peptide bond of Gly-266-Cys-267. Water-mediated hydrogen bonds from the O3 atom of the Glc moiety are formed with the nitrogen atom of the peptide bond of Cys-267-Thr-268. The Gal and Glc moieties of the Lac unit of LNT show higher B-factors than the Gal and GlcNAc moieties of the LNB unit of LNT, suggesting that the reducing-end Lac unit is relatively flexible [Supplementary Table 1]. Overall, the LNT molecule bound to the catalytic pocket of the E318S mutant is recognized by an extensive hydrogen bond network and several hydrophobic interactions which involve a residue from the neighboring subunit.

### Mutational analysis

We then introduced amino acid replacements into several positions described above to evaluate their importance in the reaction process. The substrates used were *p*NP-Gal, LNB, LNT, and LNnT. The values obtained for the WT enzyme were comparable with those reported in our previous study<sup>[11]</sup>.

*Subsite -1*: Alanine substitution for Arg-121 (R121A) increased the  $K_m$  values for the four substrates by 2- to 3-fold, while it reduced the  $k_{cat}$  values by 20- to 30-fold [Table 1]. The results suggest that Arg-121 contributes more to catalysis than to substrate binding. The Y287F substitution, which presumably causes the loss of a hydrogen bond with the Gal moiety of LNT, did not cause a marked alteration in the  $K_m$  values but caused a marked reduction in the  $k_{cat}$  values by 80- to 170-fold. The effect on the  $k_{cat}$  values for *p*NP-Gal was the third largest, following the nucleophile mutation (E318A) and the acid/base catalyst mutation (E160A), indicating that this residue plays a crucial role in the catalysis.

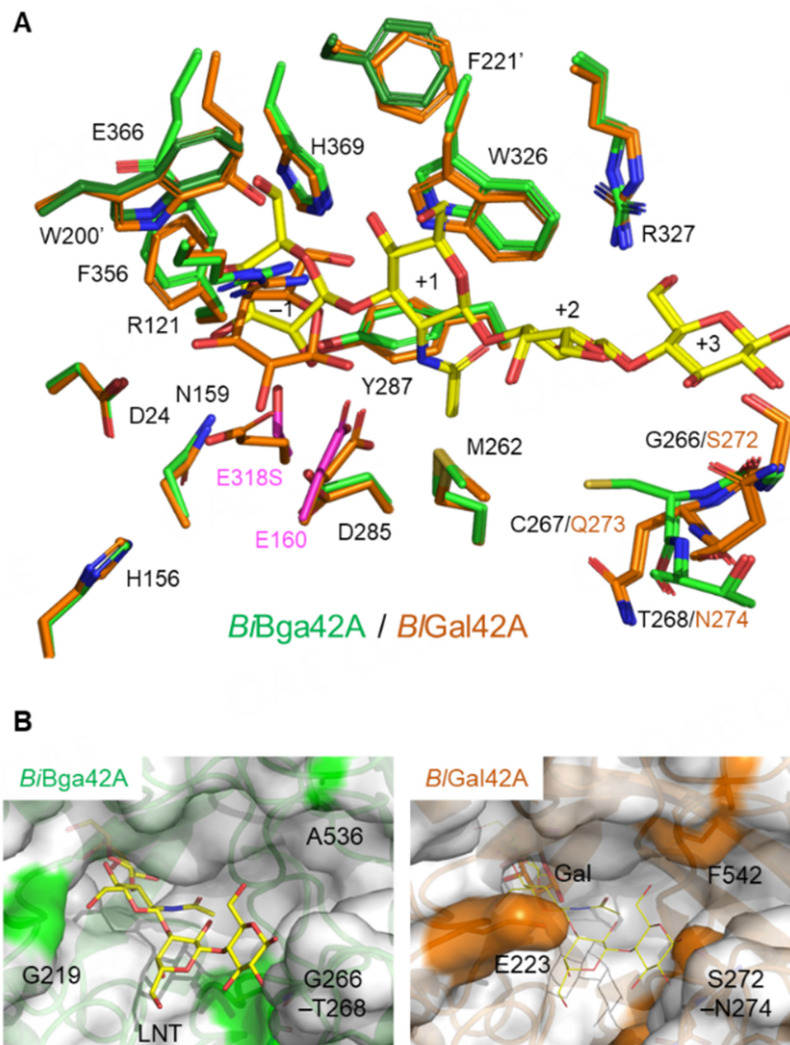
*Subsite +1*: Alanine substitution for Phe-221 (F221A), the residue from the neighboring subunit, slightly increased the  $K_m$  values for all substrates within a range less than 3-fold and decreased the  $k_{cat}$  values by 2- to 20-fold. The considerable effect was observed for LNB followed by LNT regarding the turnover, thus

demonstrating the involvement of this residue in fixation of the GlcNAc moiety during catalysis. The M262G replacement had no remarkable effects on the substrate binding by the enzyme, but considerably affected the  $k_{\text{cat}}$  values, which decreased between 30- and 100-fold. The results suggest that, in addition to its importance in the hydrophobic patch formation near subsite +1, Met-262 plays an important role in maintaining the side chain orientation of Glu-160 (the acid/base residue) properly during catalysis, as suggested by the crystal structure. The replacement of Trp-326 with alanine had various effects on activity. While it increased the  $K_m$  value by 20-fold and decreased the  $k_{\text{cat}}$  value by 13-fold for *p*NP-Gal, it increased the  $K_m$  values by 1.3- to 4-fold and decreased the  $k_{\text{cat}}$  value by 100- to 600-fold for natural substrates. The greater extent of change in the  $k_{\text{cat}}$  values for natural substrates than for *p*NP-Gal indicates the importance of the sugar-protein stacking interaction at subsite +1 in the transition state. The W326A mutation affected the activity of LNB and LNT to a similar extent.

*Subsite +3*: When Arg-327 was replaced with alanine (R327A), a slight increase in the  $K_m$  values for LNT and LNnT was observed without considerably affecting the parameters for the other substrates. Thus, we predict that the interaction between the reducing-end Glc moiety and Arg-327 may also occur between the protein and LNnT.

### **Conservation of the amino acid residues involved in the substrate recognition within structure-determined GH42 $\beta$ -galactosidases and within *BiBga42A* homologs from several *Bifidobacterium* species**

To date, crystal structures of GH42 members are reported in the apo- and/or Gal-complexed forms only. The E318S-LNT thus represents the first GH42 structure in complex with a substrate. The subsequent mutational analysis also identified several amino acid residues important for substrate recognition and catalysis. We chose all of the above-mentioned amino acid residues and examined their conservation within the GH42  $\beta$ -galactosidase members whose crystal structures have been determined<sup>[23-30]</sup>. Note that the amino acid residues that interact with the sugar (LNT) through the peptide backbone (Gly-266-Thr-268) were not considered in the conservation analysis. It should also be mentioned that structural signatures that discriminate between  $\beta$ -galactosidases and  $\alpha$ -arabinopyranosidases within GH42 members have been reported previously<sup>[31]</sup>. Figure 3A shows the conservation pattern of the selected amino acid residues. Among the 16 amino acid residues, including the two catalytic residues (E160 and E318), 12 residues were identical among the members. By reflecting distance in the phylogenetic tree [Figure 3B], Bga from *Halorubrum lacusprofundii*<sup>[29]</sup>,  $\beta$ -Gal-ase from *Marinomonus* sp.<sup>[30]</sup>, and A4- $\beta$ -Gal from *T. thermophilus*<sup>[23]</sup> have different amino acid residues at 4 positions (corresponding to residues 24, 156, 221, and 327 of *BiBga42A*). Those residues are conserved among the three enzymes except for residue 327, which interacts with the distal reducing-end Glc moiety of LNT in *BiBga42A*. Unfortunately, as substrate specificities of these three enzymes have not been reported, we are unable to discuss how these amino acid replacements are linked with specificity differences among the structure-determined GH42  $\beta$ -galactosidases. However, we were aware that there is a striking difference in the substrate specificity between *BiBga42A* and *BiGal42A*, which have identical amino acid residues at the selected 16 positions [Figures 3A and 4A] and share 62% overall amino acid sequence identity. Both enzymes efficiently hydrolyze  $\beta$ -(1 $\rightarrow$ 6)- and  $\beta$ -(1 $\rightarrow$ 3)-linkages and, to a moderate extent,  $\beta$ -(1 $\rightarrow$ 4)-linkages when disaccharides were used as substrates<sup>[12,25]</sup>. However, while *BiBga42A* can accept tetrasaccharides, *BiGal42A* shows very limited activity towards tetrasaccharides. For example, the catalytic efficiencies ( $k_{\text{cat}}/K_m$ ) of *BiBga42A* towards 3-galactobiose (Gal $\beta$ 1-3Gal) and 3-galactobiosyllactose (Gal $\beta$ 1-3Gal $\beta$ 1-3Gal $\beta$ 1-4Glc) were 422 and 842 mM<sup>-1</sup>s<sup>-1</sup>, respectively<sup>[12]</sup>. In contrast, while *BiGal42A* hydrolyses 3-galactobiose with the catalytic efficiency of 52 mM<sup>-1</sup>s<sup>-1</sup>, it was inactive on 3-galactobiosyllactose<sup>[25]</sup>. Moreover, in contrast to *BiBga42A*, which efficiently hydrolyzed LNT, *BiGal42A* was incapable of hydrolyzing LNT<sup>[12]</sup>. When the active site structures of *BiBga42A* and *BiGal42A* were superimposed, side chains of the conserved 16 residues were found to have almost the same conformation



**Figure 4.** Comparison of the active site structures between *BiBga42A* and *BiGal42A*. (A) Superimposition of the active sites of *BiBga42A* (E318S-LNT, green for protein and yellow for LNT) and *BiGal42A* (orange for protein and Gal); (B) Molecular surface of *BiBga42A* (left) and *BiGal42A* (right). Right, LNT in *BiBga42A* is superimposed as thin sticks. Surface areas of the residues contributing to shaping the different entrance sizes are colored green or orange. *BiBga42A*: a glycoside hydrolase family 42  $\beta$ -galactosidase; LNT: lacto-N-tetraose; *BiGal42A*:  $\beta$ -1,6-1,3-galactosidase.

except for Glu-366, which is displaced upon LNT binding [Figure 4A]. However, the main chain trace and the side chain conformation of Ser-272-Asn-274 in *BiGal42A* are different from those of Gly266-Thr268 in *BiBga42A*. Upon examination of the surface models of the two structures, we found that the catalytic pocket of *BiGal42A* is narrower than that of *BiBga42A* [Figure 4B]. The side chain of Ser-272 makes the pocket of *BiGal42A* slightly tighter. The protruding side chain of Glu-223 located on a loop of *BiGal42A* especially occludes the entrance of the pocket. Phe-542, which is located in the loop extending from the other side, also covers the entrance of the catalytic pocket of *BiGal42A*. *BiBga42A* substitutes Gly-219 and Ala-536 for Glu-223 and Phe-542 of *BiGal42A*, respectively. These bulky residues might interfere with the efficient access of longer oligosaccharide substrates to the catalytic site of *BiGal42A*.

We also examined the conservation of these amino acid residues among the *BiBga42A* homologs of several *Bifidobacterium* species that reside in the gastrointestinal tract (GIT) of humans and non-human mammals

[Table 2]. Genome sequence-completed, type strains of 14 species/subsp. were selected for the analysis. Twelve species/subsp. were found to possess *BiBga42A* homologs with 62%-96% identity. The lowest hit (62%) was observed for *BlGal42A*. Overall identities among the retrieved 12 homologs were between 59% and 96%, and all 16 residues involved in substrate recognition [Figure 3A] were conserved, as revealed by ClustalW analysis<sup>[44]</sup>. Interestingly, the above-mentioned Gly-219 is also conserved among the homologs including *BlGal42A* in the sequence alignment [Figure 4B and Supplementary Figure 3]. Notably, *Bl42GalA* has one amino acid insertion in the corresponding loop, which might result in the protrusion of Glu-223 at the entrance. The twelve homologs possess a glutamic acid or aspartic acid residue in the corresponding position in the sequences. The multiple alignments also revealed sequence variance at the Gly266-Thr268 and Ala-536 positions among the homologs. Further enzymatic characterization and mutational analysis are necessary to examine how these homologs efficiently hydrolyze LNT and related tetra- or longer oligosaccharides.

### Concluding remarks

GH42 enzymes, which are found exclusively in microbes, i.e., bacteria, archaea, and fungi, were previously thought to be involved in the degradation of plant-derived galactooligosaccharides released from galactans<sup>[51,52]</sup>. However, a study on *BiBga42A*, which was first described to be specific for LNT<sup>[11]</sup> but later found to act on a variety of both  $\beta$ -(1 $\rightarrow$ 3)- and  $\beta$ -(1 $\rightarrow$ 6)-linked galactosides<sup>[12]</sup>, revealed that GH42  $\beta$ -galactosidases target oligosaccharides of both plant and animal origins. A recent *in silico* analysis also demonstrated that the expression of most GH42  $\beta$ -galactosidase genes is not subject to the regulation by NagR, a global transcriptional regulator of animal host-glycan degradation genes, in the genus *Bifidobacterium*<sup>[53]</sup>. In *B. infantis*, the transcriptional levels of the three GH42 genes were not affected by the carbon sources used for cultivation, which may represent “enzymatic preparedness in gut microbes” as mentioned previously<sup>[54]</sup>.

Using *BiBga42A*, a well-characterized GH42 member, our study presents, for the first time, the structure of a GH42 enzyme complexed with its substrate. Comparison between the two liganded structures showed that the Gal moiety of LNT occupies the subsite -1 that is sterically different from that observed for the Gal in E160A/E318A-Gal and takes a different sugar conformation. The result may represent the dynamics in the catalytic cycle adopted by GH42 enzymes and may explain the lack of transglycosylation activity of anomer-retaining GH42 members. The donor sugar (Gal) bound at subsite -1 and an acceptor sugar bound at + subsite(s) may not be arranged suitably for transglycosylation to proceed. Overall, our study warrants further structural analysis of GH42 members with varied substrates and inhibitors to understand the catalytic mechanism and the molecular basis of different substrate specificities among the members, which may help us design novel prebiotics with distinct sugar linkages and compositions.

**Table 2. Conservation of BiBga42A homologs among several *Bifidobacterium* species**

Species (Genome sequence-completed type strain)	Locus tag	Identity (%)
<i>B. adolescentis</i> ATCC 15703	BAD_1603	515/690 (75%)
<i>B. angulatum</i> JCM 7096	BBAG_0066	527/691 (76%)
<i>B. animalis</i> subsp. <i>animalis</i> ATCC 25527	nr <sup>a</sup>	
<i>B. animalis</i> subsp. <i>lactis</i> DSM 10140	BALAT_0484 ( <i>B</i> Gal42A)	430/692 (62%)
<i>B. bifidum</i> JCM 1255	BBBF_1344	521/690 (76%)
<i>B. breve</i> JCM 1192	BBBR_0453	660/691 (96%)
<i>B. catenulatum</i> subsp. <i>catenulatum</i> JCM 1194	BBCT_0461	522/690 (76%)
<i>B. catenulatum</i> subsp. <i>kashiwanohense</i> JCM 15439	BBKW_0505	523/690 (76%)
<i>B. dentium</i> JCM 1195	BBDE_0630	517/690 (75%)
<i>B. eulemuris</i> DSM 100216	BE0216_04680	569/689 (83%)
<i>B. lemurum</i> DSM 28807	BL8807_06925	565/690 (82%)
<i>B. longum</i> subsp. <i>longum</i> JCM 1217	BLLJ_0443	661/691 (96%)
<i>B. pseudocatenulatum</i> JCM 1200	BBPC_0515	517/690 (75%)
<i>B. pseudolongum</i> subsp. <i>globosum</i> DSM 20092	nr	

<sup>a</sup>None retrieved. BiBga42A: a glycoside hydrolase family 42  $\beta$ -galactosidase.

## DECLARATIONS

### Acknowledgments

We thank the staff of the Photon Factory and SPring-8 for X-ray data collection, Yuuki Higashi at Ishikawa Prefectural University for technical support in protein crystallization, Toshihiko Katoh at Kyoto University for manuscript preparation, and Glycom A/S for providing LNT and LNnT.

### Authors' contributions

Conceived and designed the experiments: Katayama T, Gotoh A

Prepared the enzymes and assayed: Gotoh A, Nishimoto M, and Kitaoka M

Crystallized: Sakurama H, Katayama T

Determined the structures: Fushinobu S and Hidaka M

Analyzed the database: Sakanaka M, Katayama T

Drafted the manuscript: Gotoh A, Katayama T, Fushinobu S

Edited the manuscript and supervised the study: Katayama T, Fushinobu S

All authors discussed the data and contributed to completing the final manuscript.

### Availability of data and materials

Atomic coordinates and structure factors of WT-GOL, E160A/E318A-Gal, and E318S-LNT of the Bga42A protein from *B. infantis* ATCC15697 have been deposited in the PDB under accession numbers 8IBR, 8IBS, and 8IBT, respectively. All materials except for LNT and LNnT, which are gifts from Glycom A/S, are available upon reasonable request.

### Financial support and sponsorship

This study was partly supported by the Institute for Fermentation, Osaka (K-25-04 to TK), JSPS Research Fellowship (17J08530 to AG), and JSPS-KAKENHI (15H04481, 17K19231, and 21H02116 to TK, and 15F15091, 24380053, and 19H00929 to SF).

### Conflicts of interest

Employment of Sakanaka M at Kyoto University is in part supported by Morinaga Milk Industry Co., Ltd. The other authors declare no conflicts of interest.

### Ethical approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Copyright

© The Author(s) 2023.

## REFERENCES

1. Stewart CJ, Ajami NJ, O'Brien JL, et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* 2018;562:583-8. DOI PubMed PMC
2. Sakanaka M, Hansen ME, Gotoh A, et al. Evolutionary adaptation in fucosyllactose uptake systems supports bifidobacteria-infant symbiosis. *Sci Adv* 2019;5:eaaw7696. DOI PubMed PMC
3. Asakuma S, Hatakeyama E, Urashima T, et al. Physiology of consumption of human milk oligosaccharides by infant gut-associated bifidobacteria. *J Biol Chem* 2011;286:34583-92. DOI PubMed PMC
4. Sakanaka M, Gotoh A, Yoshida K, et al. Varied pathways of infant gut-associated *Bifidobacterium* to assimilate human milk oligosaccharides: prevalence of the gene set and its correlation with bifidobacteria-rich microbiota formation. *Nutrients* 2019;12:71. DOI PubMed PMC
5. Yamada C, Gotoh A, Sakanaka M, et al. Molecular insight into evolution of symbiosis between breast-fed infants and a member of the human gut microbiome *Bifidobacterium longum*. *Cell Chem Biol* 2017;24:515-524.e5. DOI
6. Ojima MN, Jiang L, Arzamasov AA, et al. Priority effects shape the structure of infant-type *Bifidobacterium* communities on human milk oligosaccharides. *ISME J* 2022;16:2265-79. DOI PubMed PMC
7. Engfer MB, Stahl B, Finke B, Sawatzki G, Daniel H. Human milk oligosaccharides are resistant to enzymatic hydrolysis in the upper gastrointestinal tract. *Am J Clin Nutr* 2000;71:1589-96. DOI PubMed
8. Urashima T, Asakuma S, Leo F, Fukuda K, Messer M, Ofstedal OT. The predominance of type I oligosaccharides is a feature specific to human breast milk. *Adv Nutr* 2012;3:473S-82S. DOI PubMed PMC
9. Katayama T. Host-derived glycans serve as selected nutrients for the gut microbe: human milk oligosaccharides and bifidobacteria. *Biosci Biotechnol Biochem* 2016;80:621-32. DOI PubMed
10. Garrido D, Kim JH, German JB, Raybould HE, Mills DA. Oligosaccharide binding proteins from *Bifidobacterium longum* subsp. *infantis* ;6:e17315. DOI PubMed PMC
11. Yoshida E, Sakurama H, Kiyohara M, et al. *Bifidobacterium longum* subsp. *infantis* ;22:361-8. DOI
12. Viborg AH, Katayama T, Abou Hachem M, et al. Distinct substrate specificities of three glycoside hydrolase family 42  $\beta$ -galactosidases from *Bifidobacterium longum* subsp. *infantis* ;24:208-16. DOI PubMed
13. James K, Motherway MO, Bottacini F, van Sinderen D. *Bifidobacterium breve* UCC2003 metabolises the human milk oligosaccharides lacto-N-tetraose and lacto-N-neo-tetraose through overlapping, yet distinct pathways. *Sci Rep* 2016;6:38560. DOI PubMed PMC
14. O'Connell Motherway M, Zomer A, Leahy SC, et al. Functional genome analysis of *Bifidobacterium breve* UCC2003 reveals type IVb tight adherence (Tad) pili as an essential and conserved host-colonization factor. *Proc Natl Acad Sci U S A* 2011;108:11217-22. DOI PubMed PMC
15. Wada J, Ando T, Kiyohara M, et al. *Bifidobacterium bifidum* lacto-N-biosidase, a critical enzyme for the degradation of human milk oligosaccharides with a type I structure. *Appl Environ Microbiol* 2008;74:3996-4004. DOI PubMed PMC
16. Hattie M, Ito T, Debowski AW, et al. Gaining insight into the catalysis by GH20 lacto-N-biosidase using small molecule inhibitors and structural analysis. *Chem Commun* 2015;51:15008-11. DOI
17. Sakurama H, Kiyohara M, Wada J, et al. Lacto-N-biosidase encoded by a novel gene of *Bifidobacterium longum* subspecies *longum* shows unique substrate specificity and requires a designated chaperone for its active expression. *J Biol Chem* 2013;288:25194-206. DOI PubMed PMC
18. Gotoh A, Katoh T, Sugiyama Y, et al. Novel substrate specificities of two lacto-N-biosidases towards  $\beta$ -linked galacto-N-biose-containing oligosaccharides of globo H, Gb5, and GA1. *Carbohydr Res* 2015;408:18-24. DOI PubMed
19. Suzuki R, Wada J, Katayama T, et al. Structural and thermodynamic analyses of solute-binding Protein from *Bifidobacterium longum* specific for core I disaccharide and lacto-N-biose I. *J Biol Chem* 2008;283:13165-73. DOI
20. Kitaoka M. Bifidobacterial enzymes involved in the metabolism of human milk oligosaccharides. *Adv Nutr* 2012;3:422S-9S. DOI

[PubMed PMC](#)

21. Sasaki Y, Komeno M, Ishiwata A, et al. Mechanism of cooperative degradation of gum arabic arabinogalactan protein by *Bifidobacterium longum* surface enzymes. *Appl Environ Microbiol* 2022;88:e0218721. [DOI](#) [PubMed](#) [PMC](#)
22. Sasaki Y, Yanagita M, Hashiguchi M, et al. Assimilation of arabinogalactan side chains with novel 3-O- $\beta$ -L-arabinopyranosyl- $\alpha$ -L-arabinofuranosidase in *Bifidobacterium pseudocatenulatum*. *Microbiome Res Rep* 2023;2:12. [DOI](#)
23. Hidaka M, Fushinobu S, Ohtsu N, et al. Trimeric crystal structure of the glycoside hydrolase family 42  $\beta$ -galactosidase from *Thermus thermophilus* A4 and the structure of its complex with galactose. *J Mol Biol* 2002;322:79-91. [DOI](#) [PubMed](#)
24. Godoy AS, Camilo CM, Kadowaki MA, et al. Crystal structure of  $\beta$ 1 $\rightarrow$ 6-galactosidase from *Bifidobacterium bifidum* S17: trimeric architecture, molecular determinants of the enzymatic activity and its inhibition by  $\alpha$ -galactose. *FEBS J* 2016;283:4097-112. [DOI](#) [PubMed](#)
25. Viborg AH, Fredslund F, Katayama T, et al. A  $\beta$ 1-6/ $\beta$ 1-3 galactosidase from *Bifidobacterium animalis* subsp. *lactis* ;BI-04 gives insight into sub-specificities of  $\beta$ -galactoside catabolism within Bifidobacterium. *Mol Microbiol* 2014;1024-40. [DOI](#) [PubMed](#)
26. Maksimainen M, Paavilainen S, Hakulinen N, Rouvinen J. Structural analysis, enzymatic characterization, and catalytic mechanisms of  $\beta$ -galactosidase from *Bacillus circulans* sp. *alkalophilus* ;279:1788-98. [DOI](#) [PubMed](#)
27. Solomon HV, Tabachnikov O, Feinberg H, et al. Crystallization and preliminary crystallographic analysis of GanB, a GH42 intracellular  $\beta$ -galactosidase from *Geobacillus stearothermophilus*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 2013;69:1114-9. [DOI](#) [PubMed](#) [PMC](#)
28. Fan Y, Hua X, Zhang Y, et al. Cloning, expression and structural stability of a cold-adapted  $\beta$ -galactosidase from *Rahnella* sp. R3. *Protein Expr Purif* 2015;115:158-64. [DOI](#)
29. Karan R, Mathew S, Muhammad R, et al. Understanding high-salt and cold adaptation of a polyextremophilic enzyme. *Microorganisms* 2020;8:1594. [DOI](#) [PubMed](#) [PMC](#)
30. Mangiagalli M, Lapi M, Maione S, et al. The co-existence of cold activity and thermal stability in an Antarctic GH42  $\beta$ -galactosidase relies on its hexameric quaternary arrangement. *FEBS J* 2021;288:546-65. [DOI](#) [PubMed](#)
31. Viborg AH, Katayama T, Arakawa T, et al. Discovery of  $\alpha$ -L-arabinopyranosidases from human gut microbiome expands the diversity within glycoside hydrolase family 42. *J Biol Chem* 2017;292:21092-101. [DOI](#) [PubMed](#) [PMC](#)
32. Ojima MN, Asao Y, Nakajima A, et al. Diversification of a fucosylactose transporter within the genus *Bifidobacterium*. *Appl Environ Microbiol* 2022;88:e0143721. [DOI](#) [PubMed](#) [PMC](#)
33. Nishimoto M, Kitaoka M. Practical preparation of lacto-N-biose I, a candidate for the bifidus factor in human milk. *Biosci Biotechnol Biochem* 2007;71:2101-4. [DOI](#) [PubMed](#)
34. Nihira T, Nakajima M, Inoue K, Nishimoto M, Kitaoka M. Colorimetric quantification of alpha-D-galactose 1-phosphate. *Anal Biochem* 2007;371:259-61. [DOI](#) [PubMed](#)
35. Kabsch W. XDS. *Acta Crystallogr D Biol Crystallogr* 2010;66:125-32. [DOI](#)
36. Evans PR, Murshudov GN. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* 2013;69:1204-14. [DOI](#) [PubMed](#) [PMC](#)
37. Vagin A, Teplyakov A. Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* 2010;66:22-5. [DOI](#) [PubMed](#)
38. Casañal A, Lohkamp B, Emsley P. Current developments in Coot for macromolecular model building of Electron Cryo-microscopy and Crystallographic Data. *Protein Sci* 2020;29:1069-78. [DOI](#) [PubMed](#) [PMC](#)
39. Murshudov GN, Skubák P, Lebedev AA, et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 2011;67:355-67. [DOI](#) [PubMed](#) [PMC](#)
40. Holm L, Laakso LM. Dali server update. *Nucleic Acids Res* 2016;44:W351-5. [DOI](#) [PubMed](#) [PMC](#)
41. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 2007;372:774-97. [DOI](#)
42. Henrissat B, Davies G. Structural and sequence-based classification of glycoside hydrolases. *Curr Opin Struct Biol* 1997;7:637-44. [DOI](#) [PubMed](#)
43. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 2018;35:1547-9. [DOI](#) [PubMed](#) [PMC](#)
44. Larkin MA, Blackshields G, Brown NP, et al. Clustal W and Clustal X version 2.0. *Bioinformatics* 2007;23:2947-8. [DOI](#)
45. Holm L. DALI and the persistence of protein shape. *Protein Sci* 2020;29:128-40. [DOI](#) [PubMed](#) [PMC](#)
46. Di Lauro B, Strazzulli A, Perugino G, et al. Isolation and characterization of a new family 42  $\beta$ -galactosidase from the thermoacidophilic bacterium *Alicyclobacillus acidocaldarius*: identification of the active site residues. *Biochim Biophys Acta* 2008;1784:292-301. [DOI](#) [PubMed](#)
47. Shaikh FA, Müllegger J, He S, Withers SG. Identification of the catalytic nucleophile in Family 42  $\beta$ -galactosidases by intermediate trapping and peptide mapping: YesZ from *Bacillus subtilis*. *FEBS Lett* 2007;581:2441-6. [DOI](#) [PubMed](#)
48. Cremer D, Pople JA. General definition of ring puckering coordinates. *J Am Chem Soc* 1975;97:1354-8. [DOI](#)
49. Wheatley RW, Huber RE. An allolactose trapped at the lacZ  $\beta$ -galactosidase active site with its galactosyl moiety in a (4)H3 conformation provides insights into the formation, conformation, and stabilization of the transition state. *Biochem Cell Biol* 2015;93:531-40. [DOI](#) [PubMed](#)
50. Davies GJ, Ducros VM, Varrot A, Zechel DL. Mapping the conformational itinerary of  $\beta$ -glycosidases by X-ray crystallography. *Biochem Soc Trans* 2003;31:523-7. [DOI](#) [PubMed](#)
51. Hinz SW, van den Brock LA, Beldman G, Vincken JP, Voragen AG.  $\beta$ -galactosidase from *Bifidobacterium adolescentis* DSM20083



- prefers  $\beta(1,4)$ -galactosides over lactose. *Appl Microbiol Biotechnol* 2004;66:276-84. DOI PubMed
52. Goulas T, Goulas A, Tzortzis G, Gibson GR. Comparative analysis of four  $\beta$ -galactosidases from *Bifidobacterium bifidum* NCIMB41171: purification and biochemical characterisation. *Appl Microbiol Biotechnol* 2009;82:1079-88. DOI PubMed
  53. Arzamasov AA, Nakajima A, Sakanaka M, et al. Human milk oligosaccharide utilization in intestinal bifidobacteria is governed by global transcriptional regulator NagR. *mSystems* 2022;7:e0034322. DOI PubMed PMC
  54. Sonnenburg JL, Xu J, Leip DD, et al. Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* 2005;307:1955-9. DOI