

RESEARCH ARTICLE

Crystal structure of β -L-arabinobiosidase belonging to glycoside hydrolase family 121Keita Saito¹, Alexander Holm Viborg¹, Shiho Sakamoto², Takatoshi Arakawa^{1,3}, Chihaya Yamada^{1,3}, Kiyotaka Fujita², Shinya Fushinobu^{1,3*}

1 Department of Biotechnology, The University of Tokyo, Tokyo, Japan, **2** Faculty of Agriculture, Kagoshima University, Kagoshima, Kagoshima, Japan, **3** Collaborative Research Institute for Innovative Microbiology, The University of Tokyo, Tokyo, Japan

* asfushi@mail.ecc.u-tokyo.ac.jp

Abstract

Enzymes acting on α -L-arabinofuranosides have been extensively studied; however, the structures and functions of β -L-arabinofuranosidases are not fully understood. Three enzymes and an ABC transporter in a gene cluster of *Bifidobacterium longum* JCM 1217 constitute a degradation and import system of β -L-arabinooligosaccharides on plant hydroxyproline-rich glycoproteins. An extracellular β -L-arabinobiosidase (HypBA2) belonging to the glycoside hydrolase (GH) family 121 plays a key role in the degradation pathway by releasing β -1,2-linked arabinofuranose disaccharide (β -Ara₂) for the specific sugar importer. Here, we present the crystal structure of the catalytic region of HypBA2 as the first three-dimensional structure of GH121 at 1.85 Å resolution. The HypBA2 structure consists of a central catalytic (α/α)₆ barrel domain and two flanking (N- and C-terminal) β -sandwich domains. A pocket in the catalytic domain appears to be suitable for accommodating the β -Ara₂ disaccharide. Three acidic residues Glu383, Asp515, and Glu713, located in this pocket, are completely conserved among all members of GH121; site-directed mutagenesis analysis showed that they are essential for catalytic activity. The active site of HypBA2 was compared with those of structural homologs in other GH families: GH63 α -glycosidase, GH94 chitobiose phosphorylase, GH142 β -L-arabinofuranosidase, GH78 α -L-rhamnosidase, and GH37 α,α -trehalase. Based on these analyses, we concluded that the three conserved residues are essential for catalysis and substrate binding. β -L-Arabinobiosidase genes in GH121 are mainly found in the genomes of bifidobacteria and *Xanthomonas* species, suggesting that the cleavage and specific import system for the β -Ara₂ disaccharide on plant hydroxyproline-rich glycoproteins are shared in animal gut symbionts and plant pathogens.

OPEN ACCESS

Citation: Saito K, Viborg AH, Sakamoto S, Arakawa T, Yamada C, Fujita K, et al. (2020) Crystal structure of β -L-arabinobiosidase belonging to glycoside hydrolase family 121. PLoS ONE 15(6): e0231513. <https://doi.org/10.1371/journal.pone.0231513>

Editor: Jean-Guy Berrin, Institut National de la Recherche Agronomique, FRANCE

Received: March 26, 2020

Accepted: May 18, 2020

Published: June 1, 2020

Copyright: © 2020 Saito et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files except for those submitted to PDB (ID code 6M5A).

Funding: This work was supported by a Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (to S.F., Nos. 19H00929, 15H02443, 26660083 and 24380053, and to K.F., No. 22780094). <https://www.jsps.go.jp/english/index.html> The funders had no role in

Introduction

Enzymes and metabolic pathways involved in the microbial degradation of α -linked L-arabinofuranosyl (L-Araf) residues, which are abundantly present in hemicellulosic polysaccharides of plants such as arabinoxylans and arabinans, have been extensively studied [1]. In the

study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

carbohydrate-active enzyme database [2], glycoside hydrolases (GHs) cleaving the α -L-Araf bonds, including α -L-arabinofuranosidases (EC 3.2.1.55) and endo-1,5- α -L-arabinanases (EC 3.2.1.99), are classified into GH families 2, 3, 43, 51, 54, and 62. On the contrary, studies on enzymes cleaving β -linked L-Arafs are limited, although such carbohydrate linkages have been found in hydroxyproline-rich glycoproteins (HRGPs) and peptide hormones [3–5]. Enzymes cleaving β -L-Araf bonds were found in human symbiotic gut bacteria after 2011; these enzymes are associated with carbohydrate exploitation in the human gastrointestinal tract. At first, two enzymes were discovered from a human gut bacterium, *Bifidobacterium longum*, and classified into GH121 and GH127 [6,7]. Subsequent studies on one of the major human gut bacteria, *Bacteroides thetaiotaomicron*, led to the discovery of other enzymes cleaving the β -L-Araf bonds in plant pectins; these enzymes were classified as GH137, GH142, and GH146 [8,9]. The molecular mechanism of β -L-Araf bond hydrolysis was solely studied on a GH127 β -L-arabinofuranosidase HypBA1 [10]. Although three-dimensional structures were determined for GH137, GH142, and GH146, the molecular mechanism of these enzymes was not uncovered in detail in the available literature [8,9].

Bifidobacterium species play a significant role in human health by acting with various transporters, glycosidases, and metabolic enzymes [11]. A gene cluster in the genome of *Bifidobacterium longum* JCM 1217 encodes three enzymes (Fig 1) and a transporter involved in the degradation system for plant HRGPs [6,7]. The three enzymes belong to different GH families: GH43 α -L-arabinofuranosidase (HypAA, locus tag = BLLJ_0213), GH121 β -L-arabinobiosidase (HypBA2, BLLJ_0212), and GH127 β -L-arabinofuranosidase (HypBA1, BLLJ_0211). The transporter is an ABC-type sugar importer system that consists of a substrate-binding protein (SBP, BLLJ_0208) and two transmembrane domain proteins (BLLJ_0209 and BLLJ_0210). Hydroxyproline (Hyp) residues in HRGPs, such as extensin and solanaceous lectins, are usually O-glycosylated with 3 or 4 L-Araf residues that are linked with three β -bonds and one α -bond (Ara₃-Hyp and Ara₄-Hyp, Fig 1) [12–14]. The chemical structures of Ara₃-Hyp and Ara₄-Hyp are Araf- β 1,2-Araf- β 1,2-Araf- β -Hyp and Araf- α 1,3-Araf- β 1,2-Araf- β 1,2-Araf- β -Hyp, respectively. The three enzymes in the bifidobacterial β -L-arabinooligosaccharide degradation pathway consist of two extracellular enzymes and one intracellular enzyme. The two extracellular enzymes (HypAA and HypBA2) initially act on the β -L-arabinooligosaccharides. HypAA releases the terminal L-arabinose (L-Ara) from Ara₄-Hyp by cleaving the α 1,3-bond (unpublished data), and HypBA2 liberates Araf- β 1,2-Araf (β -Ara₂) from Ara₃-Hyp [6]. The ABC transporter likely imports β -Ara₂ because the SBP (BLLJ_0208) specifically binds β -Ara₂ [15]. An intracellular enzyme, HypBA1, cleaves β -Ara₂ to release two L-Ara molecules [7].

The GH121 β -L-arabinobiosidase HypBA2 plays a key role in the degradation system by releasing the β -Ara₂ that is transferred to the specific sugar importer. GH121 currently consists

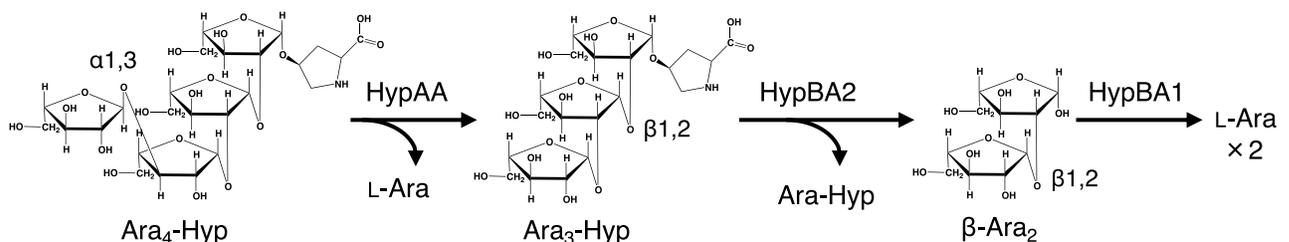


Fig 1. Glycosidic bond cleavage of the β -arabinooligosaccharide degradation system of *B. longum* JCM 1217. HypAA and HypBA2 are extracellular enzymes, and HypBA1 is an intracellular enzyme. An ABC-type transporter system specifically imports β -Ara₂ into the cell.

<https://doi.org/10.1371/journal.pone.0231513.g001>

of 292 bacterial ORFs (as of May 2020), and HypBA2 is the sole characterized member. In this study, we report the crystal structure of HypBA2 as the first three-dimensional structure of GH121.

Materials and methods

Protein expression and purification

The expression plasmid for C-terminally His-tagged proteins of CA789 (residues 33–1154) and CA1049 (residues 33–894) were constructed in our previous study [6]. Selenomethionine (SeMet)-labeled and native proteins were expressed in *Escherichia coli* BL21-CodonPlus (DE3)-RP-X and BL21-CodonPlus (DE3)-RIL (Agilent Technologies, Santa Clara, CA, USA), respectively. The transformants were grown at 37°C for 5 h in LeMaster medium (SeMet-labeled protein) [16] or LB (lysogeny broth) medium (native protein) containing 50 μ g/mL ampicillin and 34 μ g/mL chloramphenicol. Protein expression was induced by adding 0.1 mM isopropyl- β -D-thiogalactopyranoside to the medium, and the cells were further cultivated at 25°C for 20 h. The cells were harvested by centrifugation and suspended in 20 mM Tris-HCl (pH 8.0), 250 mM NaCl, and 2 mM CaCl₂ (buffer A). The cells were disrupted via sonication, and the supernatant was purified by sequential column chromatography. Ni-affinity chromatography was conducted using a HisTrap FF column (GE Healthcare, Fairfield, CT, USA) with wash and elution steps of 20 mM and 400 mM imidazole in Buffer A, respectively. The eluted protein sample was dialyzed against 20 mM Tris-HCl (pH 8.0) and 2 mM CaCl₂ (buffer B) and applied to a Mono Q 10/100 GL column (GE Healthcare) equilibrated with buffer B. The protein sample was eluted with a linear gradient of 0 to 1 M NaCl. The protein sample was concentrated by ultrafiltration (Amicon Ultra-4 centrifugal filter devices, 50,000 MWCO; Millipore, Billerica, MA, USA), and the solution was changed to buffer A. Gel filtration chromatography was conducted using a HiLoad 16/60 Superdex 200 pg column (GE Healthcare) equilibrated with buffer A at a flow rate of 1 mL/min. The purified protein was again concentrated by ultrafiltration, and the solution was changed to a buffer consisting of 5 mM Tris-HCl (pH 8.0) and 2 mM CaCl₂. Protein concentrations were determined by a BCA protein assay kit (Thermo Fisher Scientific, Waltham, MA, USA) with bovine serum albumin as the standard.

Crystallography and visualization

The protein crystals were grown at 20°C using the sitting drop vapor-diffusion method by mixing 0.5 μ L of protein solution with an equal volume of a reservoir solution. The SeMet-labeled CA789 was crystallized using a protein solution (12.5 mg/mL) and a reservoir solution containing 17% (w/v) PEG1000 and 0.1 M Tris-HCl (pH 6.6). The native CA1049 protein was crystallized using a protein solution (5 mg/mL) and a reservoir solution containing 40% (w/v) PEG400 and 0.1 M Na-acetate (pH 4.5). Crystals were flash-cooled by dipping into liquid nitrogen. X-ray diffraction data were collected at 100 K on beamlines at the Photon Factory of the High Energy Accelerator Research Organization (KEK, Tsukuba, Japan). Preliminary diffraction data were collected at SPring-8 (Hyogo, Japan). The data sets were processed using XDS [17]. The phase determination and automated model building were performed for the data of SeMet-labeled CA789 crystal using AutoSol pipeline of PHENIX [18]. The best solution of substructure search included 26 Se sites with figure of merit of 0.334 and overall score (BAYES-CC) of 48.20 ± 9.30 . After density modification, figure of merit, *R*-factor, map skew, and correlation of local root mean square density were 0.64, 0.264, 0.19, and 0.82, respectively. Automatic model building constructed 894 amino acid residues with map-model correlation coefficient and *R*/*R*_{free} values of 0.79 and 0.283/0.331, respectively. Manual model rebuilding and crystallographic refinement were performed using Coot [19] and Refmac5 [20]. The

SeMet-labeled CA789 structure was refined to R/R_{free} values of 0.249/0.309, and residues of 39–162, 176–194, 198–233, 238–281, 286–439, 447–490, 494–508, and 517–895 were built. Initial phase of the native CA1049 structure was solved by molecular replacement using Molrep [21] and the SeMet-labeled CA789 structure as a template. Molecular graphic images were prepared using PyMOL (Schrödinger, LLC, New York, NY, USA). A structural similarity search was performed using the Dali server (<http://ekhidna2.biocenter.helsinki.fi/dali/>) [22]. Sequence conservation mapping was performed using the ConSurf server [23].

Construction of site-directed mutants and enzyme assay

Site-directed mutants were constructed using PrimeSTAR Max DNA Polymerase (Takara Bio Inc., Shiga, Japan). The following primers and their complementary primers were used: 5′ – TCCATCgcgGGTGTGCTCGGCTACAAC–3′ for E383A, 5′ – TCCATCcagGGTGTGCTCGGCTACAAC–3′ for E383Q, 5′ – GGTAACgcgGCCGACGCCGTTTCCTTC–3′ for D515A, 5′ – GGTAACaacGCCGACGCCGTTTCCTTC–3′ for D515N, 5′ – CAAAACgcgTTCTGGAACGAGGACAAC–3′ for E713A, and 5′ – CAAAACcagTTCTGGAACGAGGACAAC–3′ for E713Q (mutated sites shown with lowercase letters).

The activity assay was performed using *cis*-Ara₃-Hyp-DNS as the substrate. A reaction mixture containing 50 μ M Ara₃-Hyp-DNS, 5 mM Na-acetate buffer (pH 5.5), and HypBA2 protein was incubated at 37°C for 1 h. Two μ M of the reaction mixture was spotted onto a thin layer chromatography (TLC) Silica Gel 60 F₂₅₄ plate (Merck, Darmstadt, Germany). Subsequently, the TLC plate was developed by a 2:1:1 solvent mixture (v/v/v) of ethyl acetate/acetic acid/water. Fluorescence of the dansyl (DNS) group was detected under a UV lamp (365 nm).

Sequence analysis and molecular phylogeny

269 GH121 sequences were extracted from the CAZy database (January 6th, 2020) and preliminarily aligned using MAFFT [24]. The multiple sequence alignment was inspected in Jalview [25], and all sequences were cut according to the observed boundaries of the conserved GH121 module. These modules were then realigned with MAFFT G-INS-i (iterative refinement using pairwise Needleman–Wunsch global alignments). A maximum likelihood phylogenetic tree was estimated with RAxML [26] using 100 bootstrap replicates and visualized with iTOL [27]. Short and high bootstrap-supported branches in the phylogenetic tree were collapsed, and major features and number of sequences were displayed.

Results

Identification and structure determination of the catalytic domain

HypBA2 is a multi-domain protein with 1943 amino acids (aa) (Fig 2A). Presence of the N-terminal signal sequence and the C-terminal transmembrane region indicate that this enzyme is anchored to outside of the cell of the gram-positive bacterium (*B. longum*). In our previous study, the N-terminal half of BLLJ_0212 was identified as the conserved region of GH121, and a deletion analysis indicated that a region of residues 33–1051 showed full activity in the presence of 1 mM Ca²⁺ [6]. Protein purification of HypBA2 deletants was conducted in the presence of 2 mM CaCl₂ throughout the steps (see Materials and Methods). We tried to crystallize various deletion constructs of HypBA2 and succeeded in solving the crystallographic structure using a SeMet-derivative of a construct of residues 33–1154 (CA789) by the single-wavelength anomalous dispersion method (Fig 2A and Table 1). Automated model building and subsequent manual model building resulted in a SeMet-containing protein model with many disordered regions, and we could not build a reliable polypeptide model in the C-terminal region

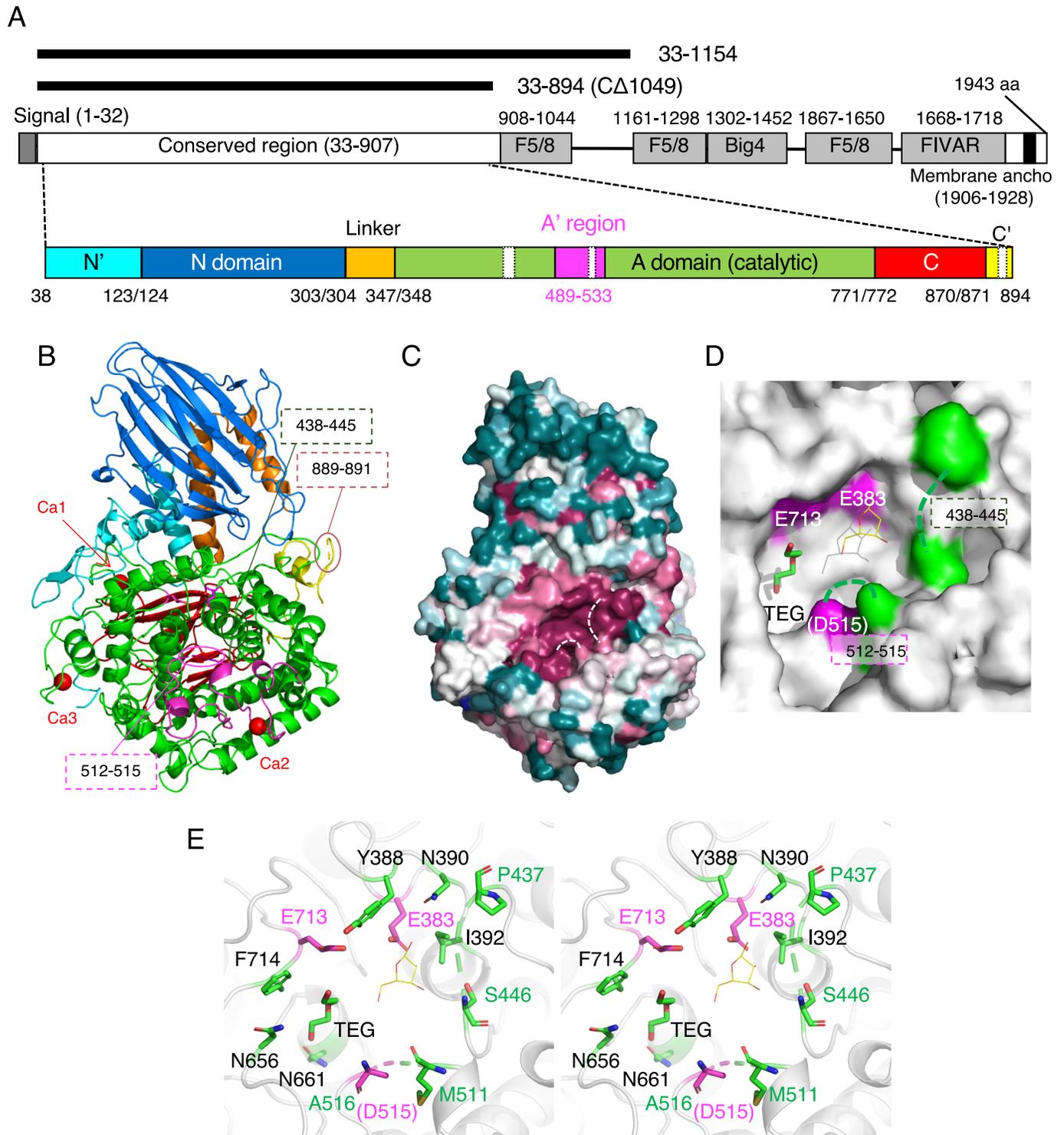


Fig 2. Structural architecture of HypBA2. (A) Schematic representations of the full-length domain structure of HypBA2 are shown in the middle. Regions and domains modeled in the crystal structure are schematically shown below. Deletion constructs used for crystallography are shown above. (B) The overall structure of the C Δ 1049 construct (residues 33–894) of HypBA2. (C) Sequence conservation mapping on the molecular surface. Amino acid sequence conservation among GH121 is colored with red (high), white (middle), and blue (low). (D) Molecular surface representation of the active site pocket in the catalytic domain. Glu383, Asp515, and Glu713 are conserved putative catalytic residues of GH121 (shown in magenta). Because Asp515 is disordered, the neighboring residue (Ala516) is shown. A triethylene glycol molecule (TEG) bound to HypBA2 and a β -L-Araf molecule bound to GH142 BT_1020 are shown as green sticks and thin yellow lines, respectively. (E) Stereoview of the active site of HypBA2. (A–E) Disordered regions are indicated with dotted lines.

<https://doi.org/10.1371/journal.pone.0231513.g002>

Table 1. Data collection and refinement statistics of the crystallography of HypBA2.

Data set	SeMet protein (33–1154)	Native protein (33–894)
Data collection ^a		
Beamline	NW12A	BL1A
Wavelength (Å)	0.9791	1.1000
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell (Å)	$a = 76.10, b = 100.60, c = 141.16$	$a = 81.53, b = 88.62, c = 127.92$
Resolution (Å)	50.0–2.20	44.31–1.85
	(2.24–2.20)	(1.89–1.85)
Total reflections	1,641,170	542,053
Unique reflections	106,306 (5,342)	79,741 (4,473)
Completeness (%)	100.0 (100.0)	100.0 (100.0)
Multiplicity	29.4 (29.0)	6.8 (7.0)
Mean $I/\sigma(I)$ ^b	28.6 (4.4)	9.1 (1.3)
R_{merge}	0.119 (0.667)	0.123 (1.405)
$CC_{1/2}$	0.981 (0.928)	0.940 (0.634)
Anomalous completeness (%)	100.0 (100.0)	
Anomalous multiplicity	15.4 (15.0)	
Refinement		
Resolution (Å)		44.35–1.85
No. of reflections		75,729
$R_{\text{work}}/R_{\text{free}}$		0.158/0.185
Number of atoms		7,337
RMSD from ideal values		
Bond lengths (Å)		0.011
Bond angles (°)		1.633
Ramachandran plot (%)		
Favored/allowed/outlier		97.4/2.5/0.1
PDB code		6M5A

^aValues in parentheses are for the highest resolution shell.

<https://doi.org/10.1371/journal.pone.0231513.t001>

(residues 896–1154) due to ambiguous electron density (See [Materials and Methods](#)). In our previous study, a construct of residues 33–894 (C Δ 1049, see [Fig 2A](#)) showed about 16% activity compared to the full-length enzyme [6]. Our re-analysis using the purified samples of the C Δ 789 and C Δ 1049 deletants demonstrated that both of them exhibited the catalytic activity of cleaving Ara₃-Hyp-DNS to release Ara-Hyp-DNS ([Fig 3](#), left). Therefore, we also crystallized the native protein of C Δ 1049. The purified protein samples of C Δ 789 and C Δ 1049 migrated as a single band on SDS-PAGE that was consistent with their calculated molecular masses (124,555 and 96,854 Da, respectively), and gel filtration experiments suggested that the recombinant proteins are both monomeric in solution ([S1 Fig](#)). The crystal structure of the native (unlabeled) C Δ 1049 protein was determined at 1.85 Å resolution ([Table 1](#)). The refined crystal structure of C Δ 1049, which contains the core catalytic region of GH121, comprises residues 38–894, and three short regions (residues 438–445, 512–515, and 889–891) were disordered ([Fig 2A](#)). Two residues from a C-terminal His₆-tag (LE of LEHHHHHH) were visible in the crystal structure and modeled as residues 895 and 896. Extensive attempts at co-crystallization and soaking experiments with β -Ara₂ (the reaction product) were undertaken, but no electron densities were found in the putative active site.

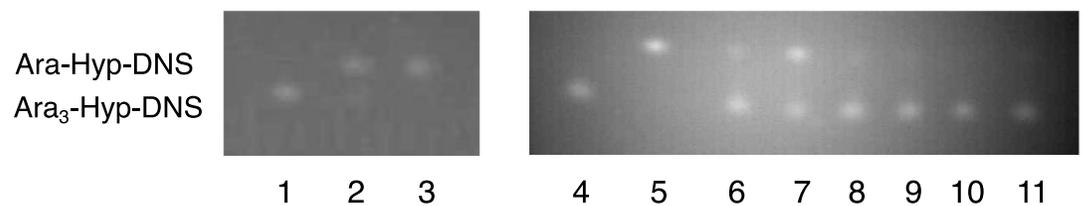


Fig 3. TLC analysis of the activities of deletion constructs and site-directed mutants towards Ara₃-Hyp-DNS. The reaction was performed in a solution containing 50 μ M Ara₃-Hyp-DNS (substrate) and 50 mM Na-acetate buffer (pH 5.5). After development, the fluorescence of the dansyl group was detected. (Left) The substrate was incubated either without (lane 1) or with the deletion constructs (lanes 2 and 3, 0.45 mg/ml protein) at 30°C for 20 min. Lane 2, C Δ 1049; lane 3, C Δ 789. (Right) The substrate was incubated either without (lane 4), with C Δ 1049 (lane 5, 2.5 mg/ml protein), or with site-directed point mutants of C Δ 1049 (lanes 6–11, 2.5 mg/ml protein) at 37°C for 1 h. Lane 6, D515A; lane 7, D515N; lane 8, E713A; lane 9, E713Q; lane 10, E383A; lane 11, E383Q.

<https://doi.org/10.1371/journal.pone.0231513.g003>

Crystal structure of GH121 HypBA2

The overall structure of C Δ 1049 consists of a mostly unstructured N-terminal region (residues 38–123, cyan), a N-domain (124–303, marine), a linker region with two α -helices (304–347, orange), a catalytic (α/α)₆ barrel domain (348–771, green), a C-domain (772–870, red), and a mostly unstructured C-terminal region with an α -helix (871–894, yellow; Fig 2B). The N- and C-domains adopt a β -sandwich fold. The N-domain is similar to an existing CATH domain (Superfamily 2.70.98.50 with Dali Z score \sim 11) [28] while the C-domain shows no similarity to established protein folds (Dali Z score $<$ 4). A Dali structural similarity search analysis of the whole structure showed that the catalytic region of GH121 is primarily similar to the GH142 protein (BT_1020, Table 2). Subsequent hits were GH enzymes belonging to GH63,

Table 2. Result of structural similarity search using the Dali server.

PDB ID ^a	Z score	RMSD (Å)	LALI ^b	Identity (%)	Protein name ^c	Organism	Activity	Family	Reference ^d
5mqr (A)	31.4	2.6	459	19	BT_1020	<i>Bacteroides thetaiotaomicron</i>	β -L-Arabinofuranosidase	GH142	[8]
5ca4 (A)	21.4	4.6	547	11	YgjK	<i>Escherichia coli</i>	α -Glycosidase	GH63	[30]
2z07 (A)	20.4	2.9	315	9	TTHA0978	<i>Thermus thermophilus</i>	(Uncharacterized)	GH63	TBP
3cih (A)	20.3	4.3	441	10	BT_1001	<i>Bacteroides thetaiotaomicron</i>	α -L-Rhamnosidase	GH78	[8]
3w5m (A)	19.4	3.8	427	13	SaRha78A	<i>Streptomyces avermitilis</i>	α -L-Rhamnosidase	GH78	[31]
1v7v (A)	19.4	4.2	581	9	ChBP	<i>Vibrio proteolyticus</i>	Chitobiose phosphorylase	GH94	[32]
4wvb (A)	19.0	2.9	315	9	Tt8MGH	<i>Thermus thermophilus</i>	Mannosylglycerate hydrolase	GH63	[33]
6gsz (A)	19.0	3.5	395	14	α -Rha	<i>Aspergillus terreus</i>	α -L-Rhamnosidase	GH78	[34]
2jjb (B)	19.0	3.4	341	12	Tre37A	<i>Escherichia coli</i>	α,α -Trehalase	GH37	[35]
5ohz (C)	18.7	3.7	331	8	MhGgH	<i>Mycobacterium hassiacum</i>	Glucosylglycerate hydrolase	GH63	[36]
4zlf (A)	18.1	4.5	469	10	CBAP	<i>Saccharophagus degradans</i>	Cellobionic acid phosphorylase	GH94	[37]
3qde (A)	18.1	4.2	585	10	CtCBP	<i>Clostridium thermocellum</i>	Cellobiose phosphorylase	GH94	[38]
3afj (A)	18.1	4.2	584	11	CgCBP	<i>Cellvibrio gilvus</i>	Cellobiose phosphorylase	GH94	[39]
4j5t (A)	18.0	3.7	444	9	GluI	<i>Saccharomyces cerevisiae</i>	Processing α -glucosidase I	GH63	[40]

The whole HypBA2 structure including all domains were used.

^aChain ID is shown in parentheses.

^bNumber of aligned residues.

^cTTHA0978 and Tt8MGH are the identical protein.

^dTBP: to be published.

<https://doi.org/10.1371/journal.pone.0231513.t002>

GH78, GH94, and GH37, and all of them adopt a catalytic $(\alpha/\alpha)_6$ barrel domain (discussed below). A structural similarity search using only the barrel domain showed that the closest structural homologs are the GH142, GH63, and GH37 enzymes (S1 Table). Interestingly, all of the structurally similar GH families are inverting enzymes, whereas GH121 HypBA2 is a retaining enzyme because it exhibited transglycosylation activity [6]. The catalytic barrel region of a structural homolog in GH63 (α -glycosidase YgjK from *E. coli*) is designated as the “A-domain,” and an extra structural unit within this domain is designated as the “A'-region” [29]. The A'-region is a long insertion between the fifth and sixth helices in the twelve helices of the $(\alpha/\alpha)_6$ barrel scaffold. In HypBA2, a similar insertion corresponding to the A'-region is also present (magenta in Fig 2A and 2B, residues 489–533).

The crystal structure of C Δ 1049 contains three calcium ions (Fig 2B, red spheres; and S2 Fig). The first Ca²⁺ (Ca1) is located between the linker region and the catalytic domain and coordinated by the side chains of Asp345, Asp349, and Asp352, the main chain carbonyl of Thr346, and two water molecules. The second Ca²⁺ (Ca2) is located in the A'-region and coordinated by the side chains of Asn497, Asn499, Asn501, Asp505, and Asp532, and the main chain carbonyl of Leu503. The third Ca²⁺ (Ca3) is located in a C-terminal region of the catalytic domain and coordinated by the side chains of Asp724, Asn726, Asp728, and Asp735, the main chain carbonyl of Val730, and one water molecule. The coordination distances range from 2.3 to 2.5 Å. The CheckMyMetal server indicated that the coordination distances and geometry of Ca1-3 are typical of calcium binding sites of proteins [41].

Active site structure and site-directed mutant analysis

The degree of amino acid sequence conservation of GH121 members mapped on the molecular surface of HypBA2 clearly illustrated highly conserved residues in a pocket of the catalytic domain (Fig 2C, red). Three conserved acidic residues (Glu383, Asp515, and Glu713) of GH121 (discussed below) and two short disordered regions (438–445 and 512–515) are located in the pocket (Fig 2D). Because one of the three conserved acidic residues (Asp515) was disordered in the crystal structure, the neighboring residue (Ala516) is shown with a magenta color in the figures. A triethylene glycol (TEG) molecule, which was derived from the crystallization precipitant, was bound in the pocket (a green stick in Fig 2D). A superimposed β -L-Araf molecule bound to GH142 BT_1020 is shown as thin yellow lines (discussed below), illustrating a possible subsite -1 area in the pocket. Glu383 and Glu713 are located near the superimposed β -L-Araf molecule (Fig 2E). Ala516 is located in a relatively far position, but Asp515 can approach the catalytic area with the flexibility of the disordered region.

Then, we constructed point mutants of C Δ 1049 by site-directed mutagenesis. As shown in Fig 3 (right), mutations of Glu383 and Glu713 (E383A, E383Q, E713A, and E317Q) completely abolished the activity. On the other hand, mutants of Asp515 showed very weak activity, and substitution with Asn (D515N) appeared to exhibit higher activity than an Ala mutant (D515A).

Sequence analysis of GH121

We conducted a protein sequence analysis of the 272 GH121 entries listed on the CAZy database in January 2020. After the deletion of 3 duplicates, 269 entries were analyzed by MAFFT multiple sequence alignment, and 56 of them are shown in S3 Fig. A phylogenetic tree (cladogram) indicated that HypBA2 is in a cluster of 22 bifidobacterial sequences (Fig 4A). 186 sequences from *Xanthomonas* species formed a very large group with little or no sequence diversity, 175 of them being in a zero branch between them. A HypBA2 homolog from *Xanthomonas euvesicatoria* (XeHypBA2) [42] was also included in this cluster. The full amino

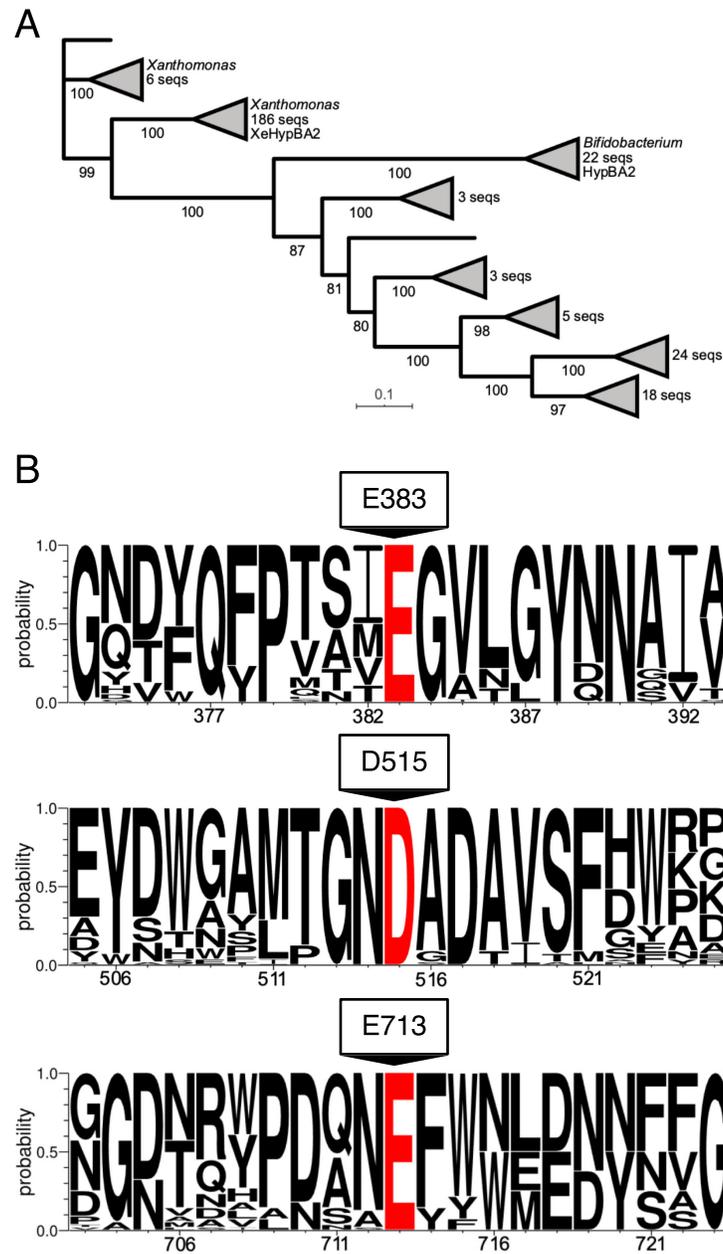


Fig 4. A phylogenetic tree (A) and sequence conservation around the selected residues (B) of GH121. (A) The maximum likelihood phylogenetic tree was generated using 269 GH121 catalytic domain sequences. Bootstrap values based on 100 replicates are shown. Short and high bootstrap support branches are collapsed as triangles. (B) Sequence logo around Glu383, Asp515, and Glu713 indicated that these residues are completely conserved in GH121.

<https://doi.org/10.1371/journal.pone.0231513.g004>

acid sequence alignment revealed that the amino acid residues at the Ca2 site are essentially conserved in GH121, whereas those at Ca1 and Ca3 are not conserved (S3 Fig). Sequence logos around the three acidic residues (E383, D515, and E713, Fig 4B) illustrated that all of them are completely conserved in GH121, and nearby sequences are also highly conserved. Therefore, the structural feature of the active site pocket, the site-directed mutagenesis

analysis, and the comprehensive sequence analysis strongly suggested that the three acidic residues play important roles in the catalytic process and/or substrate binding of HypBA2 and GH121 enzymes.

Structural comparison with other GH families

Then, we examined possible catalytic residues of GH121 by comparing its structure with structural homologs. The overall structure of HypBA2 (Fig 5A) was compared with representative structures of GH63 (Fig 5B, α -glycosidase YgjK from *E. coli*) [30], GH94 (Fig 5C, chitobiose

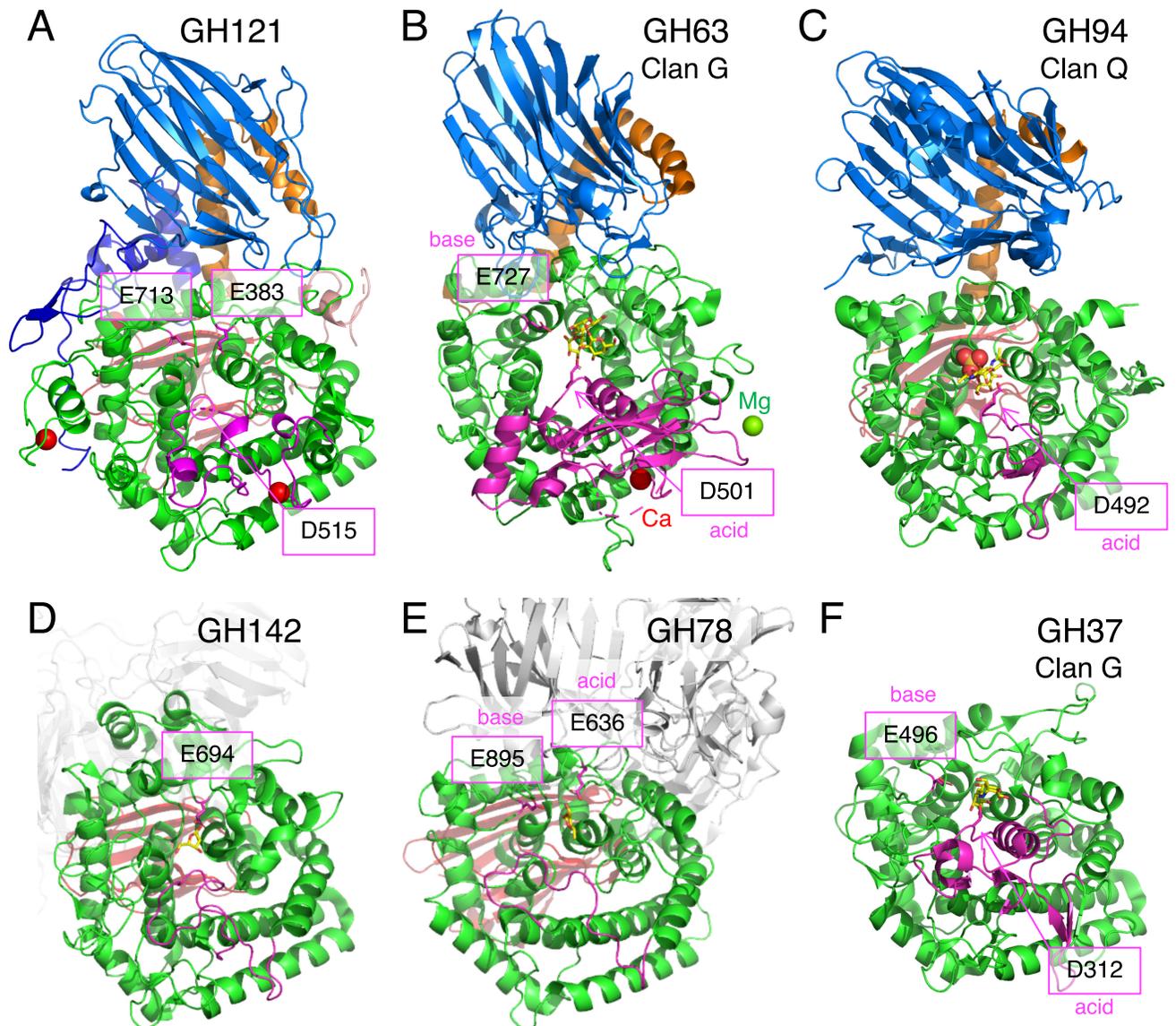


Fig 5. Structural comparison of HypBA2 with structural homologs belonging to other GH families. Overall structures of (A) GH121 HypBA2, (B) GH63 α -glycosidase YgjK from *E. coli* (PDB ID 5CA3), (C) GH94 chitobiose phosphorylase ChBP from *V. proteolyticus* (1V7X), (D) GH142 β -L-arabinofuranosidase BT_1020 from *B. thetaiotaomicron* (5MQS), (E) GH78 α -L-rhamnosidase SaRha78A from *S. avermitilis* (3W5N), and (F) GH37 α,α -trehalase Tre37A from *E. coli* (2JF4) are shown. Catalytic residues are indicated with a magenta color. The A'-region (insertion between the fifth and sixth helices of the barrel) is also shown with a magenta color.

<https://doi.org/10.1371/journal.pone.0231513.g005>

phosphorylase ChBP from *Vibrio proteolyticus*) [32], GH142 (Fig 5D, β -L-arabinofuranosidase BT_1020 from *B. thetaiotaomicron*) [8], GH78 (Fig 5E, α -L-rhamnosidase SaRha78A from *Streptomyces avermitilis*) [31], and GH37 (Fig 5F, α,α -trehalase Tre37A from *E. coli*) [43]. The GH63 and GH94 enzymes share multi-domain structures containing the N-domain (marine in Fig 5), the linker region (orange), and the catalytic barrel domain (green) with GH121. GH94, GH142, and GH78 have a β -sandwich domain similar to the C-domain (red) behind the barrel domain. It is noteworthy that GH63 has a calcium ion (Fig 5B, red sphere) at a similar position to the conserved Ca²⁺ site of HypBA2. GH37 only has a catalytic barrel domain as a structurally conserved element. The catalytic general base and acid residues for the anomer-inverting glycosidic bond hydrolysis in GH63, GH78, and GH37 have been identified in previous reports [30,31,43]. The general base residue of all of these families is located at the position corresponding to Glu713 in HypBA2 (Fig 5B, 5E and 5F). The general acid residue of GH63 and GH37 is located in the A'-region (magenta), corresponding to the position of Asp515 in HypBA2, while the general acid residue of GH78 appears to be located at a position corresponding to Glu383 in HypBA2. GH94 enzymes are inverting phosphorylases that use an inorganic phosphate as a nucleophile [32,39]. A sulfate molecule (phosphate analog) and the general acid residue of GH94 are located at positions corresponding to Glu713 and Asp515 in HypBA2, respectively (Fig 5C). The catalytic residue of GH142 was not identified, but a conserved residue (Glu694) near the β -L-Araf molecule in the putative catalytic pocket was identified (Fig 5D) [32].

Fig 6A shows a superimposition of GH142 β -L-arabinofuranosidase BT_1020 (thick sticks) and HypBA2 (thin sticks) at the active site. Glu383 in HypBA2 is overlaid well with the putative catalytic residue of BT_1020 (Glu694, red sticks). Despite the ambiguous electron density and low crystallographic resolution (3.0 Å) [8], we consider that the position of the β -L-Araf molecule indicates the approximate location of the subsite -1 in BT_1020. Glu694 forms bidentate hydrogen bonds with the O1 and O2 hydroxyls of the β -L-Araf molecule. Tyr702, Gly805, Trp943, and His1019 are also located near the ligand, but they are not conserved in GH121 at all. As shown in Fig 6B, the catalytic base and acid residues of GH63 enzyme (YgkK) correspond to Glu713 and Asp515 in HypBA2. The position of Asp501 (general acid) in YgkK is located above the subsite -1 sugar as an “anti” protonator according to the definition by Nerinckx et al. [44]. The catalytic base residue in GH78 (Glu895 in SaRha78A) also corresponds to Glu713 in HypBA2 (Fig 6C). However, the location of the general acid residue of GH78 (Glu636) does not correctly match the position of Glu383 in HypBA2 but rather is located significantly “above” it. It is noteworthy that the general acid residues of GH63, GH78, and GH37 are all “anti” positioned (see the “syn/anti lateral protonation” page in CAZypedia, <https://www.cazypedia.org>) [45]. In comparison with GH142, GH63, and GH78 enzymes, Glu383 of GH121 appears to form direct interactions with the subsite -1 sugar. This residue may have a role of ‘anchor’ that ensures nucleophilic and acid/base catalysis on the glycosidic bond by proper positioning of the substrate.

Discussion

HypBA2 is a multi-domain protein, and we revealed the three-dimensional structure of the N-terminal catalytic region in this study (Fig 2A). The C-terminal half of this protein consists of F5/8 type C and Big4 domains, and a FIVAR-transmembrane region. HypBA2 is suggested to be a membrane-anchored extracellular enzyme because the FIVAR-transmembrane region is usually involved in association with the bacterial cell wall. Similar molecular architectures were found in other bifidobacterial enzymes involved in human milk oligosaccharides and mucin degradation systems [46–48]. The F5/8 type C domain (Pfam 00754) is referred to as

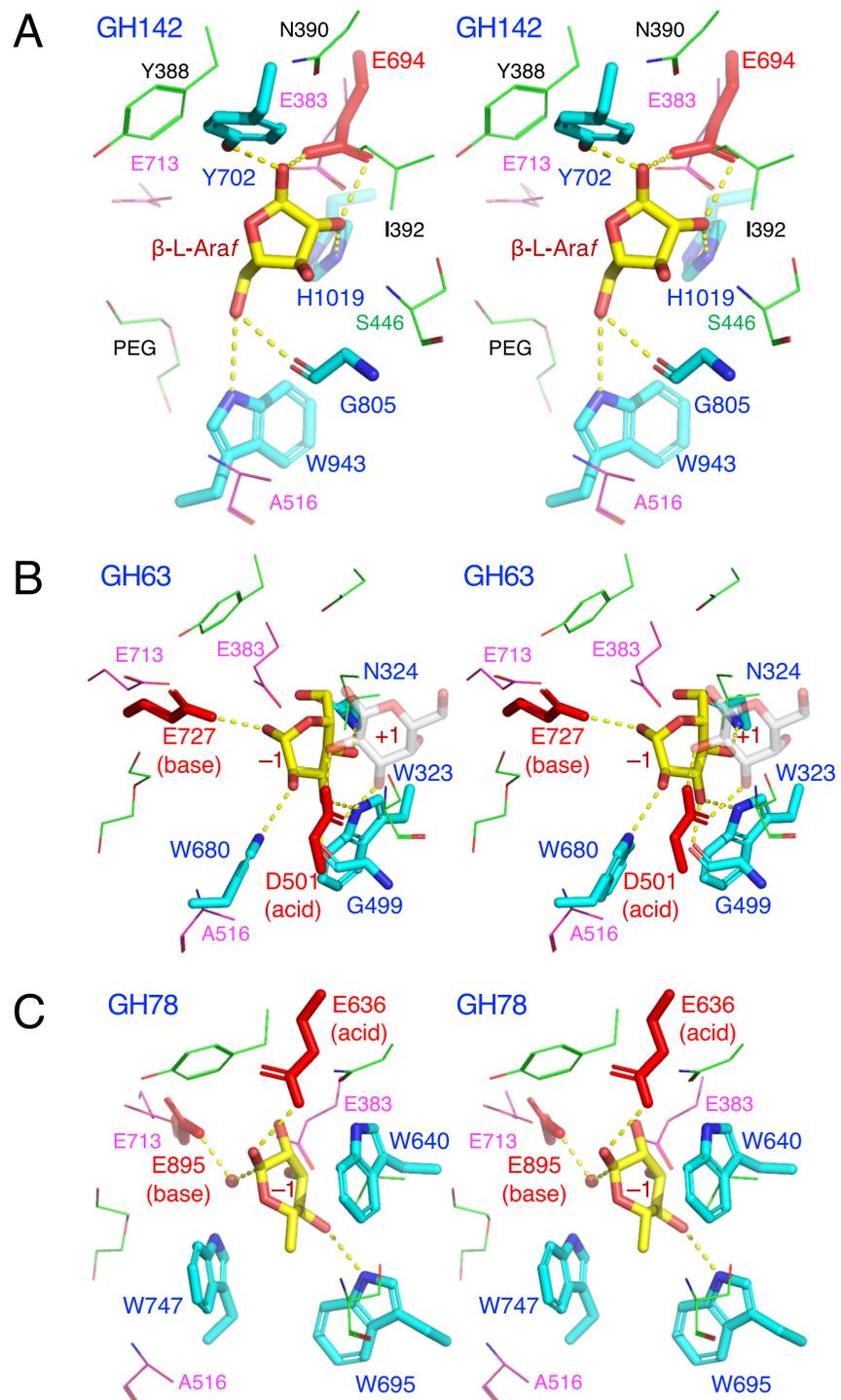


Fig 6. Stereoview of the active site of structural homologs belonging to other GH families. (A) GH142 β -L-arabinofuranosidase BT_1020 (cyan) complexed with β -L-Araf (yellow), (B) GH63 α -glycosidase YgjK (cyan) complexed with glucose (yellow) and lactose (galactose at subsite +1 is shown as transparent grey sticks), and (C) GH78 α -L-rhamnosidase SaRha78A (cyan) complexed with α -L-rhamnose (yellow). The catalytic residues of BT_1020, YgjK, and SaRha78A are shown as red sticks. Residues in the active site of HypBA2 are superimposed as magenta (putative catalytic residues) or thin green lines.

<https://doi.org/10.1371/journal.pone.0231513.g006>

carbohydrate-binding module (CBM) family 32 that is generally involved in galactose binding, and several CBM32 members also bind calcium [49,50]. Therefore, the significantly reduced and calcium-dependent activity of the C-terminally deleted construct of HypBA2 suggest that the three F5/8 type C domains may have substrate-binding and stabilizing functions [6].

Based on structural comparisons, we suggest that Glu713 and Asp515 are the likely catalytic nucleophile and acid/base residues of HypBA2, respectively. This hypothesis assumes a proper positioning of the putative acid/base residue (Asp515) on substrate binding. The catalytic acid/base residue of GH29 1,3-1,4- α -L-fucosidase AfcB from *B. longum* subsp. *infantis* suggestively exhibited a significant movement on the substrate binding due to the flexibility of a loop carrying that residue [51]. However, this hypothesis is solely based on the structural comparison with other inverting GH families and a simple activity assay of the site-directed mutants. Further analyses, such as determination of ligand complex structures and detailed kinetic measurements, will identify the catalytic residues.

HypBA2 is an exo-type glycoside hydrolase that releases a disaccharide (β -Ara₂). This mode of action is similar to other extracellular glycosidases of bifidobacteria involved in the degradation of human milk oligosaccharides and mucin O-glycans, *vis.* GH20 and GH136 lacto-*N*-biosidases [47,48] and GH101 endo- α -*N*-acetylgalactosaminidase [46]. The molecular surface representation of HypBA2 delineates a pocket that can accommodate the β -Ara₂ disaccharide (Fig 2D). The TEG molecule in the pocket appears to indicate the location of subsite -2. The CASTp server [52] calculated that the surface-accessible volume and area of the pocket are 960 Å³ and 685 Å², respectively. The pocket size may be overestimated due to the two disordered regions. Given the hypothesis that one of the catalytic residues (Asp515) is located in the shorter disordered region (residues 512–515), the longer disordered region (residues 438–445) may cover and cap the active site pocket on substrate binding as in the case of GH127 HypBA1 [10]. Interestingly, Pro437, Gly438, and Trp443 are completely conserved in GH121 members (S3 Fig), suggesting loop flexibility near the Gly residue and a sugar-aromatic stacking interaction by the Trp side chain with the substrate.

In this study, we presented the first three-dimensional view of a GH121 enzyme and demonstrated that it has some structural relationship with several GH families, including GH63 and GH142. Since we could not find the closest structural and functional relative of GH121 among the GHs, the molecular evolution of GH121 enzymes are still enigmatic. Interestingly, most of the ~290 GH121 members are found only in bifidobacteria (animal gut symbionts, 22 sequences) and *Xanthomonas* species (plant pathogens, ~190 sequences, Fig 4A). In contrast, GH127 and GH146 β -L-arabinofuranosidases (monosaccharide-releasing exo-enzymes), which were classified together as DUF1680 in Pfam [53], are widely distributed among various bacteria [7]. Bifidobacteria and *Xanthomonas* species apparently have no biological relationships with distinct biological niches. Bifidobacteria are representative species of gut microbes that are beneficial to human health while *Xanthomonas* species contain pathogens that cause bacterial spots on a wide variety of plant species. The genes for β -L-arabinooligosaccharide-degradation enzymes in *X. euvesicatoria* (*xehypBA1-BA2-AA*) are not involved in either pathogenicity or non-host resistance reactions [42]. However, the distribution of GH121 genes suggests that bifidobacteria and *Xanthomonas* species possibly benefit from the disaccharide-releasing enzyme and the disaccharide-specific importer system on their biological niches in order to live on special types of plant glycans, such as Ara₃-Hyp β -L-arabinooligosaccharides on HRGPs.

Supporting information

S1 Fig. Gel filtration chromatogram and SDS-PAGE of CA789 and CA1049. (A) Chromatogram of molecular mass markers; thyroglobulin (670 kDa), γ -globulin (158 kDa), ovalbumin

(44 kDa), myoglobin (17 kDa), and vitamin B₁₂ (1.35 kDa). (B) Chromatogram (left) and SDS-PAGE (right) of SeMet-labeled C Δ 789. (C) Chromatogram (left) and SDS-PAGE (right) of native C Δ 1049. The monomeric fraction (peak 2) was used for crystallization.

(DOCX)

S2 Fig. Calcium binding sites. Distances (\AA) of metal-coordination and water-mediated hydrogen bonds at Ca1 (A), Ca2 (B), and Ca3 (C) sites in the C Δ 1049 structure are shown.

(DOCX)

S3 Fig. Full amino acid sequence alignment of GH121 entries. From 272 entries of GH121, 56 were selected. HypBA2 is shown at the top (BAJ34647.1). Residues involved in the calcium coordination with their side chain and main chain atoms are indicated by inverted and white-back characters, respectively. Conserved Pro437, Gly438, and Trp443 residues are also indicated.

(DOCX)

S1 Table. Result of structural similarity search using the barrel domain.

(DOCX)

S1 Data.

(PDF)

Acknowledgments

We thank Dr. Akihiro Ishiwata for the helpful discussions, Mr. Masayuki Miyake for the preliminary crystallization screenings, and the staff of SPring-8 and KEK-PF for the X-ray data collection.

Author Contributions

Conceptualization: Kiyotaka Fujita, Shinya Fushinobu.

Data curation: Takatoshi Arakawa, Kiyotaka Fujita, Shinya Fushinobu.

Formal analysis: Alexander Holm Viborg.

Funding acquisition: Takatoshi Arakawa, Chihaya Yamada, Kiyotaka Fujita, Shinya Fushinobu.

Investigation: Keita Saito, Alexander Holm Viborg, Shiho Sakamoto, Takatoshi Arakawa.

Project administration: Kiyotaka Fujita, Shinya Fushinobu.

Resources: Takatoshi Arakawa, Chihaya Yamada.

Supervision: Kiyotaka Fujita, Shinya Fushinobu.

Validation: Takatoshi Arakawa, Kiyotaka Fujita, Shinya Fushinobu.

Writing – original draft: Alexander Holm Viborg, Shinya Fushinobu.

Writing – review & editing: Keita Saito, Alexander Holm Viborg, Shiho Sakamoto, Takatoshi Arakawa, Chihaya Yamada, Kiyotaka Fujita, Shinya Fushinobu.

References

1. Shallom D, Shoham Y. Microbial hemicellulases. *Current Opinion in Microbiology*. 2003. pp. 219–228. [https://doi.org/10.1016/S1369-5274\(03\)00056-0](https://doi.org/10.1016/S1369-5274(03)00056-0)

2. Garron ML, Henrissat B. The continuing expansion of CAZymes and their families. *Current Opinion in Chemical Biology*. 2019. pp. 82–87. <https://doi.org/10.1016/j.cbpa.2019.08.004> PMID: 31550558
3. Kieliszewski MJ, Lamport DTA. Extensin: repetitive motifs, functional sites, post-translational codes, and phylogeny. *The Plant Journal*. 1994. pp. 157–172. <https://doi.org/10.1046/j.1365-313X.1994.05020157.x> PMID: 8148875
4. Kieliszewski MJ, Showalter AM, Leykam JF. Potato lectin: a modular protein sharing sequence similarities with the extensin family, the hevein lectin family, and snake venom disintegrins (platelet aggregation inhibitors). *Plant J*. 1994; 5: 849–861. <https://doi.org/10.1046/j.1365-313X.1994.5060849.x> PMID: 8054990
5. Matsubayashi Y. Posttranslationally modified small-peptide signals in plants. *Annu Rev Plant Biol*. 2014; 65: 385–413. <https://doi.org/10.1146/annurev-arplant-050312-120122> PMID: 24779997
6. Fujita K, Sakamoto S, Ono Y, Wakao M, Suda Y, Kitahara K, et al. Molecular cloning and characterization of a β -L-arabinobiosidase in *Bifidobacterium longum* that belongs to a novel glycoside hydrolase family. *J Biol Chem*. 2011; 286: 5143–5150. <https://doi.org/10.1074/jbc.M110.190512> PMID: 21149454
7. Fujita K, Takashi Y, Obuchi E, Kitahara K, Suganuma T. Characterization of a novel β -L-arabinofuranosidase in *Bifidobacterium longum*: Functional elucidation of a DUF1680 protein family member. *J Biol Chem*. 2014; 289: 5240–5249. <https://doi.org/10.1074/jbc.M113.528711> PMID: 24385433
8. Ndeh D, Rogowski A, Cartmell A, Luis AS, Baslé A, Gray J, et al. Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature*. 2017; 544: 65–70. <https://doi.org/10.1038/nature21725> PMID: 28329766
9. Luis AS, Briggs J, Zhang X, Farnell B, Ndeh D, Labourel A, et al. Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic Bacteroides. *Nat Microbiol*. 2018; 3: 210–219. <https://doi.org/10.1038/s41564-017-0079-1> PMID: 29255254
10. Ito T, Saikawa K, Kim S, Fujita K, Ishiwata A, Kaeohip S, et al. Crystal structure of glycoside hydrolase family 127 β -L-arabinofuranosidase from *Bifidobacterium longum*. *Biochem Biophys Res Commun*. 2014; 447: 32–37. <https://doi.org/10.1016/j.bbrc.2014.03.096> PMID: 24680821
11. Sakanaka M, Gotoh A, Yoshida K, Odamaki T, Koguchi H, Xiao JZ, 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*. 2020. <https://doi.org/10.3390/nu12010071> PMID: 31888048
12. Lamport DTA, Miller DH. Hydroxyproline arabinosides in the plant kingdom. *Plant Physiol*. 1971; 48: 454–456. <https://doi.org/10.1104/pp.48.4.454> PMID: 16657818
13. Akiyama Y, Mori M, Kato K. ^{13}C -NMR analysis of hydroxyproline arabinosides from *Nicotiana tabacum*. *Agric Biol Chem*. 1980; 44: 2487–2489. <https://doi.org/10.1271/bbb1961.44.2487>
14. Ashford D, Desai NN, Allen AK, Neuberger A, O'Neill MA, Selvendran RR. Structural studies of the carbohydrate moieties of lectins from potato (*Solanum tuberosum*) tubers and thorn-apple (*Datura stramonium*) seeds. *Biochem J*. 1982; 201: 199–208. <https://doi.org/10.1042/bj2010199> PMID: 7082284
15. Miyake M, Terada T, Shimokawa M, Sugimoto N, Arakawa T, Shimizu K, et al. Structural analysis of β -L-arabinobiose-binding protein in the metabolic pathway of hydroxyproline-rich glycoproteins in *Bifidobacterium longum*. *FEBS J*. 2020; in press. <https://doi.org/10.1111/febs.15315> PMID: 32246585
16. LeMaster DM, Richards FM. ^1H - ^{15}N heteronuclear NMR studies of *Escherichia coli* thioredoxin in samples isotopically labeled by residue type. *Biochemistry*. 1985; 24: 7263–7268. <https://doi.org/10.1021/bi00346a036> PMID: 3910099
17. Kabsch W. XDS. *Acta Crystallogr Sect D Biol Crystallogr*. 2010; 66: 125–132. <https://doi.org/10.1107/S0907444909047337> PMID: 20124692
18. Adams PD, Afonine P V., Bunkóczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr Sect D Biol Crystallogr*. 2010; 66: 213–221. <https://doi.org/10.1107/S0907444909052925> PMID: 20124702
19. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr Sect D Biol Crystallogr*. 2010; 66: 486–501. <https://doi.org/10.1107/S0907444910007493> PMID: 20383002
20. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr Sect D Biol Crystallogr*. 2011; 67: 355–367. <https://doi.org/10.1107/S0907444911001314> PMID: 21460454
21. Vagin A, Teplyakov A. Molecular replacement with MOLREP. *Acta Crystallogr Sect D Biol Crystallogr*. 2010; 66: 22–25. <https://doi.org/10.1107/S0907444909042589> PMID: 20057045
22. Holm L, Laakso LM. Dali server update. *Nucleic Acids Res*. 2016; 44: W351–W355. <https://doi.org/10.1093/nar/gkw357> PMID: 27131377

23. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* 2016; 44: W344–W350. <https://doi.org/10.1093/nar/gkw408> PMID: 27166375
24. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol.* 2013; 30: 772–780. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690
25. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics.* 2009; 25: 1189–1191. <https://doi.org/10.1093/bioinformatics/btp033> PMID: 19151095
26. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014; 30: 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033> PMID: 24451623
27. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 2016; 44: W242–W245. <https://doi.org/10.1093/nar/gkw290> PMID: 27095192
28. Dawson NL, Lewis TE, Das S, Lees JG, Lee D, Ashford P, et al. CATH: An expanded resource to predict protein function through structure and sequence. *Nucleic Acids Res.* 2017; 45: D289–D295. <https://doi.org/10.1093/nar/gkw1098> PMID: 27899584
29. Kurakata Y, Uechi A, Yoshida H, Kamitori S, Sakano Y, Nishikawa A, et al. Structural insights into the substrate specificity and function of *Escherichia coli* K12 YgjK, a glucosidase belonging to the glycoside hydrolase family 63. *J Mol Biol.* 2008; 381: 116–128. <https://doi.org/10.1016/j.jmb.2008.05.061> PMID: 18586271
30. Miyazaki T, Nishikawa A, Tonozuka T. Crystal structure of the enzyme-product complex reveals sugar ring distortion during catalysis by family 63 inverting α -glucosidase. *J Struct Biol.* 2016; 196: 479–486. <https://doi.org/10.1016/j.jsb.2016.09.015> PMID: 27688023
31. Fujimoto Z, Jackson A, Michikawa M, Maehara T, Momma M, Henrissat B, et al. The structure of a *Streptomyces avermitilis* α -L-Rhamnosidase reveals a novel carbohydrate-binding module CBM67 within the six-domain arrangement. *J Biol Chem.* 2013; 288: 12376–12385. <https://doi.org/10.1074/jbc.M113.460097> PMID: 23486481
32. Hidaka M, Honda Y, Kitaoka M, Nirasawa S, Hayashi K, Wakagi T, et al. Chitobiose phosphorylase from *Vibrio proteolyticus*, a member of glycosyl transferase family 36, has a clan GH-L-like (α / α)(6) barrel fold. *Structure.* 2004; 12: 937–47. <https://doi.org/10.1016/j.str.2004.03.027> PMID: 15274915
33. Miyazaki T, Ichikawa M, Iino H, Nishikawa A, Tonozuka T. Crystal structure and substrate-binding mode of GH63 mannosylglycerate hydrolase from *Thermus thermophilus* HB8. *J Struct Biol.* 2015. <https://doi.org/10.1016/j.jsb.2015.02.006> PMID: 25712767
34. Pachi P, Škerlová J, Šimčíková D, Kotik M, Křenková A, Mader P, et al. Crystal structure of native α -L-rhamnosidase from *Aspergillus terreus*. *Acta Crystallogr Sect D Struct Biol.* 2018; 74: 1078–1084. <https://doi.org/10.1107/S2059798318013049> PMID: 30387766
35. Cardona F, Parmeggiani C, Faggi E, Bonaccini C, Gratteri P, Sim L, et al. Total syntheses of casuarine and its 6-O- α -glucoside: Complementary inhibition towards glycoside hydrolases of the GH31 and GH37 families. *Chem—A Eur J.* 2009; 15: 1627–1636. <https://doi.org/10.1002/chem.200801578> PMID: 19123216
36. Cereija TB, Alarico S, Lourenço EC, Manso JA, Ventura MR, Empadinhas N, et al. The structural characterization of a glucosylglycerate hydrolase provides insights into the molecular mechanism of mycobacterial recovery from nitrogen starvation. *IUCrJ.* 2019; 6: 572–585. <https://doi.org/10.1107/S2052252519005372> PMID: 31316802
37. Nam YW, Nihira T, Arakawa T, Saito Y, Kitaoka M, Nakai H, et al. Crystal structure and substrate recognition of cellobionic acid phosphorylase, which plays a key role in oxidative cellulose degradation by microbes. *J Biol Chem.* 2015. <https://doi.org/10.1074/jbc.M115.664664> PMID: 26041776
38. Bianchetti CM, Elsen NL, Fox BG, Phillips GN. Structure of cellobiose phosphorylase from *Clostridium thermocellum* in complex with phosphate. *Acta Crystallogr Sect F Struct Biol Cryst Commun.* 2011. <https://doi.org/10.1107/S1744309111032660> PMID: 22102229
39. Hidaka M, Kitaoka M, Hayashi K, Wakagi T, Shoun H, Fushinobu S. Structural dissection of the reaction mechanism of cellobiose phosphorylase. *Biochem J.* 2006; 398: 37–43. <https://doi.org/10.1042/BJ20060274> PMID: 16646954
40. Barker MK, Rose DR. Specificity of processing α -glucosidase I is guided by the substrate conformation: Crystallographic and in silico studies. *J Biol Chem.* 2013; 288: 13563–13574. <https://doi.org/10.1074/jbc.M113.460436> PMID: 23536181

41. Zheng H, Cooper DR, Porebski PJ, Shabalin IG, Handing KB, Minor W. CheckMyMetal: A macromolecular metal-binding validation tool. *Acta Crystallogr Sect D Struct Biol*. 2017; 73: 223–233. <https://doi.org/10.1107/S2059798317001061> PMID: 28291757
42. Nakamura M, Yasukawa Y, Furusawa A, Fuchiwaki T, Honda T, Okamura Y, et al. Functional characterization of unique enzymes in *Xanthomonas euvesicatoria* related to degradation of arabinofurano-oligosaccharides on hydroxyproline-rich glycoproteins. *PLoS One*. 2018; 13: e0201982. <https://doi.org/10.1371/journal.pone.0201982> PMID: 30092047
43. Gibson RP, Gloster TM, Roberts S, Warren RAJ, Storch De Gracia I, García Á, et al. Molecular basis for trehalase inhibition revealed by the structure of trehalase in complex with potent inhibitors. *Angew Chemie—Int Ed*. 2007; 46: 4115–4119. <https://doi.org/10.1002/anie.200604825> PMID: 17455176
44. Nerinckx W, Desmet T, Piens K, Claeysens M. An elaboration on the syn-anti proton donor concept of glycoside hydrolases: Electrostatic stabilisation of the transition state as a general strategy. *FEBS Lett*. 2005; 579: 302–312. <https://doi.org/10.1016/j.febslet.2004.12.021> PMID: 15642336
45. The CAZypedia Consortium. Ten years of CAZypedia: a living encyclopedia of carbohydrate-active enzymes. *Glycobiology*. 2018; 28: 3–8. <https://doi.org/10.1093/glycob/cwx089> PMID: 29040563
46. Suzuki R, Katayama T, Kitaoka M, Kumagai H, Wakagi T, Shoun H, et al. Crystallographic and mutational analyses of substrate recognition of endo- α -N-acetylgalactosaminidase from *Bifidobacterium longum*. *J Biochem*. 2009; 146: 389–98. <https://doi.org/10.1093/jb/mvp086> PMID: 19502354
47. Ito T, Katayama T, Hattie M, Sakurama H, Wada J, Suzuki R, et al. Crystal structures of a glycoside hydrolase family 20 lacto-*N*-biosidase from *Bifidobacterium bifidum*. *J Biol Chem*. 2013; 288: 11795–11806. <https://doi.org/10.1074/jbc.M112.420109> PMID: 23479733
48. Yamada C, Gotoh A, Sakanaka M, Hattie M, Stubbs KA, Katayama-Ikegami A, 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. <https://doi.org/10.1016/j.chembiol.2017.03.012> PMID: 28392148
49. Ficko-Blean E, Boraston AB. The interaction of a carbohydrate-binding module from a *Clostridium perfringens* N-acetyl- β -hexosaminidase with its carbohydrate receptor. *J Biol Chem*. 2006; 281: 37748–37757. <https://doi.org/10.1074/jbc.M606126200> PMID: 16990278
50. Boraston AB, Ficko-Blean E, Healey M. Carbohydrate recognition by a large sialidase toxin from *Clostridium perfringens*. *Biochemistry*. 2007; 46: 11352–11360. <https://doi.org/10.1021/bi701317g> PMID: 17850114
51. Sakurama H, Fushinobu S, Hidaka M, Yoshida E, Honda Y, Ashida H, et al. 1,3–1,4- α -L-Fucosynthase that specifically introduces Lewis a/x antigens into type-1/2 chains. *J Biol Chem*. 2012. <https://doi.org/10.1074/jbc.M111.333781> PMID: 22451675
52. Tian W, Chen C, Lei X, Zhao J, Liang J. CASTp 3.0: Computed atlas of surface topography of proteins. *Nucleic Acids Res*. 2018; 46: W363–W367. <https://doi.org/10.1093/nar/gky473> PMID: 29860391
53. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. *Nucleic Acids Res*. 2019; 47: D427–D432. <https://doi.org/10.1093/nar/gky995> PMID: 30357350