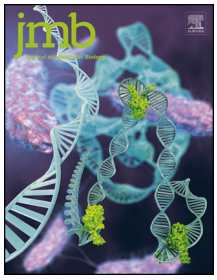


N-glycan Utilization by *Bifidobacterium* Gut Symbionts Involves a Specialist β-Mannosidase

Rosa Lorizolla Cordeiro [1,2], Renan Augusto Siqueira Pirolla [1], Gabriela Felix Persinoti [1], Fábio Cesar Gozzo [3], Priscila Oliveira de Giuseppe [1] and Mario Tyago Murakami [1]

1 - Brazilian Bioethanol Science and Technology Laboratory (CTBE), *Brazilian Center for Research in Energy and Materials (CNPEM), Campinas, SP, Brazil*

2 - Graduate Program in Functional and Molecular Biology, *Institute of Biology, University of Campinas, Campinas, SP, Brazil*

3 - Dalton Mass Spectrometry Laboratory, *Institute of Chemistry, University of Campinas, Campinas, SP, Brazil*

Correspondence to Priscila Oliveira de Giuseppe and Mario Tyago Murakami: *Brazilian Bioethanol Science and Technology Laboratory (CTBE), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas, Sao Paulo 13083-970, Brazil. priscila.giuseppe@ctbe.cnpem.br, mario.murakami@ctbe.cnpem.br*

https://doi.org/10.1016/j.jmb.2018.12.017

Edited by Thomas J. Smith

Abstract

Bifidobacteria represent one of the first colonizers of human gut microbiota, providing to this ecosystem better health and nutrition. To maintain a mutualistic relationship, they have enzymes to degrade and use complex carbohydrates non-digestible by their hosts. To succeed in the densely populated gut environment, they evolved molecular strategies that remain poorly understood. Herein, we report a novel mechanism found in probiotic Bifidobacteria for the depolymerization of the ubiquitous 2-acetamido-2-deoxy-4-*O*-(β-D-mannopyranosyl)-D-glucopyranose (Man-β-1,4-GlcNAc), a disaccharide that composes the universal core of eukaryotic *N*-glycans. In contrast to *Bacteroidetes,* these Bifidobacteria have a specialist and strain-specific β-mannosidase that contains three distinctive structural elements conferring high selectivity for Man-β-1,4-GlcNAc: a lid that undergoes conformational changes upon substrate binding, a tryptophan residue swapped between the two dimeric subunits to accommodate the GlcNAc moiety, and a Rossmann fold subdomain strategically located near to the active site pocket. These key structural elements for Man-β-1,4-GlcNAc specificity are highly conserved in *Bifidobacterium* species adapted to the gut of a wide range of social animals, including bee, pig, rabbit, and human. Together, our findings uncover an unprecedented molecular strategy employed by *Bifidobacteria* to selectively uptake carbohydrates from *N*-glycans in social hosts.

© 2019 Elsevier Ltd. All rights reserved.

Introduction

Gut microbiota coevolved with human hosts over thousands of years, cooperating to shape our physiology, metabolism, and immunity [1–3]. Bifidobacteria are among the first microorganisms to colonize the human intestine, corresponding to about 70% of the newborn gut bacteria [4]. They can be acquired during natural childbirth and through breastfeeding, being present in the vagina and breast milk [5]. Although the relative abundance of *Bifidobacterium* species declines as people get older [4], this first microbiota assemblage is considered to play fundamental roles in human health, both in infancy and at later stages of life.

Bifidobacteria have beneficial effects on human health, being therefore considered probiotics. They seem to modulate the host immune system, protect against infectious diseases, and have been indicated for the treatment of disorders such as intestinal inflammatory disease (Crohn's disease), diarrhea, colitis, and obesity [6–8]. Recent studies have also shown the use of *Bifidobacterium* strains for the alternative treatment of different types of cancer, depression, diabetes, and other diseases of worldwide importance [9–12]. However, the molecular

© 2019 Elsevier Ltd. All rights reserved.

mechanisms involved in these health-promoting effects are still largely unresolved [13].

Most *Bifidobacterium* species inhabit the transverse and descending colon, where nutrient availability is scarce. To be competitive in this niche, they evolved specific adaptations reflected, for example, in different profiles of glycoside hydrolase genes to degrade and utilize complex glycans non-digested by the host [14]. Some of the molecular strategies employed by *Bifidobacterium* strains to deconstruct human milk oligosaccharides [15,16], *O*-glycans [17], galactooligosaccharides [18], and plant-derived carbohydrates [19,20] have been elucidated with atomic detail. Nevertheless, other mechanisms, such as those related to *N*-glycan utilization, are still elusive.

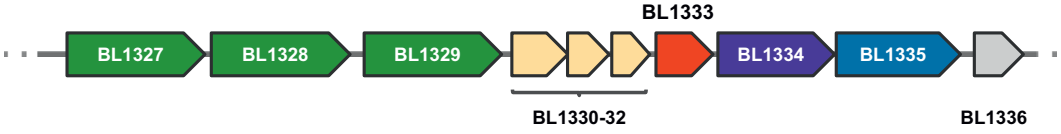
Mannosylated glycans are important sources of energy for many saccharolytic bacteria, being mannose easily used in the glycolytic pathway via isomerization of mannose-6-phosphate to fructose-6-phosphate. Moreover, the mannose and GlcNAc units released from these glycans can serve, respectively, as building blocks for the synthesis of bacterial exopolysaccharides and cell wall peptidoglycan. For some *Bifidobacterium* species, *N*-glycans promote the rapid and selective growth [21], indicating that they have a fully adapted molecular machinery for the utilization of these complex carbohydrates.

In this work, we demonstrate a novel mechanism for the cleavage of a disaccharide found in the common core of eukaryotic *N*-glycans [(2-acetamido-2-deoxy-4-*O*-(β-D-mannopyranosyl)-D-glucopyranose (Man-β-1,4-GlcNAc)]. This cleavage is performed by a specialist β-mannosidase conserved in several *Bifidobacterium* species present in the intestinal tract of diverse animals such as honey bees and humans. Phylogenetic analyses revealed that the evolution of *Bl*Man5B gene in *Bifidobacterium* spp. is remarkably complex, bringing novel insights into the genomic plasticity of *Bifidobacterium* species and the molecular strategies they evolved to become gut symbionts.

Results

A specialist β-mannosidase for the disaccharide Man-β-1,4-GlcNAc is present in the *N*-glycan PULpolysaccharide utilization locus of *Bifidobacterium longum*

The capacity of some *Bifidobacterium* species to grow on the presence of *N*-glycans has been associated with the existence of polysaccharide utilization loci (PUL) dedicated to degrade, import, and catabolize mannosylated glycans (such as BL1327-BL1336 in *B. longum* subsp. *longum* NCC2705) [22]. This PUL in *B. longum* consists of three α-mannosidases (BL1327–29), three ABC transporter components



	GH family	Predicted function (RefSeq)	Predicted subcellular localization (CELLO)	Signal peptide (SignalP)
	GH38	α-mannosidase	Cytoplasmic	No
	-	ABC transporter	Extracellular, transmembrane, transmembrane	Yes, no, no
	GH5	Hypothetical protein	Cytoplasmic	No
	GH31	Pentosidase or hexosidase	Cytoplasmic	No
	GH85	N-acetylglucosaminidase *	Extracellular	Yes
	-	LacI transcriptional regulator	Cytoplasmic	No

Fig. 1. The gene BL1333 (*Bl*Man5B) belongs to an *N*-glycan PUL identified in the genome from *B. longum* subsp. *longum* NCC2705. Arrows represent gene locus color-coded according to the respective function. The gray box highlights the target of this study. The asterisk at *N*-acetylglucosaminidase indicates the only enzyme whose function has been experimentally validated so far [23]. The subcellular localization and presence of signal peptide were predicted from the amino acid sequences using CELLO v.2.5 [24] and SignalP v. 4.1. [25], respectively.

(BL1330–32), an *N*-acetylglucosaminidase (BL1335), an α-glucosidase (BL1334), and a glycoside hydrolase with unknown specificity annotated as a hypothetical protein (BL1333) (Fig. 1).

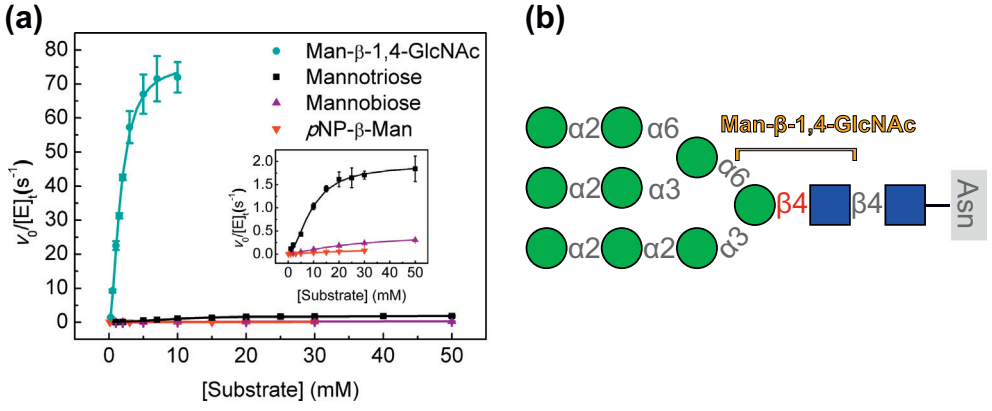
To decipher the role of the BL1333 gene (named here as *Bl*Man5B) in the *N*-glycan depolymerization, it was overexpressed in *Escherichia coli* and purified to homogeneity. The enzyme activity was tested against a range of polymeric and synthetic substrates of which only *p*-nitrophenyl-β-D-mannopyranoside (*p*NP-β-Man) was slowly cleaved, indicating that this protein is a β-mannosidase. To determine the kinetic rates, enzymatic reactions were conducted with *p*NP-β-Man concentrations ranging from 0.1 to 30 mM, under optimum pH (6.0) and temperature conditions (50 °C) determined for the enzyme (Fig. 2a). However, substrate saturation was not reached, preventing reliable estimation of k_{cat} and K_m, and indicating that substrate interaction with positive subsite(s) might be crucial for substrate affinity.

In eukaryotic *N*-glycans, β-linkages involving a non-reducing mannosyl residue are only found in the subcomponent Man-β-1,4-GlcNAc (Fig. 2b). Thus, we hypothesized that the disaccharide Man-β-1,4- could be the physiological substrate of *Bl*Man5B. Indeed, kinetic analysis using mass spectrometry (Supplementary File 1) revealed the clear preference for Man-β-1,4-GlcNAc (k_{cat} = 76.7 ± 1.98 s^{-1} and K_m = 1.73 ± 0.08 mM) compared to mannobiose (k_{cat} of 0.44 ± 0.01 s^{-1} and K_m 26.3 ± 1.55 mM) and mannotriose (k_{cat} of 1.93 ± 0.08 s^{-1} and K_m 9.10 ± 0.66 mM) (Fig. 2a, c). *Bl*Man5B displayed 2215-fold higher specificity constant (k_{cat}/K_m) for Man-β-1,4-GlcNAc compared to mannobiose, indicating a high preference for this disaccharide derived from *N*-glycan degradation.

*Bl*Man5B crystal structure reveals a dimeric arrangement and a new module in GH enzymes

The crystal structure of *Bl*Man5B was solved using single isomorphous replacement with anomalous scattering (SIRAS) method (Table 1). Experimental phasing was necessary due to the lack of significant sequence similarity to structures available at Protein Data Bank (https://www.rcsb.org/).

*Bl*Man5B structure comprises a $(\beta/\alpha)_8$ barrel domain, which is typical for GH5 family members, followed by a C-terminal module composed of three parallel β-strands interleaved with helices and loops



(c)

Substrate	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1} mM^{-1})
Man-β-1,4-GlcNAc	1.73 ± 0.08	76.7 ± 1.98	44.3
Mannotetraose	B. D.	B. D.	B. D.
Mannotriose	9.10 ± 0.66	1.93 ± 0.08	0.21
Mannobiose	26.3 ± 1.55	0.44 ± 0.01	0.02
*p*NP-β-Man	N. D.	N. D.	N. D.

Fig. 2. *Bl*Man5B is a β-mannosidase with high specificity for Man-β-1,4-GlcNAc. (a) Kinetic curves of *Bl*Man5B activity against *p*NP-β-Man, mannobiose, mannotriose, and Man-β-1,4-GlcNAc. A zoomed view of mannooligosaccharide curves are shown in the inset. (b) Representation of a *N*-glycan oligosaccharide, highlighting the preferential substrate of *Bl*Man5B, Man-β-1,4-GlcNAc, a subproduct of *N*-glycan decomposition. (c) K_m, k_{cat} and k_{cat}/K_m parameters against Man-β-1,4-GlcNAc, mannotriose, and mannobiose. The substrates *p*NP-β-man and mannotetraose were also assayed, but kinetic parameters could not be determined in these cases. N. D. = not determined, B. D. = below detection.

Table 1. Data collection and refinement statistics

	Iodine derivative	Native	GlcNAc complex (soaking)	E257A mutant	E257A mutant-GlcNAc (soaking)	E257A mutant-GlcNAc (co-crystal)
Data collection						
Wavelength (Å)	1.46	1.46	1.46	1.46	1.46	1.46
Resolution range (Å)	48.14–2.10 (2.23–2.10)	45.82–1.98 (2.05–1.98)	41.03–1.9 (1.97–1.9)	42.31–1.8 (1.87–1.8)	48.58–1.75 (1.81–1.75)	43.49–2.5 (2.59–2.5)
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell	77.24 100.27 172.17	76.57 101.92 1 71.59	69.89 101.38 173.01	75.40 102.24 171.28	69.80 101.25 172.94	75.83 100.88 171.63633
Total reflections	824,612 (100,238)	460,166 (25,348)	623,272 (57,773)	631,100 (60,336)	743,882 (60,705)	274,879 (21,325)
Unique reflections	150,689 (24,286)	92,507 (8,536)	97,470 (9,601)	122,705 (12,014)	123,201 (11,848)	44,664 (3,952)
Multiplicity	5.5 (4.1)	5.0 (3.0)	6.4 (6.0)	5.1 (5.0)	6.0 (5.1)	6.2 (5.4)
Completeness (%)	99.9 (99.8)	98.21 (91.89)	99.9 (99.7)	99.6 (98.9)	99.2 (96.6)	96.1 (86.7)
Mean *I*/sigma(*I*)	11.71 (1.70)	13.51 (1.04)	11.13 (0.91)	12.47 (0.91)	13.05 (0.90)	9.13 (0.85)
R-meas	0.101 (0.896)	0.085 (1.644)	0.135 (2.102)	0.092 (2.045)	0.092 (1.941)	0.166 (1.645)
$CC_{1/2}$	0.998 (0.771)	0.999 (0.55)	0.998 (0.482)	0.999 (0.633)	0.999 (0.545)	0.997 (0.61)
Heavy atom sites	16					
FOM	0.595					
Pseudo-free CC	0.664					
Refinement						
Reflections used in refinement		92,493 (8,525)	97,446 (9,596)	122,691 (11,999)	123,096 (11,836)	44,621 (3,948)
Reflections used for *R*-free		4,527 (445)	4,830 (414)	6,113 (614)	6,028 (593)	2,182 (184)
R-work		0.166 (0.345)	0.178 (0.323)	0.184 (0.407)	0.172 (0.301)	0.227 (0.370)
R-free		0.189 (0.363)	0.202 (0.364)	0.203 (0.439)	0.194 (0.332)	0.255 (0.359)
Number of non-hydrogen atoms		7,191	7,397	7,187	7,295	6,702
Macromolecules		6,595	6,754	6,592	6,662	6,637
Ligands		48	28		44	30
Solvent		548	615	595	589	35
Protein residues		838	854	836	847	850
RMS(bonds) (Å)		0.013	0.005	0.010	0.008	0.004
RMS(angles) (Å)		1.38	1.14	1.28	1.22	1.05
Ramachandran favored (%)		97.24	97.41	97.60	97.39	97.16
Ramachandran allowed (%)		2.76	2.59	2.40	2.61	2.84
Ramachandran outliers (%)		0.00	0.00	0.00	0.00	0.00
Rotamer outliers (%)		1.60	0.71	0.73	1.16	0.58
Clashscore		1.84	2.18	2.32	1.37	6.05
Average *B*-factor ($Å^2$)		42.71	34.31	44.27	34.32	63.90
Macromolecules		42.24	33.59	44.04	33.64	63.88
Ligands		55.50	49.61		37.08	75.50
Solvent		47.31	41.60	46.77	41.77	57.87
Number of TLS groups		2	2	2	2	2

(Fig. 3a, b). Structural comparisons against CATH database revealed that this C-terminal module is a three-layer (aba) sandwich with Rossmann fold topology and belongs to the Superfamily 3.40.50.12330, with a single member characterized so far (3c97A01) [26]. This member corresponds to the fragment of the response regulator receiver domain of a signal transduction histidine kinase from

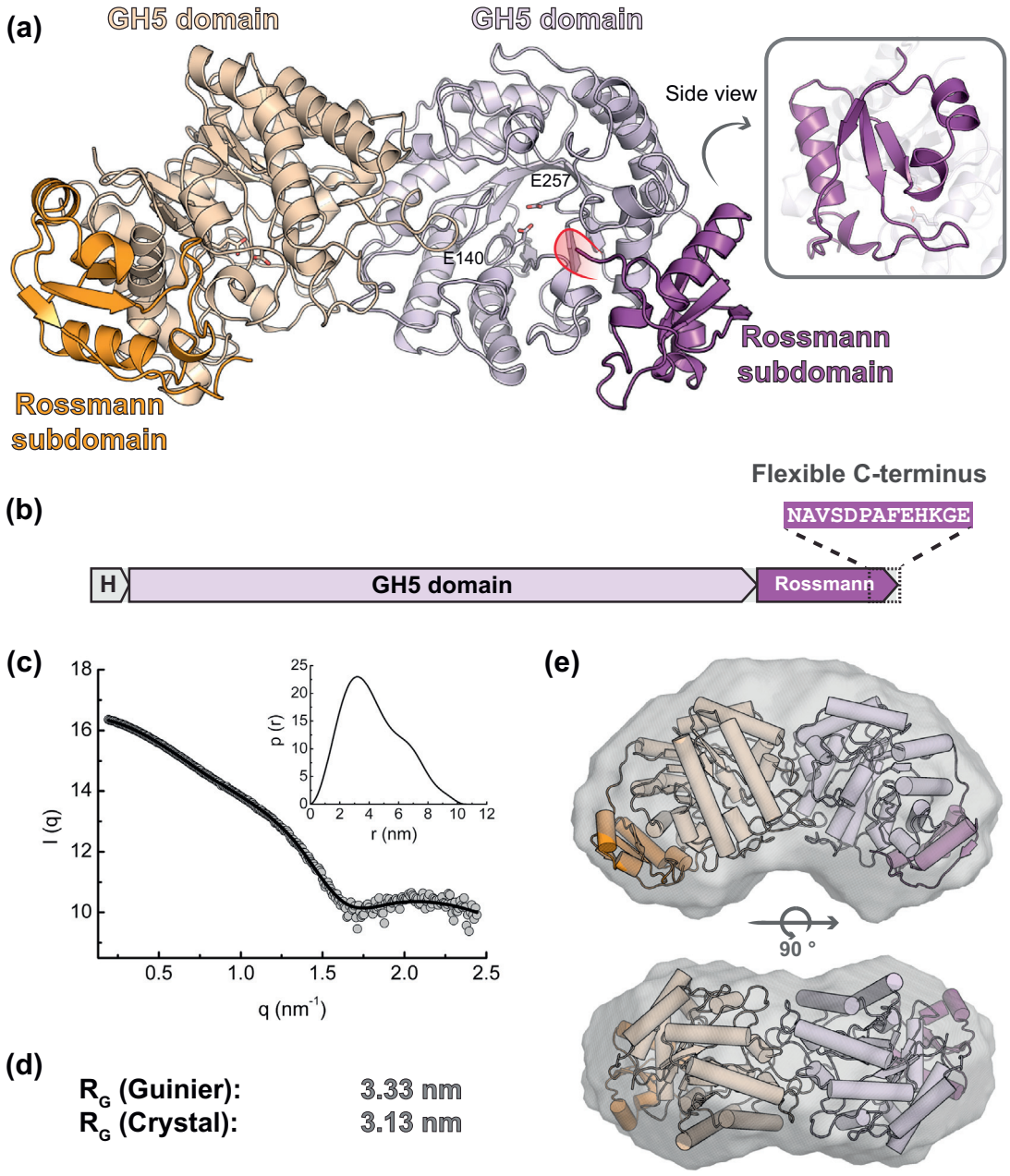


Fig. 3. *Bl*Man5B is a dimeric enzyme with unique features. (a) Cartoon representation of *Bl*Man5B, which comprises a GH5 domain followed by a Rossmann fold subdomain (dark purple) and a C-terminal extension not modeled in the crystal structure, probably heading toward the active site (red). The catalytic residues E140 (acid–base) and E257 (nucleophile) are shown in sticks. (b) Domain scheme of the recombinant *Bl*Man5B, showing the N-terminal 6xHis tag (H), GH5 domain, Rossmann fold subdomain, and the flexible C-terminal region (NAVSDPAFEHKGE). (c) Experimental SAXS data ($I(q)$ *versus* q) shown in gray circles. The theoretical scattering curve from the model ensemble is shown in black, calculated from $P(r)$ function (inset), where P is the pair-distance distribution function and r is the distance vector. (d) Radius of gyration (R_G) estimated from Guinier region of SAXS curve, or from crystallographic data. (e) Superposition of the envelope generated from SAXS data to the crystallographic data of *Bl*Man5B, in two views. Yellow and purple represent each subunit of the dimer.

Aspergillus oryzae (PDB ID: 3C97) and displays 11.6% sequence identity to the C-terminal domain of *Bl*Man5B, superimposing with a RMSD of 2.9 Å for 43 aligned Cα atoms. The last 12 C-terminal residues (^{418}AVSDPAFEHKGE^{429}) of *Bl*Man5B protein were not observed in the electron density map, possibly due to a high intrinsic disorder of this region in the crystal structure.

Two protein molecules were found in the asymmetric unit of *Bl*Man5B crystals. Bioinformatics analysis using PISA server [27] indicated that these molecules could form a stable dimer in solution (contact surface area = 5710 Å^2, ΔG^{int} = −14.5 kcal/mol, ΔG^{diss} = 15.8 kcal/mol). To confirm this prediction, we performed small-angle X-ray scattering (SAXS) experiments, which revealed that the size (R_g = 3.33 nm) and shape (low-resolution molecular envelope) of *Bl*Man5B in solution agree with the crystallographic dimer (Fig. 3c–e).

The active site is formed by three distantly related structural elements

Structural comparisons revealed that *Bl*Man5B conserves the classical Koshland double-displacement mechanism of retaining enzymes [28], with the acid–base (Glu140) and nucleophile (Glu257) side chains placed at 3.9 Å apart from each other (Fig. 4a). Since the native enzyme was solved in complex with glycerol, and extensive attempts to obtain the crystal structure of enzyme-mannose, enzyme-substrate or mutant E257A-substrate failed, we compared the glycerol coordinates to that of the mannosyl residue bound into −1 subsite of *Rm*Man5B, from *Rhizomucor miehei* [29] (Fig. 4b). According to this analysis, five *Bl*Man5B residues, conserved in GH5 enzymes, can interact with a mannosyl moiety via hydrogen bonds (Asn139, Glu140, and Glu257) to O2 or stacking interactions with the sugar ring (Trp293). Another

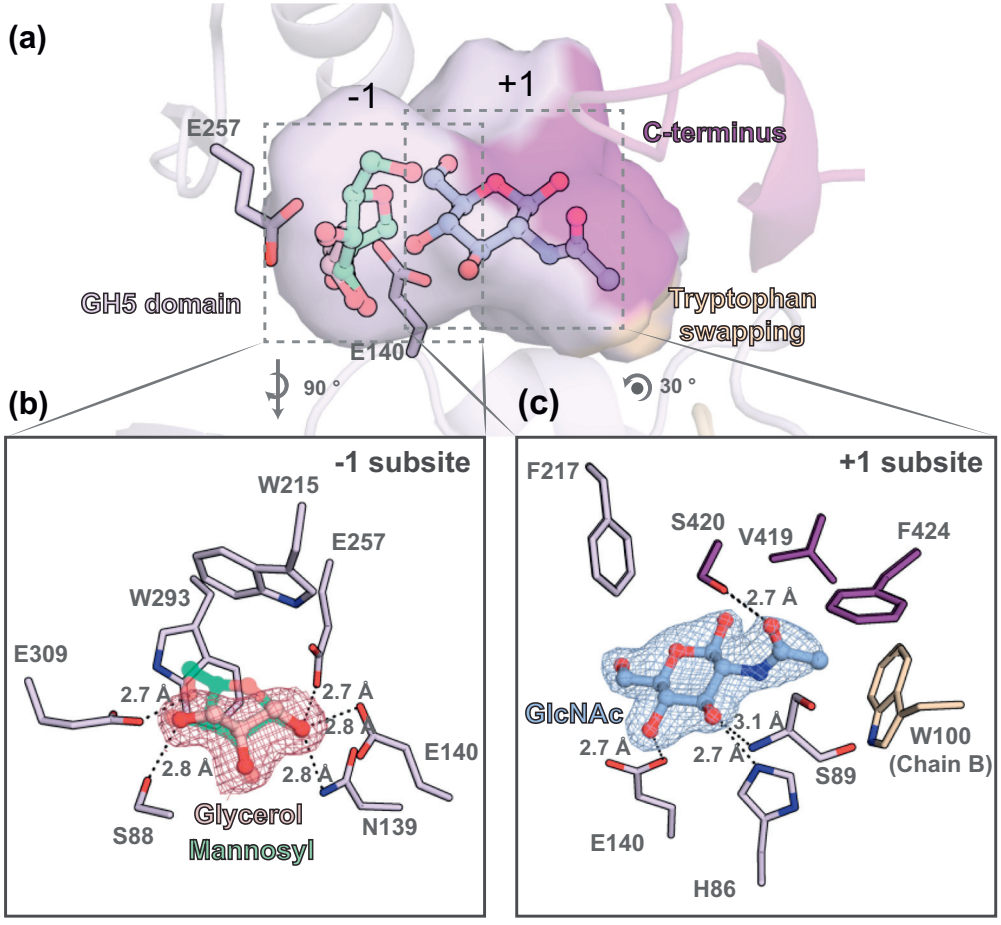


Fig. 4. *Bl*Man5B active site forms a cage that precisely fits the disaccharide Man-β-1,4-GlcNAc. (a) The active site comprises two subsites, −1 and + 1, which accommodates a mannosyl and a GlcNAc moiety, respectively. Catalytic pocket is delimited by residues from the GH5 domain (light purple), the C-terminal extension (dark purple) and by a swapped tryptophan from the adjacent protomer (yellow). (b) Interactions at −1 subsite in complex with glycerol (pink carbons), superposed to a mannosyl molecule (green carbons) in order to estimate interactions formed during enzymatic reactions. (c) Interactions of +1 subsite residues with GlcNAc (light blue carbons).

three residues (Ser88, Trp215, and Glu309), conserved in GH5_18 subfamily, are also in position to interact with the mannosyl moiety. Ser88 side chain is 2.8 Å far from glycerol O3 (corresponds to mannosyl O4), Trp215 possibly stacks with C5 and C6 mannosyl atoms, and Glu309 likely interacts with the OH-6 group of the mannosyl residue (Fig. 4b).

Complexes with GlcNAc were obtained using native enzyme and E257A mutant. The crystal structure of enzyme–GlcNAc complexes revealed that GlcNAc recognition at +1 subsite involves three distinct elements: (i) residues from the $(\beta/\alpha)_8$ barrel domain, which are conserved in GH5_18 and interact with the glycosidic ring of the hexosamine; (ii) a flexible C-terminal extension from the Rossmann fold module, which is stabilized in presence of GlcNAc and interacts with its acetamide portion; and (iii) a tryptophan residue that is swapped between the dimer subunits and performs CH–π interaction with the acetamide portion of GlcNAc (Fig. 4c).

Three residues from the barrel domain interacts with GlcNAc. Phe217 makes hydrophobic interactions with the glycosidic ring, whereas His86 and Ser89 bind to GlcNAc-O3 via side-chain (His86) or main-chain (Ser89) hydrogen bonds. H86A and F217A mutations abolished the enzyme activity against Man-β-1,4-GlcNAc, showing that His86 and Phe217 residues are essential for *Bl*Man5B activity. Residues from the flexible C-terminal region interact with the acetamide portion of GlcNAc via hydrophobic contacts (Val419 and Phe424) and hydrogen bonds (Ser420), with the assistance of Trp100 from the other subunit (Fig. 4c).

The C-terminal extension plays a role in substrate specificity

When GlcNAc is incorporated into *Bl*Man5B crystals, by soaking or co-crystallization, it binds to the +1 subsite and the flexible C-terminal region is stabilized over the active site, acting as a lid (Fig. 5a, b). This conformational change is induced by interactions between three residues of the lid and the acetamide moiety of GlcNAc (Fig. 5c). In crystal soaking experiments, only one active site of the dimer was occupied with GlcNAc, probably due to molecular packing restraints. In co-crystallization experiments, GlcNAc occupied both active site pockets, inducing the lid closure in both protomers (Fig. 5c).

To investigate the role of this region in enzyme activity, we constructed the ΔN417–E429 mutant, which lacks the last 13 flexible residues from the C-terminus. Kinetic assays of ΔN417–E429 mutant against Man-β-1,4-GlcNAc showed k_{cat} of 11.98 ± 0.31 s^{-1} and K_m of 3.92 ± 0.19 mM (Fig. 5d). The lower k_{cat} compared to the wild-type enzyme points to a loss of 84% in the catalytic performance, whereas the 2-fold increase in K_m reflects a lower affinity to the substrate, indicating that stabilization of the GlcNAc moiety by the C-terminal extension is critical for the full enzyme activity against the substrate Man-β-1,4-GlcNAc.

Tryptophan swapping dictates substrate specificity

In *Bl*Man5B, a tryptophan residue (Trp100) is swapped at the dimer interface and interacts with the acetamide group of GlcNAc. Structure comparison of *Bl*Man5B with the homologous *Rm*Man5B (GH5_7) indicates that Trp100 positioning might be an adaptation for GlcNAc recognition, since it disfavors the interaction with mannosyl residues to allow the accommodation of the larger GlcNAc moiety in relation to a mannosyl residue (Fig. 6a). In *Rm*Man5B, the opposite occurs. A residue equivalent to Trp100 (Tyr145) delineates the +1 subsite from its own chain, forming a steric barrier for GlcNAc binding (Fig. 6b).

To probe the role of Trp100 in enzyme activity, we built a mutant in which this tryptophan was replaced by an alanine residue (W100A). The mutant W100A was capable to fold and preserved the quaternary structure of native enzyme, according to CD (Fig. 6c) and dynamic light scattering (DLS) assays (Rh = 4.04 ± 0.47 nm, polydispersity = 11.6%). However, a kinetic curve of mutant W100A activity against Man-β-1,4-GlcNAc showed a 3-fold higher K_m (5.95 ± 0.45 mM) compared to the wild-type enzyme (Fig. 6d), indicating that the CH–π interaction of Trp100 with the acetamide group contributes to substrate affinity. Regarding the catalytic turnover, mutant W100A displayed a k_{cat} 3652-fold lower than the native enzyme against Man-β-1,4-GlcNAc (0.021 ± 0.01 s^{-1}), demonstrating the importance of Trp100 side chain for the catalytic turnover as well. Besides the contribution of Trp100 to substrate binding, a local damage to the active site caused by the lack of a large and hydrophobic side chain in the W100A mutant might also have contributed to the drastic reduction of enzyme activity.

Man-β-1,4-GlcNAc specificity is highly conserved in GH5_18 subfamily

All members from the subfamily GH5_18 are proteins produced by Actinobacteria species, including *Bifidobacterium* sp., *Streptomyces* sp., and *Cutibacterium* sp. as major representative groups. To estimate whether the functional and structural features found in *Bl*Man5B extend to the entire subfamily, we made a multiple sequence alignment of GH5_18 proteins and investigated the conservation of key structural elements identified in this study across the members (Supplementary File 2).

The amino acid residues of *Bl*Man5B -1 subsite, which recognize the mannosyl moiety of the substrate,

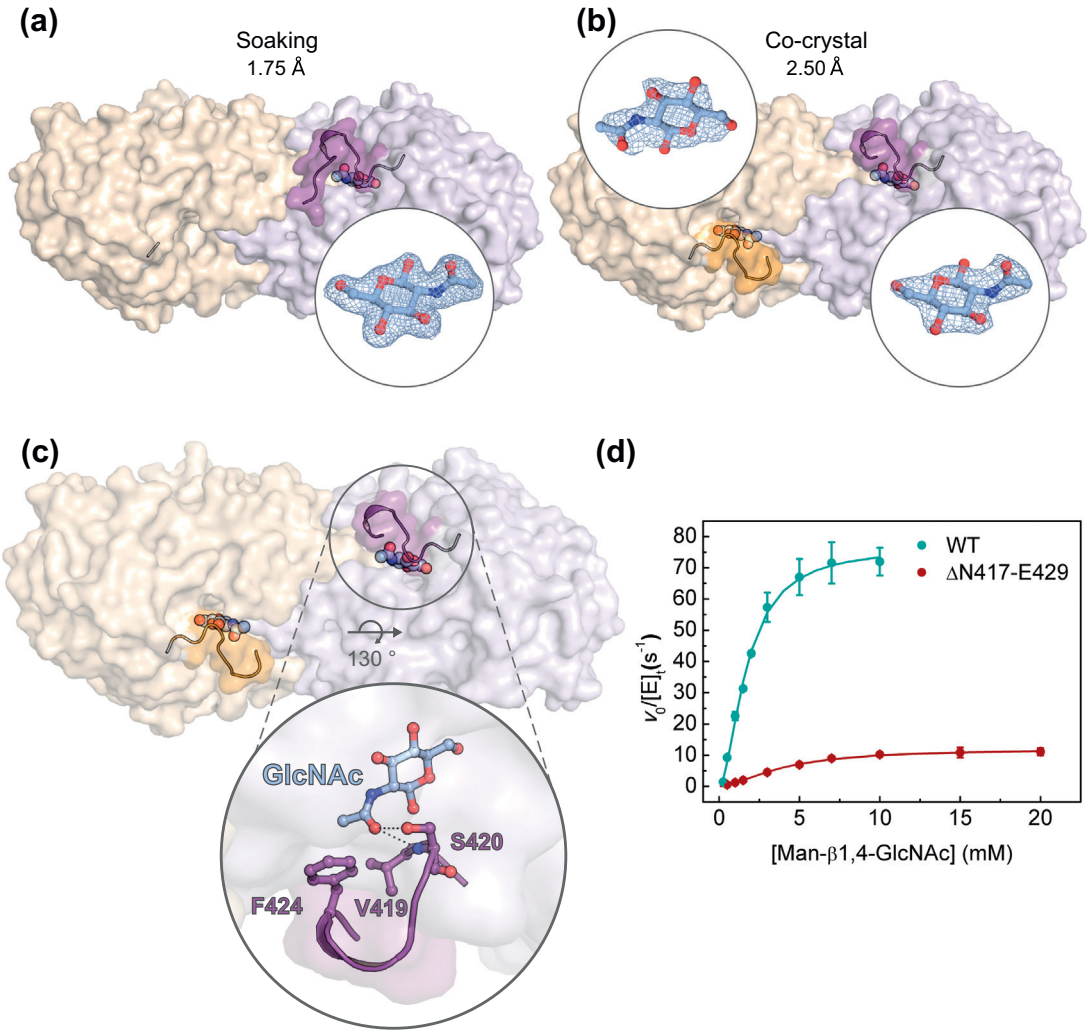


Fig. 5. The C-terminal extension acts as a flexible lid that plays a key role in GlcNAc recognition, substrate affinity, and catalytic turnover. (a) *Bi*Man5B E257A crystal structure soaked with GlcNAc (light blue, detail), in which only one active pocket was occupied by GlcNAc and the C-terminal extension is closed on the active site, acting as a lid (dark regions, cartoon representation). The wild-type and E257A mutant did not show structural differences to each other, except for better resolution of E257A crystal structure. (b) *Bi*Man5B E257A crystal structure co-crystallized with GlcNAc, in which the active pocket of both protomers is occupied by GlcNAc. (c) *Bi*Man5B crystal structure co-crystallized with GlcNAc, showing in detail residues that interact with the NAc portion of GlcNAc. (d) Kinetic curves of *Bi*Man5B native (blue) and ΔN417–E429 mutant (red) activity against Man-β-1,4-GlcNAc. Error bars represent standard deviation of the mean for three replicates.

are conserved between the analyzed sequences, supporting that GH5_18 is a β-mannosidase subfamily. At the +1 subsite, the residues $His86^{chain\ A}$, $Trp100^{chain\ B}$, and $Phe217^{chain\ A}$, which interact with the GlcNAc moiety of the substrate, are conserved in at least 95% of the analyzed sequences (total = 176), indicating that most GH5_18 members are able to recognize Man-β-1,4-GlcNAc as substrate. On the other hand, the C-terminal extension, which also plays a role in GlcNAc recognition, is highly variable between Actinobacteria and might provide distinct properties for the enzymes from genera other than *Bifidobacterium*.

Insights into the evolution of GH5_18 genes in *Bifidobacterium* spp.

From 305 available *Bifidobacterium* genomes, 130 have homologs of the gene encoding *Bi*Man5B (Fig. 7a), including species isolated from non-human primates (black lemur: *B. eulemuris*, marmoset: *B. callitrichos*, *B. hapali*, and *B. reuteri*), pigs (*B. longum* subsp. *suis*), chicken (*B. pullorum*), rabbit (*B. saecullarum*), and bees (*B. asteroides*, *B. bohemicum*, and *B. bombi*). These genes are most prevalent in strains of *B. breve* (found in 100% of 51 strains), *B. longum* subsp. *infantis* (100% of

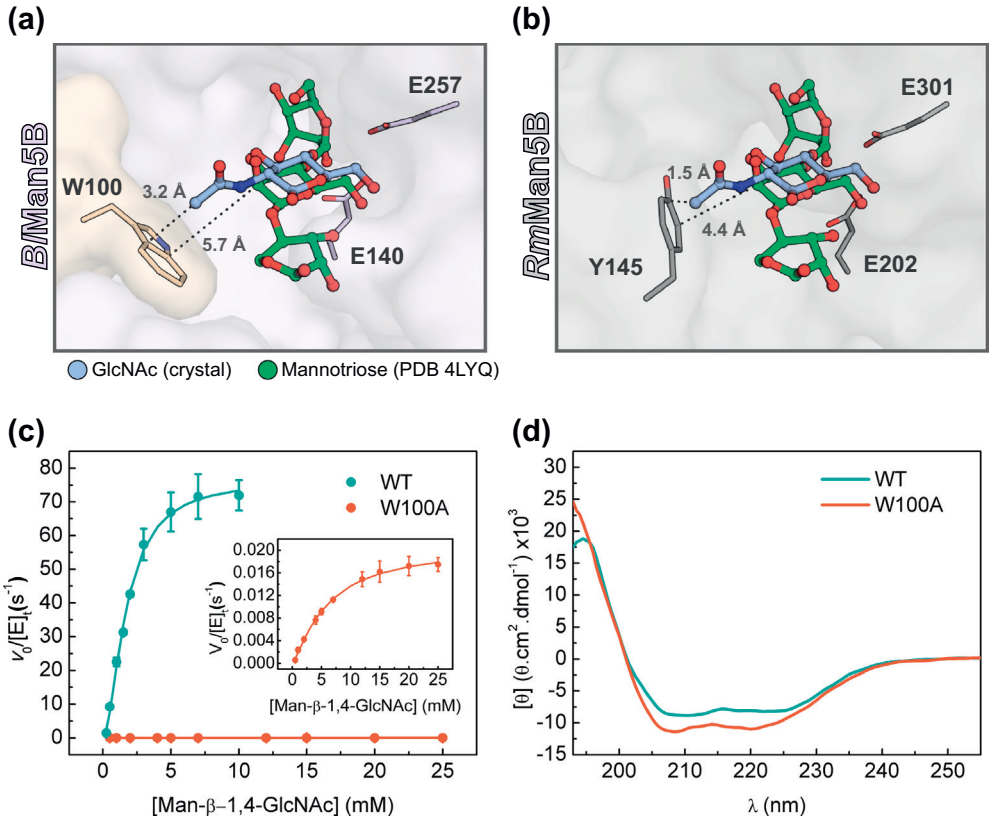


Fig. 6. A tryptophan swapping is crucial for *Bl*Man5B activity. (a) *Bl*Man5B structure in complex with GlcNAc (light blue carbons) superposed to mannotriose (green carbons) from *Rm*Man5B, a generalist GH5 β-mannosidase (PDB ID: 4LYQ). (b) *Rm*Man5B in complex with mannotriose superposed to *Bl*Man5B-GlcNAc. Note that in *Bl*Man5B (A), the residue W100 is too far to interact with the second manosyl moiety of mannotriose (5.7 Å) but in a proper position to interact with the acetamide group of GlcNAc (3.2 Å). In *Rm*Man5B, the residue Y145 offers a steric clash to GlcNAc binding (1.5 Å) but can interact with the second mannosyl residue of mannotriose (4.4 Å). (c) Kinetic curves of *Bl*Man5B native (blue) and W100A mutant (salmon) activity against Man-β-1,4-GlcNAc. A zoomed view of W100A curve is shown in the inset. (d) CD analysis of purified W100A mutant (salmon) and native enzyme (blue) as a control.

18 strains), and *B. longum* subsp. *longum* (57% of 67 strains), which are (sub)species commonly found in human intestinal microbiotas at different stages of human life [5]. Interestingly, the species *B. bifidum*, *B. adolescentis*, *B. dentium*, *B. animalis* subsp. *lactis,* and *B. catenulatum* lack any copy of the gene encoding *Bl*Man5B, supporting that the presence of this enzymatic tool is strain-specific in Bifidobacteria and might be related to the adaptation to particular nutrient niches in the gut.

To gather insight into the evolution of GH5_18 genes in *Bifidobacterium* spp., we constructed a phylogenetic tree based on a multiple sequence alignment of GH5_18 enzymes (Fig. 7b). This phylogenetic reconstruction revealed several inconsistencies with the species tree, showing that evolution of GH5_18 in *Bifidobacterium* included processes other than speciation (Fig. 7). Gene-tree reconciliation analyses using Notung [30] indicate the presence of 14 duplication events and 79 losses during the evolutionary history of GH5_18 genes in *Bifidobacterium* spp. Despite this high rate of predicted gene duplications events, the amino acid residues that determine the substrate specificity of *Bl*Man5B are conserved in the analyzed sequences, indicating that these genes share the same function even in distant-related *Bifidobacterium* species.

Discussion

In this work, we deciphered a missing part of the mechanistic puzzle related to *N*-glycan utilization by probiotic gut symbionts Bifidobacteria. The so far obscure mechanism concerns the cleavage of the disaccharide Man-β-1,4-GlcNAc, which in some *Bacteroides* spp. is performed by mannoside phosphorylases (GH130) [31], whereas in Bifidobacteria is made by a mannoside hydrolase.

Here, we revealed that the enzyme in charge for breaking down this intermediate product of *N*-glycan decomposition in Bifidobacteria belongs to the GH5_18

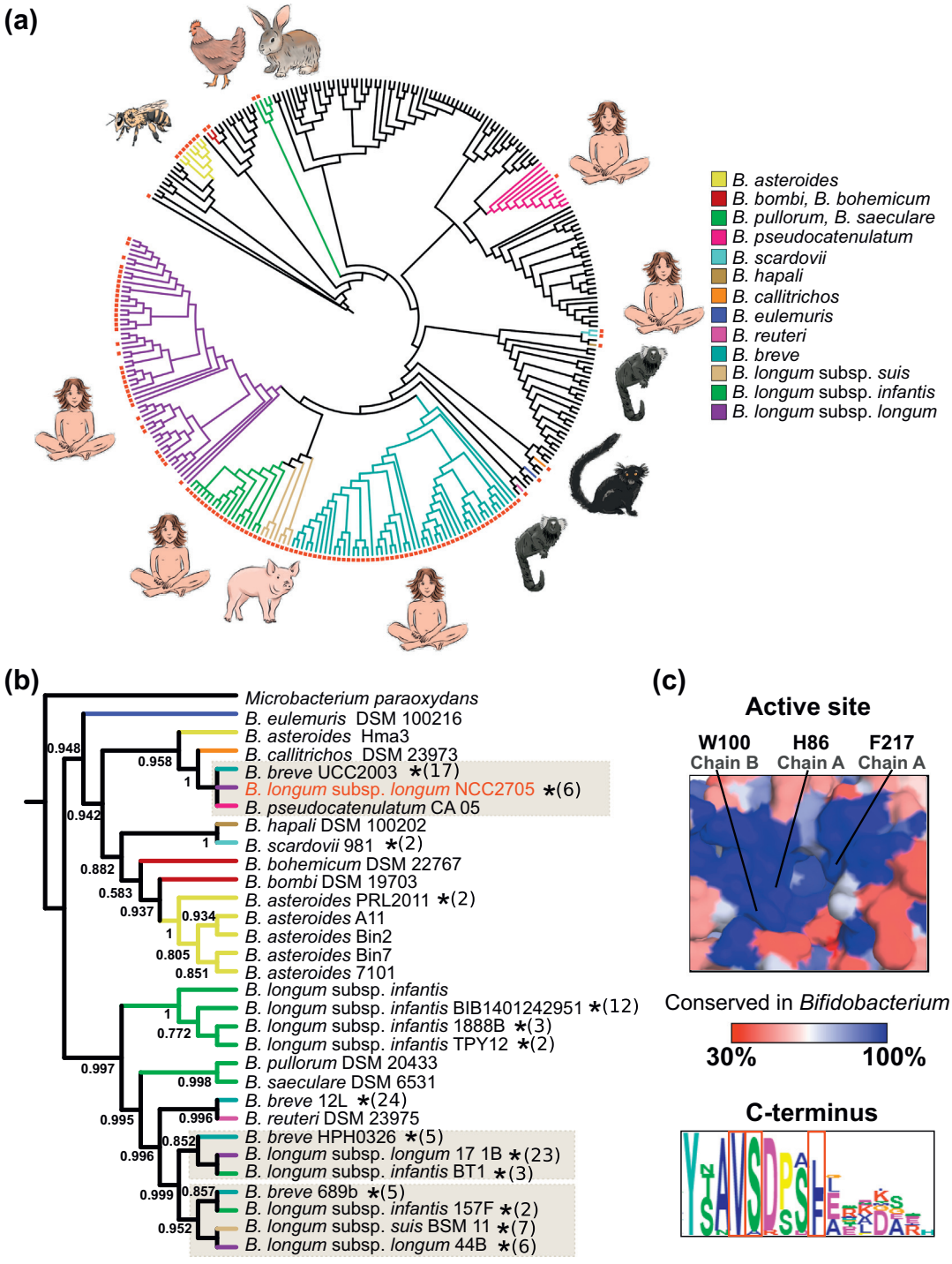


Fig. 7. Evolution of *Bl*Man5B in *Bifidobacterium* spp. (a) Phylogenetic tree of *Bifidobacterium* (sub)species, representing 305 available genomes. Red squares indicate strains containing orthologs for *Bl*Man5B coding gene. (b) Phylogenetic tree of *Bl*Man5B gene and its orthologs found in *Bifidobacterium* genomes. Sequences sharing 99% identity are represented in the same leaf, and *Bl*Man5B is highlighted in red. Support values based on the Shimodaira–Hasegawa test with 1000 bootstrap replicates are shown next to the branches (except for those with values below 0.5). (c) Structure of active site, showing amino acid conservation across *Bifidobacterium* genera, colored from red (<30% identity) to blue (100% identity), and conservation at C-terminus. The amino acid residues that interact to substrate at +1 subsite (W100, H86 and F217) are conserved in all *Bifidobacterium* representants, but the C-terminus is variable between homologs.

subfamily and is a β-mannosidase specialized in Man-β-1,4-GlcNAc cleavage. Unlike the more generalists GH1 and GH2 β-mannosidases, and even the GH5 β-mannosidase from *R. miehei*, *Bl*Man5B displays a high specificity for the disaccharide Man-β-1,4-GlcNAc. GH1 and GH2 β-mannosidases play a role in *N*-glycan decomposition in many organisms, but the enzymes described so far recognize manno-oligosaccharides as well as polymeric carbohydrates, sometimes acting as exo-mannanases [32,33]. In such enzymes, substrate recognition at +1 subsite is performed by stacking interactions with the pyranose ring of the carbohydrate, a common feature shared between hexoses and hexosamines. For this reason, their active pockets accommodate either manno-oligosaccharides or Man-β-1,4-GlcNAc [33]. However, in *Bl*Man5B, carbohydrate recognition at +1 subsite mainly occurs at acetamide group of GlcNAc, an exclusive feature of aceto-hexosamines.

The unique functional role of *Bl*Man5B reflects unique structural features. Dimerization is a prerequisite to shape the active-site pocket, strategically positioning Trp100 from the other subunit to interact with the acetamide group of Man-β-1,4-GlcNAc. Upon substrate binding, a flexible C-terminal extension from the Rossmann subdomain folds over the GlcNAc moiety, acting as a lid for the active site. The perfect fit of Man-β-1,4-GlcNAc into the active-site pocket of *Bl*Man5B confers to this enzyme a high specificity toward Man-β-1,4-GlcNAc, avoiding cross-reactions with manno-oligosaccharides and allowing a high efficiency of Man-β-1,4-GlcNAc hydrolysis during *N*-glycan depolymerization.

A previous study showed that the endo-*N*-acetylglucosaminidase EndoBB (BL1335), which belongs to the same PUL of *Bl*Man5B, cleaves the GlcNAc-β-1,4-GlcNAc linkage and releases the *N*-glycans from proteins before downstream depolymerization of the *N*-glycan core [23]. According to our structural data, the active site of *Bl*Man5B is uncapable to accommodate *N*-glycan structures containing α-mannosylations or decorations appended to the Man-β-1,4-GlcNAc disaccharide. Moreover, *Bl*Man5B lacks a signal peptide and is predicted to be cytoplasmic, supporting a role downstream the action of the extracellular enzyme EndoBB. Thus, we suggest that *Bl*Man5B catalyzes the last step of *N*-glycan degradation cascade in *B. longum*, after the release of *N*-glycans from glycoproteins by EndoBB, the internalization of the *N*-glycan by the ABC transporter, and the cleavage of *N*-glycan branches and decorations by the other enzymes encoded by this *N*-glycan PUL, also predicted to be cytoplasmic (Fig. 1).

Phylogenetic analysis indicates that the evolution of Man5B gene in Bifidobacteria was marked by several duplication and loss events, evidencing a high genomic plasticity that rendered this gene a strain-specific attribute. The presence of this gene correlates with the capability of *B. longum* subsp. *infantis* to use *N*-glycans as a sole carbon source, whereas its absence correlates with the lack of growth of *Bifidobacterium animalis* subsp. *lactis* when provided with this carbon source [21]. Homologs of *Bl*Man5B are found in *Bifidobacterium* species adapted to the gut of a wide range of social animals, including bee, pig, rabbit, and human. In such bacteria, key structural elements for Man-β-1,4-GlcNAc specificity are highly conserved, suggesting that these genes are under selective pressure emerging from a nutrient niche shared between distantly related hosts.

Together, our findings illustrate how nature can sculpt highly specific enzymes by coupling unconventional modules to versatile catalytic domains and using multimerization to customize the active site. Moreover, our studies contribute to a better understand of the molecular strategies used by *Bifidobacterium* strains to colonize the gut of social animals as distant as bees, chickens and humans.

Materials and Methods

Gene cloning, site-directed mutagenesis, and expression

The *B. longum* BL1333 gene encoding *Bl*Man5B was obtained by GeneScript (Piscataway, NJ) in pET28a vector, containing an N-terminal 6xHistag. A truncated construct (with C-terminal region removed) ΔN417–429 was amplified by PCR from BL1333 gene in pET28a using primers: Bl18_F1 (5′-GGGCATATGAAGTTCGGCGTTAAC-3′) and Bl_ΔN417_R (5′-TTTTCTCGAGTTAATACGCGT GGCCCGCCAC-3′). The DNA products were cloned into the NdeI and XhoI sites of pET28a-TEV. Site-directed mutagenesis was carried out using the Kit QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA, USA), with BL1333 as the template and specific primers, generating E257A, H86A, F217A, and W100A mutants in pET28a.

Protein expression

The gene coding for *Bl*Man5B and its mutants inserted into pET28a and pET28a-TEV were transformed in *E. coli* BL21 (DE3) cells (Agilent Technologies). Transformed cells were grown in 500 mL of LB medium and kanamycin at 37 °C and 200 rpm. When optical density reached 0.6–0.9, temperature was decreased to 20 °C, expression was induced by the addition of 0.5 mM IPTG, and cells were grown for further 16 h. Cells were harvested by centrifugation at 7000*g* for 30 min at 4 °C. Pellets were resuspended in lysis buffer [20 mM imidazole, 500 mM NaCl, 20 mM

sodium phosphate (pH 7.4), and 1 mM PMSF]. Cell suspensions were sonicated on ice using 8 pulses of 15 s in 30% amplitude separated by 45-s intervals. The lysed cells were centrifuged at 30,913*g* for 50 min at 4 °C. Lysates were filtered through a syringe 0.45-μm filter.

Protein purification

Wild-type enzyme and mutants were purified by immobilized metal ion affinity chromatography followed by size-exclusion chromatography. After cell lysis and centrifugation, the filtered supernatant was applied to a pre-equilibrated 5 mL HiTrap Chelating HP column (GE Healthcare) coupled to an ÄKTA FPLC system (GE Healthcare). The column was washed at 1 mL/min with 2 column volumes of buffer A [500 mM NaCl and 20 mM sodium phosphate (pH 7.4)] and then eluted with a linear gradient of imidazole in buffer B [500 mM imidazole, 500 mM NaCl and 20 mM sodium phosphate (pH 7.4)] in 20 column volumes. Protein was detected with an online detector monitoring at 280 nm and eluted fractions were collected and analyzed by SDS-PAGE. Fractions containing the enzyme were pooled, concentrated using Amicon Ultra-15 centrifugal filter devices (Merck), and loaded onto a Superdex 200 16/60 Hiload column (GE Healthcare) pre-equilibrated with phosphate buffer [150 mM NaCl and 20 mM sodium phosphate (pH 7.4)] using a flow of 1 mL/min. Fractions were analyzed with SDS-PAGE and DLS, and monodisperse samples were concentrated using an Amicon Ultra-4 centrifugal filter device (Merck). Protein concentration was determined by measuring the sample absorbance at 280 nm using the protein molar absorptivity coefficient (ε) [34].

Enzyme assays using DNS and pNP-derivatives

Reactions to detect activity against polymeric carbohydrates were performed using DNS (3,5-dinitrosalicylic acid) method [35]. Mannan (10 mg/mL), glucomannan (5 mg/mL), galactomannan (5 mg/mL), carboxymethyl cellulose (10 mg/mL), arabinan (100 mg/mL), and xylan (20 mg/mL) were tested. Assays were performed in 200 μL of reactions containing 100 μg of enzyme at pH 5.0 and 8.0 at 37 °C for 24 h. The reactions were heat-neutralized at 95 °C for 5 min. To 100 μL of the solution was added 100 μL of DNS and resuspended at 95 °C for 10 min. The final volume was transferred to a plate and monitored at 540 nm. For the screening of pNP-derivatives substrates, we used 4-nitrophenyl-α-D-galactopyranoside, 4-nitrophenyl-α-D-glucopyranoside, 4-nitrophenyl-α-D-mannopyranoside, 4-nitrophenyl-α-D-xylopyranoside, 4-nitrophenyl-α-L-arabinofuranoside, 4-nitrophenyl-α-L-arabinopyranoside, 4-nitrophenyl-β-D-cellobioside, 4-nitrophenyl-β-D-galactopyranoside, 4-nitrophenyl-β-D-glucopyranoside, 4-nitrophenyl-β-D-mannopyranoside, and 4-nitrophenyl-β-D-xylopyranoside. The reactions (110 μL) were incubated at 37 °C for 1 h, 600 rpm, and contained 5.45 mM substrate and 30 μg enzyme in McIlvaine buffer at pH 5.0 and 8.0. All the reactions were neutralized with 110 μL of saturated sodium tetraborate solution and monitored at 400 nm on an Infinite 200 PRO plate reader (Tecan Life Sciences).

Enzyme assays using pNP-β-Man

The reactions to determine *Bl*Man5B optimal pH and temperature were carried out in McIlvaine buffer using 4-nitrophenyl-β-D-mannopyranoside (pNP-β-Man) (Sigma) as substrate. Reactions were performed in a total volume of 110 μL containing 30 μg of enzyme and 5.45 mM pNP-β-Man and McIlvaine buffer (pH from 3.0 to 8.0; varying 0.5 between each measure) at 37 °C under 700 rpm. For native enzyme and W100A mutant, reactions were incubated for 30 min; and for ΔN417–E429 mutant, 20 min. After determined optimum pH (6.0), reactions were performed varying temperature (30, 40, 45, 50, 55, 60, and 70 °C) using the same conditions described above. Steady-state enzyme kinetics were carried out at 50 °C in McIlvaine buffer (pH 6.0) in a final volume of 110 μL using 30 μg enzyme and pNP-β-Man at a concentration ranging from 0.1 to 30 mM. The reactions were incubated for 30 min for *Bl*Man5B and 20 min for *Bl*Man5B ΔN417–E429. Kinetic parameters were obtained by sigmoidal fitting using Hill equation (Origin8.1 software). Inhibition studies were performed using 100 mM monosaccharides (glucose, galactose, mannose, arabinose, and xylose) and 8 mM pNP-β-man, in reactions containing 36 μg *Bl*Man5B and incubated for 30 min at 50 °C. All the reactions were neutralized with 110 μL of saturated sodium tetraborate solution and monitored at 400 nm on an Infinite 200 PRO plate reader (Tecan Life Sciences). The reactions were performed in triplicate, and statistical analysis was performed using one-tailed Student's *t* test (Origin8.1 software). Calibration curves were constructed with 4-nitrophenol (Sigma).

Enzyme kinetics using mass spectrometry

Man-β-1,4-GlcNAc was purchased from CarboSynth, whereas mannobiose, mannotriose, and xylobiose were purchased from Megazyme. All the mass spectra were analyzed on a Waters Synapt HDMS (Waters Co.) equipment, at V mode, ESI(+) at 3 kV and 130 °C in the source, *m*/*z* scan 150–500, direct infusion of 60 s, and acquisition of 1 Hz. All the reactions were performed at 50 °C in McIlvaine buffer (pH 6.0). For reactions containing Man-β-1,4-GlcNAc as substrate, we used 3 μg/mL of *Bl*Man5B native, 30 μg/mL *Bl*Man5B ΔN417–E429, and 600 μg/mL

*Bl*Man5B W100A, in 5, 10, and 20 min of reaction, respectively. Man-β-1,4-GlcNAc concentrations ranged from 0.25 and 10 mM for *Bl*Man5B native, from 0.5 to 20 mM for *Bl*Man5B ΔN417–E429, and from 0.5 to 25 mM for *Bl*Man5B W100A. Reactions with H86A and F217A mutants were carried out using the same conditions of native enzyme, but only at 3 mM of substrate to determine relative activity rates of enzymes. For mannobiose as substrate, we used 300 μg/mL of *Bl*Man5B native in reactions of 20 min, with mannobiose concentrations ranging from 1 to 50 mM. Relative rates were determined for *Bl*Man5B ΔN417–E429, H86A, and F217A mutants in 30 mM of mannobiose. Forty microliters of methanol was added to all the reactions to stop them. A fixed concentration of xylobiose was added to the quenched reactions to serve as an internal standard. Calibration curves were constructed for GlcNAc and mannose for quantifying product formation and determination of the reaction rate. Kinetic parameters (k_{cat} and K_m) were obtained by sigmoidal fitting using Hill equation (Origin8.1 software). Representative mass spectra are shown in Supplementary File 1.

Dynamic light scattering

Hydrodynamic radius (Rh) of *Bl*Man5B and its mutants were determined by DLS acquired using a Zetasizer Nano ZS (Malvern Instruments) equipment with a 633-nm laser in a quartz cell with a scattering angle of 90°. All the samples were measured in phosphate buffer [150 mM NaCl, 20 mM sodium phosphate (pH 7.4)] at 20 °C, in a concentration range of 1 to 6 mg/mL. Data were analyzed using Zetasizer Nano software.

Small-angle X-ray scattering

SAXS data were collected at the SAXS1 beamline at Brazilian Synchrotron Light Laboratory (LNLS, Brazil). Scattering patterns were captured on a Pilatus 300 K (Dectris) at 888-mm distance, in a q range from 0.13 to 4.97 nm^{-1}. Sample preparation of *Bl*Man5B for SAXS experiments included the removal of 6xHistag using 150 U thrombin in 600 μL reaction containing *Bl*Man5B at 22 mg/mL for 16 h at 4 °C, followed by size-exclusion chromatography. Monodisperse samples were prepared in phosphate buffer [150 mM NaCl and 20 mM sodium phosphate (pH 7.4)] in three concentrations: 4, 6, and 8 mg/mL. Ten images were recorded for each sample using an exposure time of 10 and 0.1 s of waiting time between them. As radiation damage was detected since image 2, only the first image of each sample was used for data processing using the ATSAS program suite [36]. The radius of gyration (R_g) was determined from the Guinier equation and by indirect Fourier transform method using the PRIMUSQT package [37]. GNOM [38] was used to calculate the particle distance distribution function $p(r)$. Twenty *ab initio* low-resolution models were generated using DAMMIN [39] program. Models probabilities were analyzed using DAMAVER [40], outliers were excluded, and the remaining models were aligned for a final model construction. The final low-resolution model and *Bl*Man5B crystal structure were superimposed using the SUPCOMB program [41].

Crystallization

Crystals of *Bl*Man5B and *Bl*Man5B E257A were grown by hanging-drop vapor diffusion method at 18 °C. Crystals were generated by mixing 1 μL of protein (12–16 mg/mL) with 1 μL crystallization solution [100 mM sodium citrate (pH 4.5) and 12%–16% PEG 3000), equilibrated against 200 μL reservoir solution. *Bl*Man5B-GlcNAc and *Bl*Man5B E257A-GlcNAc complexes were obtained by soaking crystals with 100 mM GlcNAc added to the mother liquor and by co-crystallization, mixing 16 mg/mL *Bl*Man5B E257A with 19 mM GlcNAc. This mixture was submitted to crystallization assays at the same conditions described above.

Data collection, structure determination, and refinement

X-ray diffraction data were collected at the MX2 beamline from LNLS (Campinas, Brazil) equipped with PILATUS 2 M detector (Dectris). Derivative crystals were obtained by soaking in the mother liquor supplemented with 500 mM NaI for 30 s. Crystals were cryoprotected by soaking with the reservoir solution supplemented with 25% glycerol for *Bl*Man5B native and derivative crystals, and 20% PEG 400 for *Bl*Man5B E257A and GlcNAc-bound complexes. Data sets were indexed, integrated, and scaled using the XDS package [42]. The structure was determined by single isomorphous replacement with anomalous scattering (SIRAS) using the native and iodine-derivative data. The heavy atom sites were determined by SHELXD [43] with the hkl2map interface, and the phases were calculated with SHELXE [43]. The initial model was built using the Bucanner program [44] and manually adjusted using COOT [45]. The crystal structures of *Bl*Man5B E257A and GlcNAc-bound complexes were solved by molecular replacement method using *Bl*Man5B native coordinates as search model in PHASER program [46]. Generated models were refined with REFMAC5 [47] and Phenix [48]. The final structures were validated using MOLPROBITY [49].

Phylogenetic analysis

Phylogenetic analyses were conducted using 305 *Bifidobacterium* genomes available at NCBI on August 2018, chosen based on EzBioCloud taxonomy

database [50], and the genome of *Microbacterium paraoxydans* as an outgroup. A set of 92 single-copy marker genes were used to perform a multi-locus phylogenetic analysis using the UBCG (up-to-date Bacterial Core Gene) pipeline [51]. Briefly, UBCG corresponding sequences were extracted from each of the 306 genomes, individually aligned, concatenated, and used to estimate the species phylogeny based on the maximum likelihood method and the WAG + CAT model of amino acid evolution implemented in FastTree 2.1.8 [52] software. The group of orthologous sequences of the GH5_18 subfamily was extracted from the same set of genomes using a reimplementation of OrthoMLC, the software FastOrtho. The set of GH5_18 sequences was clustered at 99% of identity using cd-hit 4.6 [53]. One representative sequence for each species in each cd-hit cluster was aligned using Maft v7.299b [54] and used to estimate the gene tree using the maximum likelihood method and the WAG + CAT model of amino acid evolution implemented in FastTree 2.1.8. Local support values of each split in the tree were estimated based on the Shimodaira–Hasegawa test [55] with 1000 bootstrap replicates, as implemented in FastTree 2.1.8. Gene tree-species tree reconciliation was estimated using Notung 2.9 [56] software using both Duplication-Loss (DL) Model and Duplication-Transfer-Loss (DTL) Model, although no feasible solutions were found using the last one.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2018.12.017.

Accession numbers

Coordinates and structure factors for *Bl*Man5B-glycerol, *Bl*Man5B-GlcNAc (co-crystal), BlMan5B-GlcNAc (soaking), E257A mutant-GlcNAc (soaking), and E257A mutant have been deposited in the Protein Data Bank under accession codes 6MP2, 6MOY, 6MPA, 6MP7, and 6MPC. Other data are available from the corresponding authors upon request.

Acknowledgments

We would like to thank Brazilian Synchrotron Light Laboratory (LNLS) and Brazilian Biosciences National Laboratory (LNBio) for the provision of time on the MX2 and SAXS1 beamlines, and ROBOLAB facility. We thank Esther Lorizolla Cordeiro for providing illustrations for figure preparation. This work was supported by Grant No. 2015/26982-0 from São Paulo Research Foundation (FAPESP) (to M.T. M.). RLC was supported by FAPESP Grant No. 2016/00740-2.

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

Received 13 November 2018;
Received in revised form 18 December 2018;
Accepted 24 December 2018
Available online 11 January 2019

Keywords:
GH5 family;
Bifidobacterium longum;
N-acetylglucosamine;
N-glycan;
molecular mechanism

Abbreviations used:
DLS, dynamic light scattering; Man-β-1,4-GlcNAc, 2-acetamido-2-deoxy-4-*O*-(β-D-mannopyranosyl)-D-glucopyranose; PUL, Polysaccharide utilization loci; *p*NP-β-Man, *p*-nitrophenyl-β-D-mannopyranoside; SAXS, Small-angle X-ray Scattering.

References

[1] L.V. Hooper, D.R. Littman, A.J. Macpherson, The gut microbiota interactions between the microbiota and the immune system, Science (80-) (2012) https://doi.org/10.1126/science.1223490.
[2] C.M. Guinane, P.D. Cotter, Role of the gut microbiota in health and chronic gastrointestinal disease: understanding a hidden metabolic organ, Ther. Adv. Gastroenterol. 6 (2013) 295–308, https://doi.org/10.1177/1756283X13482996.
[3] A.H. Moeller, A. Caro-Quintero, D. Mjungu, A.V. Georgiev, E.V. Lonsdorf, M.N. Muller, A.E. Pusey, M. Peeters, B.H. Hahn, H. Ochman, Cospeciation of gut microbiota with hominids, Science 353 (2016) 380–382, https://doi.org/10.1126/science.aaf3951.
[4] S. Arboleya, C. Watkins, C. Stanton, R.P. Ross, Gut bifidobacteria populations in human health and aging, Front. Microbiol. 7 (2016), 1204. https://doi.org/10.3389/fmicb.2016.01204.
[5] C. Milani, F. Turroni, S. Duranti, G.A. Lugli, L. Mancabelli, C. Ferrario, D. van Sinderen, M. Ventura, Genomics of the genus *Bifidobacterium* reveals species-specific adaptation to the glycan-rich gut environment, Appl. Environ. Microbiol. 82 (2016) 980–991, https://doi.org/10.1128/AEM.03500-15.
[6] A. O'Callaghan, D. van Sinderen, Bifidobacteria and their role as members of the human gut microbiota, Front. Microbiol. 7 (2016), 925. https://doi.org/10.3389/fmicb.2016.00925.
[7] P.R. Marteau, M. de Vrese, C.J. Cellier, J. Schrezenmeir, Protection from gastrointestinal diseases with the use of probiotics, Am. J. Clin. Nutr. 73 (2001) 430s–436s, https://doi.org/10.1093/ajcn/73.2.430s.
[8] J.L. Round, S.K. Mazmanian, The gut microbiota shapes intestinal immune responses during health and disease, Nat. Rev. Immunol. 9 (2009) 313–323, https://doi.org/10.1038/nri2515.

[9] S. Roy, G. Trinchieri, Microbiota: a key orchestrator of cancer therapy, Nat. Rev. Cancer 17 (2017) 271–285, https://doi.org/10.1038/nrc.2017.13.
[10] T.C. Fung, C.A. Olson, E.Y. Hsiao, Interactions between the microbiota, immune and nervous systems in health and disease, Nat. Neurosci. 20 (2017) 145–155, https://doi.org/10.1038/nn.4476.
[11] Y. Wang, L.H. Kasper, The role of microbiome in central nervous system disorders, Brain Behav. Immun. (2014) https://doi.org/10.1016/j.bbi.2013.12.015.
[12] H.-S. Lye, C.-Y. Kuan, J.-A. Ewe, W.-Y. Fung, M.-T. Liong, The improvement of hypertension by probiotics: effects on cholesterol, diabetes, renin, and phytoestrogens, Int. J. Mol. Sci. 10 (2009) 3755–3775, https://doi.org/10.3390/ijms10093755.
[13] F. Turroni, C. Milani, D. Van Sinderen, M. Ventura, Chapter 12—Bifidobacteria: Ecology and Coevolution With the Host, Elsevier Inc., 2018 https://doi.org/10.1016/B978-0-12-805060-6.00012-0.
[14] A.V. Chaplin, B.A. Efimov, V.V. Smeianov, L.I. Kafarskaia, A.P. Pikina, A.N. Shkoporov, Intraspecies genomic diversity and long-term persistence of *Bifidobacterium longum*, PLoS One (2015) https://doi.org/10.1371/journal.pone.0135658.
[15] C. Yamada, A. Gotoh, M. Sakanaka, M. Hattie, K.A. Stubbs, A. Katayama-Ikegami, J. Hirose, S. Kurihara, T. Arakawa, M. Kitaoka, S. Okuda, T. Katayama, S. Fushinobu, Molecular insight into evolution of symbiosis between breast-fed infants and a member of the human gut microbiome *Bifidobacterium longum*, Cell Chem. Biol. 24 (2017) 515–524.e5, https://doi.org/10.1016/j.chembiol.2017.03.012.
[16] D.A. Sela, Y. Li, L. Lerno, S. Wu, A.M. Marcobal, J.B. German, X. Chen, C.B. Lebrilla, D.A. Mills, An infant-associated bacterial commensal utilizes breast milk sialyloligosaccharides, J. Biol. Chem. 286 (2011) 11909–11918, https://doi.org/10.1074/jbc.M110.193359.
[17] M. Sato, D. Liebschner, Y. Yamada, N. Matsugaki, T. Arakawa, S.S. Wills, M. Hattie, K.A. Stubbs, T. Ito, T. Senda, H. Ashida, S. Fushinobu, The first crystal structure of a family 129 glycoside hydrolase from a probiotic bacterium reveals critical residues and metal cofactors, J. Biol. Chem. 292 (2017) 12126–12138, https://doi.org/10.1074/jbc.M117.777391.
[18] A.H. Viborg, F. Fredslund, T. Katayama, S.K. Nielsen, B. Svensson, M. Kitaoka, L. Lo Leggio, M. Abou Hachem, A β1–6/β1–3 galactosidase from *Bifidobacterium animalis* subsp. *lactis* Bl-04 gives insight into sub-specificities of β-galactoside catabolism within *Bifidobacterium*, Mol. Microbiol. 94 (2014) 1024–1040, https://doi.org/10.1111/mmi.12815.
[19] A. Bujacz, M. Jedrzejczak-Krzepkowska, S. Bielecki, I. Redzynia, G. Bujacz, Crystal structures of the apo form of β-fructofuranosidase from *Bifidobacterium longum* and its complex with fructose, FEBS J. 278 (2011) 1728–1744, https://doi.org/10.1111/j.1742-4658.2011.08098.x.
[20] T. Ito, K. Saikawa, S. Kim, K. Fujita, A. Ishiwata, S. Kaeothip, T. Arakawa, T. Wakagi, G.T. Beckham, Y. Ito, S. Fushinobu, Crystal structure of glycoside hydrolase family 127 Î²-l-arabinofuranosidase from *Bifidobacterium longum*, Biochem. Biophys. Res. Commun. 447 (2014) 32–37, https://doi.org/10.1016/j.bbrc.2014.03.096.
[21] S. Karav, A. Le Parc, J.M. Leite Nobrega de Moura Bell, S.A. Frese, N. Kirmiz, D.E. Block, D. Barile, D.A. Mills, Oligosaccharides released from milk glycoproteins are selective growth substrates for infant-associated Bifidobacteria, Appl. Environ. Microbiol. 82 (2016) 3622–3630, https://doi.org/10.1128/AEM.00547-16.
[22] J.L. Sonnenburg, C.T.L. Chen, J.I. Gordon, Genomic and metabolic studies of the impact of probiotics on a model gut symbiont and host, PLoS Biol. 4 (2006), e413. https://doi.org/10.1371/journal.pbio.0040413.
[23] D. Garrido, C. Nwosu, S. Ruiz-Moyano, D. Aldredge, J.B. German, C.B. Lebrilla, D.A. Mills, Endo-β-*N*-acetylglucosaminidases from infant gut-associated Bifidobacteria release complex *N*-glycans from human milk glycoproteins, Mol. Cell. Proteomics 11 (2012) 775–785, https://doi.org/10.1074/mcp.M112.018119.
[24] C.S. Yu, Y.C. Chen, C.H. Lu, J.K. Hwang, Prediction of protein subcellular localization, Proteins 64 (2006) 643–651, https://doi.org/10.1002/prot.21018.
[25] T.N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, SignalP 4.0: discriminating signal peptides from transmembrane regions, Nat. Methods 8 (2011) 785–786, https://doi.org/10.1038/nmeth.1701.
[26] N.L. Dawson, T.E. Lewis, S. Das, J.G. Lees, D. Lee, P. Ashford, C.A. Orengo, I. Sillitoe, CATH: an expanded resource to predict protein function through structure and sequence, Nucleic Acids Res. 45 (2017) D289–D295, https://doi.org/10.1093/nar/gkw1098.
[27] E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state, J. Mol. Biol. 372 (2007) 774–797, https://doi.org/10.1016/j.jmb.2007.05.022.
[28] F. Barras, M. Bauzan, J. Rouvier, C. Geyb, A. Heyraudb, B. Henrissatb, Stereochemistry of the Hydrolysis Reaction Catalyzed by Endoglucanase % from Erwiniu Chrysanthemi, https://ac-els-cdn.ez106.periodicos.capes.gov.br/001457939280183H/1-s2.0-001457939280183H-main.pdf?_tid=9e44dd2f-28b5-4df1-a959-152dcf140b8e&acdnat=1536094294_20c083b081081e5e266c4d75df9480e4 1992, Accessed date: 4 September 2018.
[29] P. Zhou, Y. Liu, Q. Yan, Z. Chen, Z. Qin, Z. Jiang, Structural insights into the substrate specificity and transglycosylation activity of a fungal glycoside hydrolase family 5 β-mannosidase, Acta Crystallogr. D Biol. Crystallogr. 70 (2014) 2970–2982, https://doi.org/10.1107/S1399004714019762.
[30] M. Stolzer, H. Lai, M. Xu, D. Sathaye, B. Vernot, D. Durand, Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees, Bioinformatics 28 (2012) i409–i415, https://doi.org/10.1093/bioinformatics/bts386.
[31] S. Ladevèze, L. Tarquis, D.A. Cecchini, J. Bercovici, I. André, C.M. Topham, S. Morel, E. Laville, P. Monsan, V. Lombard, B. Henrissat, G. Potocki-Véronèse, Role of glycoside phosphorylases in mannose foraging by human gut bacteria, J. Biol. Chem. 288 (2013) 32370–32383, https://doi.org/10.1074/jbc.M113.483628.
[32] M.N. Domingues, F.H.M. Souza, P.S. Vieira, M.A.B. de Morais, L.M. Zanphorlin, C.R. dos Santos, R.A.S. Pirolla, R.V. Honorato, P.S.L. de Oliveira, F.C. Gozzo, M.T. Murakami, Structural basis of exo-β-mannanase activity in the GH2 family, J. Biol. Chem. (2018)https://doi.org/10.1074/jbc.RA118.002374 (jbc.RA118.002374).
[33] L.E. Tailford, V.A. Money, N.L. Smith, C. Dumon, G.J. Davies, H.J. Gilbert, Mannose foraging by *Bacteroides thetaiotaomicron*: structure and specificity of the beta-mannosidase, BtMan2A, J. Biol. Chem. 282 (2007) 11291–11299, https://doi.org/10.1074/jbc.M610964200.
[34] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein identification and

analysis tools on the ExPASy server, Proteomics Protoc. Handb, Humana Press, Totowa, NJ 2005, pp. 571–607, https://doi.org/10.1385/1-59259-890-0:571.
[35] G.L. Miller, Use of dinitrosalicylic acid reagent for determination of reducing sugar, Anal. Chem. 31 (1959) 426–428, https://doi.org/10.1021/ac60147a030.
[36] D. Franke, M.V. Petoukhov, P.V. Konarev, A. Panjkovich, A. Tuukkanen, H.D.T. Mertens, A.G. Kikhney, N.R. Hajizadeh, J.M. Franklin, C.M. Jeffries, D.I. Svergun, N.N. Semenov, ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions, J. Appl. Crystallogr. 50 (2017) 1212–1225, https://doi.org/10.1107/S1600576717007786.
[37] P.V. Konarev, V.V. Volkov, A.V. Sokolova, M.H.J. Koch, D.I. Svergun, IUCr, *PRIMUS*: a Windows PC-based system for small-angle scattering data analysis, J. Appl. Crystallogr. 36 (2003) 1277–1282, https://doi.org/10.1107/S0021889803012779.
[38] D.I. Svergun, IUCr, determination of the regularization parameter in indirect-transform methods using perceptual criteria, J. Appl. Crystallogr. 25 (1992) 495–503, https://doi.org/10.1107/S0021889892001663.
[39] D.I. Svergun, Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing, Biophys. J. 76 (1999) 2879–2886, https://doi.org/10.1016/S0006-3495(99)77443-6.
[40] V.V. Volkov, D.I. Svergun, Uniqueness of *ab initio* shape determination in small-angle scattering, J. Appl. Crystallogr. 36 (2003) 860–864, https://doi.org/10.1107/S0021889803000268.
[41] M.B. Kozin, D.I. Svergun, IUCr, automated matching of high- and low-resolution structural models, J. Appl. Crystallogr. 34 (2001) 33–41, https://doi.org/10.1107/S0021889800014126.
[42] W. Kabsch, Integration, scaling, space-group assignment and post-refinement, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 133–144, https://doi.org/10.1107/S0907444909047374.
[43] G.M. Sheldrick, IUCr, experimental phasing with *SHELXC/D/E* : combining chain tracing with density modification, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 479–485, https://doi.org/10.1107/S0907444909038360.
[44] K. Cowtan, The *Buccaneer* software for automated model building. 1. Tracing protein chains, Acta Crystallogr. D Biol. Crystallogr. 62 (2006) 1002–1011, https://doi.org/10.1107/S0907444906022116.
[45] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Biological Crystallography Features and development of Coot, (n.d.). doi:https://doi.org/10.1107/S0907444910007493.
[46] A.J. McCoy, R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, R.J. Read, IUCr, *Phaser* crystallographic software, J. Appl. Crystallogr. 40 (2007) 658–674, https://doi.org/10.1107/S0021889807021206.
[47] G.N. Murshudov, A.A. Vagin, E.J. Dodson, IUCr, refinement of macromolecular structures by the maximum-likelihood method, Acta Crystallogr. D Biol. Crystallogr. 53 (1997) 240–255, https://doi.org/10.1107/S0907444996012255.
[48] P.D. Adams, P.V. Afonine, G. Bunkóczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.-W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, P.H. Zwart, *PHENIX* : a comprehensive Python-based system for macromolecular structure solution, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 213–221, https://doi.org/10.1107/S0907444909052925.
[49] V.B. Chen, W.B. Arendall, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson, *MolProbity* : all-atom structure validation for macromolecular crystallography, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 12–21, https://doi.org/10.1107/S0907444909042073.
[50] S.H. Yoon, S.M. Ha, S. Kwon, J. Lim, Y. Kim, H. Seo, J. Chun, Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies, Int. J. Syst. Evol. Microbiol. 67 (2017) 1613–1617, https://doi.org/10.1099/ijsem.0.001755.
[51] S.-I. Na, Y.O. Kim, S.-H. Yoon, S. Ha, I. Baek, J. Chun, UBCG: up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction, J. Microbiol. 56 (2018) 280–285, https://doi.org/10.1007/s12275-018-8014-6.
[52] M.N. Price, P.S. Dehal, A.P. Arkin, FastTree 2—approximately maximum-likelihood trees for large alignments, PLoS One 5 (2010), e9490. https://doi.org/10.1371/journal.pone.0009490.
[53] W. Li, A. Godzik, Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences, Bioinformatics 22 (2006) 1658–1659, https://doi.org/10.1093/bioinformatics/btl158.
[54] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, Mol. Biol. Evol. 30 (2013) 772–780, https://doi.org/10.1093/molbev/mst010.
[55] H. Shimodaira, M. Hasegawa, Multiple comparisons of log-likelihoods with applications to phylogenetic inference, Mol. Biol. Evol. 16 (1999) 1114–1116, https://doi.org/10.1093/oxfordjournals.molbev.a026201.
[56] K. Chen, D. Durand, M. Farach-Colton, NOTUNG: a program for dating gene duplications and optimizing gene family trees, J. Comput. Biol. 7 (2000) 429–447, https://doi.org/10.1089/106652700750050871.