

RESEARCH ARTICLE

Structural-Functional Analysis Reveals a Specific Domain Organization in Family GH20 Hexosaminidases

Cristina Val-Cid, Xevi Biarnés, Magda Faijes*, Antoni Planas

Laboratory of Biochemistry, Institut Químic de Sarrià, Universitat Ramon Llull, Barcelona, Spain

* magda.faijes@iqs.edu



OPEN ACCESS

Citation: Val-Cid C, Biarnés X, Faijes M, Planas A (2015) Structural-Functional Analysis Reveals a Specific Domain Organization in Family GH20 Hexosaminidases. PLoS ONE 10(5): e0128075. doi:10.1371/journal.pone.0128075

Academic Editor: Elena Papaleo, University of Copenhagen, DENMARK

Received: July 2, 2014

Accepted: April 23, 2015

Published: May 29, 2015

Copyright: © 2015 Val-Cid 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.

Funding: This work was supported in part by grants NANO3BIO (FP7-KBBE-2013-7; # 613931) from the European Commission and GLYCOZYMES (BIO2013-49022-C2-1-R) from MEC, Spain. C. Val-Cid acknowledges financial support from the Fundación Instituto Danone (2012-2013). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Hexosaminidases are involved in important biological processes catalyzing the hydrolysis of N-acetyl-hexosaminy residues in glycosaminoglycans and glycoconjugates. The GH20 enzymes present diverse domain organizations for which we propose two minimal model architectures: Model A containing at least a non-catalytic GH20b domain and the catalytic one (GH20) always accompanied with an extra α -helix (GH20b-GH20- α), and Model B with only the catalytic GH20 domain. The large *Bifidobacterium bifidum* lacto-N-biosidase was used as a model protein to evaluate the minimal functional unit due to its interest and structural complexity. By expressing different truncated forms of this enzyme, we show that Model A architectures cannot be reduced to Model B. In particular, there are two structural requirements general to GH20 enzymes with Model A architecture. First, the non-catalytic domain GH20b at the N-terminus of the catalytic GH20 domain is required for expression and seems to stabilize it. Second, the substrate-binding cavity at the GH20 domain always involves a remote element provided by a long loop from the catalytic domain itself or, when this loop is short, by an element from another domain of the multidomain structure or from the dimeric partner. Particularly, the lacto-N-biosidase requires GH20b and the lectin-like domain at the N- and C-termini of the catalytic GH20 domain to be fully soluble and functional. The lectin domain provides this remote element to the active site. We demonstrate restoration of activity of the inactive GH20b-GH20- α construct (model A architecture) by a complementation assay with the lectin-like domain. The engineering of minimal functional units of multidomain GH20 enzymes must consider these structural requirements.

Introduction

N-acetyl- β -hexosaminidases are widely distributed in nature participating in many biological processes [1,2]. They catalyze the hydrolysis of N-acetyl-hexosaminy residues in glycoproteins, glycosaminoglycans, chitin, and glycolipids, and are classified in CAZY families GH3, GH20 and GH84 (Carbohydrate Active Enzymes database, www.cazy.org [3]). GH20 hexosaminidases are found in crustaceans, insects and fungi, participating in the degradation of chitin, and in

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

bacteria, where they are involved in the catabolism of oligosaccharides that serve as nitrogen and carbon sources. From the biotechnological point of view, hexosaminidases, together with chitinases, are important enzymes for industrial chitin degradation, the most abundant polysaccharide after cellulose [2]. They are also powerful insect and fungal control agents as biopesticides. In humans, the two lysosomal hexosaminidases HexA and HexB are responsible for the hydrolysis of GlcNAc and GalNAc residues of glycosphingolipids such as ganglioside GM2 and their dysfunction is linked to neurodegenerative disorders known as Tay-Sachs and Sandhoff diseases [4,5]. In addition, some bacterial hexosaminidases contribute to the virulence of some infections in humans as in the case of *Actinobacillus actinomycetemcomitans* that colonizes the human oral cavity and causes periodontitis and *Streptococcus pneumoniae* infections [2].

GH20 enzymes comprise not only β -N-acetylhexosaminidases (EC.3.2.1.52), catalyzing the removal of terminal non-reducing N-acetylhexosamine residues, but also lacto-N-biosidases (EC.3.2.1.140), hydrolyzing the terminal lacto-N-biosyl residues from the non-reducing end of oligosaccharides such as human milk oligosaccharides (HMOs). HMOs play an important role in the development of intestinal flora and in the modulation of the immune system of breast-fed infants [6–8]. HMOs are constituted by two core tetrasaccharide structures: the lacto-N-tetraose (type 1) characterized by lacto-N-biose (Gal β 1,3GlcNAc) and lactose units linked by a β -1,3 glycosidic bond, and lacto-neo-N-tetraose (type 2) formed by β -1,3-linked N-acetylglucosamine (Gal β 1,4GlcNAc) and lactose. They are elongated to different isomers of lacto-N-hexa-, octa- and decaose and then, decorated with fucose and sialic residues via α -1,2/3/4 linkages and α -2,3/6 respectively [9,10]. Bifidobacteria have a unique metabolism on HMO resulting in their predominant proliferation in the colon of breast-fed infants [7]. Specifically, *Bifidobacterium bifidum* metabolizes type 1 human milk oligosaccharides by the action of extracellular sialidases, fucosidases and lacto-N-biosidase.

To date, the structures of thirteen GH20 enzymes are known. Their domain organization is diverse and several accompanying domains are present apart from the catalytic GH20 domain. In addition, some of them exist as dimeric proteins. We aimed at defining the minimal functional domain architecture for enzymatic activity in these GH20 family enzymes. We chose the large *B. bifidum* lacto-N-biosidase (LnbB) as a model protein because of its complex domain organization among GH20 enzymes and its significance in the metabolism of human milk oligosaccharides. The LnbB gene encodes for a membrane-anchored extracellular enzyme of 1112 amino acid residues. Recently, it was cloned and the recombinant full length protein was characterized confirming that this enzyme catalyzes the hydrolysis of the tetrasaccharide Gal β 1,3GlcNAc β 1,3Gal β 1,4Glc to lacto-N-biose (Gal β 1,3GlcNAc) and lactose (Fig 1) [11,12]. LnbB is a retaining exo-enzyme that acts by substrate-assisted catalysis, which is a common feature of GH20 enzymes.

By means of structural and sequence analysis of GH20 enzymes of known structure, we have established two levels of domains organization in this family, and further engineered LnbB in different truncated forms according to these models. We report the structural requirements for functionality, of this and the other GH20 enzymes, to obtain a minimal functional unit that retains the enzymatic activities. The importance of a remote element in the active site such as loop 2 seen in some GH20 enzymes is analyzed in all structures. It is shown that the isolated GH20 catalytic domain in LnbB is inactive and requires a remote element provided by

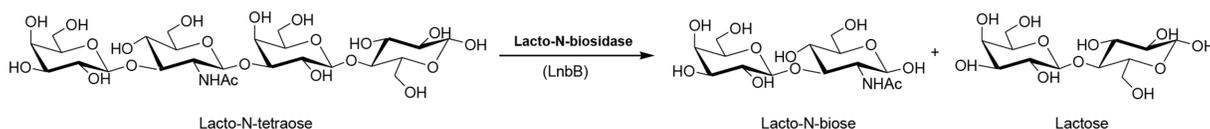


Fig 1. Hydrolytic reaction of lacto-N-tetraose catalyzed by *B. bifidum* lacto-N-biosidase.

doi:10.1371/journal.pone.0128075.g001

the adjacent lectin domain for activity, as observed in a complementation experiment. This structural and functional analysis provides new insights for further enzyme engineering for biotechnological and biomedical applications.

Materials and Methods

Cloning, expression and purification of Lacto-N-biosidase enzyme from *Escherichia coli*

Two synthetic genes of Lacto-N-biosidase from *Bifidobacterium bifidum* (LnbB), codon-optimized for the expression in *Escherichia coli* were produced by GenScript (GenScript, NJ, USA). The LnbB gene (37–1064) was cloned into the pET24b plasmid (Novagen, Madison, WI, USA) using the restriction sites NdeI and XhoI, and the construct A gene (40–528) into the pET28a+ vector using NdeI and Bpu1102. Constructs B (176–497) and B α (176–528) genes were obtained by PCR reaction using construct A as template, and constructs D (40–776) and F (546–776) genes using the full length LnbB gene. The following forward and reverse primers were used (restriction site are underlined): P1 (5' -GGAATTCCCATATGAAGCCGAAATATAAAGAACGT-3') and P2 (5' -TAATAGCTCAGCGGCGTCAAC-3') for construct B; P1 and P3 (5' -TAA TAGCTCAGCACGGGAGTCACTCCA-3') for construct B α ; P4 (5' -GCGTATCCATGGCTGATGACTCCGCAGCAGGCTA-3') and P5 (5' -GCGTATCTCGAGCAGGCTGCCGGTCAGC-3') for construct D, and P6 (5' -GCGTATCCATGGTGGATGCGGGTATC-3') and P5 for construct F. All constructs were cloned into the pET28a+ vector. *E. coli* DH5 cells were transformed and positive transformants were verified by DNA sequencing.

Expression and purification of the full length protein and constructs were carried out as follows: recombinant cultures of *E. coli* BL21 (DE3) *star* cells were grown following an autoinduction protocol [13] in LB medium with 30 $\mu\text{g}\cdot\text{mL}^{-1}$ kanamycin, 25mM Na₂HPO₄, 25mM K₂HPO₄, 50mM NH₄Cl, 5mM Na₂SO₄, 2mM MgSO₄, 0.5% glycerol, 0.05% glucose and 0.2% lactose at 250 rpm and 30°C for 24 h. The full length protein and constructs A, B and B α were also obtained using 1 mM IPTG induction after 16 h of growth at 37°C in LB medium with kanamycin achieving similar results. Cells were harvested by centrifugation (4000g, 15 min, 4°C), resuspended in buffer A (Na₂HPO₄ 20mM, NaCl 150mM, pH 7.5) and disrupted in a cell disrupter (20000 psi). After centrifugation (20000g, 4°C, 2 h), the supernatant was loaded into a 1mL Ni²⁺-charged HiTrap chelating column chromatography. Proteins were purified to homogeneity by a linear gradient of Buffer B (Na₂HPO₄ 20 mM, NaCl 150 mM, imidazole 500mM, pH 7.5). Collected fractions were analyzed by SDS-PAGE and fractions containing the proteins were combined and dialyzed against buffer A with a 30 kDa cut-off membrane. All proteins were purified further using an XK16/60 Superdex 200 column (GE Healthcare), pre-equilibrated with buffer A. Fractions were pooled, concentrated with a Centricon YM-10 unit (Merk Millipore) and stored at 4°C. Protein concentrations were determined spectrophotometrically by the BCA method measuring A590 nm using BSA as standard [14].

Protein size was determined using dynamic light scattering in a Nano ZS Nanosizer (Malvern Instruments Ltd., UK) with a laser light wavelength of 632.8 nm and a scattering angle of 173 degrees. Temperature was set at 25°C. Protein solutions in buffer A were measured without previous dilution. Full length protein showed a single peak for monodisperse particles of 10 nm and PDI 0.30, which is consistent with monomeric form.

Kinetics of the hydrolase activity

Lacto-N-biosidase activity assays with *p*-nitrophenyl β -lacto-N-bioside (Toronto Research Chemicals Inc.) were performed at 30°C in a volume of 0.1 mL using 96-wells microtiter plates

measuring p-nitrophenol concentration at 405 nm. Enzymatic reactions were performed in a reaction mixture containing 0.25 mM substrate and 50mM of citrate-phosphate buffer (pH 4.5). After pre-incubation for 5 min, reactions were initiated by addition of the enzyme (20 nM of full length protein and construct D, 20 nM-4 μ M of construct A). At regular time intervals, reactions were quenched by the addition of 0.15 mL of quenching buffer (0.5 M of glycine, pH 10) and p-nitrophenol release was measured by absorbance at 400 nm with a microplate-reader MRX (Cultek). Triplicate measurements at each concentration were performed.

Complementation assays

Enzymatic reactions with constructs A and F were performed using standard conditions in a Bravo liquid handling Robot (Agilent). After pre-incubation of construct A (50 nM) and F (25 nM-500 nM) for 5 minutes, addition of the substrate initiated the reaction. Triplicate measurements at each concentration were performed. The complex of constructs A and F was also characterized using gel filtration chromatography. The mixture of construct A (42 μ M) with excess of construct F (245 μ M) was loaded to the Superdex 200 column pre-equilibrated with buffer A and calibrated with protein standards. Fractions were further analyzed by SDS-PAGE and their hydrolase activity was determined. This purification was performed by duplicate.

Enzymatic activities of construct A (50 nM) in the presence of BSA at 1:1, 1:4 and 1:10 ratio were analyzed using standard conditions.

GH20 structures and sequences retrieval and analysis

The list of currently available 3D structures of GH20 enzymes was retrieved from CAZY database (<http://www.cazy.org>, [3]). Protein structures were downloaded from the Protein Data Bank (<http://www.pdb.org>). Protein sequences were retrieved from UniProt (<http://www.uniprot.org>). Assignment of protein domains were obtained from different databases (Pfam, Superfamily, SCOP, PSI-BLAST and CAZY) and the consensus domain organizations are reported in [Table 1](#). Protein structures were analysed and visualized with VMD [15]. Structure alignment was performed with MultiSeq [16] restricting the alignment region to GH20b and GH20 domains only. Sequence alignment was performed with PROMALS [17] with default options. Only the sequence fragments covering GH20b, GH20 domains and extra α -helix were considered.

Results and Discussion

1. Structural domains organization of GH20 proteins

According to CAZY classification [3], GH20 family is mainly composed of both prokaryotic and eukaryotic β -N-acetylhexosaminidases with different substrate specificities. Using data integration from different databases (Pfam, Superfamily, SCOP, PSI-BLAST and CAZY databases), we report in [Table 1](#) the consensus domains organization in GH20 proteins of known structure. The catalytic activity in this family of enzymes is assigned to the GH20 domain (PFAM code PF00728) which typically folds into a $(\beta/\alpha)_8$ -barrel topology. Except for dispersin B from *A. actinomycetemcomitans* that presents only a single GH20 domain, the rest of structures show the catalytic domain accompanied by several domains with quite diverse functionalities: a non-catalytic domain, commonly named as GH20b, which is conserved in most GH20 enzymes although with unknown function, several lectin domains, carbohydrate binding domains, and other domains of unknown function.

Due to this diversity in domain organization, different questions come out: are all the accompanying domains essential for activity? Would the GH20 catalytic domain alone be

Table 1. Domain organization of GH20 β -N-acetylhexosaminidases.

Model A					
Enzyme	Organism	Length (aa)	PDB	Domains ¹	GH20b-GH20- α location
Monomeric					
β -N-acetylhexosaminidase (SpHex)	<i>Streptomyces plicatus</i>	506	1HP5[A]	<u>GH20b-GH20-α</u>	19–494
β -N-acetylhexosaminidase (ScHex)	<i>Streptomyces coelicolor</i> A3(2)	535	4C7G[A]	<u>GH20b-GH20-α</u>	11–511
β -N-acetylhexosaminidase (Hex1T)	<i>Paenibacillus</i> sp.	978	3GH5[A]	<u>GH20b-GH20-α-Lectin-CBD</u>	15–493
Chitobiase (SmCHB)	<i>Serratia marcescens</i>	885	1QBB[A]	<u>CHB_HEX-GH20b-GH20-α-CHB_HEX_C</u>	215–814
Lacto-N-biosidase (LnbB)	<i>Bifidobacterium bifidum</i>	1112	4JAW[A,B]	<u>GH20b-GH20-α-Lectin-CBM32-Ig like</u>	34–515
β -N-acetylhexosaminidase (NahA)*	<i>Arthrobacter aurescens</i>	540	3RCN[A]	<u>GH20b-GH20-</u>	5–506
β -N-acetylhexosaminidase (Bf3009*)	<i>Bacteroides fragilis</i>	518	4PYS[A,B]	<u>GH20b-GH20-α</u>	21–498
Heterodimeric					
β -N-acetylhexosaminidase (HexA)	<i>Homo sapiens</i>	529/556	2GK1 [A,B,C,D,E,F,G]	<u>GH20b-GH20-α</u>	α :23–510 β :56–540
Homodimeric					
β -N-acetylhexosaminidase (HexB)	<i>Homo sapiens</i>	556/556	1NP0[A,B]	<u>GH20b-GH20-α</u>	56–540
β -N-acetylhexosaminidase (OfHex1)	<i>Ostrinia furnacalis</i>	593/593	3OZO[A]	<u>GH20b-GH20-α</u>	65–574
β -N-acetylhexosaminidase (GcnA)	<i>Streptococcus gordonii</i>	627/627	2EPN[A,B]	<u>GH20b-GH20-α-Domain III</u>	2–414
Model B					
Enzyme	Organism	Length (aa)	PDB	Domains ¹	GH20 domain location
Monomeric					
β -1,6-N acetylglucosaminidase Dispersin B (DspB)	<i>Aggregatibacter actinomycetemcomitans</i>	361	1YHT[A]	<u>GH20</u>	20–342
β -N-acetylhexosaminidase (StrH)	<i>Streptococcus pneumoniae</i> R6	1312	3RPM[A,B]	<u>GH20-GH20-G5-G5</u>	190–538 635–972
				<u>GH20-GH20-G5-G5</u>	190–538
				<u>GH20-GH20-G5-G5</u>	635–972
	<i>Streptococcus pneumoniae</i> TIGR4	1312	2YL8[A] 2YL9[A,B,C,D]	<u>GH20-GH20-G5-G5</u>	190–538 635–972

Data obtained using integration from different databases (Pfam, Superfamily, SCOP, PSI-BLAST and CAZY databases). PDB structure accession code and chain identifier used for structural analysis are indicated and domains actually present in the PDB structure are underlined.

¹ **Domains:** domain nomenclature according to the consensus domain organization by integrating Pfam, Superfamily, SCOP, PSI-BLAST and CAZY databases. Preferentially CAZY annotation is used, if available. CBD: carbohydrate binding domain, CBM32: carbohydrate binding module family 32, CHB_HEX: chitobiose hexosaminidase accompanying domain in the N-terminus, CHB_HEX_C: chitobiose hexosaminidase C-terminal accompanying domain, Lectin: lectin-like domain according to PSI-BLAST, Ig like: Immunoglobulin-like fold domain, Domain III: C-terminal domain needed to dimerize according to authors, G5: accompanying domain of carbohydrate metabolism enzymes.

*: Oligomerization state in solution is unknown

doi:10.1371/journal.pone.0128075.t001

functional? Which is the importance of GH20b? Several studies on the structure–function relationships at the active site have been performed [18–22] but only few have addressed the structural organization of these enzymes [23–25].

Regardless of the number of accompanying domains, we propose two different models of organization (Table 1): Model A, comprising proteins with at least both GH20b and GH20 domains, which is characterized by a GH20b-GH20- α architecture, where GH20 is always accompanied by GH20b at the N-terminus and followed by an extra α -helix after the GH20 (β/α)₈-barrel, and Model B with proteins that only present the GH20 domain without GH20b.

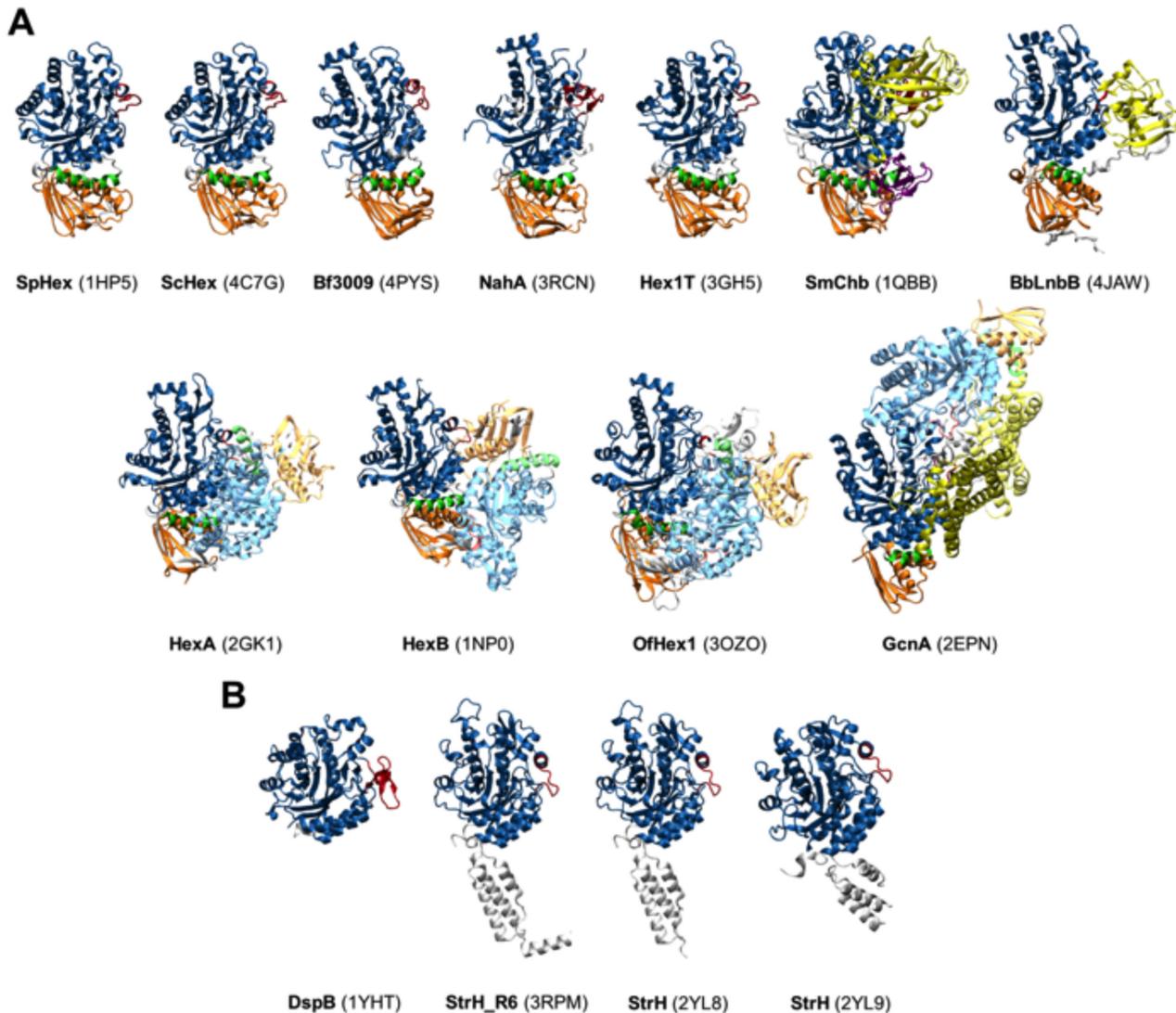


Fig 2. Crystal structures of GH20 β -N-acetylhexosaminidases. (A) Structures according to model A architecture. (B) Structures according to model B architecture. GH20b domains are colored in orange, GH20 domains in blue, α -helix in green, and the rest of accompanying domains in yellow and purple. Loop 2 from GH20 domain is in red. Dimer counterparts follow the same coloring scheme but with lower intensity.

doi:10.1371/journal.pone.0128075.g002

To date, eleven of the thirteen GH20 enzymes structures fit into Model A (Fig 2A). As summarized in Table 1, five of them are monomeric proteins in solution (SpHex, ScHex, Hex1T, SmCHB, and LnbB), whereas other four are dimeric proteins (HexA, Hex B, OfHex1, and GcnA). The oligomerization state of the *Arthrobacter aureescens* (NahA) and *Bacteroides fragilis* (Bf3009) enzymes remains unknown.

For the monomeric proteins, SpHex and ScHex structures, which share 94% sequence identity, comprise the GH20b domain followed by the catalytic GH20 domain [26,27]. Hex1T consists of four domains, GH20b and GH20, and domains III and IV which are described as concanavalin A-like lectin and cellulose-binding domains respectively, according to Superfamily database (Supfam). The Hex1T crystal structure is a truncated form (residues 1–502), with only the GH20b and GH20 domains [25]. SmCHB is also constituted by four domains, where the N-terminal domain (CHB_HEX) is annotated as a carbohydrate binding module although its

function remains unknown, and the C-terminal domain (CHB_HEX_C) as an immunoglobulin-like fold domain (Supfam). The central domains correspond to GH20b and GH20 domains [21]. LnbB is formed by five domains. Domains I and II are the GH20b and GH20 ones, Domain III (residues 546–700) is similar to a lectin domain according to a PSI-Blast search, Domain IV (784–932) has sequence homology to CBM32 (Supfam), and domain V (962–1041) is described as a Immunoglobulin-like domain in Pfam. The recent crystal structure of this enzyme corresponds to a C-terminal truncated form (41–663) with the GH20b, GH20 and part of the lectin-like domain [23]. Finally, NahA and Bf3009, with unknown oligomerization states but included in the group of monomeric proteins in Table 1, present only the GH20b and GH20 domains.

For the dimeric proteins, each subunit of HexA, HexB [28,29] and OfHex1 [30] have only the two domains, GH20b and GH20, while the GcnA structure contains an additional domain III, which is predominantly α -helical and is reported to form a significant part of the dimer interface [31]. In general, these accompanying domains are non-catalytic domains involved in binding and substrate specificity but still the mechanism by which this occurs remains uncertain [32,33].

The remaining two GH20 enzymes present structures that are assigned to Model B (Fig 2B): dispersin B from *Aggregatibacter actinomycetemcomitans* (DspB) and the β -hexosaminidases from *Streptococcus pneumoniae* (StrH). Both are monomeric proteins in solution where DspB presents only one domain, the unique GH20 domain [34], and StrH consists of four domains, where the N-terminus is a tandem repeat of two GH20 domains, designated as GH20-1 and GH20-2, and the C-terminus is a tandem repeat of two domains tentatively designated as G5 with unknown function (Pfam). Truncated forms containing a single GH20-1 and GH20-2 domains of *S. pneumoniae* were crystallized [24]. The GH20-1 and GH20-2 domains are quite similar although specific features confer differential substrate specificities.

2. *B. bifidum* lacto-N-biosidase truncated forms

The full length protein was truncated according to the two different models of structure organization we propose, and three different truncated proteins were designed: construct A, B and B α (Fig 3).

Construct A (residues 40–528) follows the architecture organization of Model A enzymes and is composed of the GH20b and GH20 domains always followed by the extra α -helix. This construct would be expected to be active since several enzymes of model A exhibit just these two domains alone in the wild-type protein (such as SpHex structure and HexA, HexB and OfHex1 subunits). Furthermore, a truncated form of Hex1T (encoding the GH20b-GH20- α domains alone) is reported to be active [25]. However, a previous work by Ito *et al.* (2013) in which different LnbB truncated forms were designed, revealed that a truncated protein (37–520) equivalent to construct A exhibits no catalytic activity. A longer truncated protein (37–663) composed by GH20b-GH20- α and a part of the lectin-like domain was reported to be as active as the full length. These data indicate that the GH20b-GH20- α architecture is not the minimal functional unit for LnbB, as it is for the rest of the model A GH20 enzymes. Further studies are presented to determine the details of the minimum requirement for LnbB to be catalytically active.

Construct B (residues 176–497) mimics Model B enzymes and is formed by the single catalytic GH20 domain alone. Lastly, construct B α (residues 176–528) was designed to tackle the open question whether the accompanying GH20b domain can be removed in Model A enzymes. This construct encodes the catalytic GH20 domain with the extra α -helix at the C-terminus. Both constructs, B and B α , were designed for the study of the behaviour of the GH20 catalytic domain in model A proteins.

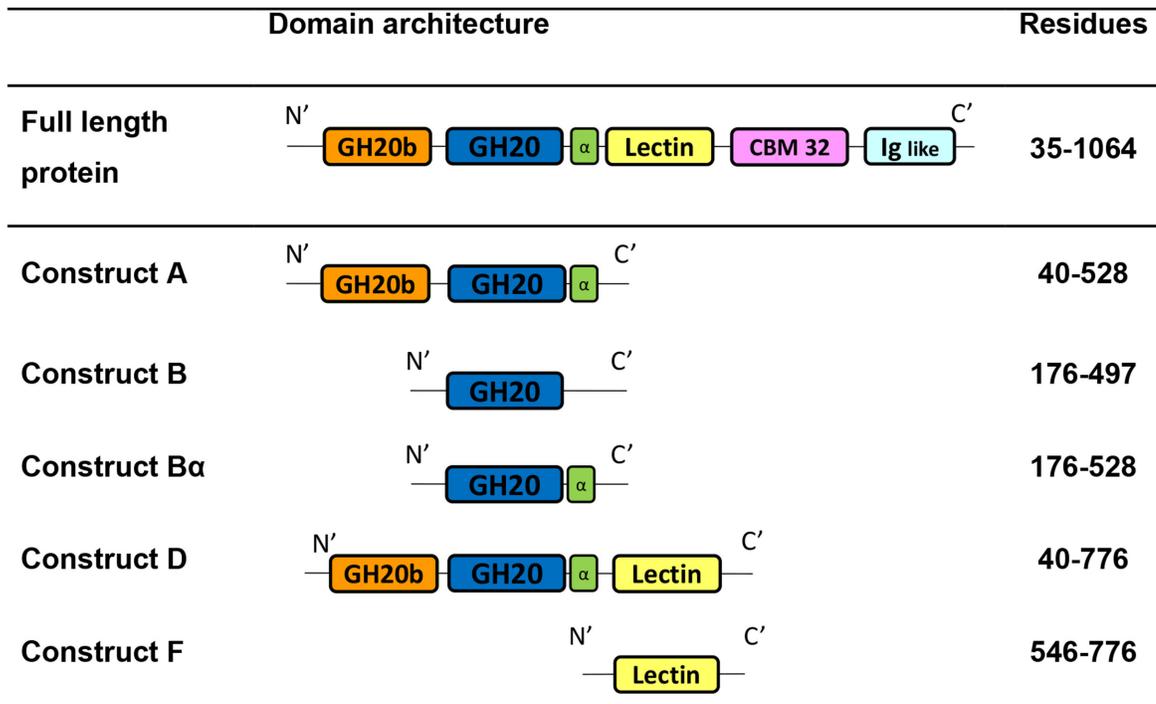


Fig 3. Truncated forms of *B.bifidum* Lacto-N-biosidase.

doi:10.1371/journal.pone.0128075.g003

Engineered proteins of LnbB enzyme were cloned and expressed in *E. coli*. The full length protein (37–1064), without the signal peptide and membrane anchor regions, was expressed as C-terminal His₆-Tagged protein. Constructs A, B and B α were expressed as N-terminal His₆-Tagged proteins. Cultures were grown at 30°C for 24 h with autoinduction medium. The full length protein and construct A were obtained as soluble intracellular proteins with yields of 16 and 27 mg/L of culture respectively. SDS-PAGE analysis and MALDI-TOF mass spectrometry confirmed the theoretical 112 kDa of the full length protein and 57 kDa of Construct A. Further characterization by gel filtration chromatography showed that the full length protein is monomeric with particle size of about 10 nm by dynamic light scattering, whereas Construct A is mainly dimeric in solution (Fig 4). The full length protein was also expressed as N-terminal His₆-Tagged protein to evaluate the effect of the N-terminal tag. Both, C-terminal and N-terminal His₆-Tagged full length proteins, were soluble and had the same specific activity (data not shown).

On the other hand, constructs B and B α were only present in the insoluble pellet and no soluble protein was achieved after different expression conditions such as autoinduction or IPTG induction at different temperatures (data not shown). This seems to indicate that the single GH20 catalytic domain of LnbB, with or without the extra α -helix, is not stable enough and yields insoluble aggregates. Thus, GH20b domain at the N-terminus of GH20 is required for LnbB expression and stability. This domain is present in all model A enzymes being the relative orientation of GH20b and GH20 catalytic domain the same in all structures (Fig 2). The number of β -strands of the GH20b domain varies (for example five in GcnA, six in HexA and HexB and seven in SpHEX and LnbB) but the interface between both domains is well conserved: two loops (residues 54–60 of loop 1 and residues 127–130 of loop 4 in LnbB) and three helices (residues 82–99, 146–163 and the extra α -helix of the GH20 domain which is oriented toward an hydrophobic pocket). GH20b domain is not only found in GH20 enzymes but also in some GH84 hexosaminidases and GH67 glucuronidases annotated in Pfam as GH20b and GH67N

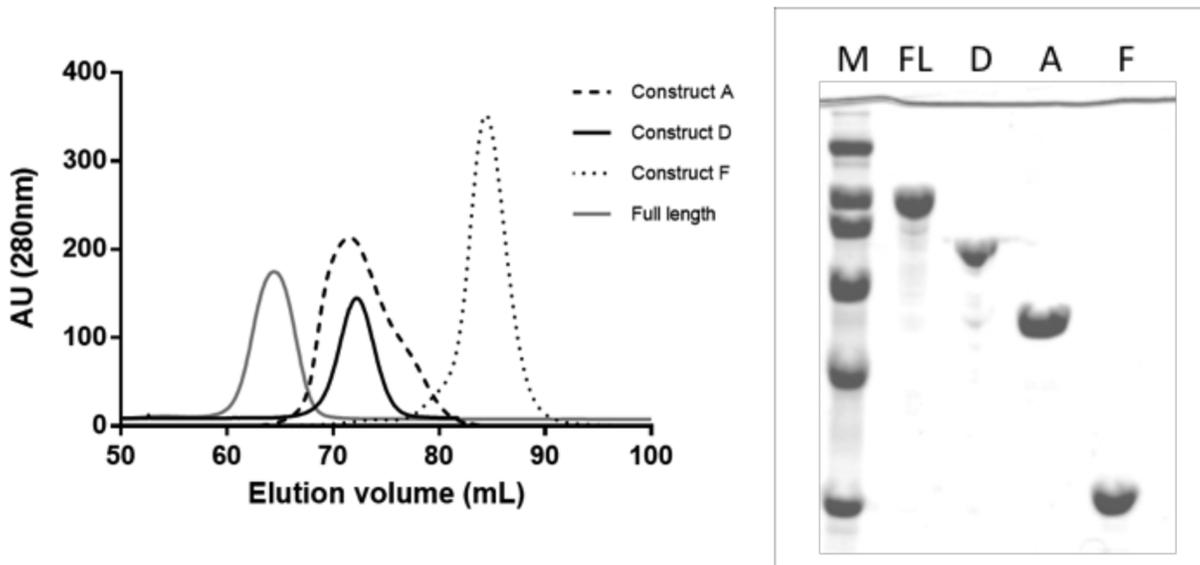


Fig 4. Size exclusion chromatography of the full length protein and different constructs of Lacto-N-biosidase from *B.bifidum*. (A) Chromatogram of the full length protein (FL) and constructs A, D and F. (B) SDS-PAGE of the eluted proteins. M: broad range protein molecular weight marker (200, 116.2, 97.4, 66.2, 45, 31 kDa).

doi:10.1371/journal.pone.0128075.g004

[35], respectively, always located at the N-terminus of the catalytic domain. Considering this scenario, we conclude that GH20b-GH20- α architecture of model A enzymes cannot be reduced to that of model B.

The hydrolytic activity of the soluble constructs was assayed using p-nitrophenyl β -lacto-N-bioside as substrate at pH 4.5 and 30°C. Full length protein showed a specific activity of $27.2 \pm 3.5 \text{ s}^{-1}$, but construct A was inactive, as expected according to previous studies [23]. The GH20b-GH20- α architecture of LnbB is enough to guarantee protein expression, but this architecture does not correspond to the minimal functional unit maybe requiring other accompanying domains from the C-terminus for the enzyme to be active.

3. Structural requirements in the GH20 domain for functionality

While this minimal GH20b-GH20- α architecture assures the activity of SpHex and Hex1T [25,36] and also of dimeric proteins such as HexA, HexB and OfHex1 [30,37,38], it is not enough for LnbB. The crystal structure of LnbB (residues 41–663) in complex with the cyclic intermediate analogue lacto-N-biose-thiazoline [23] consisting of three out of the five domains (GH20b, GH20 and part of the lectin-like domain) was active. The accompanying lectin domain resulted to be the missing element required in our construct A to be fully active. We have carefully analyzed the interaction of the catalytic GH20 domain with other C-terminus domains in all crystallized protein structures of model A (Fig 5). Exploring the structure, important structural features can be assigned. The lectin domain interacts with the catalytic GH20 domain, providing a remote element that folds on the active site (Fig 5A). This element (residues 570–578) consists of a beta-turn, in which Leu574 completes the substrate binding cavity favouring a hydrophobic interaction with the hydroxymethyl group of the galactose moiety of the substrate. In addition, Leu574 is positioned between two conserved amino acids Gln190 and Asp467 of subsites -1 and -2 forming further interactions between this turn and the active site (Fig 5A). For instance, through a water mediated hydrogen bond, the backbone nitrogen atom of Leu574 favours the orientation of Gln190 towards the O4 of the galactose unit of the

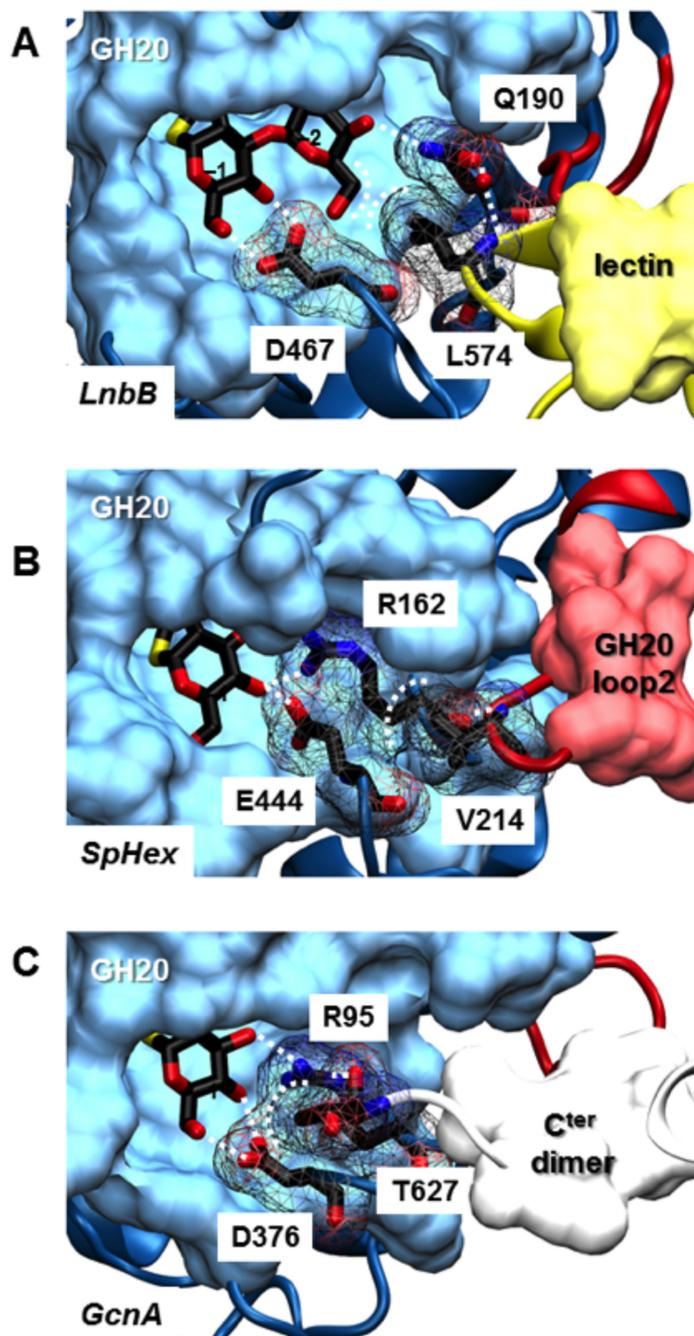


Fig 5. Remote elements and their key residues that complement the active site of GH20 β -N-acetylhexosaminidases. (A) *B. bifidum* lacto-N-biosidase. (B) *S. plicatus*. (C) *S. gordonii* N-acetylhexosaminidases. Dot lines represent interactions between residues of the remote element and the active site as indicated in [Table 2](#).

doi:10.1371/journal.pone.0128075.g005

disaccharide ligand. Engineered forms of LnbB in which part of the C-terminus of the lectin domain were even shorter [23] or completely removed as in our construct A, showed an important decrease in activity of the enzyme, indicating that the interactions we observe in the structure between the remote lectin domain and the substrate binding site are essential for catalysis.

Table 2. Remote element and conserved interactions with the active site of model A GH20 β -N-acetylhexosaminidases.

Enzyme	Loop 2 length	Remote element	GH20 conserved Arg/Asp or Glu at subsite -1	Remote element interactions with conserved Arg/Asp	
				non-polar interactions	polar interactions
SpHex	Long G209-T223	Loop 2	R162	V214	V214
			E444	V214	-
ScHex	Long G238-T252	Loop 2	R191	V243	V243
			E473	V243	-
NahA	Long G193-T219	Loop 2	R146	-	R196
			E456	-	-
Bf3009	Long G200-S214	Loop 2	R153	-	Q209
			D450	-	-
Hex1T	Long G217-T231	Loop 2	R170	V222	V222
			E443	V222	-
SmCHB	Long G396-S425	Loop 2	R349	Y413	R399
			E739	Y413	-
LnbB	Short T231-T236	Lectin G570-T578	Q190*	L574	S233
			D467	L574	-
HexA	Short α G225- α T235	C-terminus subunit β A543-N552	α R178	β A548	β G549
			α E462	β Y547	-
HexA	Short β G258- β T267	C-terminus subunit α A514-E523	β R211	α V519	α G520
			β E491	α V519	-
HexB	Short G258-T267	C-terminus dimer A514-E523	R211	A548	G549
			E491	Y547	-
OfHex1	Short G267-T276	C-terminus dimer A577-S594	R220	P582	E583
			E526	P582	-
GcnA	Short G136-T143	C-terminus dimer S622-T627	R95	-	T627
			D376	T627	-

*Residue in LnbB substituting the function of the conserved arginine but in subsite -2.

doi:10.1371/journal.pone.0128075.t002

Interestingly, we have detected that this pattern is reproduced in all proteins of model A: there is always a remote element assisting in the definition of the substrate binding cavity by means of direct interactions with conserved amino acids directly involved in protein—substrate interactions. Table 2 summarizes the exact localization of these remote elements for each member of model A and the interactions it forms with conserved active site residues. In general, this remote element is provided by a long loop of the GH20 domain itself (loop 2). In the cases where this loop is too short, the remote element comes from an accompanying domain (such as the lectin domain in the case of LnbB) or from the C-terminus of the other monomer in the case of dimeric proteins. In detail, structures of SpHex, ScHex, NahA, Bf3009, SmCHB and Hex1T present this remote element coming from the extended loop 2 of the GH20 domain itself (Fig 5B). This loop positions a hydrophobic residue, such as Val214 in SpHex (equivalent to Leu574 in LnbB), into the second shell of the subsite -1 cavity just behind the conserved arginine and aspartate in all known GH20 hexosaminidases, Arg 162 and Asp444 (equivalent to Gln190 and Asp467 in LnbB, see alignment in Fig 6). These amino acid residues can form hydrophobic interactions between their side chains and also, a hydrogen bond between the

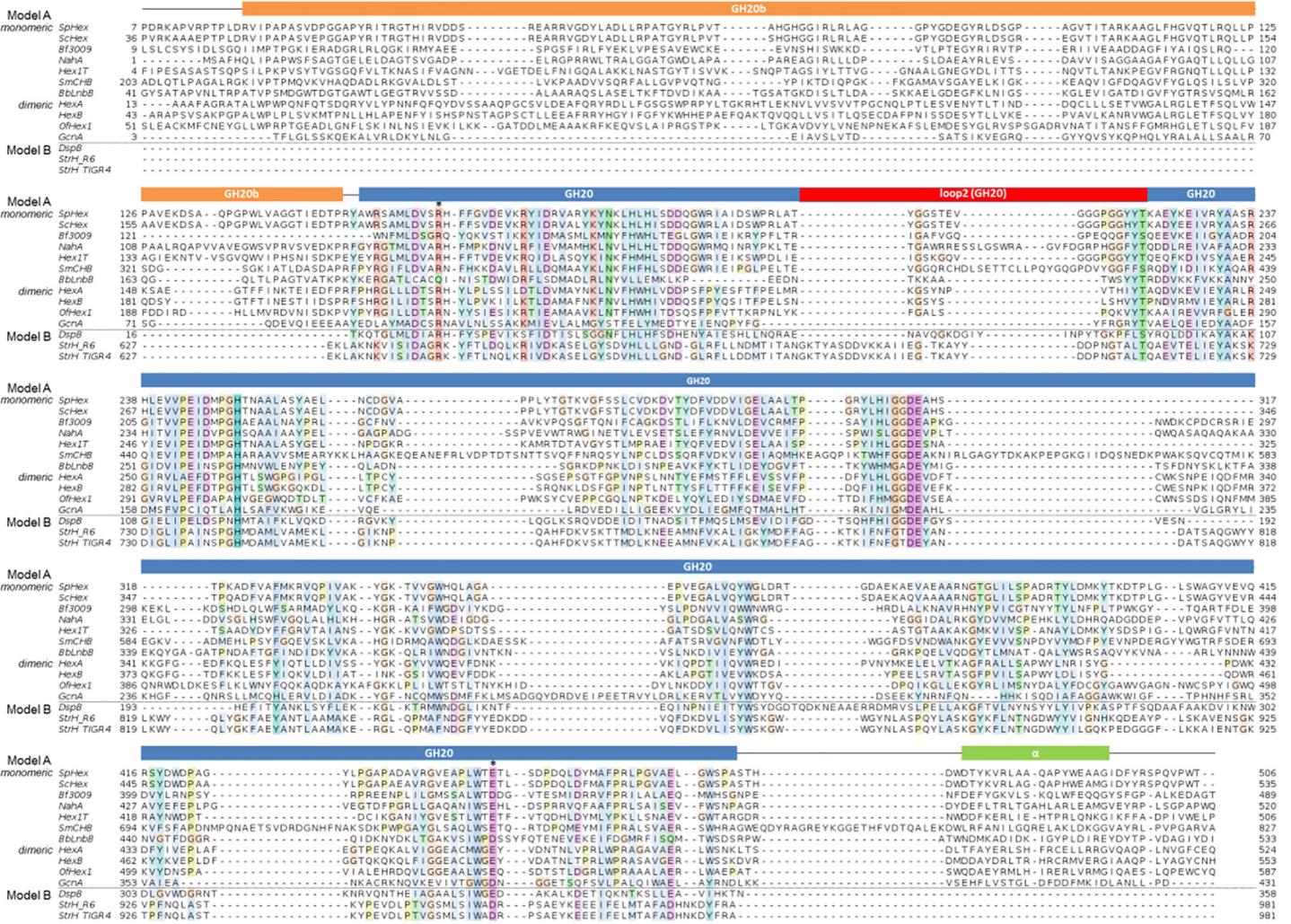


Fig 6. Multiple sequence alignment of β -N-acetylhexosaminidases of known structure from family GH20. Domains distribution is shown as colored boxes on top. Accompanying domains are not shown. Conserved arginine and glutamate/aspartate sites are marked with a star. Sequences were downloaded from UniProt entry names: SpHex (O85361_STRPL), ScHex (Q9L068), Bf3009 (Q5LAT3) NahA (A1RBZ5_ARTAT), Hex1T (D2KW09_9BACL), SmChb (CHB_SERMA), LnbB (B3TL6_BIFBI), HexA (HEXA_HUMAN), HexB (HEXB_HUMAN), OfHex1 (Q06GJ0_OSTFU), GcnA (Q6ST21_STRGN), DspB (Q840G9_AGGAC), StrH_R6 (Q8DR16_STRR6), StrH_TIGR4 (STRH_STRPN). Alignment was performed with PROMALS [17]. Conserved positions are coloured according to ClustalW colour scheme.

doi:10.1371/journal.pone.0128075.g006

backbone NH group of Val214 and the carbonyl group of the backbone of Arg162 (Fig 5B). These interactions between residues of the remote element and the conserved arginine and aspartate of subsite -1, are present in all the structures with long loop 2 (Table 2). ScHex presents Val243 of loop2 next to Arg191 and Asp473 in the same way. NahA and SmChb show the long-loops 2 as seen in the alignment. They provide the remote elements next to the corresponding arginines, Arg146 and Arg349, and aspartates, Asp456 and Asp739 respectively. One arginine of the remote element, Arg196 in NahA and Arg399 in SmChb, is hydrogen-bonded to the backbone of the active site arginine. Similarly, Bf3009 presents Gln209 of loop2 next to Arg153 and Asp450. In Hex1T, Val222 from the remote element is positioned behind Arg170 and Asp443 and the backbone NH group of this valine donates a hydrogen bond to the arginine carbonyl oxygen atom. Interestingly, a truncated form of Hex1T with only the GH20b and GH20 domains has been reported to be active [25]. Consistently with our explanations,

this truncated form does not affect the definition of the active site, because this construct keeps the long loop 2 in the GH20 domain.

By contrast, when loop 2 is short, this remote element comes from the accompanying domains, as in LnbB, or from the C-terminus domains of the other monomer in dimeric structures, such as GcnA, HexA, HexB and OfHex1. For example, in the homodimeric structure of GcnA, domain III forms a significant part of the dimer interface [31]. The C-terminus of domain III of chain B (622–627) winds back into the structure and contacts the catalytic GH20 of chain A (Fig 5C). Specifically, the C-terminus amino acid residue Thr627 projects into subsite -1 and it is adjacent to the conserved arginine and aspartate (Arg95, Asp376) and is part of the active-site pocket of the dimer partner. Besides, the carboxyl group of this C-terminal Thr627 and the Arg95 side chain are at less than 3 Å distance, establishing electrostatic interactions. As in LnbB, any truncated form of this C-terminus domain III is believed to be inactive. In the dimeric structures of HexA, HexB and OfHex1 [28–30], dimerization occurs exclusively between the GH20 catalytic domains. Dimerization is essential for catalysis since residues of one subunit structurally complete the active site of the other one. For example, Tyr456 of subunit β is part of the active site of subunit α in HexA. In addition, the C-terminus of subunit β provides this remote element very close to the subsite-binding cavity of subunit α and vice versa (Table 2). Particularly, the hydrophobic residue Ala548 (subunit β) of this element in HexA forms part of the second shell of the cavity behind the conserved Arg178 (subunit α) and Asp462 (subunit α), and the backbone nitrogen atom of Gly549 (subunit β) interacts by hydrogen bond with the carbonyl oxygen atom of Arg178 (subunit α). The equivalent interactions between subunits are also seen in HexB and OfHex1 crystal structures (Table 2).

Therefore, all GH20 structures of model A clearly show the presence of a remote element folding on the substrate-binding cavity providing amino acid residues that either complete the active site, as in the extended subsite -1 of the *S. gordonii* hexosaminidase or in subsite -2 of the LnbB, or participate in the second shell of this cavity behind the conserved arginine and aspartate residues in subsite -1 of hexosaminidases.

To provide experimental evidence of the requirement of this remote element in LnbB, a complementation assay was designed. If Construct A (composed of the GH20b and GH20-α domains) is inactive because of the lack of the remote element, would addition of the lectin domain (as a single protein) reconstitute the active site and restore activity? Toward this end, Construct F (the single lectin domain, Fig 3) to assay complementation with Construct A, as well as Construct D (Fig 3), which contains the non-catalytic GH20b domain, the catalytic GH20 domain and the lectin domain, were expressed. The C-terminal His₆-Tagged proteins were purified and analyzed by SDS-PAGE and MALDI-TOF, confirming the corresponding theoretical molecular masses of 83 and 28 kDa for constructs D and F, respectively. Gel filtration chromatography showed that both proteins were monomers in solution (Fig 4). Construct D was active with the same specific activity than the full length enzyme on p-nitrophenyl β-lactobioside (27 s⁻¹). Therefore, both C-terminal domains in LnbB, the CBM32 and Ig like domains are not important for the catalytic activity, and construct D is the minimal functional unit of LnbB. Next, a complementation experiment was assayed by adding Construct F (the single lectin domain) on the inactive Construct A. Activity was restored in a concentration-dependent manner when increasing the concentrations of Construct F (Fig 7). At a molar ratio of 1:6 activity levels off at 11% of the hydrolytic activity of the full length enzyme. The interaction between constructs A and F at a 1:6 molar ratio was analyzed by gel filtration chromatography (Fig 8). Fractions L1–L2 correspond to the excess of Construct F, the lectin domain with molecular weight of 28 kDa, and fractions C1–C6 to proteins with molecular weight of 74 kDa that could be the dimer of construct A (as in Fig 4) and the complex between A and F with theoretical molecular mass of 85 kDa. The hydrolase activity of Fractions C1–C6 and the presence of

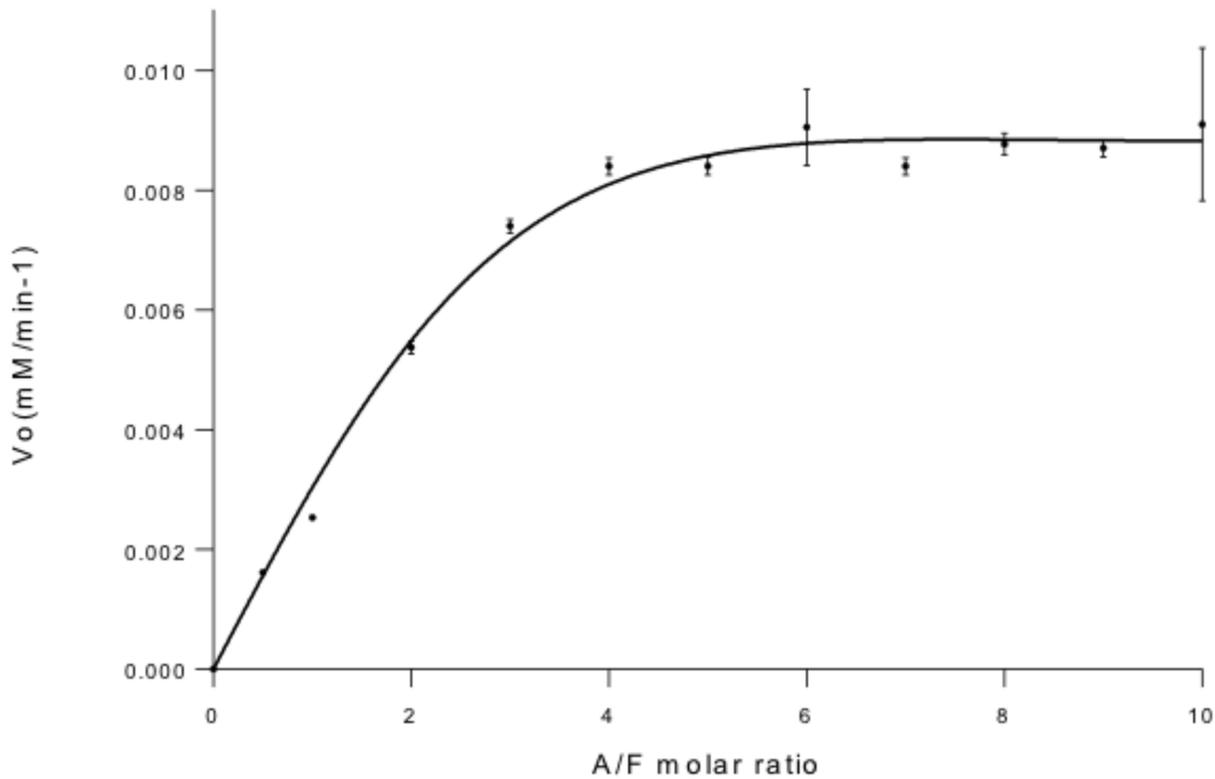


Fig 7. Complementation assay of construct A with construct F at different molar ratios. Conditions: 50 nM construct A, 25 nM-500 nM construct F, 0.25 mM p-nitrophenyl β -lacto-N-bioside, 25 mM citrate-25 mM phosphate buffer, pH 4.5, 30°C. The relative standard deviation did not overcome 2% except at 6 and 10 molar ratio that was lower than 15%.

doi:10.1371/journal.pone.0128075.g007

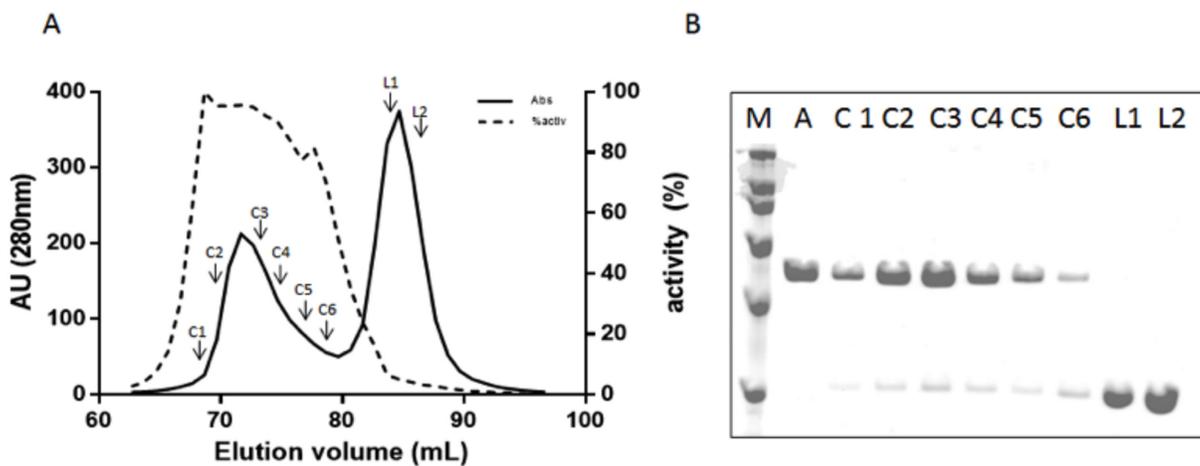


Fig 8. Size exclusion chromatography of the complementation assay in a ratio 1:6 of constructs A and F. (A) Chromatogram of the eluted fractions: absorbance at 280nm (-), hydrolase activity (•••). (B) SDS-PAGE of different fractions. M: broad range protein molecular weight marker (200, 116.2, 97.4, 66.2, 45, 31 kDa); A: construct A Standard, fractions C1–C6, fractions L1–L2. The relative standard deviations for the molecular weights and activities were less than 2% and 15% respectively.

doi:10.1371/journal.pone.0128075.g008

lectin in these fractions seen in SDS-PAGE (Fig 8) demonstrate that the complex between A and F is formed and active. As it is observed in Fig 8, not all construct A protein is complexed with construct F. By densitometry analysis we determined that the lectin represents 10% of total protein in the mixture. Assuming a 1:1 interaction, 10% of Construct A is in complex with Construct F and therefore the specific activity of the complex is similar to that of the full length protein. To exclude non-specific interactions, complementation assay was performed with Construct A and BSA at different concentration ratios and no activity recovery was found (data not shown).

Conclusions

This structural-functional analysis allows us to come to different conclusions. First, we identified two levels of domain organization in GH20 enzymes: model A with GH20b-GH20- α architecture and model B with the single catalytic GH20 domain alone. As LnbB constructs show, model A enzymes cannot be reduced to model B, since GH20b domain is a structural requirement in the N-terminus of GH20 to assure protein stability. Secondly, GH20b-GH20- α architecture is not always the minimal functional unit. For the LnbB enzyme here studied, the isolated GH20b-GH20- α domain is inactive and requires the lectin domain which provides an essential loop to shape the active site. Based on these results and analysis of the structures of GH20 enzymes, we propose a broader mechanism by which important interactions in the substrate binding cavity of the catalytic GH20 domain are provided by a remote element in all structures, which must be preserved to ensure a proper definition of the active site. The remote element can directly be provided by a long loop 2 of the GH20 domain itself. In this case, GH20b-GH20- α can act as a minimal functional unit, and large multidomain enzymes can be engineered to reduced forms with a minimal model A architecture. In other circumstances, when loop 2 is short, proteins dimerize or present an accompanying domain in the C-terminus of GH20 that provide this remote element. In the latter cases, protein engineering of these enzymes must be designed maintaining the accompanying domains. The proposed model can stimulate further studies on other members of the superfamily.

Author Contributions

Conceived and designed the experiments: CVC MF XB AP. Performed the experiments: CVC MF XB. Analyzed the data: CVC MF XB AP. Contributed reagents/materials/analysis tools: CVC MF XB AP. Wrote the paper: CVC MF XB AP.

References

1. Liu T, Yan J, Yang Q. Comparative biochemistry of GH3, GH20 and GH84 beta-N-acetyl-D-hexosaminidases and recent progress in selective inhibitor discovery. *Curr Drug Targets*. 2012; 13: 512–525. PMID: [22280348](#)
2. Slamova K, Bojarova P, Petraskova L, Kren V. beta-N-acetylhexosaminidase: what's in a name...? *Biotechnol Adv*. 2010; 28: 682–693. doi: [10.1016/j.biotechadv.2010.04.004](#) PMID: [20438826](#)
3. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The carbohydrate-active enzymes database (CAZy): an expert resource for glycomics. *Nucl Acids Res*. 2008; 37: D233–238. doi: [10.1093/nar/gkn663](#) PMID: [18838391](#)
4. Liu T, Chen L, Ma Q, Shen X, Yang Q. Structural Insights into Chitinolytic Enzymes and Inhibition Mechanisms of Selective Inhibitors. *Curr Pharm Des*. 2013.
5. Rountree JS, Butters TD, Wormald MR, Boomkamp SD, Dwek RA, Asano N et al. Design, synthesis, and biological evaluation of enantiomeric beta-N-acetylhexosaminidase inhibitors LABNAc and DABNAc as potential agents against Tay-Sachs and Sandhoff disease. *ChemMedChem*. 2009; 4: 378–392. doi: [10.1002/cmdc.200800350](#) PMID: [19145603](#)
6. Bode L. Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology*. 2012; 22: 1147–1162. doi: [10.1093/glycob/cws074](#) PMID: [22513036](#)

7. Garrido D, Kim JH, German JB, Raybould HE, Mills DA. Oligosaccharide binding proteins from *Bifidobacterium longum* subsp. *infantis* reveal a preference for host glycans. *PLoS One*. 2011; 6: e17315. doi: [10.1371/journal.pone.0017315](https://doi.org/10.1371/journal.pone.0017315) PMID: [21423604](https://pubmed.ncbi.nlm.nih.gov/21423604/)
8. Sela DA, Mills DA. Nursing our microbiota: molecular linkages between bifidobacteria and milk oligosaccharides. *Trends Microbiol*. 2010; 18: 298–307. doi: [10.1016/j.tim.2010.03.008](https://doi.org/10.1016/j.tim.2010.03.008) PMID: [20409714](https://pubmed.ncbi.nlm.nih.gov/20409714/)
9. Kobata A. Structures and application of oligosaccharides in human milk. *Proc Jpn Acad Ser B Phys Biol Sci*. 2010; 86: 731–747. PMID: [20689231](https://pubmed.ncbi.nlm.nih.gov/20689231/)
10. Kunz C, Rudloff S, Baier W, Klein N, Strobel S. Oligosaccharides in human milk: structural, functional, and metabolic aspects. *Annu Rev Nutr*. 2000; 20: 699–722. PMID: [10940350](https://pubmed.ncbi.nlm.nih.gov/10940350/)
11. Hattie M, Debowski AW, Stubbs KA. Development of tools to study lacto-N-biosidase: an important enzyme involved in the breakdown of human milk oligosaccharides. *Chembiochem*. 2012; 13: 1128–1131. doi: [10.1002/cbic.201200135](https://doi.org/10.1002/cbic.201200135) PMID: [22514018](https://pubmed.ncbi.nlm.nih.gov/22514018/)
12. Wada J, Ando T, Kiyohara M, Ashida H, Kitaoka M, Yamaguchi M et al. *Bifidobacterium bifidum* lacto-N-biosidase, a critical enzyme for the degradation of human milk oligosaccharides with a type 1 structure. *Appl Environ Microbiol*. 2008; 74: 3996–4004. doi: [10.1128/AEM.00149-08](https://doi.org/10.1128/AEM.00149-08) PMID: [18469123](https://pubmed.ncbi.nlm.nih.gov/18469123/)
13. Studier FW. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif*. 2005; 41: 207–234. PMID: [15915565](https://pubmed.ncbi.nlm.nih.gov/15915565/)
14. Smith PK, Krohn RI, Hermanson GT, Mallia AK, Gartner FH, Provenzano MD et al. Measurement of protein using bicinchoninic acid. *Anal Biochem*. 1985; 150: 76–85. PMID: [3843705](https://pubmed.ncbi.nlm.nih.gov/3843705/)
15. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*. 1996; 14: 33–38. PMID: [8744570](https://pubmed.ncbi.nlm.nih.gov/8744570/)
16. Roberts E, Eargle J, Wright D, Luthey-Schulten Z. MultiSeq: unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics*. 2006; 7: 382. PMID: [16914055](https://pubmed.ncbi.nlm.nih.gov/16914055/)
17. Pei J, Grishin NV. PROMALS: towards accurate multiple sequence alignments of distantly related proteins. *Bioinformatics*. 2007; 23: 802–808. PMID: [17267437](https://pubmed.ncbi.nlm.nih.gov/17267437/)
18. Pluvinage B, Higgins MA, Abbott DW, Robb C, Dalia AB, Deng L et al. Inhibition of the pneumococcal virulence factor StrH and molecular insights into N-glycan recognition and hydrolysis. *Structure*. 2011; 19: 1603–1614. doi: [10.1016/j.str.2011.08.011](https://doi.org/10.1016/j.str.2011.08.011) PMID: [22078560](https://pubmed.ncbi.nlm.nih.gov/22078560/)
19. Sumida T, Stubbs KA, Ito M, Yokoyama S. Gaining insight into the inhibition of glycoside hydrolase family 20 exo-beta-N-acetylhexosaminidases using a structural approach. *Org Biomol Chem*. 2012; 10: 2607–2612. doi: [10.1039/c2ob06636j](https://doi.org/10.1039/c2ob06636j) PMID: [22367352](https://pubmed.ncbi.nlm.nih.gov/22367352/)
20. Williams SJ, Mark BL, Vocadlo D, James MNG, Withers SG. Aspartate 313 in the *Streptomyces plicatus* Hexosaminidase Plays a Critical Role in Substrate-assisted Catalysis by Orienting the 2-Acetamido Group and Stabilizing the Transition State. *Biol Chem*. 2002; 277 (42): 40055–40065. PMID: [12171933](https://pubmed.ncbi.nlm.nih.gov/12171933/)
21. Prag G, Papanikolaou Y, Tavlas G, Vorgias CE, Petratos K, Oppenheim AB. Structures of chitobiase mutants complexed with the substrate Di-N-acetyl-d-glucosamine: the catalytic role of the conserved acidic pair, aspartate 539 and glutamate 540. *J Mol Biol*. 2000; 300: 611–617. PMID: [10884356](https://pubmed.ncbi.nlm.nih.gov/10884356/)
22. Mark BL, Vocadlo DJ, Knapp S, Triggs-Raine BL, Withers SG, James MN. Crystallographic evidence for substrate-assisted catalysis in a bacterial beta-hexosaminidase. *J Biol Chem*. 2001; 276: 10330–10337. PMID: [11124970](https://pubmed.ncbi.nlm.nih.gov/11124970/)
23. 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. doi: [10.1074/jbc.M112.420109](https://doi.org/10.1074/jbc.M112.420109) PMID: [23479733](https://pubmed.ncbi.nlm.nih.gov/23479733/)
24. Jiang YL, Yu WL, Zhang JW, Frolet C, Di Guilmi AM, Zhou CZ et al. Structural basis for the substrate specificity of a novel beta-N-acetylhexosaminidase StrH protein from *Streptococcus pneumoniae* R6. *J Biol Chem*. 2011; 286: 43004–43012. doi: [10.1074/jbc.M111.256578](https://doi.org/10.1074/jbc.M111.256578) PMID: [22013074](https://pubmed.ncbi.nlm.nih.gov/22013074/)
25. Sumida T, Ishii R, Yanagisawa T, Yokoyama S, Ito M. Molecular cloning and crystal structural analysis of a novel beta-N-acetylhexosaminidase from *Paenibacillus* sp. TS12 capable of degrading glycosphingolipids. *J Mol Biol*. 2009; 392: 87–99. doi: [10.1016/j.jmb.2009.06.025](https://doi.org/10.1016/j.jmb.2009.06.025) PMID: [19524595](https://pubmed.ncbi.nlm.nih.gov/19524595/)
26. Mark BL, Vocadlo DJ, Zhao D, Knapp S, Withers SG, James MN. Biochemical and structural assessment of the 1-N-azasugar GalNAc-isofagomine as a potent family 20 beta-N-acetylhexosaminidase inhibitor. *J Biol Chem*. 2001; 276: 42131–42137. PMID: [11522797](https://pubmed.ncbi.nlm.nih.gov/11522797/)
27. Thi NN, Offen WA, Shareck F, Davies GJ, Doucet N. Structure and activity of the *Streptomyces coelicolor* A3(2) beta-N-acetylhexosaminidase provides further insight into GH20 family catalysis and inhibition. *Biochemistry*. 2014; 53: 1789–1800. doi: [10.1021/bi401697j](https://doi.org/10.1021/bi401697j) PMID: [24559145](https://pubmed.ncbi.nlm.nih.gov/24559145/)
28. Lemieux MJ, Mark BL, Cherney MM, Withers SG, Mahuran DJ, James MN. Crystallographic structure of human beta-hexosaminidase A: interpretation of Tay-Sachs mutations and loss of GM2 ganglioside hydrolysis. *J Mol Biol*. 2006; 359: 913–929. PMID: [16698036](https://pubmed.ncbi.nlm.nih.gov/16698036/)

29. Mark BL, Mahuran DJ, Cherney MM, Zhao D, Knapp S, James MN. Crystal structure of human beta-hexosaminidase B: understanding the molecular basis of Sandhoff and Tay-Sachs disease. *J Mol Biol.* 2003; 327: 1093–1109. PMID: [12662933](#)
30. Liu T, Zhang H, Liu F, Wu Q, Shen X, Yang Q. Structural determinants of an insect beta-N-Acetyl-D-hexosaminidase specialized as a chitinolytic enzyme. *J Biol Chem.* 2011; 286: 4049–4058. doi: [10.1074/jbc.M110.184796](#) PMID: [21106526](#)
31. Langley DB, Harty DW, Jacques NA, Hunter N, Guss JM, Collyer CA. Structure of N-acetyl-beta-D-glucosaminidase (GcnA) from the endocarditis pathogen *Streptococcus gordonii* and its complex with the mechanism-based inhibitor NAG-thiazoline. *J Mol Biol.* 2008; 377: 104–116. doi: [10.1016/j.jmb.2007.09.028](#) PMID: [18237743](#)
32. Cuskin F, Flint JE, Gloster TM, Morland C, Basle A, Henrissat B et al. How nature can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic specificity. *Proc Natl Acad Sci U S A.* 2012; 109: 20889–20894. doi: [10.1073/pnas.1212034109](#) PMID: [23213210](#)
33. Drickamer K, Fadden AJ. Genomic analysis of C-type lectins. *Biochem Soc Symp.* 2002;59–72. PMID: [12655774](#)
34. Ramasubbu N, Thomas LM, Rangunath C, Kaplan JB. Structural analysis of dispersin B, a biofilm-releasing glycoside hydrolase from the periodontopathogen *Actinobacillus actinomycetemcomitans*. *J Mol Biol.* 2005; 349: 475–486. PMID: [15878175](#)
35. Nurizzo D, Nagy T, Gilbert HJ, Davies GJ. The structural basis for catalysis and specificity of the *Pseudomonas cellulosa* alpha-glucuronidase, GlcA67A. *Structure.* 2002; 10: 547–556. PMID: [11937059](#)
36. Mark BL, Wasney GA, Salo TJ, Khan AR, Cao Z, Robbins PW et al. Structural and functional characterization of *Streptomyces plicatus* beta-N-acetylhexosaminidase by comparative molecular modeling and site-directed mutagenesis. *J Biol Chem.* 1998; 273: 19618–19624. PMID: [9677388](#)
37. Gravel RA, Clarke JTR, Kaback MM, Mahuran D, Sandhoff K, Suzuki K. In: *The metabolic basis of inherited disease*, (Eds. Scriver CV, Beaudet AL, Sly WS, and Valle D). New York: McGraw-Hill, 1995. 2839–79.
38. Mahuran D, Novak A, Lowden JA. The lysosomal hexosaminidase isozymes. *Isozymes Curr Top Biol Med Res.* 1985; 12: 229–288. PMID: [3886595](#)