



Review

Structural Insights into the Molecular Evolution of the Archaeal Exo- β -D-Glucosaminidase

Shouhei Mine ^{1,*} and Masahiro Watanabe ²

¹ Biomedical Research Institute (BMD), National Institute of Advanced Industrial Science and Technology (AIST), 1-8-31 Midorigaoka, Ikeda, Osaka 563-8577, Japan

² Research Institute for Sustainable Chemistry (ISC), AIST, 3-11-32 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-0046, Japan; masa-watanabe@aist.go.jp

* Correspondence: s-mine@aist.go.jp; Tel.: +81-72-751-9628

Received: 5 April 2019; Accepted: 16 May 2019; Published: 18 May 2019



Abstract: The archaeal exo- β -D-glucosaminidase (GlmA), a thermostable enzyme belonging to the glycosidase hydrolase (GH) 35 family, hydrolyzes chitosan oligosaccharides into monomer glucosamines. GlmA is a novel enzyme in terms of its primary structure, as it is homologous to both GH35 and GH42 β -galactosidases. The catalytic mechanism of GlmA is not known. Here, we summarize the recent reports on the crystallographic analysis of GlmA. GlmA is a homodimer, with each subunit comprising three distinct domains: a catalytic TIM-barrel domain, an α/β domain, and a β 1 domain. Surprisingly, the structure of GlmA presents features common to GH35 and GH42 β -galactosidases, with the domain organization resembling that of GH42 β -galactosidases and the active-site architecture resembling that of GH35 β -galactosidases. Additionally, the GlmA structure also provides critical information about its catalytic mechanism, in particular, on how the enzyme can recognize glucosamine. Finally, we postulate an evolutionary pathway based on the structure of an ancestor GlmA to extant GH35 and GH42 β -galactosidases.

Keywords: exo- β -D-glucosaminidase; archaea; β -galactosidase; evolution

1. Introduction

Glucosamine (GlcN) has an array of biological functions and is widely used as a food additive as well as in medicines. GlcN can be enzymatically produced from chitin, which is an abundant bioresource broadly distributed in nature as a major structural component of fungal cell walls, insect exoskeletons, and crustacean shells. Chitin is a β -1,4-linked *N*-acetylglucosamine polysaccharide (GlcNAc)_n, and its enzymatic conversion to GlcN has become attractive in the chemical industry because it opens a new route for achieving sustainable glucosamine production.

The unique chitin catabolic pathway of hyperthermophilic archaea differs from the known pathways found in other organisms and has been described in *Thermococcus kodakaraensis* KOD1 [1–3]. In this pathway, chitin is first degraded into diacetylchitobiose [(GlcNAc)₂] by chitinase (ChiA) (EC 3.2.1.14), and the acetyl group of the nonreducing side of (GlcNAc)₂ is deacetylated by a deacetylase (Dac) (EC 3.5.1.105). The resulting product, GlcN-GlcNAc, is subsequently hydrolyzed into GlcN and GlcNAc by an exo- β -D-glucosaminidase (GlmA) (EC 3.2.1.165), followed by further deacetylation of the remaining GlcNAc to GlcN by Dac. These enzymes are thermostable, with an optimal temperature of ~80 °C, which is an important requisite for industrial applications since most industrial processes are conducted under harsh conditions (e.g., high temperature and pressure). Previous determination of the chemical structures of ChiA and Dac provided insights into their catalytic mechanism and adaptation to extremely high temperatures [4–10]. However, for almost 14 years after the first description of GlmA, its structure has remained unknown.

According to the Carbohydrate-Active Enzymes [CAZy] database, which bases its predictions on the amino acid sequence similarity [11], GlmA belongs to the glycoside hydrolase (GH) 35 family. The other exo- β -D-glucosaminidases found in bacteria and Eukaryota belong to the GH2 [12] and the GH9 [13] families, respectively, and they show little to no sequence similarity to GlmA. Although more than 150 GH families have been classified in the CAZy database, GlmA is almost unique in its sequence, as it presents sequence homology to both GH35 and GH42 β -galactosidases (EC 3.2.1.23) despite its lack of β -galactosidase activity [2,14]. GlmA can hydrolyze various chain lengths of chitoooligosaccharides (GlcN₂₋₆), cellobiose, and laminaribiose [2]; however, these activities have not been reported for GH35 and GH42 β -galactosidases. Strikingly, the highly conserved motifs around the catalytic residues of these β -galactosidases are not conserved in GlmA [2]. Therefore, it is impossible to predict the key amino acids involved in substrate binding and catalysis of GlmA only from sequence comparisons among these enzymes.

To address this critical question, we determined the crystal structure of GlmA_{TK} (encoded by the TK1754 gene) from *Thermococcus kodakaraensis* KOD1 [15]. The crystal structures of two proteins, which are highly homologous to GlmA_{TK}, GlmA_{Ph} (encoded by the PH0511 gene) [16] and GlmA_{Pf} (encoded by the PF0363 gene) [14], from the closely related hyperthermophiles *Pyrococcus* species *Pyrococcus horikoshii* and *Pyrococcus furiosus*, respectively, were also determined. The structure of GlmA elucidated the substrate-binding site as well as the substrate selection mechanism. It also revealed that GlmA is a structurally interesting intermediate between GH35 and GH42 β -galactosidases. Here, we review the most recent findings on the structure–function relationship of GlmA and describe the unique structural features that link it to the molecular evolution of glycoside hydrolases.

2. Structure and Thermostability of GlmA

The structure of GlmA_{Ph} was deduced using the single-wavelength anomalous dispersion of selenomethionine atoms and refined at 2.60-Å resolution (PDB 5GSL) [15]. The structure of GlmA_{Pf} and GlcN-bound GlmA_{TK} was determined at 1.75-Å resolution (PDB 6JOW, unpublished) and 1.27-Å resolution (PDB 5GSM) [15], respectively, using molecular replacement of the structure of the GlmA_{Ph} monomer as the search model. The structures of GlmA_{Ph} and GlmA_{Pf} showed little variation to that of GlmA_{TK}, as reflected in the RMSD values of 0.90 Å for 767 C α atoms and 0.74 Å for 751 C α atoms, respectively (Figure 1A). Moreover, both proteins shared high sequence identity with GlmA_{TK} (63%) and the active site architecture is fully conserved among these GlmAs (see Section 4.1). These results suggest that general aspects of these proteins, such as the structural features and the catalytic mechanisms, are very likely to be equivalent. GlmA_{Pf} has been described as a putative β -galactosidase [2,14]. However, structural analysis results indicate that it must be an exo- β -D-glucosaminidase. The highest-resolution complex structure of GlmA_{TK} is described throughout this review unless otherwise noted.

GlcN-bound GlmA_{TK} is a homodimer and each monomer (chains A and B) consists of three distinct domains (Figure 1B). Domain I (residues 1–435) is a TIM-barrel structure typical of the GH family. Generally, it contains the catalytic residues [17,18]. As expected, a single molecule of GlcN is located in the bottom of each monomer's barrel (Figure 1B). Domain II (residues 436–648) is an α/β domain involved in the dimerization process and forms an interface with the TIM-barrel domain of the other monomer. Domain III (residues 649–786) is a β 1 domain. There is no structural evidence that this domain is involved in protein activity, but it might contribute to maintaining the overall structural conformation of GlmA_{TK}. Indeed, Arg676 from this domain forms hydrogen bonds with His354 and Thr355 from the TIM-barrel domain of the neighboring polypeptide.

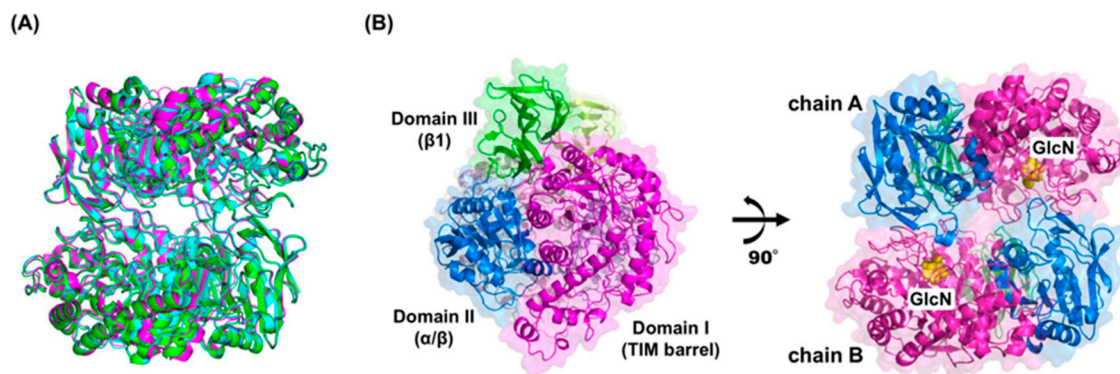


Figure 1. The overall structure of GlmA_{Tk}. (A) The structural superposition of GlmA_{Tk} (magenta), GlmA_{Ph} (cyan), and GlmA_{Pf} (green); (B) The dimer structure of GlmA_{Tk} is presented in two views. GlmA_{Tk} consists of a homodimer (chains A and B) and comprises three distinct domains (TIM-barrel: magenta, α/β : blue, $\beta 1$: green). The bound GlcN is represented by yellow van der Waals spheres. The figures were prepared using PyMOL (Schrödinger, LLC, New York, NY, USA).

To date, physical and chemical features have been proposed to explain the enhanced protein thermostability [19]. Among them, oligomerization has been considered a form of adaptation to extreme temperatures due to the increase in the number of intermolecular interactions [20–22]. For GlmA_{Tk}, the buried solvent-accessible surface area upon dimer formation is 5530 Å², i.e., 24% of the monomer surface, which is quite large for a dimerization interface. The dimer of GlmA_{Tk} is held together by numerous interactions at the subunit interface, involving 29 hydrogen bonds and 16 salt bridges per monomer and interactions between at least 36 residues at each monomer interface. Since GlmA_{Tk} exhibited its highest activity at 80 °C toward GlcN₂ with an k_{cat}/K_m value of 0.1 $\mu\text{M}^{-1} \text{s}^{-1}$ [2], this stability may be caused by the cumulative effect of these interactions, which may also contribute to the rigidity of the dimer. Recently, the cold-adapted β -galactosidase from *Paracoccus* sp. 32d, Par β DG, a member of the GH2 family, has been reported [23]. Although Par β DG forms a stable dimer, the average B-factor values of Par β DG (37.0 Å²) are much higher than those of GlmA_{Tk} (14.0 Å²), indicating that Par β DG has a high degree of flexibility in comparison to that of GlmA_{Tk}. This result shows that a global conformational rigidity of GlmA_{Tk} is indeed associated with thermostability.

3. Structural Comparison with GlmA Homologous Proteins

Bioinformatics analyses using the Dali server [24], which identifies global structural homologs, revealed that the dimer structure of GlmA_{Tk} does not resemble that of any other protein. However, the three-domain structure of the GlmA_{Tk} monomer has the same domain order as the GH42 β -galactosidase, although GlmA_{Tk} is actually a GH35 enzyme. To date, five crystal structures of GH42 β -galactosidases have been reported, including those from *Thermus thermophilus* A4- β -gal (PDB 1KWK) [25], *Bacillus circulans* sp. *alkalophilus* Bca- β -gal (PDB 3TTY) [26], *Geobacillus stearothermophilus* GanB (PDB 4OIF) [27], *Bifidobacterium animalis* BiGal42A (PDB 4UNI) [28], and *Bifidobacterium bifidum* S17 BbgII (PDB 4UZS) [29]. GlmA_{Tk} and these GH42 β -galactosidases only share 15–17% sequence identity, but their monomer structures could be superimposed with RMSD values of 2.6–3.0 Å for equivalent C α atoms, except for 80 additional residues at the C-terminal region of GlmA_{Tk} (Figure 2A,B). Yet, an important difference is present in the oligomerization state: GH42 β -galactosidases form a homotrimer; thus, their overall structures are quite different from that of GlmA_{Tk} (Figure 2C).

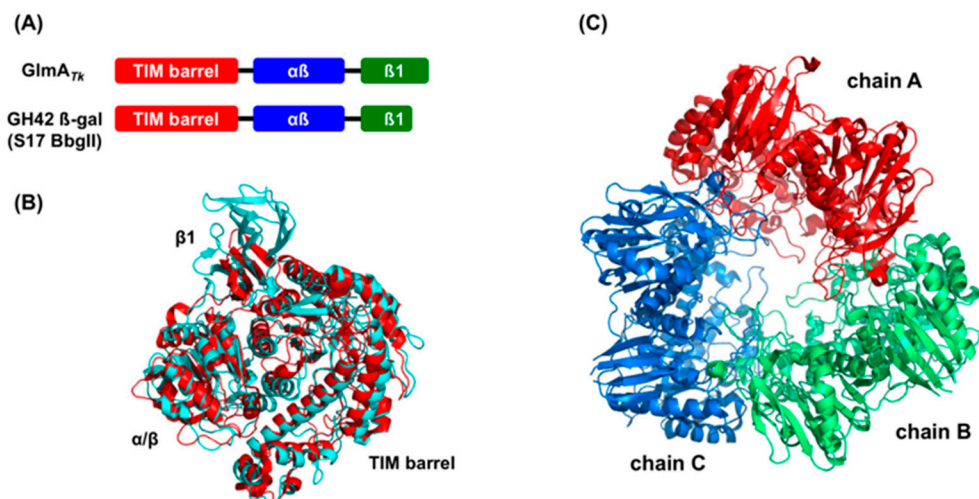


Figure 2. Structural comparison between GlmA_{Tk} and GH42 β-galactosidases. The comparison was performed with five GH42 β-galactosidases, but, for clarity, only the result of BbgII (PDB 4UZS) is shown in the figure. (A) Schematic presentation of the domain organization of GlmA_{Tk} and BbgII; (B) The superimposed models of the monomer structure of GlmA_{Tk} (cyan) and BbgII (red). The figure was drawn in the same orientation as in the left panel of Figure 1B; (C) Trimeric structure of BbgII (chain A: red, chain B: green, chain C: blue).

Distinctively, a DALI search indicated that the TIM-barrel domain of GlmA_{Tk} (residues 1–435) bears the highest structural similarity to GH35 β-galactosidases of *Aspergillus oryzae* (Ao-β-gal, PDB 4IU8) [30], *Aspergillus niger* (AnβGal, PDB 5IFP) [31], *Trichoderma reesei* (Tri-β-gal, PDB 3OGR) [32], *Penicillium* sp. (Psp-β-gal, PDB 1XC6) [33], *Homo sapiens* (Hs-β-gal, PDB 3THC) [34], *Bacillus circulans* (Bc-BgaC, PDB 4MAD) [35], and *Streptococcus pneumoniae* (Sp-BgaC, PDB 4E8C) [36]. These GH35 β-galactosidases are roughly divided into two groups: Group 1 comprises the former four β-galactosidases (Ao-β-gal, AnβGal, Tri-β-gal, and Psp-β-gal), and Group 2 is formed by the latter three (Hs-β-gal, Bc-BgaC, and Sp-BgaC). Group 1 β-galactosidases have five domains—a TIM-barrel domain and four β-domains (β1, β2, β3, and β4) (Figure 3A,B). The β-galactosidases in Group 2 show similar domain organization to those of Group 1 but lack the β1 and β2 domains (Figure 3A,C). The domain organization of GH35 β-galactosidases quite differs from that of GlmA_{Tk} (Figure 3A,D,E). However, the TIM-barrel domains are clearly superimposable, with RMSD values of 1.7–2.3 Å (Figure 3F). RMSD values for GH42 β-galactosidases were slightly improved (2.3–2.9 Å) when only the TIM-barrel domain was compared, indicating that the TIM-barrel domain of GlmA_{Tk} is more similar to those of GH35 β-galactosidases than to those of GH42 members. Indeed, a high degree of similarity within the entire catalytic centers was observed between GlmA_{Tk} and GH35 β-galactosidases, as described below. Collectively, these observations indicate that GlmA_{Tk}, GH35, and GH42 β-galactosidases are evolutionarily related.

The TIM-barrel domain and the β1 domain of GlmA_{Tk} could be superimposed onto those of Group 1 β-galactosidases (Figure 3E). This will be discussed in Section 5.

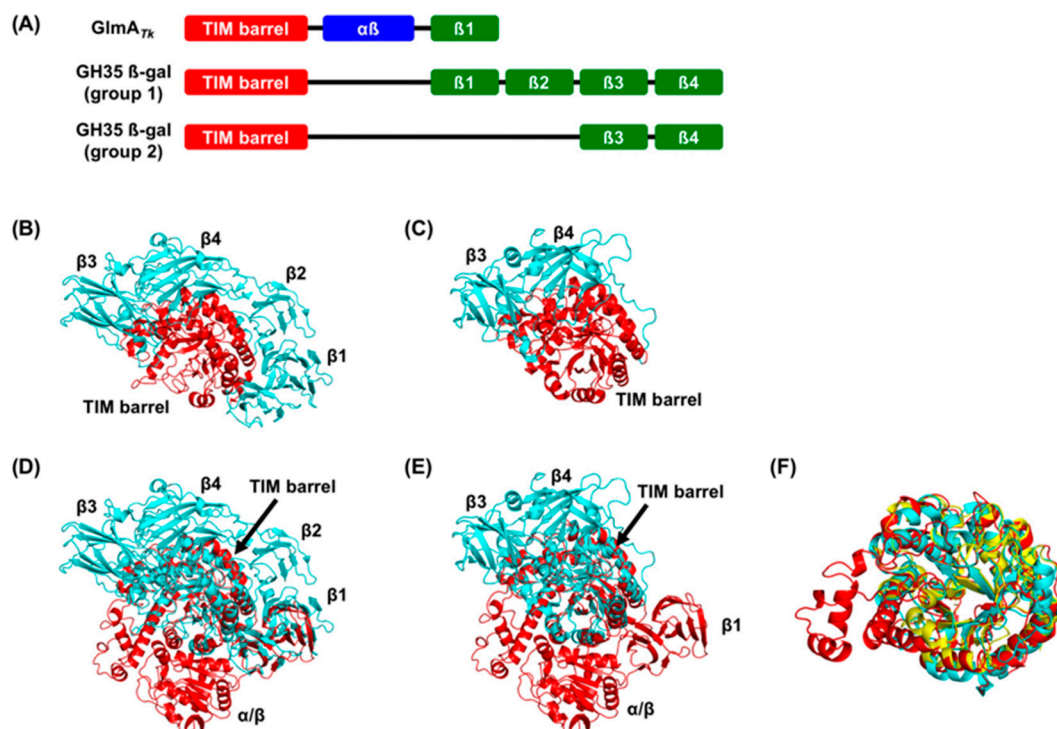


Figure 3. Structural comparison between $GlmA_{Tk}$ and GH35 β -galactosidases. Seven GH35 β -galactosidases were compared to $GlmA_{Tk}$, but, for clarity, only the results of Tri- β -gal (PDB 3OGR) and Hs- β -gal (PDB 3THC) are shown as representatives of Group 1 and 2, respectively. (A) Schematic presentation of the domain organization of $GlmA_{Tk}$, a Group 1 GH35 β -galactosidase, and Group 2 GH35 β -galactosidase; (B) The structure of Tri- β -gal (TIM-barrel: red, β 1–4: cyan); (C) The structure of Hs- β -gal (TIM-barrel: red, β 3 and β 4: cyan); (D) Superimposed models of $GlmA_{Tk}$ monomer structure (red) and Tri- β -gal (cyan); orientation, same as that in B; (E) Superimposed models of $GlmA_{Tk}$ monomer structure (red) and Hs- β -gal (cyan); orientation, same as that in C; (F) Superimposed models of the TIM-barrel domain (chain A) of $GlmA_{Tk}$ (red), Tri- β -gal (cyan), and Hs- β -gal (yellow).

4. $GlmA$ Active Site and Catalytic Mechanism

4.1. The Active-Site Architecture of $GlmA_{Tk}$: Comparison with the GH35 β -Galactosidase

We selected the structure of the galactose-bound Hs- β -gal (PDB 3THC) for further comparison with the $GlmA_{Tk}$ active site because Hs- β -gal is the best-characterized GH35 β -galactosidase to date, both structurally and biochemically [34,37].

In $GlmA_{Tk}$, a GlcN molecule is bound to each monomer in the chair conformation and it is fixed by making hydrogen bonds with eight residues. Superposition of the TIM-barrel structures of $GlmA_{Tk}$ and Hs- β -gal yields an RMSD of 1.7 Å over 292 C α atoms with 32% sequence identity and a strong structural similarity between the -1 subsites of these proteins (Figure 4A). GlcN and galactose, which are different but structurally similar molecules, are located at almost the same position. Interestingly, four of the eight substrate-binding residues in $GlmA_{Tk}$, namely, Tyr53, Glu103, Glu179, and Glu347 are present in Hs- β -gal as functionally conserved residues Tyr83, Glu129, Glu188, and Glu268, respectively (Figure 4B,C). These residues form direct hydrogen bonds with the galactose molecule similar to the GlcN-bound $GlmA_{Tk}$ structure. Gly102 of $GlmA_{Tk}$ is sterically identical to Ala128 of Hs- β -gal and the main-chain amide of each residue forms a hydrogen bond with the O3 of GlcN/galactose, indicating that this replacement is a conservative substitution. Furthermore, $GlmA_{Tk}$ Trp308 overlaps well with Hs- β -gal Tyr270 (Figure 4A–C). Tyr270 of Hs- β -gal performs two functions: it maintains the orientation of Glu268 for the hydrogen-bond catalytic reactions and contributes to the formation of the hydrophobic pocket [34]. Trp308 of $GlmA_{Tk}$ appears to perform the same function through a hydrogen

bond to Glu347 (Figure 4B). Remarkably, these residues, which are important for the recognition of GlmA_{TK}'s substrate, are structurally conserved in GlmA_{Ph} and GlmA_{Pf} (Figure 4D), but they are either composed or located differently in GH42 β -galactosidases (Figure 4E), thus excluding GlmA from the GH42 family classification.

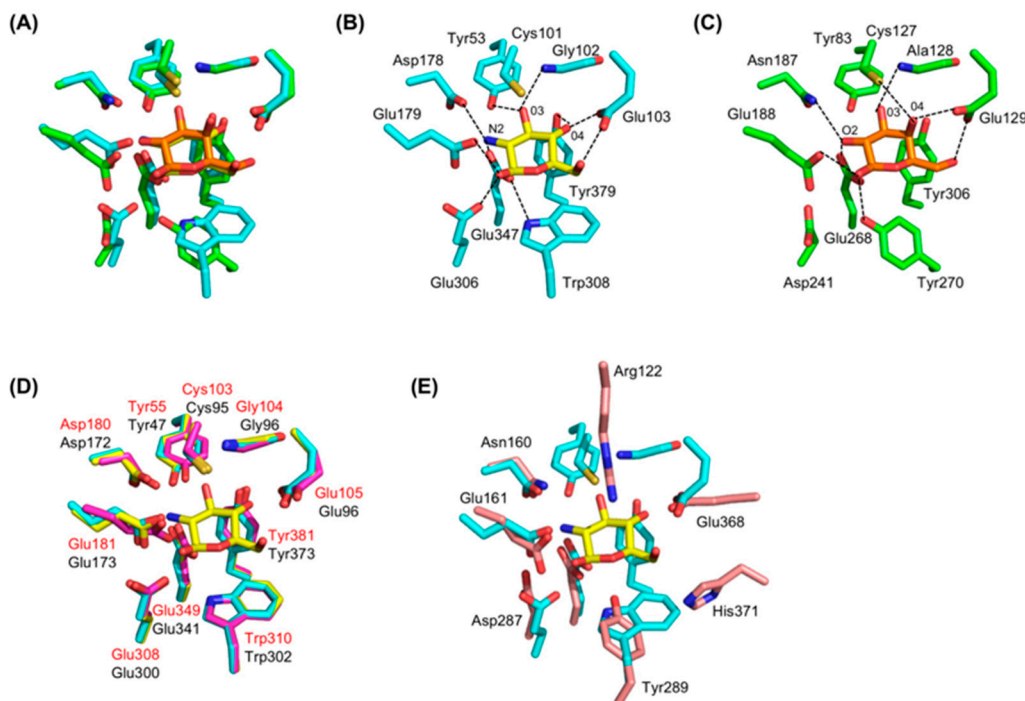


Figure 4. Comparison of the active site residues of GlmA_{TK} with those of representative enzymes of the families GH35, GH42, and other GlmAs. (A) Superposition of GlmA_{TK} (cyan sticks) and Hs- β -gal (green sticks) in complex with GlcN (yellow sticks) and galactose (orange sticks), respectively. Active site residues of GlmA_{TK} (B) and Hs- β -gal (C). Polar interactions are indicated by dashed lines; (D) Superposition of GlmA_{TK} (cyan sticks with red labels), GlmA_{Ph} (magenta sticks with black labels), and GlmA_{Pf} (yellow sticks); (E) Superposition of GlmA_{TK} (cyan sticks) and BbgII (GH42 β -galactosidase) (pink sticks with black labels). All the figures were drawn in the same orientation as in A.

4.2. GlmA Catalytic Mechanism Determined through In-Depth Crystallographic Analysis

From the structural comparison, Glu179 and Glu347 of GlmA_{TK} are sterically identical to the acid/base Glu188 and the nucleophile Glu268 of Hs- β -gal, respectively (Figure 4A, B, C). GlmA_{TK} mutations, E179Q and E347Q, resulted in dramatic activity loss [15], supporting the notion that these residues are involved in protein catalysis. Furthermore, these Glu residues are located in the β 4 and β 7 strands of the TIM-barrel domain and are separated by 4.8 Å [15]. All proteins in the GH35 family belong to a GH-A clan that comprises enzymes with two conserved catalytic Glu residues in the C-terminals of β 4 and β 7 [17]. The spatial arrangement of the two GlmA_{TK} Glu residues is in entire agreement with the structural features of the GH-A enzymes. Thus, Glu179 and Glu347 act as the acid/base residue and the nucleophilic residue, respectively, and GlmA_{TK} cleaves the glycosidic bond through a double-displacement retaining mechanism, like the other GH-A enzymes [17].

Despite the absence of sequence identity around the catalytic residues, sequence alignments show that the acid/base Glu179 of GlmA_{TK} aligns with the catalytic residues of GH35 and GH42 β -galactosidases (Figure 5A,B). The nucleophile Glu347 of GlmA_{TK} also aligns with those of GH42 β -galactosidases (Figure 5A), but not to those of GH35 β -galactosidases (Figure 5B). Instead of Glu347, GlmA_{TK} Glu306 locates at the position that corresponds to the nucleophilic residue of GH35 β -galactosidases (Figure 5B). Glu306 forms a hydrogen bond with GlcN O1 (Figure 4B) and contributes to the protein's catalytic activity [15]. However, in Hs- β -gal, this position is structurally occupied

by an Asp residue (Asp241), which is found in almost all GH35 β -galactosidases [30–34,36]. Asp241 cannot form a hydrogen bond with galactose because of its side chain orientation (Figure 4C) and, for this reason, should not be involved in the catalytic reaction. These results strongly suggest that a prediction of GlmA_{Tk}'s catalytic residues from sequence comparisons is not reliable and that the catalytic mechanisms could only be determined through in-depth crystallographic analysis.

(A)

GlmA _{Tk}	---MGKVEF-SGKRYVIDGEPVTIAGGTQLQFRRVPAD--AWKDRLLKMRAGLNTVDT-YVANNWHEPEKGSDFKGETHPQRNLVGFLELADELGFYVIRPQGYICGEWRNGGIPDWL	113
A4- β -gal	-----MLGVCYYPEHWPKE--RWKEDARRMRAGLSHVRIGEFAWALLEPEPGRLEWGLDEA-----IATLAEGLKVVLTPTATPPKWL	80
B1Gal42A	MSASTQHRRAHRWPQPLFGNDRKIMFGADYNPDQWPEP--VQDEDIRLMKQAGVNIWSLAIFSWANIEFSDGNFFFDWLDLDRV-----IDKLYKAGIADVLASATASPPMWL	103
S17 BbgII	---MSKRRKHSWPQKLGAEASRLMYGGDYNDQWPEE--VWDDDIRLMKAGVNLVSVGIFSWAKIEPEEGKYDFDWLDRA-----IDKLGKAGIADVLASATASPPMWL	100
Bca- β -gal	-----MINEKFKIWIYGGDYNPEQWKA--TMEEDMRMFLAGIDVATVNVFWSAKIQRDEVSYDFTWLDDI-----IERLTKENITLCLATSTGAHPAM	89
GanB	-----MPKYERTYTTQANFILHGGDYNPDQWLDLDRPDLQADLEMLKSHSTNTFTVGVFAWSALEPEEGYRFEWLDKV-----FDDIYRIGRVLATPSSGARPAWL	97
	acid/base	
GlmA _{Tk}	IDHEPEILAKGPNGLPRDIYYPPIITLYHPTYLEAVGEWYNAVFPVIRKYLTYNGGPIISVSDDEPSYWETIFQPFLLTYNEIITKPGGLWEKWLQNYT--LEDLRRR-----YKGF	226
A4- β -gal	VDRYPEILPVDRREGRRRRFGRRRHVCFSSPVYREARRIVT-----LLAERYGGLEAVAGFQTDNRYGCHD-----TVRCY--CPRCQEAFRGWLEARYG-TIEALNEAWGTAFWWSQRY	186
B1Gal42A	TSAPHEVLRREDEQGHVWPGARQHWRTSPVFRYALNLCR-----EMAEHYKDNPAIVSWHVGNRYGCHN-----YF-DY--SDDAVQAFREWCRRDRYG-TIDKVNAAWGTAFWWSQRL	208
S17 BbgII	TQAHPVELWKDERGDTVWPGAREHWRPTSPVFRYALNLCR-----RMAEHYKDNPAIVSWHVGNRYGCHN-----RF-DY--SDDAMRAFQKWKCKRYK-TIDAVNEAWGTAFWAQHM	205
Bca- β -gal	AKKYPDVLRVDYEGRRKFGRRHNSCPNSPTRYKAKILAG-----KLAERYKDHQPQVMHVSNRYGGYC-----Y-----CDNCEKQFRVWLKERYG-TLEALNKAWTFSWSHTF	191
GanB	SQKYPEVLRVNAARVRLQHGRRHHCFTSSVYREKTHINR-----LLAERYGDHALLMWHVSNRYGGEC-----H-----CNLCQEAFLREWLKKNYHDLALNAAWTFSWSHTY	200
	nucleophile	
GlmA _{Tk}	KDYSEIKVPTSFSEP-----LPLKIDWHFKLWMIINEYVRNIYERMAREFVPIISILDPY-----LLQVAMRHFFTYMREHNLKHVWTFEWFYSFRSSD-	316
A4- β -gal	RSFAEVLPHLTVLVA-----EPNPSHLLDYRFPASDQVRAFNRLOVEIL-RAHAPGFYTHNFMGF-----FTDLDAFALAQDLDFASWDSYPLGFTD-----LMPLPEEKL	282
B1Gal42A	NSFEELIPRYPVGGEG--NFTNPGRLLDLDFKPCSDALKEFFCAERDVL-SEVTNPIPLTNTMVSAS-----QNTLDYDDMAHEVDFVSNDDHYVTPGS-----	298
S17 BbgII	NDFSEIIPRYPYIG-DG-----NFMNPKGLLDYRFRSSDALKELYIAERDVL-ESITPGLPLTNTMVSAG-----GSLDYDDWGADEVDFVSNDDHYVTPGE-----	294
Bca- β -gal	YDWDEIVAPNALSEWSGNRTNFQGISLDYRRFQSDSLLECFKMERDEL-KRWTFPIPTVNTLMGF-----YPELDYFKWAKEMDVSWDNYPSMDTP-----FS-----	285
GanB	TDWSQIESPSPIGE-----HTIHGLNLWKRFTVDQTISSFENEIVPL-RELTPhiPITTNFMADTHDLIPFQGLDYSKFAKHLVDVSWDAYPAWHND-----WE-----	294
	nucleophile	
GlmA _{Tk}	--FKEDKLGHIYKTYGIYRHYVRKAGTPPLSIE ^T QSSLAH--TID---PTEAELLSYLPLGIPNINYLTVGGENPEGYESHNGITWVYSPVGLDGE-RPHFGVIKALSETMTS-	426
A4- β -gal	RYARTGHDPVAAFHHDLYRGV--GRGR-FWVME ^Q QPGPVNWAHPNPSAPGMVRLWTEALAHGAEVVSYFRWRQ--APFAQE ^Q MHAGL-----HRPDSAPDQ-GFFEAKRVAEELAAL	390
B1Gal42A	-----WHIDELAYSASLVDGI--SRKKPWFLME ^Q STSAVNWREINPRKEPGLIRDSMLHLAMGADAICYFQWRQ--SRSGAEK ^F HSSAM-----LPLAG-EHSQIYRDVLCALGADLDTL	402
S17 BbgII	-----AHFDEVAYAASLMDGI--SRKEPWFOME ^H STSAVNWRPINYRAEPGVVRRDLSLAQVAMGADAICYFQWRQ--SKAGAEK ^H SSM-----VPHAG-EDSQIFRDCVLCALGADLRL	398
Bca- β -gal	-----FTAM-----AHNLMRGL--KSGQPFLLME ^Q TPGVQNWQPYNSAKRPGVMRLWSYQVAHGADTVMFQ ^L LR--SVGACEK ^Y HGAV-----IEHVGHETHRVFRECALGKELQQL	386
GanB	--STADLAMKVGFINLDRS--LQKQPFLLME ^C TPSLVNW ^H KVNKAKRPGMHFLSSMQMIAHGSDSILYFQWRK--SRGSFEK ^F HGAV-----VDHNRDTSRVFQEA ^V EVGKALKKM	401
	nucleophile	
GlmA _{Tk}	--AELA-DAELRPKVAVGLYEPYALNLNGYEGLEESTDL-NEYLLGERGLFTLLAMNSPTFPAVDLEDVTLDELLSYDQLMVVYSLDFMSREVDKLVFVARGGNLVLMP ^L PRYDENL	543
A4- β -gal	-----ALPPVAQAPVA---LVFDYEAAMIYEQP--QGAENSYLGLVYLFYSALRRGLLDVDPVPPG--ASLRGYAFVAVVPSLPIVREALEAFRE---AEGPVLFGPRSGSKTETF	492
B1Gal42A	SDAGILRSKLSKARVA---IVQDIQSEWATEHTATPTQHIREN-TEPLDFAAFANRGGVADVTPPIH---AQWDTYDAVVICPVYLFSEMAERLRTFVRNNGKAFVYYSALADEHD	513
S17 BbgII	SDEGLMGT ^K TVKSKVA---VVFDEYSQWATEYITANPTQ ^Q VDHN-TEPLDWFALADNGITADVDPVY---SDWDSYEAIVLPCVYLLSETSRRVREFVANGGLFVYTYTGLSDEND	509
Bca- β -gal	GDT--ILDARSEAKVA---VMYDWE ^N WALELSSGFSIA-LNYVNEV ^H KYD ^A LYKQNIQ ^T DMISVE---EDLSKYK ^V VIAPV ^M Y ^M K ^P GAERVERFVAQ ^G Q ^T V ^T FF ^F SGV ^I NEND	495
GanB	PVDVITKE---HDFSRYK ^L LIA ^M LYL ^V SEETIARLKEFVANGGLV ^M YISGLVDEHDSGI---VGTNRPAEVA---ILYD ^W EN ^N WAL ^N DAQ ^G FAA ^T K ^R Y ^P Q ^T L ^V Q ^H Y ^R PF ^W ERDI	454
	nucleophile	
GlmA _{Tk}	EPY-----SSLKDFLGEVEREKA--RRNPRLIQFLSVSA-----EGIDRLVRNTVRGVRGGEPIAFLGKPVGAFVVRKGGGSVAVLGFRLQYTT---SHHDLHRKFW	638
A4- β -gal	QIPKELPPGFLQALLPLKVVRE---SLPPGLLEVAE-----GALGRFPLGLNREWEAPLKLPTFDQDKGALYREGRYLYLAAMPSP ^E LARLLS	581
B1Gal42A	RLHTEGWPLIGLDVVGVRIEEHCPLGTLFPGLDHL ^D VS-----NGTVVHDLADVIDAIDDTT ^V LATFEADPATGMDGRAAI-----TVHPYHEGGVAYIAGKLRGDI ^S QSLP	618
S17 BbgII	HIWLGYPGSIRDVVGVRVEEFAPMGNDMPGALDHL ^D LD-----NGTVAHDFADVITSTADTST ^V LAS ^Y K ^A ERWTGMNEVPAI-----VANGYGDGR ^T VYV ^G CR ^L GR ^Q LAK ^S LP	614
Bca- β -gal	LVTLGGYPGE ^L RNV ^M GIWAE ^E ID---ALLPGH ^Q NEI ^V LR ^Q DW ^G LRS ^Y SC ^G L ^C VD ^I HA---ETA ^E VLA ^E Y ^G AD ^Y YK---GTP ^V L-----TRN ^K FG ^N G ^S Y ^V VAS ^S PD ^A FL ^Q GL ^I A	599
GanB	LAYLGGW ^H ODLRE ^M FG ^M E ^P IETD---TLYPRDRNSV ^H YGR---SYELK ^D YAT ^V I ^K I---HAAT ^V EG ^V YED ^D FYA---DTP ^V A-----TSN ^O Y ^K G ^Q AY ^I GG ^R LE ^D OF ^H RD ^F YQ	608
	nucleophile	
GlmA _{Tk}	KLKELQGVREDFEVTNPM ^D IVL ^P MEG--KGYAYLAVTNP ^R GHPIK ^R ISY ^R GL ^E VPVLLD ^G IELK ^R RR ^T LYL ^P FG ^V RK ^G D ^V E ^V AYATAT ^L VM ^W EG ^D VL ^T FR ^N HL ^S GH ^S E ^I AL ^K GV ^E SV ^K	756
A4- β -gal	ALAAEAGLKVLSL---PEGLRLR---RRGTWVFAFN ^Y G-----PEAVEA--PA-----SEGARFL ^L GSRRV-----GPYDLA ^W VEE	644
B1Gal42A	EICALGFELDADPRAGD ^V RV ^V REQ-EDGAI ^F EFL ^N RT-----RNTVTAD ^R PA-----GDML ^I CS ^L AD ^T SD ^K VT ^L EP ^N GL ^A FR	695
S17 BbgII	AMLGSM ^L SDL--AGDGRVLR ^V ERADAAA ^S H ^F EV ^F NRT-----HEPVT ^D VE-----GEA ^I AS ^L A ^H VD ^D GRAT ^I DP ^T GV ^V LR	689
Bca- β -gal	NLCEE ^Q GK ^F LLN--TPDG ^E V ^A -ERVK ^N G ^T S ^Y L ^F VM ^N HN-----AEEM ^T FD ^A GA-----SRQ ^R DL ^L TG ^K T ^I S ^Q AT ^I PARG ^V M ^L ER	674
GanB	ELMEK ^L DL ^R VL ^F V ^K HE ^K GV ^S VQ-ARQA ^P EC ^D Y ^V IM ^N FT-----EEK ^Q AV ^V -LE-----EK ^V K ^D L ^T GE ^E IV ^G E ^I M ^L DR ^Y EV ^R VEK	684
	nucleophile	
GlmA _{Tk}	SGGKIVDGS ^D GEVLR ^I VIE ^H PG ^E YF ^E VELL	786
A4- β -gal	A-----	645
B1Gal42A	-----	695
S17 BbgII	-----	689
Bca- β -gal	A-----	675
GanB	RR-----	686

Figure 5. Cont.

(B)



Figure 5. Sequence alignment of the catalytic residues (red) of GH42 β-galactosidases (A) and GH35 β-galactosidases (B). Alignments were carried out with ClustalW [38].

4.3. The Role of Asp178

Several unique structural features of GlnA_{Tk} can provide insights into its substrate recognition mechanisms. The most important substrate-recognizing residue is Asp178, which precedes the acid/base Glu179 residue. The Asp–Glu motif replaces the conserved Asn–Glu motif (equivalent to the Asn187–Glu188 motif in Hs-β-gal) in all GH35 and GH42 β-galactosidases. Hs-β-gal Asn187 forms a hydrogen bond with the C2–OH of galactose (Figure 4C), while GlnA_{Tk} Asp178 forms a hydrogen bond with the C2–NH₂ of GlcN (Figure 4B). The pK_a values of the Asp178 carboxyl group and of the GlcN N2 group are approximately 3.7 and 7.4 [39], respectively. Therefore, at pH 6.0, at which GlnA_{Tk} activity is the highest [2], negatively charged Asp178 could interact with the protonated NH₃⁺ form of N2. To confirm this hypothesis, a D178N mutant was created and experimental results showed that it was inactive [15], implying that the charge–charge interaction is a major factor for the GlnA_{Tk} recognition of GlcN. This assumption is supported by a previous observation that GlnA_{Tk} has very weak β-glucosidase activity [2]. Glucose differs from GlcN only at the C2 substituent, which is C2–OH, and the loss of a charged interaction between Asp178 and the C2–OH of glucose should result in fundamental loss of β-glucosidase activity. On the other hand, GlcNAc differs from GlcN only at the C2 substituent, which contains a large acetoamide group. This group would sterically clash with Asp178, in accordance with GlnA_{Tk} being completely unable to hydrolyze (GlcNAc)₂ [2]. These results suggest that Asp178 is a key residue because of its ability to discriminate between substrates.

CsxA from *Amycolatopsis orientalis*, a member of the GH2 family, is the only other exo-β-D-glucosaminidase with a known structure [40,41]. In CsxA, Glu394 binds to GlcN C2–NH₂ by means of a charged interaction [40], similar to that of GlnA_{Tk}. However, Glu394 is distantly located from the Asp469 acid/base residue in the CsxA sequence, and the residue that precedes that acid/base residue is Ser468. Therefore, the use of an Asp–Glu motif to discriminate among substrates is only found in GlnA glycoside hydrolases characterized thus far.

4.4. Residue Conservation during Evolution

The other remarkable GlnA_{Tk} conserved residues are Cys101 and Tyr379, which could be superimposed onto Hs-β-gal Cys127 and Tyr306, respectively (Figure 4A–C). These two residues are conserved in almost all GH35 β-galactosidases [34,36]. GlcN and galactose differ in their chirality of O4, which is equatorial in GlcN and axial in galactose. GlnA_{Tk} Tyr379 forms a hydrogen bond (2.8 Å) with the equatorial O4 of GlcN (Figure 4B) and provides an aromatic stacking interaction with GlcN through a hydrophobic platform for the C4 side. Hs-β-gal Tyr306 also serves as a hydrophobic stacking platform to accommodate galactose. However, it cannot provide a hydrogen bond to the axial O4 of galactose because it is very distant from it (4.6 Å) (Figure 4C). Instead, Hs-β-gal Cys127 forms a hydrogen bond (3.3 Å) with the axial O4 of galactose via its thiol group (Figure 4C). Likewise, GlnA_{Tk} Cys101 is present as Hs-β-gal Cys127 counterpart, but its location is too far (4.8 Å) to form a hydrogen bond with the equatorial O4 of GlcN (Figure 4B). In brief, GlnA_{Tk} and GH35 β-galactosidases possess residues that could form hydrogen bonds with axial and equatorial O4 forms in the glycosidic substrate. The presence of Cys and Tyr residues is regarded as a remnant of evolution. As far as we know, such residues have never been seen in different functional glycoside hydrolases in the course of evolution, and they constitute a compelling link to the molecular evolution of these enzymes.

4.5. GlnA Dimer Structure Influences Substrate Specificity

As shown in Figures 1B and 6B, the active sites of GlnA_{Tk} are located within a deep pocket that intrudes into the core of the TIM-barrel domain of each monomer. Such active sites may act independently and their entrances, which are the only means of passage for substrates and products, are created by a reciprocal donation of each monomer. That is, the 3₁₀-helix of the α/β-domain of chain A interacts with the TIM-barrel domain of chain B, narrowing the active site entrance. As a result, the distance of subsite -1 from the active site entrance is approximately 20 Å (Figure 6A,B), which may

contribute to a size selection. Indeed, GlmA_{TK} showed higher activity against GlcN₂ (approximately 12 Å in length) and its activity decreased in proportion to the length of the chitooligosaccharide chain [2]. These results suggest that dimer formation enables GlmA_{TK} to form an active site with an appropriate shape for binding smaller substrates.

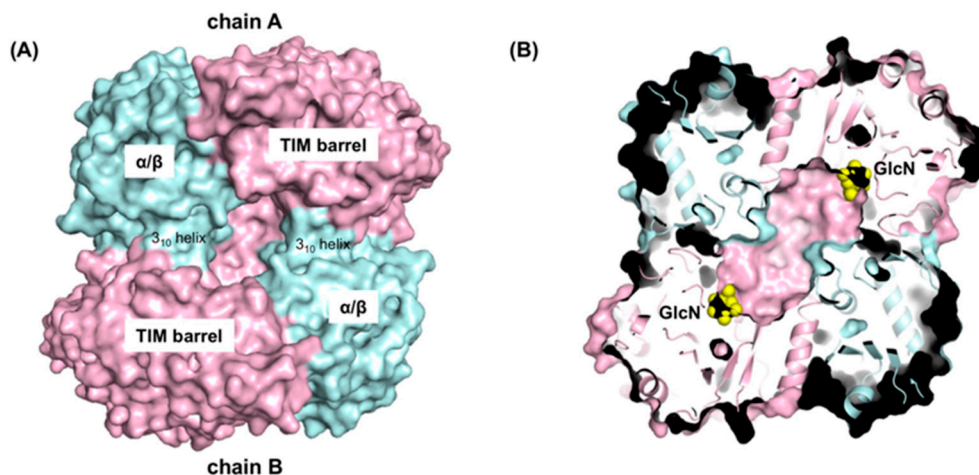


Figure 6. The dimerization interface and the deep active site pocket of GlmA_{TK}. (A) Surface representation of the GlmA_{TK} dimer (TIM-barrel: light pink, α/β : cyan). The figure was drawn from the same orientation as in the right panel of Figure 1; (B) Section drawing of the GlmA_{TK} dimer containing GlcN molecules (yellow) in the active site. The figure was drawn in the same orientation as in A.

5. Molecular Evolution of GlmAs and β -Galactosidases

The crystal structure of GH35 GlmA_{TK} shows structural similarities to both GH35 and GH42 β -galactosidases. Briefly, the monomeric structure of GlmA_{TK}, which comprises the TIM-barrel domain, the α/β domain, and the β 1 domain, is similar to that of GH42 β -galactosidases, whereas in the TIM-barrel domain, the key amino acids involved in substrate binding and catalysis at subsite -1 are highly conserved between GlmA_{TK} and GH35 β -galactosidases. As previously stated, the sequence of GlmA_{TK} bears homology to sections of GH35 and GH42 β -galactosidases [2]. Therefore, these β -galactosidases may have evolved from GlmA_{TK} via gene duplication, truncation, or domain insertion.

GlmA_{TK} and GH42 β -galactosidases are active as a dimer and a trimer, respectively. They have a cleft-type active site in their monomeric forms, which is suitable for binding to long-chain polysaccharides. Despite having different oligomerization states, both enzymes change the active site from the cleft to the pocket-type upon oligomerization to better accommodate smaller substrates. Moreover, Juers et al. reported other features that reduce the size of the active site [42]: a lengthening loop at the end of the TIM-barrel domain that partially fills in the active site cleft; and the addition of extra domains on either side of the active site cleft. During molecular evolution, lengthening loops would be more efficient than oligomerization or the addition of domains, but it seems that evolutionary selection gives priority to function over efficiency [42]. For GH42 β -galactosidases, trimer formation is essential to exhibit high enzymatic activity as well as to ensure size-based substrate specificity [25,27,28]. Thus, the use of GlmA_{TK}'s monomer structure frameworks (i.e., domain organization) might be necessary for fulfilling its functions via trimer formation. However, the substrate-binding residues of GlmA_{TK} are not well conserved in GH42 β -galactosidases (Figure 4E) and the underlying evolutionary selection pressure that led to this diversity in the active site remains unknown.

In GH35 β -galactosidases, the original substrate-binding residues of GlmA_{TK} are highly conserved, and those with some conservative substitutions (e.g., Trp308 of GlmA_{TK} is substituted in Hs- β -gal by Tyr270) and the catalytic machinery were retained, whereas the reaction specificity has evolved toward β -galactosidase activity. In other words, the GH35 β -galactosidase could have evolved from

ancestral GlmA_{Tk} to be able to recognize galactose through a subtle change of residues around subsite -1 . Indeed, a single residue, GlmA_{Tk} Asp178, replaces the conserved Asn residue in the GH35 β -galactosidase and plays an essential role in the recognition of GlcN. In addition, GlmA_{Tk} Cys101 and Tyr379, which are well conserved and similarly spatially located in the GH35 β -galactosidase, have the potential for forming hydrogen bonds with either the axial (galactose) or equatorial (GlcN) forms of O4 in the glycosidic substrate, respectively. This further supports the proposed evolutionary approach. The underlying mechanisms could be driven by constraints in the availability of different substrates in the organism's habitat or in the ability to survive [43,44]. Therefore, the change in GlmA_{Tk} substrate specificity might have developed under the positive constraint of galactose presence.

Although the GlmA_{Tk} substrate-binding scaffold is almost entirely retained in GH35 β -galactosidases, the domain organization is different (Figure 3A). It has been suggested that evolutionary pathways can be tracked at the structural level [25,45,46]. Moreover, Matthews et al. proposed the evolutionary path of β -galactosidase from *Escherichia coli* (Ec- β -gal), an enzyme that belongs to the GH2 family and is the best studied β -galactosidase [42,47]. Ec- β -gal is a homotetramer, its monomer structure can be divided into five domains, and is built around the TIM-barrel with the remaining four domains similar to the Group 1 GH35 β -galactosidase. First, Ec- β -gal may have evolved from a much smaller enzyme, such as a single TIM-barrel domain, that cleaves long polysaccharides. Second, during the modulation of the substrate specificity, additional domains may have been added. Based on this scenario, we speculate that an early GH35 β -galactosidase ancestor with a structure similar to that of Group 1 β -galactosidases (Ao- β -gal, An β Gal, Tri- β -gal, and Psp- β -gal) may have first arisen from GlmA_{Tk} via deletion of the α/β domain, which could accommodate extended substrates (Figure 7). This hypothesis is supported by the observation that the TIM-barrel domain and the $\beta 1$ domain of GlmA_{Tk} could be superimposed onto those of Group 1 β -galactosidases (Figure 3E). Subsequently, addition of the $\beta 2$, $\beta 3$, and $\beta 4$ domains could then have conferred the substrate specificity on the enzymes. The extended loop from the $\beta 3$ domain especially not only contains the substrate specificity determinant residue, but also constitutes a part of the active site pocket [36]. Although the functions of the $\beta 1$, $\beta 2$, and $\beta 4$ domains remain unknown, they seem to stabilize the complete structure of Group 1 β -galactosidases. The final step—deletion of the $\beta 1$ and $\beta 2$ domains—could have led to the appearance of Group 2 β -galactosidases (Hs- β -gal, Bc-BgaC, and Sp-BgaC) (Figure 7). Interestingly, Group 2 β -galactosidases form dimers, whereas other β -galactosidases function as monomeric enzymes, suggesting that a deletion of the $\beta 1$ and $\beta 2$ domains may be needed for dimer formation. In other words, Group 2 β -galactosidases may form a dimer to compensate the instability caused by the deletion of the $\beta 1$ and $\beta 2$ domains.

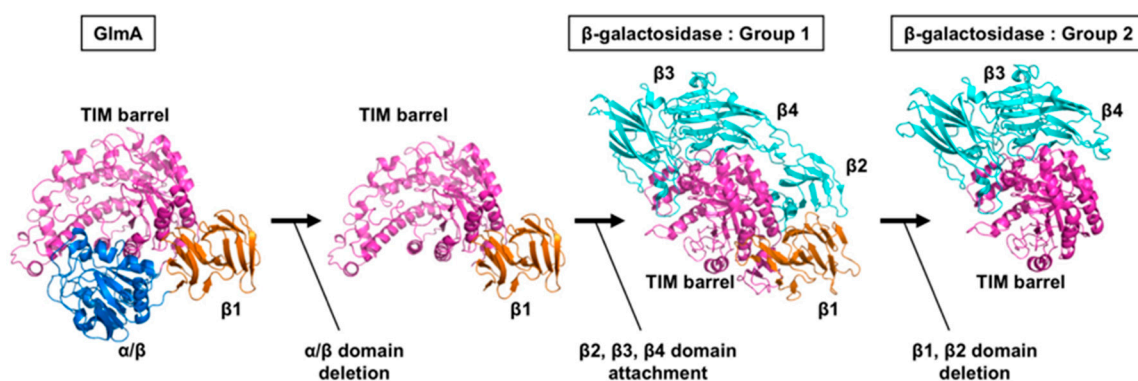


Figure 7. Hypothetical evolutionary pathway from GlmA to GH35 β -galactosidases. TIM-barrel domain, α/β : domain, $\beta 1$ domain, and $\beta 2$ – $\beta 4$ domains are colored magenta, blue, orange, and cyan, respectively.

6. Conclusions

GlmA_{Tk} , GH35, and GH42 β -galactosidases belong to the same GH-A clan. A clan is a group of families that show significant similarities in the tertiary structure as well as conservation of catalytic

residues and mechanisms, and its members are therefore considered to have common ancestry [11]. Accordingly, our findings presented here strongly suggest that GlnA_{TK} is a common ancestor of both GH35 and GH42 β -galactosidases.

Author Contributions: S.M. and M.W. wrote and edited the manuscript.

Funding: Synchrotron experiments were performed at SPring-8 under the approval of the Japan Synchrotron Radiation Research Institute under proposal numbers 2014B6953, 2014B6903, 2015A6546, 2015A6559, 2015B6559, 2015B6546, 2016A6645, 2016A6657, 2016B6645, 2016B6657, 2017A6755, 2017A6767, 2017B6755, and 2017B6767.

Acknowledgments: We thank the beamline staff at SPring-8 BL44XU for their technical assistance during data collection.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tanaka, T.; Fukui, T.; Fujiwara, S.; Atomi, H.; Imanaka, T. Concerted action of diacetylchitobiose deacetylase and exo- β -D-glucosaminidase in a novel chitinolytic pathway in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J. Biol. Chem.* **2004**, *279*, 30021–30027. [[CrossRef](#)] [[PubMed](#)]
2. Tanaka, T.; Fukui, T.; Atomi, H.; Imanaka, T. Characterization of an exo- β -D-glucosaminidase involved in a novel chitinolytic pathway from the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J. Bac.* **2003**, *185*, 5175–5181. [[CrossRef](#)]
3. Aslam, M.; Horiuchi, A.; Simons, J.R.; Jha, S.; Yamada, M.; Odani, T.; Fujimoto, R.; Yamamoto, Y.; Gunji, R.; Imanaka, T.; et al. Engineering of a hyperthermophilic archaeon, *Thermococcus kodakarensis*, that displays chitin-dependent hydrogen production. *Appl. Environ. Microbiol.* **2017**, *83*, e00280-17. [[CrossRef](#)] [[PubMed](#)]
4. Mine, S.; Nakamura, T.; Sato, T.; Ikegami, T.; Uegaki, K. Solution structure of the chitin-binding domain 1 (ChBD1) of a hyperthermophilic chitinase from *Pyrococcus furiosus*. *J. Biochem.* **2014**, *155*, 115–122. [[CrossRef](#)] [[PubMed](#)]
5. Mine, S.; Niiyama, M.; Hashimoto, W.; Ikegami, T.; Koma, D.; Ohmoto, T.; Fukuda, Y.; Inoue, T.; Abe, Y.; Ueda, T.; et al. Expression from engineered *Escherichia coli* chromosome and crystallographic study of archaeal *N,N'*-diacetylchitobiose deacetylase. *FEBS J.* **2014**, *281*, 2584–2596. [[CrossRef](#)]
6. Nakamura, T.; Mine, S.; Hagihara, Y.; Ishikawa, K.; Ikegami, T.; Uegaki, K. Tertiary structure and carbohydrate recognition by the chitin-binding domain of a hyperthermophilic chitinase from *Pyrococcus furiosus*. *J. Mol. Biol.* **2008**, *381*, 670–680. [[CrossRef](#)]
7. Nakamura, T.; Mine, S.; Hagihara, Y.; Ishikawa, K.; Uegaki, K. Structure of the catalytic domain of the hyperthermophilic chitinase from *Pyrococcus furiosus*. *Acta Crystallogr. Sect. F* **2007**, *63*, 7–11. [[CrossRef](#)]
8. Hanazono, Y.; Takeda, K.; Niwa, S.; Hibi, M.; Takahashi, N.; Kanai, T.; Atomi, H.; Miki, K. Crystal structures of chitin binding domains of chitinase from *Thermococcus kodakarensis* KOD1. *FEBS Lett.* **2016**, *590*, 298–304. [[CrossRef](#)] [[PubMed](#)]
9. Nakamura, T.; Niiyama, M.; Hashimoto, W.; Ida, K.; Abe, M.; Morita, J.; Uegaki, K. Multiple crystal forms of *N,N'*-diacetylchitobiose deacetylase from *Pyrococcus furiosus*. *Acta Crystallogr. Sect. F* **2015**, *71*, 657–662. [[CrossRef](#)]
10. Nakamura, T.; Yonezawa, Y.; Tsuchiya, Y.; Niiyama, M.; Ida, K.; Oshima, M.; Morita, J.; Uegaki, K. Substrate recognition of *N,N'*-diacetylchitobiose deacetylase from *Pyrococcus horikoshii*. *J. Struct. Biol.* **2016**, *195*, 286–293. [[CrossRef](#)] [[PubMed](#)]
11. Henrissat, B.; Bairoch, A. Updating the sequence-based classification of glycosyl hydrolases. *Biochem. J.* **1996**, *316*, 695–696. [[CrossRef](#)] [[PubMed](#)]
12. Cote, N.; Fleury, A.; Dumont-Blanchette, E.; Fukamizo, T.; Mitsutomi, M.; Brzezinski, R. Two exo- β -D-glucosaminidases/exochitosanases from actinomycetes define a new subfamily within family 2 of glycoside hydrolases. *Biochem. J.* **2006**, *394*, 675–686. [[CrossRef](#)] [[PubMed](#)]
13. Honda, Y.; Shimaya, N.; Ishisaki, K.; Ebihara, M.; Taniguchi, H. Elucidation of exo- β -D-glucosaminidase activity of a family 9 glycoside hydrolase (PBPR0520) from *Photobacterium profundum* SS9. *Glycobiology* **2011**, *21*, 503–511. [[CrossRef](#)]

14. Kaper, T.; Verhees, C.H.; Lebbink, J.H.; van Lieshout, J.F.; Kluskens, L.D.; Ward, D.E.; Kengen, S.W.; Beerthuyzen, M.M.; de Vos, W.M.; van der Oost, J. Characterization of β -glycosylhydrolases from *Pyrococcus furiosus*. *Methods Enzymol.* **2001**, *330*, 329–346. [[PubMed](#)]
15. Mine, S.; Watanabe, M.; Kamachi, S.; Abe, Y.; Ueda, T. The structure of an Archaeal β -glucosaminidase provides insight into glycoside hydrolase evolution. *J. Biol. Chem.* **2017**, *292*, 4996–5006. [[CrossRef](#)] [[PubMed](#)]
16. Liu, B.; Li, Z.; Hong, Y.; Ni, J.; Sheng, D.; Shen, Y. Cloning, expression and characterization of a thermostable exo- β -D-glucosaminidase from the hyperthermophilic archaeon *Pyrococcus horikoshii*. *Biotechnol. Lett.* **2006**, *28*, 1655–1660. [[CrossRef](#)]
17. Henrissat, B.; Callebaut, I.; Fabrega, S.; Lehn, P.; Mornon, J.P.; Davies, G. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 7090–7094. [[CrossRef](#)]
18. Reardon, D.; Farber, G.K. The structure and evolution of α/β barrel proteins. *FASEB J.* **1995**, *9*, 497–503. [[CrossRef](#)]
19. Pucci, F.; Rooman, M. Physical and molecular bases of protein thermal stability and cold adaptation. *Curr. Opin. Struct. Biol.* **2017**, *42*, 117–128. [[CrossRef](#)]
20. Akiba, T.; Nishio, M.; Matsui, I.; Harata, K. X-ray structure of a membrane-bound β -glycosidase from the hyperthermophilic archaeon *Pyrococcus horikoshii*. *Proteins* **2004**, *57*, 422–431. [[CrossRef](#)]
21. Walden, H.; Bell, G.S.; Russell, R.J.; Siebers, B.; Hensel, R.; Taylor, G.L. Tiny TIM: A small, tetrameric, hyperthermostable triosephosphate isomerase. *J. Mol. Biol.* **2001**, *306*, 745–757. [[CrossRef](#)]
22. Fraser, N.J.; Liu, J.W.; Mabbitt, P.D.; Correy, G.J.; Coppin, C.W.; Lethier, M.; Perugini, M.A.; Murphy, J.M.; Oakeshott, J.G.; Weik, M.; et al. Evolution of protein quaternary structure in response to selective pressure for increased thermostability. *J. Mol. Biol.* **2016**, *428*, 2359–2371. [[CrossRef](#)]
23. Rutkiewicz-Krotewicz, M.; Pietrzyk-Brzezinska, A.J.; Sekula, B.; Cieslinski, H.; Wierzbicka-Wos, A.; Kur, J.; Bujacz, A. Structural studies of a cold-adapted dimeric β -D-galactosidase from *Paracoccus* sp. 32d. *Acta Crystallogr. D* **2016**, *72*, 1049–1061. [[CrossRef](#)] [[PubMed](#)]
24. Holm, L.; Sander, C. Mapping the protein universe. *Science* **1996**, *273*, 595–603. [[CrossRef](#)]
25. Hidaka, M.; Fushinobu, S.; Ohtsu, N.; Motoshima, H.; Matsuzawa, H.; Shoun, H.; Wakagi, T. Trimeric crystal structure of the glycoside hydrolase family 42 β -galactosidase from *Thermus thermophilus* A4 and the structure of its complex with galactose. *J. Mol. Biol.* **2002**, *322*, 79–91. [[CrossRef](#)]
26. Maksimainen, M.; Paavilainen, S.; Hakulinen, N.; Rouvinen, J. Structural analysis, enzymatic characterization, and catalytic mechanisms of β -galactosidase from *Bacillus circulans* sp. *alkalophilus*. *FEBS J.* **2012**, *279*, 1788–1798. [[CrossRef](#)]
27. Solomon, H.V.; Tabachnikov, O.; Lansky, S.; Salama, R.; Feinberg, H.; Shoham, Y.; Shoham, G. Structure-function relationships in Gan42B, an intracellular GH42 β -galactosidase from *Geobacillus stearothermophilus*. *Acta Crystallogr. D* **2015**, *71*, 2433–2448. [[CrossRef](#)] [[PubMed](#)]
28. Viborg, A.H.; Fredslund, F.; Katayama, T.; Nielsen, S.K.; Svensson, B.; Kitaoka, M.; Lo Leggio, L.; Abou Hachem, M. A β 1-6/ β 1-3 galactosidase from *Bifidobacterium animalis* subsp. *lactis* Bl-04 gives insight into sub-specificities of β -galactoside catabolism within *Bifidobacterium*. *Mol. Microbiol.* **2014**, *94*, 1024–1040. [[CrossRef](#)] [[PubMed](#)]
29. Godoy, A.S.; Camilo, C.M.; Kadowaki, M.A.; Muniz, H.D.; Espirito Santo, M.; Murakami, M.T.; Nascimento, A.S.; Polikarpov, I. Crystal structure of β 1 \rightarrow 6-galactosidase from *Bifidobacterium bifidum* S17: Trimeric architecture, molecular determinants of the enzymatic activity and its inhibition by α -galactose. *FEBS J.* **2016**, *283*, 4097–4112. [[CrossRef](#)] [[PubMed](#)]
30. Maksimainen, M.M.; Lampio, A.; Mertanen, M.; Turunen, O.; Rouvinen, J. The crystal structure of acidic β -galactosidase from *Aspergillus oryzae*. *Int. J. Biol. Macromol.* **2013**, *60*, 109–115. [[CrossRef](#)] [[PubMed](#)]
31. Rico-Diaz, A.; Ramirez-Escudero, M.; Vizoso-Vazquez, A.; Cerdan, M.E.; Becerra, M.; Sanz-Aparicio, J. Structural features of *Aspergillus niger* β -galactosidase define its activity against glycoside linkages. *FEBS J.* **2017**, *284*, 1815–1829. [[CrossRef](#)] [[PubMed](#)]
32. Maksimainen, M.; Hakulinen, N.; Kallio, J.M.; Timoharju, T.; Turunen, O.; Rouvinen, J. Crystal structures of *Trichoderma reesei* β -galactosidase reveal conformational changes in the active site. *J. Struct. Biol.* **2011**, *174*, 156–163. [[CrossRef](#)]

33. Rojas, A.L.; Nagem, R.A.; Neustroev, K.N.; Arand, M.; Adamska, M.; Eneyskaya, E.V.; Kulminskaya, A.A.; Garratt, R.C.; Golubev, A.M.; Polikarpov, I. Crystal structures of β -galactosidase from *Penicillium* sp. and its complex with galactose. *J. Mol. Biol.* **2004**, *343*, 1281–1292. [[CrossRef](#)] [[PubMed](#)]
34. Ohto, U.; Usui, K.; Ochi, T.; Yuki, K.; Satow, Y.; Shimizu, T. Crystal structure of human β -galactosidase: Structural basis of Gm1 gangliosidosis and morquio B diseases. *J. Biol. Chem.* **2012**, *287*, 1801–1812. [[CrossRef](#)] [[PubMed](#)]
35. Henze, M.; You, D.J.; Kamerke, C.; Hoffmann, N.; Angkawidjaja, C.; Ernst, S.; Pietruszka, J.; Kanaya, S.; Elling, L. Rational design of a glycosynthase by the crystal structure of β -galactosidase from *Bacillus circulans* (BgaC) and its use for the synthesis of *N*-acetylglucosamine type 1 glycan structures. *J. Biotechnol.* **2014**, *191*, 78–85. [[CrossRef](#)]
36. Cheng, W.; Wang, L.; Jiang, Y.L.; Bai, X.H.; Chu, J.; Li, Q.; Yu, G.; Liang, Q.L.; Zhou, C.Z.; Chen, Y. Structural insights into the substrate specificity of *Streptococcus pneumoniae* $\beta(1,3)$ -galactosidase BgaC. *J. Biol. Chem.* **2012**, *287*, 22910–22918. [[CrossRef](#)]
37. Suzuki, H.; Ohto, U.; Higaki, K.; Mena-Barragan, T.; Aguilar-Moncayo, M.; Ortiz Mellet, C.; Nanba, E.; Garcia Fernandez, J.M.; Suzuki, Y.; Shimizu, T. Structural basis of pharmacological chaperoning for human β -galactosidase. *J. Biol. Chem.* **2014**, *289*, 14560–14568. [[CrossRef](#)]
38. Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680. [[CrossRef](#)]
39. Beecher, C.N.; Larive, C.K. ^1H and ^{15}N NMR Characterization of the Amine Groups of Heparan Sulfate Related Glucosamine Monosaccharides in Aqueous Solution. *Anal. Chem.* **2015**, *87*, 6842–6848. [[CrossRef](#)]
40. Van Bueren, A.L.; Ghinet, M.G.; Gregg, K.; Fleury, A.; Brzezinski, R.; Boraston, A.B. The structural basis of substrate recognition in an exo- β -D-glucosaminidase involved in chitosan hydrolysis. *J. Mol. Biol.* **2009**, *385*, 131–139. [[CrossRef](#)] [[PubMed](#)]
41. Pluvinaige, B.; Ghinet, M.G.; Brzezinski, R.; Boraston, A.B.; Stubbs, K.A. Inhibition of the exo- β -D-glucosaminidase CsxA by a glucosamine-configured castanospermine and an amino-australine analogue. *Org. Biomol. Chem.* **2009**, *7*, 4169–4172. [[CrossRef](#)]
42. Juers, D.H.; Huber, R.E.; Matthews, B.W. Structural comparisons of TIM barrel proteins suggest functional and evolutionary relationships between β -galactosidase and other glycohydrolases. *Protein Sci.* **1999**, *8*, 122–136. [[CrossRef](#)]
43. Qian, W.; Zhang, J. Genomic evidence for adaptation by gene duplication. *Genome Res.* **2014**, *24*, 1356–1362. [[CrossRef](#)] [[PubMed](#)]
44. Bai, Y.; Gangoiti, J.; Dijkstra, B.W.; Dijkhuizen, L.; Pijning, T. Crystal structure of 4,6- α -glucanotransferase supports diet-driven evolution of GH70 enzymes from α -amylases in oral bacteria. *Structure* **2017**, *25*, 231–242. [[CrossRef](#)]
45. Eklof, J.M.; Shojania, S.; Okon, M.; McIntosh, L.P.; Brumer, H. Structure-function analysis of a broad specificity *Populus trichocarpa* endo- β -glucanase reveals an evolutionary link between bacterial licheninases and plant XTH gene products. *J. Biol. Chem.* **2013**, *288*, 15786–15799. [[CrossRef](#)]
46. Gangoiti, J.; Pijning, T.; Dijkhuizen, L. The *Exiguobacterium sibiricum* 255-15 GtfC enzyme represents a novel glycoside hydrolase 70 subfamily of 4,6- α -glucanotransferase enzymes. *Appl. Environ. Microbiol.* **2016**, *82*, 756–766. [[CrossRef](#)] [[PubMed](#)]
47. Matthews, B.W. The structure of *E. coli* β -galactosidase. *C. R. Biol.* **2005**, *328*, 549–556. [[CrossRef](#)]

