The Structure of an Archaeal β-Glucosaminidase Provides Insight into Glycoside Hydrolase Evolution*

Received for publication, November 7, 2016, and in revised form, January 6, 2017. Published, JBC Papers in Press, January 27, 2017, DOI 10.1074/jbc.M116.766535

Shouhei Mine†,2, Masahiro Watanabe†,1,3, Saori Kamachi§,1, Yoshito Abe§, and Tadashi Ueda†

From the †Biomedical Research Institute (BMD), National Institute of Advanced Industrial Science and Technology (AIST), 1-8-3 Midorigaoka, Ikeda, Osaka 563-8577, the ‡Research Institute for Sustainable Chemistry (ISIC), AIST, 3-11-32 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-0046, and the §Laboratory of Protein Structure, Function and Design, Graduate School of Pharmaceutical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan

Edited by Gerald W. Hart

The role of exo-β-D-glucosaminidase in the chitin catabolic pathway of hyperthermophilic archaea has been defined. The degradation of chitin into diacetylchitobiose (GlcNAC₂) is initiated by chitinase (ChiA) (EC 3.2.1.14), and this product is then deacetylated at its nonreducing GlcNac residue by deacetylase (Dac) (EC 3.5.1.-) (2). The resulting product, GlcN-GlcNac, is subsequently hydrolyzed into GlcN and GlcNac by exo-β-D-glucosaminidase, and the remaining GlcNac is further deacetylated to GlcN by Dac (2, 3). To understand these enzymes’ catalysis and adaptation to extreme high temperature, we had previously determined the structures of ChiA (4–6) and Dac (7); however, the structure of exo-β-D-glucosaminidase remained unknown. To date, two exo-β-D-glucosaminidases from hyperthermophilic archaea, which are called GlmA, have been described: GlmA₂₇₉ from Thermococcus kodakaraensis KOD1 (3) and GlmA₉₈ from Pyrococcus horikoshii (8). The sequence identity between GlmA₂₇₉ and GlmA₉₈ is 63%, and both enzymes show the same substrate specificities and exist as dimers in solution, suggesting that their tertiary structures and catalytic mechanisms are probably identical.

GlmA belongs to the GH35 subfamily of the GH-A superfamily, which is the largest GH superfamily and contains 19 subfamilies. All members of this superfamily include a TIM-barrel fold as a catalytic domain that contains two carboxylic acids that function as an acid/base catalyst (9, 10). Most characterized GH35 enzymes are β-galactosidases (EC 3.2.1.23), which hydrolyze the β(1–3) and β(1–4) galactosyl bonds in oligosaccharides. Interestingly, the sequence of GelmA has homology with parts of GH35 and GH42 β-galactosidases, although GelmA does not exhibit β-galactosidase activity (3). The highly conserved motifs around the catalytic residues of these β-galactosidases are not conserved in GelmA (3). Furthermore, GelmA₂₇₉ was found to exhibit weak β-glucosidase activity in addition to its major β-glucosaminidase activity (3). The only determined structure of an exo-β-D-glucosaminidase among GH-As is that from the bacteria Amycolatopsis orientalis (CsxA), a member of the GH2 subclass (11). However, GelmA is

*This work was supported in part by Grant-in-Aid for Scientific Research 25450143 from the Japan Society for the Promotion of Science (to S. M.). The authors declare that they have no conflicts of interest with the content of this article.

The authors declare that this article has essential experimental data available at https://www.wwpdb.org.

1 Both authors contributed equally to this work.

2 To whom correspondence may be addressed. Tel.: 81-72-751-9549; Fax: 81-775-561-4809; E-mail: s-mine@aist.go.jp.

3 To whom correspondence may be addressed. Tel.: 81-82-420-8285; Fax: 81-82-423-7820; E-mail: masa-watanabe@aist.go.jp.

4 The abbreviations used are: GlcN, glucosamine; GH, glycosidase hydrolase; GlmAᵢᵢ, archaeal exo-β-D-glucosaminidase from T. kodakaraensis KOD1; GlmAᵢᵢ, archaeal exo-β-D-glucosaminidase from P. horikoshii; RMSD, root mean square deviation; SeMet, selenomethionine; Bistris propane, 1,3-bis [tris(hydroxymethyl)methylamino]propane; PDB, Protein Data Bank.

Chitin is a polysaccharide consisting of β-1,4-linked N-acetylglucosamine (GlcNAC). It is a major constituent of fungal cell walls, the exoskeletons of insects, and the shells of crustaceans. Glucosamine (GlcN),† which is derived from the hydrolysis of deacetylated chitin (chitosan), has a variety of biological functions and, thus, has been used as a food additive and in medicines. Exo-β-D-glucosaminidase (EC 3.2.1.165) catalyzes the hydrolysis of the β(1–4) linkage of chitosan oligosaccharides to remove a GlcN residue from the non-reducing termini, and retaining enzymes of this type are classified into glycoside hydrolase (GH) subfamilies 2 and 35 (according to the Carbohydrate Active Enzymes (CAZy) database (1)). This enzyme is found in bacteria and archaea and has been thoroughly investigated because of its ability to produce monomeric GlcN.

The structure of an archaeal exo-β-D-glucosaminidase (GlmA) is a dimeric enzyme that hydrolyzes chitosan oligosaccharides into monomer glucosamines. GlmA is a member of the glycosidase hydrolase (GH)-A superfamily-subfamily 35 and is a novel enzyme in terms of its primary structure. Here, we present the crystal structure of GlmA in complex with glucosamine at 1.27 Å resolution. The structure reveals that a monomeric form of GlmA shares structural homology with GH42 β-galactosidases, whereas most of the spatial positions of the active site residues are identical to those of GH35 β-galactosidases. We found that upon dimerization, the active site of GlmA changes shape, enhancing its ability to hydrolyze the smaller substrate in a manner similar to that of homotrimeric GH42 β-galactosidase. However, GlmA can differentiate glucosamine from galactose although it shares with GH35 and GH42 similar to that of homotrimeric GH42 β-galactosidase.

The atomic coordinates and structure factors (codes 5GSL and 5GSM) have been deposited in the Protein Data Bank (http://wwpdb.org/).
distinct from CsxA in its substrate specificity and oligomerization state (3, 11), and it shows low sequence similarity. These results suggest that GlmAs might have a unique active site structure that does not resemble that of CsxA.

To further clarify the existing knowledge regarding these enzymes, we determined the structures of GlmA<sub>alg</sub> and GlmA<sub>T</sub> that link the molecular evolution of GH35 and GH42 β-galactosidases in GH-A.

**Results**

**Overall Structure**—The structure of GlmA<sub>alg</sub> was solved using the single-wavelength anomalous dispersion of selenomethionine and refined at 2.6 Å resolution (Table 1). The structure of GlmA<sub>T</sub> in complex with its reaction product, GlcN, was determined at 1.27 Å resolution using molecular replacement with the structure of the GlmA<sub>alg</sub> monomer as the search model (Table 1). GlmA<sub>alg</sub> and GlmA<sub>T</sub> showed almost identical tertiary structures, with a root mean square deviation (RMSD) of 0.90 Å for 775 Ca atoms (Fig. 1A). Therefore, we describe the highest-resolution structure of GlmA<sub>T</sub> throughout this report, unless otherwise noted. In the structure of GlmA<sub>T</sub>, the asymmetric unit contains two polypeptides (chains A and B) corresponding to a non-crystallographic 2-fold axis (Fig. 1B). This dimer assembly is consistent with the data obtained from gel filtration chromatography, suggesting that the crystal structure of GlmA<sub>T</sub> corresponds to the biologically relevant form of the protein. The monomer structure of GlmA<sub>T</sub> contains three distinct domains (Fig. 1B). The N-terminal domain is a (β/α)<sub>3</sub> barrel, or TIM-barrel, domain (residues 1–435). It contains the catalytic machinery and is a common structure among glycoside hydrolases. The second domain is an α/β fold domain with central β-sheets and α-helices (residues 436–648). The C-terminal domain is a β-fold domain with antiparallel β-sheets (residues 649–786). These three domains of chains A and B form extensive interactions with all of the domains of the other chain. Because the fraction of buried atoms and their interactions (salt bridges, ion networks, and hydrogen bonds) represent features that might be important for thermostability (12), we explored the dimer interface of GlmA<sub>T</sub> using PISA (Protein Interfaces, Surfaces, Assemblies) (13) software. Our analysis showed that a large surface area of 5530 Å<sup>2</sup> is buried in the structure and that 29 hydrogen bonds and 16 salt bridges are created upon dimer formation. These findings indicate that both monomers are intimately associated, which might contribute to the high thermostability of GlmA<sub>T</sub>.

**Structural Comparison with Other Glycoside Hydrolases**—Structural similarity searches using the DALI server with the whole protein revealed that the dimer structure of GlmA<sub>T</sub> does not resemble those of any others. However, surprisingly, the monomer structure of GlmA<sub>T</sub> shares substantial similarity with other glycoside hydrolases, such as the archaeal exo-β-D-glucosaminidase complex (Fig. 1C). The overall structure of GlmA that link the molecular evolution of GH35 and GH42 β-galactosidases in GH-A.

**TABLE 1**

| Data collection and refinement statistics | GlmA<sub>alg</sub> | GlmA<sub>alg</sub> SeMet | GlmA<sub>T</sub> GlcN complex |
|-----------------------------------------|----------------|----------------|--------------------------|
| Space group                            | 222           | 222           | 222                      |
| Cell dimensions (Å)                    | 4997, 4997, 4997 | 4997, 4997, 4997 | 4997, 4997, 4997          |
| a, b, c (Å)                            | 149.84, 149.78, 149.78 | 149.84, 149.78, 149.78 | 149.84, 149.78, 149.78 |
| ω, β, γ (%)                            | 90, 90, 90    | 90, 90, 90    | 90, 90, 90                |
| Resolution (Å)                         | 1.29, 1.29     | 1.29, 1.29     | 1.29, 1.29                |
| Flack (‰)                              | 0.9900         | 0.9900         | 0.9900                    |
| Wavelength                             | 0.91000        | 0.91000        | 0.91000                   |
| Resolution (Å)                         | 0.9000         | 0.9000         | 0.9000                    |
| Flack (‰)                              | 0.9799         | 0.9799         | 0.9799                    |
| Wavelength                             | 0.9000         | 0.9000         | 0.9000                    |

**FIGURE 1. The overall structure of GlmA.** A, the structural superposition of GlmA<sub>alg</sub> (cyan) and GlmA<sub>T</sub> (orange) is shown as a ribbon diagram. B, the dimer structure of GlmA<sub>T</sub> is presented in two views. GlmA<sub>T</sub> consists of a homodimer (chains A and B) and comprises three distinct domains (TIM-barrel (red), α/β (blue), and β (green)). The bound GlcN is shown as a sphere. Arrow, N-terminal β-sheet (residues 4–13).
with GH42 β-galactosidases, whereas the TIM-barrel domain is found to be more similar to the architecture of GH35 β-galactosidases rather than to that of GH42; this observation will be explained below.

First, despite low sequence similarities (15–17%), the three domain structures of monomer GlmA29 could be readily superimposed on four GH42 β-galactosidases with Z scores >25 and RMSD values of 2.6–3.0 Å for equivalent Ca atoms, except for 80 residues of the C-terminal region of GlmA29 (Fig. 2A). This result indicates that these β-galactosidases are spatially homologous to GlmA29 at the level of individual domain folds and domain orientation, although their quaternary structures differ from that of GlmA29 (Fig. 3A) and belong to different GH families within the GH-A family. In contrast, the residues in the TIM-barrel domain of GlmA29 show low structural similarities with four GH35 β-galactosidases: Thermus thermophilus A4-β-gal (Z = 25.5; RMSD = 3.0 Å; PDB code 1KWK (40)), Bacillus circulans sp. Bca-β-gal (Z = 27.8; RMSD = 3.0 Å; PDB code 3TTY (41)), Geobacillus stearothermophilus Gal42B (Z = 26.2; RMSD = 2.7 Å; PDB code 4OIF (24)), and Bifidobacterium animalis BIgA2A (Z = 29.4; RMSD = 2.6 Å; PDB 4UNJ (42)). The structures are colored as follows: GlmA29 (red), A4-β-gal (light yellow), Bca-β-gal (cyan), and BIgA2A (light green). The figure is drawn from the same orientation as the right panel of Fig. 18. The right panel shows the sequence alignment around the catalytic residues (red) of the GH42 β-galactosidases from ClusterA (43). Superimposed models of the TIM-barrel domain (chain A) of GlmA29 and GH35 β-galactosidases. A DALI search was conducted on the residues in the TIM-barrel domain of GlmA29 (residues 1–435) only, which resulted in a top hit consisting of four GH35 β-galactosidases: Homo sapiens Hs-β-gal (Z = 38.3; RMSD = 1.8 Å; PDB code 3THC (19)), T. reesei Tri-β-gal (Z = 36.8; RMSD = 2.2 Å; PDB code 3OGR (14)), Penicillium sp. Psp-β-gal (Z = 36.7; RMSD = 2.3 Å; PDB code 1XC6 (20)), and Streptococcus pneumoniae BgaC (Z = 36.2; RMSD = 2.0 Å; PDB code 4ESC (21)). The TIM-barrel structures are colored as follows: GlmA29 (red), Hs-β-gal (light yellow), Tri-β-gal (light blue), Psp-β-gal (cyan), and BgaC (light green). The right panel shows the sequence alignment around the catalytic residues (red) of the GH35 β-galactosidases from ClusterA (43). Chain A and chain B are quite similar, with an overall RMSD of 0.43 Å over 786 Ca atoms; thus, superposition of chain B against GH42 β-galactosidases and the TIM-barrel domains of GH35 β-galactosidases showed almost the same RMSD values as those of chain A. Accordingly, for simplicity, only the results of chain A are shown in the figure.

Interestingly, the superimposition of the TIM-barrel domain of GlmA29 with Tri-β-gal revealed a high degree of structural similarity between the −1 subsites of these proteins (Fig. 4B), although their substrates are different. Both the GlcN and galactose molecules adopt a chair conformation with their C1 hydroxyl group (O1) in the β-anomer configuration. They sit in almost the same position and form direct hydrogen bonds with eight residues. Four of the eight residues involved in direct substrate binding in GlmA29 (Try53 (interacting with O3), Glu103 (O4, O6), Glu779 (O1), and Glu347 (N2)) could be superimposed onto Tyr96 (O3), Glu142 (O4, O6), Glu200 (O1), and Glu298 (O2) of Tri-β-gal, respectively, resulting in almost identical protein-carbohydrate interactions with no substantial differences in the interatomic distances (Fig. 4, C and D). Gly102 of GlmA29 forms a hydrogen bond with O3 (2.9 Å) of GlcN via its main-chain amide, whereas the structurally equivalent residue of Tri-β-gal is Ala141, which fulfills the same function through a hydrogen bond to O3 (2.9 Å) of galactose (Fig. 4, C and D). Therefore, this replacement is a conservative substitution. In addition, Thr308 of GlmA29, which is involved in a hydrophobic stacking interaction with the planar face of the GlcN moiety, overlaps well with Tyr260 of Tri-β-gal (Fig. 4B). The component important for the recognition of the GlmA29 Substrate was not conserved.
in GH42 β-galactosidase (data not shown), despite the structural similarity of their monomers.

Among the substrate-binding residues described above, Glu179 and Glu347 of GlmA<sub>Tk</sub> are supposed to be catalytic residues, and the steric counterparts of Tr<sub>-</sub>-β-gal are the acid/base Glu200 and the nucleophile Glu298, respectively (Fig. 4E). Glu179 forms a hydrogen bond with O1 (2.7 Å) and is oriented toward the glycosidic oxygen, whereas Glu347 forms a hydrogen bond with N2 (2.8 Å) and is positioned to serve as a catalytic nucleophile (Fig. 4, C and E). Consistent with the predicted roles of these residues, the mutation E347Q virtually inactivated the enzyme, whereas the mutation E179Q retained less than ~3% of residual hydrolysis activity (Fig. 5, A and B) (Table 2). In the measurement performed here, acetate was produced as one of the reaction products (Fig. 5A) and could act as a nucleophile (15–17). Therefore, the retained activity of E179Q might be ascribed to the chemical rescue of acetate. In addition, Glu179 and Glu347 are located in the β4- and β7-strands of the TIM-barrel domain, respectively, and the average distance between the oxygen atoms of these residues is 4.8 Å. This is consistent with the common structural features of the retaining enzyme in GH-A (9). Thus, these results strongly suggest that GlmA<sub>Tk</sub> hydrolyzes its substrate in a double displacement retaining mechanism using the acid/base residue Glu179 and nucleophilic residue Glu347 similarly to the GH35 β-galactosidases characterized thus far. The catalytic center, which is almost entirely conserved, indicates the close evolutionary relationship between these enzymes.

Based on their sequence alignment (Fig. 2, A and B), the acid/base residue Glu179 of GlmA<sub>Tk</sub> aligns with those of GH35 and GH42 β-galactosidases. The nucleophile Glu347 of GlmA<sub>Tk</sub> also aligns with those of GH42 β-galactosidases. However, it could not be aligned with those of GH35 β-galactosidases. Together with the fact that the highly conserved motifs around the catalytic residues of GH35 and GH42 β-galactosidases are not conserved in GlmA<sub>Tk</sub> (3), these results indicate that the locations of the catalytic residues predicted based on the sequence comparisons are uncertain and unreliable. Our structure determination of GlmA<sub>Tk</sub> combined with structure-guided mutagenesis studies facilitated identifying the catalytic residues accurately.

The Discrimination of GlcN from Galactose by GlmA<sub>Tk</sub>—Despite the high structural similarities at the active site, radical differences were observed in the rest of the substrate-binding residues of GlmA<sub>Tk</sub>: Asp178, Tyr379, and Glu306. The former two residues correlated with the chemical structure of GlcN. GlcN and galactose differ in terms of the substituent at C2 and the chirality of C4 and C6. The major difference is the substituent at C2, which is an amine group (N2) in GlcN and a hydroxyl group (O2) in galactose. Asn199 of Tri<sub>-</sub>-β-gal, which precedes the acid/
base residue Glu200, forms a hydrogen bond (2.9 Å) with O2 (Fig. 4, D and F), and this Asn-Glu motif is highly conserved in GH35 and GH42 β-galactosidases (Fig. 2, A and B). The equivalent motif in GlmA_{Tk} is Asp178-Glu179 (the acid/base), and Asp178 forms a hydrogen bond (2.7 Å) with N2 (Fig. 4, C and F). The pK_{a} of N2 in GlcN is reported to be 7.4 (18); therefore, at pH 6.0, at which the activity of GlmA_{Tk} is maximized (3), the N2 of GlcN will be in its protonated NH_{3}^{+} form. Additionally, the side chain of Asp178 will be negatively charged based on its average pK_{a} of ~3.7. To confirm the importance of the acidic character of the carboxyl group of Asp178, we mutated Asp178 to asparagine. As a result, the D178N mutant dramatically lost its catalytic activity (Fig. 5B) (Table 2), implying that charge-charge complementarity is indispensable for the interaction between Asp178 and the N2 of GlcN. The necessity of this interaction is also supported by a previous report indicating that...
FIGURE 5. The activities of wild-type and mutant GlmAs. A, mechanism of action of GlmA hydrolase using Dac deacetylase. First, (GlcNAc)₂ is deacetylated at its nonreducing GlcNAc residue by Dac. The resulting product, GlcN-GlcNAc, is then hydrolyzed into GlcN and GlcNAc by GlmA, and the remaining GlcNAc is further deacetylated to GlcN by Dac. Peaks I and II are those shown in B and C. B and C, representative NMR spectra of the methyl proton regions. The left spectra in both B and C correspond to the reaction mixtures of Dac with (GlcNAc)₂. The other spectra represent the subsequent reaction with GlmAₜk (wild type, D178N, E179Q, E306Q, and E347Q) (B) and GlmAₚh (wild type, D180N, E181Q, E308Q, and E349Q) (C) as their counterparts in GlmAₜk, shown in the right panel of Fig. 2, A and B, in the presence of Dac. Peaks I and II are assigned to the N-acetyl group of GlcN-GlcNAc and the methyl group of the resulting acetic acid, respectively, here and in the reaction scheme shown in A.
GlmA$_{Tk}$ has very weak $\beta$-glucosidase activity (3). Chemically, glucose differs from GlcN only at the C2 of the pyranose ring, which contains a hydroxyl group (O2), indicating that the absence of a charged interaction between Asp$_{278}$ and O2 of glucose should cause a profound loss of $\beta$-glucosidase activity. Additionally, GlmA$_{Tk}$ could not hydrolyze (GlcNac)$_2$ at all (3). GlcNAc also differs from GlcN only at the C2 substituent, which is replaced by a bulky acetoamido group; thus, the presence of GlcNAc at the $\alpha$-1 subsite cannot be tolerated because of its steric clash with Asp$_{178}$.

An additional difference between GlcN and galactose is the chirality of O4, which is equatorial in GlcN and axial in galactose. Tyr$_{379}$ of GbmA$_{Tk}$ forms a hydrogen bond (2.8 Å) with the equatorial O4 of GlcN (Fig. 4, C and G) and also serves as the lateral face of the hydrophobic pocket to accommodate GlcN. Surprisingly, Tyr$_{379}$ could be superimposed onto Tyr$_{342}$ of Tri-$\beta$-gal (Figs. 4, B and G), which is a strictly conserved residue in GH35 $\beta$-galactosidases. Tyr$_{342}$ packs against the C4 atom of galactose in a similar manner as Tyr$_{379}$ of GbmA$_{Tk}$; however, its position is too distant (4.6 Å) to form a hydrogen bond with the axial O4 of galactose. Instead of Tyr$_{342}$, Asn$_{140}$ of Tri-$\beta$-gal is in a suitable position to form a hydrogen bond (2.8 Å) with the axial O4 (Fig. 4, D and G). Likewise, Asn$_{140}$ could be structurally superimposed onto Cys$_{101}$ of GbmA$_{Tk}$ (Fig. 4, B and G), and other GH35 $\beta$-galactosidases, such as BgaC and Hs-$\beta$-gal, also have Cys residues at the same position. In BgaC, the counterpart Cys$_{366}$ forms a hydrogen bond with the axial O4 of galactose via its thiol group (19). Therefore, Cys$_{101}$ of GbmA$_{Tk}$ is supposed to be a conservative substitution. However, it is located 4.8 Å away from the equatorial O4 of GlcN, preventing hydrogen bond formation. Briefly, GbmA$_{Tk}$ and GH35 $\beta$-galactosidases possess residues with the potential to form hydrogen bonds with the axial and equatorial forms of O4 in the glycosidic substrate, respectively, thereby contributing to the recognition of GlcN or galactose. Therefore, we defined these Cys (Asn) and Tyr residues as "evolutionary heritage residues." To our knowledge, this is the first time such a heritage was seen in different functional glycoside hydrolases.

The other unique substrate-binding residue of GbmA$_{Tk}$ is Glu$_{306}$, which forms a hydrogen bond (2.9 Å) with O1 (Fig. 4, C and H) and is important for maximal catalytic activity; this was confirmed by determining that the mutation E306Q decreased the enzymatic activity $\sim$2.5-fold (Fig. 5B) (Table 2). In Tri-$\beta$-gal, Asp$_{258}$ is located at this position (Fig. 4, B and H); however, it cannot form a hydrogen bond with the O1 in galactose because of its side chain orientation (at a distance of $\sim$4.6 Å). Because the configuration of O1 of GlcN and galactose is a $\beta$-anomer, Glu$_{306}$ may contribute to transition state stabilization rather than substrate recognition. In contrast, as described above, the O6 hydroxyl group of GlcN and galactose form hydrogen bonds with Glu$_{103}$ of GbmA$_{Tk}$ and Glu$_{142}$ of Tri-$\beta$-gal, respectively, despite the different chirality of the O6 hydroxyl group (Fig. 4, C and D). However, Tyr$_{364}$ of Tri-$\beta$-gal forms an additional hydrogen bond with the equatorial O6 of galactose (Fig. 4D), whereas the corresponding residue is absent in GbmA$_{Tk}$. Tyr$_{364}$ is well conserved in other GH35 $\beta$-galactosidases (14, 19–21), indicating that it contributes to the recognition of galactose, but not in a major way.

These data indeed suggest that Asp$_{178}$, Glu$_{306}$, and Tyr$_{379}$ of GbmA$_{Tk}$ play an important role in the recognition or stabilization of the GlcN molecule. However, as explained above, Asp$_{178}$ is supposed to be the most important residue responsible for the recognition of GlcN.

Additionally, the sequence alignment between GbmA$_{Tk}$ and GbmA$_{ph}$ illustrated that the catalytic residues, together with the other key residues of GbmA$_{Tk}$ discussed above, were strictly conserved in GbmA$_{ph}$ (data not shown) and are consistent with the mutagenesis analyses of GbmA$_{ph}$ (Fig. 5C) (Table 2). These findings suggest that their catalytic mechanisms and substrate profiles are probably identical.

**Table 2. Catalytic activity of GbmA and mutants**

| Enzyme     | Relative activity $^a$ (%) |
|------------|---------------------------|
| GbmA$_{Tk}$ Wild type | 100       |
| D178N      | 0           |
| E179Q      | 3           |
| E306Q      | 40          |
| E347Q      | 0           |
| GbmA$_{ph}$ Wild type | 100       |
| D180N      | 0           |
| E181Q      | 10          |
| E308Q      | 15          |
| E349Q      | 0           |

$^a$ The activities of the mutants are calculated from the relative peak I area against that of the wild type.

**The Dimer Structure Influences Substrate Specificity**—In the TIM-barrel domain, one of the ends of the $\beta$-barrel is closed by the N-terminal $\beta$-sheet (residues 4–13) (Fig. 1B), whereas the other end, termed the "catalytic face" (22), is buried within a deep and narrow pocket upon dimer formation. The most notable feature of the dimer interface is that the C$_{\alpha}$-domain protrudes toward the catalytic face of the adjacent monomer, interacting via some hydrogen bonds and a salt bridge. These interactions involve Arg$_{563}$, Asn$_{565}$, and Arg$_{567}$ of the $\alpha$-$\beta$-domain and Asp$_{132}$, Tyr$_{134}$, Tyr$_{135}$, and Glu$_{188}$ of the TIM-barrel domain (Fig. 6A). This interaction decreased the size of the active site entrance, and the $\alpha$-$\beta$-domain and the TIM-barrel domain of chains A and B create one large cavity, which abuts the active site pockets of both monomers (Fig. 6A). As previously stated, the opposite side of the catalytic face is closed; thus, this cavity is the only means of entry or egress for the substrate or product. The depth of the active site pocket from the center of the cavity is $\sim$20 Å (Fig. 6B), which could restrict the access of lengthy substrates. Consistently, GbmA$_{Tk}$ showed higher activity against GlcN$_2$ ($\sim$12 Å in length) than against longer molecules (N$_{2-6}$) (3). These results suggest that dimer formation is essential for GbmA$_{Tk}$ to exhibit enzymatic activity and to form an active site with an appropriate shape. Conversely, CsxA, which is the only other exo-$\beta$-$d$-glucosaminidase with a known structure, can hydrolyze oligomeric substrates ranging from GlcN$_2$ to GlcN$_6$ with similar efficiencies (11). CsxA is a member of the GH2 family of the GH-A group and functions as a monomeric enzyme with its active site easily accessible to the solvent. However, its domain organization differs substantially from that of GbmA$_{Tk}$ (data not shown). This structural feature of CsxA is suitable for longer molecules.
Discussion

The structure of GlmA\textsubscript{Tk} provides new insights into the structural composition and substrate recognition mechanisms of different enzymes and, thus, their molecular evolution. Briefly, a monomeric form of GlmA\textsubscript{Tk} shares substantial structural similarity with GH42 \(\beta\)-galactosidases, whereas a high number of conserved active site residues are shared with GH35 \(\beta\)-galactosidase, allowing GlmA\textsubscript{Tk} to discriminate glucosamine from galactose based on a subtle difference in the structure of \(\beta\)-galactosidase bound to galactose. Indeed, Asp\textsuperscript{178} of GlmA\textsubscript{Tk} plays an essential role in the discrimination of GlcN from galactose, whereas the equivalent in GH35 \(\beta\)-galactosidase is an Asn residue. To the best of our knowledge, this is the first observation of such a high degree of conservation within the entire catalytic centers of different enzymes. In addition, the evolutionary heritage residues, which have the potential to form hydrogen bonds with the axial and equatorial forms of O4 in the glycosidic substrate, respectively, are an interesting finding that emphasizes the high evolutionary conservation of these enzymes. These structural features strongly suggest that GlmA is a common ancestor of these \(\beta\)-galactosidases, as discussed below.

The active sites of glycoside hydrolases are classified into three types: cleft type, tunnel type, and pocket type (23). Both GlmA and GH42 \(\beta\)-galactosidases have a cleft-type active site in their monomeric forms; however, the shape of the active site changes to a pocket type upon oligomerization, which can better accommodate smaller substrates (24) (Fig. 6). Thus, oligomerization is a key factor for size-based substrate specificity and the high stability of these proteins. \(\beta\)-Galactosidase may have evolved from a prototypical single TIM-barrel domain with a cleft- or tunnel-type active site, and then, during the subsequent process of modifying the active site to prefer a smaller substrate, extra domains were added to change the active site from a cleft to a pocket type (25). As described above, the monomer structure and a part of the sequence of GlmA show similarity to those of GH42 \(\beta\)-galactosidase, suggesting that GH42 \(\beta\)-galactosidase might have emerged from the evolutionary branch that originated from GlmA in the oldest organisms, archaea, and then differentiated into other members of the glycoside hydrolase family. Additionally, the frameworks of their monomer structure (i.e. the domain organization) might be suitable or necessary for oligomerization. However, the substrate-binding residues of GH42 enzymes are not conserved in GlmA (data not shown), excluding GlmA from being classified into the GH42 family, and the underlying evolutionary selection pressures that led to this diversity in the active site remain unknown. In contrast, the residues within the active site pocket are well conserved between GlmA and GH35 \(\beta\)-galactosidase, suggesting that GH35 \(\beta\)-galactosidase evolved from archaeal exo-\(\beta\)-D-glucosaminidase through gene duplication. Enzyme substrate ambiguity is probably the starting point for the evolution of divergent enzymes through gene duplication (26, 27). Consistently, GlmA\textsubscript{Tk} exhibits broad substrate specificity, showing weak hydrolytic activities toward various \(\beta\)-disaccharides (such as cellobiose and laminaribiose) in addition to its major \(\beta\)-glucosaminidase activity (3). Thus, the promiscuous activities of GlmA\textsubscript{Tk} might have developed through mutations that affected the subsequent functional adaptation of the newly emergent \(\beta\)-galactosidases (which favored \(\beta\)-galactoside) while retaining the original substrate-binding residues and the catalytic machinery. However, the highly conserved active site residues of both GlmA and GH35 \(\beta\)-galactosidase indicate that there has been relatively weak evolutionary pressure on the catalytic center to convert the enzyme to perform different functions. As described before, GlmA and GH35 and GH42 \(\beta\)-galactosidases belong to the same GH-A “superfamily.” A superfamily is a group that shows significant similarities in the tertiary structure together with conservation of the catalytic residues and mechanism, and its members are therefore considered to have a common ancestry (28). Accordingly, our finding that GlmA shares structural and mechanistic features with both the GH35 and GH42 \(\beta\)-galactosidases strongly suggested that GlmA is a common ancestor of these \(\beta\)-galactosidases.

Taken together, our results suggest that GH35 and GH42 \(\beta\)-galactosidases have evolved by taking advantage of the structural features of GlmA. The structural information
Structure of the Archaeal β-Glucosaminidase Complex

reported here for GlmA could be used to design a new enzyme, such as a thermostable β-galactosidase or β-glucosidase, by subtly changing the active site residues in GlmA.

Experimental Procedures

Protein Expression and Purification—The genes encoding GlmA<sub>Ph</sub> (residues 1–778, GenBank<sup>TM</sup> accession number PH_RS02375) and GlmA<sub>Tk</sub> (residues 1–786, GenBank<sup>TM</sup> accession number AB100422) were codon-optimized for expression in <i>E. coli</i> and synthesized (Eurofins NMG Operon). The constructs for GlmA<sub>Ph</sub> and GlmA<sub>Tk</sub> were cloned into a pCold-II vector (Takara Bio) and a PET-32b vector (Merck Millipore), respectively, with an N-terminal PreScission protease cleavage site followed by a hexahistidine tag using the NdeI and EcoRI restriction sites. The resulting vectors were transformed into <i>E. coli</i> Rosetta (DE3)pLysS (Merck Millipore). For GlmA<sub>Ph</sub> expression, the cells were grown at 37 °C in lysogeny broth (LB) medium until an A<sub>600 nm</sub> of 0.5 was reached, and then the cultivation was continued for 24 h at 15 °C. The overexpression of GlmA<sub>Tk</sub> was induced with 1.0 mM isopropyl-β-D-galactoside (IPTG) and 80 °C. For protein purification, the cells were disrupted by sonication in buffer A (50 mM Tris-HCl, 0.5 M NaCl, pH 8.0). The cell debris and insoluble proteins were removed by centrifugation at 15,000 × g for 30 min after heat treatment at 80 °C for 30 min. The supernatant was loaded onto a nickel-nitrilotriacetic acid (GE Healthcare) column equilibrated with buffer A, and the bound protein was eluted with buffer A containing 50 mM imidazole. Subsequently, the eluted protein was dialyzed against 50 mM Tris-HCl (pH 8.0) in the presence of PreScission protease at 4 °C overnight. The cleaved tag was removed with a second nickel-nitrilotriacetic acid purification, and the flow-through was further purified using a HiTrap Q HP column (GE Healthcare) with a linear gradient of 0–1.0 M NaCl in 50 mM Tris-HCl (pH 8.0). The further purified protein was then subjected to a HiLoad 26/600 Superdex 200 preparation grade column (GE Healthcare) equilibrated with buffer (20 mM Tris-HCl, 150 mM NaCl, pH 8.0), concentrated to 10 mg/ml, and stored at −80 °C.

The point mutants of GlmA<sub>Ph</sub> and GlmA<sub>Tk</sub> were generated using site-directed mutagenesis. All mutant proteins were expressed and purified in the same manner as the wild-type proteins. The expression levels and isolated yields of these mutants were comparable with those of the wild-type enzyme.

Crystallography—After many crystallization trials of GlmA<sub>Ph</sub> and GlmA<sub>Tk</sub>, only GlmA<sub>Ph</sub> produced crystals. The crystals of GlmA<sub>Ph</sub> were grown at 20 °C using the sitting drop vapor diffusion method in 100 mM Tris-HCl (pH 8.5) and 20% (w/v) polyethylene glycol (PEG) 1000 with a protein/reservoir volume ratio of 1:1. Crystals appeared within a week. Therefore, we prepared selenomethionine-substituted GlmA<sub>Ph</sub> according to the procedure described under “Protein Expression and Purification,” except the cells were grown in LeMaster broth (29). The selenomethionine-incorporated protein was crystallized using the same conditions as for the native protein. Next, the co-crystallization of GlmA<sub>Ph</sub> and GlmA<sub>Tk</sub> with GlcN (10, 50, and 100 mM) was performed using commercial crystallization screens. However, we obtained crystals for only the GlmA<sub>Tk</sub>-GlcN complex, not the GlmA<sub>Ph</sub>-GlcN complex. Among the several sets of conditions attempted for GlmA<sub>Tk</sub>-GlcN complex crystallization, well diffracting crystals of the GlmA<sub>Tk</sub>-GlcN complex were obtained via sitting drop vapor diffusion at 20 °C in 100 mM sodium propionate-sodium cacodylate-Bistris propane buffer (pH 7.0), 25% (w/v) PEG 1500, and 50 mM GlcN. Crystals of GlmA<sub>Ph</sub> (SeMet and native) and GlmA<sub>Tk</sub>-GlcN were cryoprotected with 15% (v/v) glycerol and 25% (w/v) PEG 400, respectively, and subsequently flash-frozen in liquid nitrogen. All data sets were collected on a BL44XU instrument at SPring-8 (Harima, Japan) with a MX300HE detector (Rayonix) under a cryostream at 90 K, and the data were processed and scaled using the HKL-2000 program suite (30). Data sets for native and SeMet GlmA<sub>Ph</sub> were collected at wavelengths of 0.9 and 0.97898 Å, respectively, with a single-wavelength anomalous dispersion of selenium atoms (peak). The initial phases of the GlmA<sub>Ph</sub> structure were determined at 3.5 Å resolution from the data of SeMet GlmA<sub>Ph</sub> using Phenix (31), and then the native data were phase-extended to 2.6 Å. The GlmA<sub>Ph</sub> model was built using Phasix. Further model building was performed with Coot (32). The structure was refined by CNS (Crystallography and Nuclear Magnetic Resonance (NMR) System) (33) and REFMAC (Refinement of Macromolecular Structures) (34) with rigid body refinement. Data sets of the GlmA<sub>Tk</sub>-GlcN complex were also collected at a wavelength of 0.9 Å. The initial phases of the GlmA<sub>Tk</sub>-GlcN complex model were solved using Phaser (35) with the structure of GlmA<sub>Ph</sub> as a search model and then the model building and refinement and the addition of water molecules were performed using ARP/wARP (36). Further model building and refinements were performed using Coot and REFMAC with individual anisotropic B-factor value refinement, respectively. The final models of the GlmA<sub>Ph</sub> and GlmA<sub>Tk</sub>-GlcN complexes were validated by MolProbity (37).

Enzymatic Activities of GlmA<sub>Tk</sub> and GlmA<sub>Ph</sub> Mutants—The hydrolase activities of the GlmAs were detected using <sup>1</sup>H NMR spectroscopy. NMR experiments were conducted at 35 °C on a Varian Inova (600 MHz) (Palo Alto, CA) equipped with a z-gradient, triple-resonance TR probe. The chemical shifts were referenced to an internal standard; 4,4-dimethyl-4-silapentane-1-sulfonic acid. The substrate GlcN-GlcNac for the GlmAs was prepared as follows. 1.6 mM (GlcNac)<sub>2</sub> was incubated with 5 μM Dac for 10 min in 50 mM potassium phosphate buffer (pH 6.0) containing 10% (w/v) D<sub>2</sub>O. After the first reaction was completed, the subsequent reaction was initiated by adding wild-type GlmAs or mutants to a final concentration of 5 μM in the presence of Dac, and the spectra were collected after 150 min. One of the resulting products, GlcNac, was immediately deacetylated to GlcN by Dac; thus, the NMR spectrum of the N-acetyl group of GlcN-GlcNac was expected to disappear when the GlmA reaction was complete (Fig. 5A). The water signal was suppressed using the Watergate pulse sequence (38). The one-dimensional <sup>1</sup>H signals consisted of 8,192 sampling points covering a spectral width of 15 ppm. The relaxation delay was set at 1 s, and 64 scans were accumulated for each spectrum. This acquisition was repeated every 1.45 s. The chemical shifts were assigned as described previously (7).
Structure of the Archaeal β-Glucosaminidase Complex

15. Wang, Q., Trimbur, D., Graham, R., Warren, R. A., and Withers, S. G. (1995) Identification of the acid/base catalyst in Agrobacterium faecalis β-glucosidase by kinetic analysis of mutants. Biochemistry 34, 14554–14562

16. Richard, I. P., Huber, R. E., Heo, C., Amyes, T. L., and Lin, S. (1996) Structure-reactivity relationships for β-galactosidase (Escherichia coli, lac Z). 4. Mechanism for reaction of nucleophiles with the galactosyl-enzyme intermediates of E461G and E461Q β-galactosidases. Biochemistry 35, 12387–12401

17. Schwarz, A., Brecker, L., and Nidetzky, B. (2007) Acid-base catalysis in Leuconostoc mesenteroides sucrose phosphorylase probed by site-directed mutagenesis and detailed analysis of kinetic type of wild-type and Glu337 → Gln mutant enzymes. Biochem. J. 403, 441–449

18. Beecher, C. N., and Larive, C. K. (2015) 1H and 13C NMR characterization of the amine groups of heparan sulfate related glucosamine monosaccharides in aqueous solution. Anal. Chem. 87, 6842–6848

19. Ohto, U., Usui, K., Ochi, T., Yuki, K., Satow, Y., and Shimizu, T. (2012) Crystal structure of human β-galactosidase: structural basis of Gm1 ganglosidosis and mogroiso B diseases. J. Biol. Chem. 287, 1801–1812

20. Rojas, A. L., Nagem, R. A., Neustroev, K. N., Arand, M., Adamska, M., Eneyaskaya, E. V., Kulminskaya, A. A., Garratt, R. C., Golubev, A. M., and Polikarpov, I. (2004) Crystal structures of β-galactosidase from Pencillium sp. and its complex with galactose. J. Mol. Biol. 343, 1281–1292

21. Cheng, W., Wang, L., Jiang, Y. L., Bai, X. H., Chu, J., Li, Q., Yu, G., Liang, Q. L., Zhou, C. Z., and Chen, Y. (2012) Structural insights into the substituent specificity of Streptococcus pneumoniae β(1,3)-galactosidase BgaC. J. Biol. Chem. 287, 22910–22918

22. Sterner, R., and Höcker, B. (2005) Catalytic versatility, stability, and evolution of the βαβ-barrel enzyme fold. Chem. Rev. 105, 4038–4055

23. Davies, G., and Henriessat, B. (1995) Structures and mechanisms of glycosyl hydrodrolases. Structure 3, 853–859

24. Solomon, H. V., Tabachnikov, O., Lansky, S., Salama, R., Feinberg, H., Shoham, Y., and Shoham, G. (2015) Structure-function relationships in Gar42B, an intracellular GH42 β-galactosidase from Geobacillus stearothermophilus. Acta Crystallogr. D Biol. Crystallogr. 71, 2433–2448

25. Juers, D. H., Huber, R. E., and Matthews, B. W. (1999) Structural comparisons of TIM barrel proteins suggest functional and evolutionary relationships between β-galactosidase and other glycosyrdrolases. Protein Sci. 8, 122–136

26. Khersonsky, O., and Tawfik, D. S. (2010) Enzyme promiscuity: a mechanistic and evolutionary perspective. Annu. Rev. Biochem. 79, 471–505

27. Pandya, C., Farelli, J. D., Dunaway-Mariano, D., and Allen, K. N. (2014) Enzyme promiscuity: engine of evolutionary innovation. J. Biol. Chem. 289, 30229–30236

28. Henriessat, B., and Bairoch, A. (1996) Updating the sequence-based classification of glycosyl hydrodrolases. Biochem. J. 316, 695–696

29. Hendrickson, W. A., Horton, J. R., and LeMaster, D. M. (1990) Selenomethionyl proteins produced for analysis by multilavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. EMBO J. 9, 1665–1672

30. Terwilliger, T. C., and Berendzen, J. (1999) Automated MAD and MIR structure solution. Acta Crystallogr. D Biol. Crystallogr. 55, 849–861

31. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Readman, D. C., et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. 66, 213–221

32. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132

33. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. 54, 905–921
34. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* 53, 240–255
35. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674
36. Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* 3, 1171–1179
37. Word, J. M., Lovell, S. C., LaBean, T. H., Taylor, H. C., Zalis, M. E., Presley, B. K., Richardson, J. S., and Richardson, D. C. (1999) Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. *J. Mol. Biol.* 285, 1711–1733
38. Pietto, M., Saudek, V., and Sklenár, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* 2, 661–665
39. Holm, L., and Sander, C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* 233, 123–138
40. Hidaka, M., Fushinobu, S., Ohtsu, N., Motoshima, H., Matsuzawa, H., Shoun, H., and Wakagi, T. (2002) 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.* 322, 79–91
41. Maksimainen, M., Paavilainen, S., Hakulinen, N., and Rouvinen, J. (2012) Structural analysis, enzymatic characterization, and catalytic mechanisms of β-galactosidase from *Bacillus circulans* sp. alkalophilus. *FEBS J.* 279, 1788–1798
42. Viborg, A. H., Fredslund, F., Katayama, T., Nielsen, S. K., Svensson, B., Kitaoka, M., Lo Leggio, L., and Abou Hachem, M. (2014) A β1–6/β1–3 galactosidase from *Bifidobacterium animalis* subsp. lactis BI-04 gives insight into sub-specificities of β-galactoside catabolism within *Bifidobacterium*. *Mol. Microbiol.* 94, 1024–1040
43. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948