G\textsubscript{M1} gangliosidoses and Morquio B are autosomal recessive lysosomal storage diseases associated with neurodegenerative disorder or dwarfism and skeletal abnormalities, respectively. These diseases are caused by deficiencies in the lysosomal enzyme \(\beta\)-galactosidase (\(\beta\)-Gal), which lead to accumulations of the \(\beta\)-Gal substrates, \text{G}_{\text{M1}}\text{galactoside}, and keratan sulfate. \(\beta\)-Gal is an exoglycosidase that catalyzes the hydrolysis of terminal \(\beta\)-linked galactose residues. This study shows the crystal structures of human \(\beta\)-Gal in complex with its catalytic product galactose or with its inhibitor 1-deoxygalactonojirimycin. Human \(\beta\)-Gal is composed of a catalytic TIM barrel domain and two \(\beta\)-domains. The mechanisms by which these two disorders are caused have not been fully clarified. From patients with different ethnic backgrounds, a variety of mutations have been identified (9). \(\beta\)-Gal activity is almost lost in patients with severe infantile \text{G}_{\text{M1}}, (type I) (<1% of normal), whereas in juvenile or adult \text{G}_{\text{M1}} patients (type II and III), residual activities of \(\beta\)-Gal were detectable but very weak (<9% of normal). This indicates that the residual enzyme activity is inversely correlated with the severity of the clinical types of this disease (6, 10).

Human \(\beta\)-Gal cDNA encodes 677 amino acid residues, which include an N-terminal 23-amino acid secretion signal (11, 12) (Fig. 1A). The enzyme is synthesized as an 88-kDa precursor and is transferred to the lysosomal compartment (13), where it is processed to the 64-kDa mature enzyme by proteolytic cleavage of the C-terminal region (14, 15). The C-terminal fragment remains associated with the N-terminal fragment (16). Human \(\beta\)-Gal forms a multienzyme complex with protective protein/cathepsin A and neuraminidase. Complex formation with other proteins is important for the proper processing and activity of \(\beta\)-Gal (17).

Human \(\beta\)-Gal is classified into glycoside hydrolase (GH) family 35 based on its amino acid sequence similarities (18). It is predicted to have the TIM barrel domain, which is typical for GH family members (19, 20). Human \(\beta\)-Gal is a retaining glycosidase in which the product retains the same stereochemistry as the starting substrate due to a double-displacement reaction mechanism whereby two consecutive nucleophilic attacks on the anomeric carbon lead to overall retention of the anomeric configuration (13). A pair of carboxylic acids is necessary for this reaction; one carboxylic acid acts as a catalytic nucleophile, and the other acts as an acid/base catalyst. In human \(\beta\)-Gal, Glu-268 has been identified as the catalytic nucleophile, and Glu-188 is a candidate for the acid/base catalytic (22). These residues are located in the fourth and seventh \(\beta\)-strand of the
Crystal Structure of Human β-Galactosidase

TIM barrel domain, respectively. For this reason, human β-Gal belongs to the 4/7 superfamily (23).

Alternatively spliced products of human β-Gal gene have been reported. Two mRNA lengths exist in the transcripts of the human β-Gal gene in human fibroblast cells; one is human β-Gal, and the other is the elastin binding receptor (24, 25). To date, the structures of several β-Gal enzymes have been reported, including those from Arthrobacter sp. C2-2 (C221 β-Gal) (26), Escherichia coli (EC β-Gal) (27), Thermus thermophilus A4 (A4 β-Gal) (28), Bacteroides thetaiotaomicron (BT β-Gal) (PDB ID 3D3A), Penicillium sp. (Psp β-Gal) (29), Trichoderma reesei (Tr β-Gal) (30), and Sulfolobus solfataricus (Sul β-Gal) (31). Psp β-Gal (Eukaryota), Tr β-Gal (Eukaryota), and S. solfataricus β-Gal (Archaea) are non-bacterial, but no β-Gal structures from mammals have so far been reported. Studies are needed to elucidate the effects of the various pathogenic mutations in β-Gal in terms of the protein structure as well as to gain insights into the structural characterization and catalytic mechanism of β-Gal. Here we report the first crystal structures of human β-Gal in complex with its catalytic product galactose and with its inhibitor 1-deoxygalactonojirimycin (DGJ).

EXPERIMENTAL PROCEDURES

Protein Expression, Purification, and Crystallization—The details of protein expression, purification, and crystallization are published (32). Briefly, human β-Gal (residues 24–677) was expressed in yeast Pichia pastoris and was purified to homogeneity. In the course of purification, polysaccharide moieties attached to the protein were trimmed off by endoglycosidase Hf treatment. For crystallization, β-Gal was subjected to limited proteolysis with bovine trypsin and was further purified. Crystallization experiments were performed with sitting-drop vapor-diffusion methods at 4 °C. Crystals of the DGJ complex were obtained with the reservoir solution containing 20% (w/v) PEG3350, 0.2 M ammonium sulfate, and 100 mM Tris HCl (pH 8.0). To prepare crystals of the β-Gal-galactose complex, crystals of the β-Gal-DGJ complex were transferred to and incubated overnight in the mother solution (25% PEG3350, 0.2 M ammonium sulfate, 100 mM Tris HCl (pH 8.0)) supplemented with 200 mM galactose.

Data Collection and Structure Determination—Diffraction datasets were collected at beamline NW12A of the Photon Factory (Tsukuba, Japan) under cryogenic conditions at 95 K. The crystals of the galactose complex were cryoprotected with the mother solution supplemented with 200 mM galactose and 15% ethylene glycol. The crystals of the β-Gal-DGJ complex were cryoprotected with the mother solution supplemented with 1 mM DGJ and 15% ethylene glycol. The datasets were processed with the HKL2000 package (33), and further analyses were carried out using the CCP4 suite (34). Structures of human β-Gal were determined with the molecular replacement method by using the program Molrep (35) and the coordinates of BT β-Gal (PDB code 3D3A) that has 32% sequence identity with human β-Gal. The structural models of the β-Gal-galactose complex and the β-Gal-DGJ complex were refined at a 1.8-Å resolution with stepwise cycles of a manual model-building using program COOT (36) and of restrained refinement using REFMAC (37) until the R factor was converged.

The final β-Gal-galactose complex structural model contains 4 ligand molecules, 4 chloride ions, 8 sulfate ions, 8 ethylene glycol molecules, 16 N-acetylgalcosamin (GlcNAc) residues, and 2244 water molecules. The β-Gal-DGJ complex structural model contains the same nonprotein atoms as the β-Gal-galactose complex, except that it contains 2185 water molecules.

The qualities of the final models were evaluated with PROCHECK (38). The most favored and the additionally allowed regions in the Ramachandran plot were 95.0 and 4.2%, respectively, for the β-Gal-galactose complex. Residues in the disallowed regions included Lys-498 (chains A-D), Asp-508 (chains A-D), Cys-127 (chains A-D), Glu-186 (chains A-D), Asn-458 (chains A and B), and Ser-611 (chain A) in the β-Gal-galactose complex; these residues were well fitted to the electron density maps. The β-Gal-DGJ complex had nearly the same quality as the β-Gal-galactose complex.

The refinement statistics are summarized in Table 1. Figures showing structures were prepared with PyMOL (39). The coordinates of the β-Gal-galactose complex and β-Gal-DGJ complex have been deposited, respectively, with the Protein Data Bank (codes 3THC and 3THD).

G_{M1} Ganglioside and Morquio B Disease Mutations—G_{M1} ganglioside and Morquio B disease mutations were obtained from the UniProt data base (40). The accessible surface areas (ASAs) shown in Table 2 were calculated by using ArealMol in the CCP4 suite (34). The ASA values in Table 2 are for chain A of the β-Gal-galactose complex.
| Disease phenotype | Mutation | Location* | ASA (Å²) | Importance in β-galactosidase | Proposed effect of mutations | Disease phenotype | Mutation | Location* | ASA (Å²) | Importance in β-galactosidase | Proposed effect of mutations |
|------------------|---------|----------|---------|-----------------------------|-------------------------------|------------------|---------|----------|---------|-----------------------------|-------------------------------|
| GM1G1            | P10L    | signal sequence | 45.1    | H-bonds to L207 and A301    |                              | P546L            | Inter   | 17       | Tightly buried, restricted phs, little room for Leu |                              |
|                  | R49C    | Surface   | 3.3     | H-bonds to H56, E128, E129, and N213 |                              | K578R            | Inter   | 2.8      | Tightly buried, ion pair to E478, little room for Arg |                              |
|                  | R592H   | Intra     | 0.8     | H-bonds to E268, Y83, Q184, and Q81 |                              | R592H            | Inter   | 0        | Tightly buried, ion pair to E478, H-bonds to N318 and M480 |                              |
|                  | R121S   | Inter     | 0.4     | H-bonds to Y125, E186, E149, and D153 |                              | R592H            | Inter   | 5.4      | Tightly buried, ion pair to E478, H-bonds to N318 and M480 |                              |
|                  | M132T   | Intra     | 0       | No room for side chain |                              | R592H            | Surface  | 5.4      | Tightly buried, ion pair to E478, H-bonds to N318 and M480 |                              |
|                  | G134V   | Intra     | 0       | No room for side chain |                              | R592H            | Surface  | 5.4      | Tightly buried, ion pair to E478, H-bonds to N318 and M480 |                              |
|                  | P136S   | Intra     | 1.5     | Partially buried |                              | R81W              | Intra   | 1        | Tightly buried, stacked on Y64 |                              |
|                  | L147R   | Intra     | 2.3     | Partially buried |                              | R81W              | Intra   | 2.3      | Tightly buried, stacked on Y64 |                              |
|                  | R148S   | Intra     | 0.8     | Ion pair to E186, H-bonds to N187, and stacked on Y125 |                              | R201C            | Surface  | 118.9    | Ion pair to D198 |                              |
|                  | D151V/Y | Surface   | 18.8    | Ion pair to Y125, H-bonds to L146 and D153 |                              | R201C            | Surface  | 27.8     | Stacked between F252 and R256 |                              |
|                  | L162S   | Intra     | 0       | Partially buried |                              | L201S            | Surface  | 27.8     | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | L173P   | Intra     | 1.6     | Partially buried |                              | Y333H            | Ligand   | 5.8      | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | Q164R   | Intra     | 0.1     | H-bonds to R211, E186, N187, and T220 |                              | R592H            | Inter   | 5.4      | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | G190D   | Surface   | 37.3    | Restricted phs in turn, no room for Asp, restricted phs in turn, no room for Asp |                              | R592H            | Inter   | 5.4      | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | Y199C   | Intra     | 0       | No room for side chain |                              | R592H            | Surface  | 142.5    | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | R201C   | Surface   | 118.9   | Ion pair to D198 |                              | R592H            | Surface  | 142.5    | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | R208C   | Surface   | 57      | H-bonds to G212 and V215 |                              | R592H            | Surface  | 142.5    | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | V216A   | Intra     | 12      | Partially buried |                              | R592H            | Surface  | 142.5    | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | T239M   | Intra     | 0       | No room for side chain |                              | R592H            | Surface  | 142.5    | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | V240M   | Intra     | 0       | No room for side chain |                              | R592H            | Surface  | 142.5    | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | Q255H   | Intra     | 0       | No room for side chain |                              | R592H            | Surface  | 142.5    | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | G272D   | Intra     | 2       | Partially buried, no room for Asp |                              | R592H            | Surface  | 142.5    | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | H281Y   | Surface   | 36.8    | H-bonds to D275 |                              | R592H            | Surface  | 142.5    | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | Y316C   | Inter     | 4.4     | H-bonds to E517 |                              | R592H            | Surface  | 142.5    | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | N318H   | Inter     | 2.4     | H-bonds to M480 and R590 |                              | R592H            | Surface  | 142.5    | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | T329I   | Inter     | 3.9     | H-bonds to R590 |                              | R592H            | Surface  | 142.5    | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | D332E/N | Surface   | 16.6    | H-bonds to N334 and Q339 |                              | R592H            | Surface  | 142.5    | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | K346N   | Surface   | 8.3     | Ion pair to D334, H-bonds to T271, D84, and D334 |                              | R592H            | Surface  | 142.5    | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | Y347C   | Intra     | 0       | Tightly buried, H-bond to K71 |                              | R592H            | Surface  | 142.5    | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | 377–381 | missing   | 3.2     | Partially buried, H-bonds to T507 and D591 |                              | R592H            | Surface  | 142.5    | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | T420P   | Intra     | 3.2     | Partially buried, little room for Arg |                              | R592H            | Surface  | 142.5    | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | L422R   | Intra     | 3.5     | Partially buried, H-bonds to K493 and Q493 |                              | R592H            | Surface  | 142.5    | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | D441N   | Intra     | 4.1     | Partially buried, H-bonds to K493 and Q493 |                              | R592H            | Surface  | 142.5    | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | R442Q   | Inter     | 19.7    | Partially buried, H-bonds to E478 and E520 |                              | R592H            | Surface  | 142.5    | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | R482H   | Inter     | 0       | Partially buried, H-bonds to E478 and E520 |                              | R592H            | Surface  | 142.5    | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | R491N   | Surface   | 0       | Restricted phs in turn, no room for Asp |                              | R592H            | Surface  | 142.5    | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | G491Y   | Surface   | 0       | Little room for Tyr |                              | R592H            | Surface  | 142.5    | Tightly buried, restricted phs in turn, no room for Asp |                              |
|                  | Q494C   | Intra     | 0       | Restricted phs in turn, no room for Asp |                              | R592H            | Surface  | 142.5    | Tightly buried, restricted phs in turn, no room for Asp |                              |

* Ligand, residues involved in ligand recognition; inter, protein core residues in the interdomain; intra, protein core residues in the intradomain; surface, residues in the protein surface.
RESULTS

Identification of Trypsinized β-Gal—After trypsinization of the recombinant β-Gal, we obtained 50- and 20-kDa fragments that correspond to the N- and C-terminal domains of β-Gal, respectively (32). We performed N-terminal peptide sequencing of these fragments. The N-terminal amino acid sequence of the 50-kDa fragment was Asp-Ala-Thr-Gln-Arg, indicating that the signal sequence of β-Gal was cleaved after Arg-25. The first amino acid should have been Asn, but it was probably not read correctly due to glycosylation. On the other hand, the N-terminal sequence of the 20-kDa fragment was Asp-Ser-Gly-His-His, indicating that the 20-kDa fragment started at Asp-531 and was cleaved after Arg-530. We also conducted mass spectrometry on the C-terminal 20-kDa fragment (Fig. 1B). The peak at m/z 15025.1 is in good agreement with the theoretical values of 15138.8 (residues 531–660), confirming that 17 residues from the C-terminal end of the 20-kDa fragment had been removed. We could not obtain mass spectrometry results for the 50-kDa fragment because of technical problems, but the 50-kDa fragment should be composed of residues 26–530, because Arg-530 was visible in the electron density map. Taken together, the results indicate that the 50- and 20-kDa fragments corresponded to the N-terminal domain (residues 26–530) and C-terminal domain (residues 531–660) of β-Gal, respectively. It has been reported that the precursor of human β-Gal is proteolytically cleaved between Ser-543—Ser-544 and that the C-terminal domain (residues 531–660) of β-Gal corresponded to the N-terminal domain (residues 26–530) and together, the results indicate that the 50- and 20-kDa fragments because Arg-530 was visible in the electron density map. Taken together, the results indicate that the 50- and 20-kDa fragments corresponded to the N-terminal domain (residues 26–530) and C-terminal domain (residues 531–660) of β-Gal, respectively. It has been reported that the precursor of human β-Gal is proteolytically cleaved between Ser-543—Ser-544 and that the large domain is further processed at Arg-530 (16). Thus, the trypsinized β-Gal used in this study closely corresponds to the mature form of β-Gal (Fig. 1A).

Structure of Human β-Gal—The deglycosylated and trypsinized forms of human β-Gal in complex with its inhibitor, DGJ, were crystallized in the monoclinic space group P2₁. Initial crystals were obtained in the form of the DGJ complex with the cocystalization method. Crystals of the complex with galactose were obtained by replacement of DGJ with galactose by the soaking method.

All four molecules in the asymmetric units of the β-Gal-galactose and β-Gal-DGJ complexes were well superimposed, with root mean square deviation (r.m.s.d.) values ranging from 0.2 to 0.4 Å. The β-Gal-galactose complex showed a strong overall agreement with the β-Gal-DGJ complex, with r.m.s.d. values from 0.1 to 0.3 Å. Therefore, the structure of the β-Gal-galactose complex (chain A) is described throughout this report, unless otherwise stated.

The crystallographic asymmetric unit contained four β-Gal molecules (molecules A to D). Molecules A and B were composed of two polypeptide fragments of residues 29–530 and 545–647. Molecules C and D were composed of two polypeptide fragments of residues 29–527 and 545–647. Residues 531–544 were not visible in the electron density maps of the four β-Gal monomers in asymmetric units; therefore, these residues were not modeled.

The crystal structure showed that Arg-530 and Asn545 are separated by about 40 Å, and they point in opposite directions. Therefore, it seems that these residues were connected by a flexible loop before trypsinization. Residues after Ser-647 were not visible in the electron density map.

The overall structure of the β-Gal monomer was flat, with dimensions of ~75 × 50 × 30 Å. β-Gal was folded into three domains: the TIM barrel domain, also called (α/β)₈ (residues 1–359); β-domain 1 (residues 397–514); β-domain 2 (residues 545–647) (Fig. 1, A and C, and 2A). The TIM barrel domain formed an eight-stranded α/β barrel structure that is responsible for catalysis and is characteristic of the GH family (19). One galactose or DGJ was located in the bottom of the barrel for the β-Gal-galactose or β-Gal-DGJ complex, respectively. The β-domain 1 and β-domain 2 were composed of four- and six-stranded β-sheets, respectively. The TIM barrel domain and β-domain 1 were connected with a loop region of ~50 Å (residues 360–396, hereafter referred to as the TIM-β1 loop). This loop was across β-domain 2, and residues 372–384 in the loop region formed an additional β-strand with β-domain 2.

Human β-Gal possessed seven potential N-glycosylation sites (Asn-26, Asn-247, Asn-464, Asn-498, Asn-542, Asn-545, and Asn-555). Based on the electron density, four GlcNACs for Asn-247, Asn-464, Asn-498, and Asn-555 were modeled in each β-Gal monomer. Asn-247 was located in the TIM barrel domain, Asn-464 and Asn-498 were in β-domain 1, and Asn-555 was in β-domain 2. All of these residues were found to point away from the ligand binding pocket. Therefore, the glycosylations presumably mediate the secretion and protection of β-Gal rather than its ligand binding. Asn-26 and Asn-542 were not included in the model due to a lack of electron density. The electron density around Asn-545 was found to be too poor to make a sugar chain. We were unable to determine whether these residues were glycosylated.

Based on the electron density, we modeled one chloride ion, two sulfate ions, and two ethylene glycol molecules in each β-Gal monomer. The chloride ion was located just beneath the ligand binding pocket and was surrounded by Tyr-83, Tyr-306, Arg-121, Gln-81, His-56, and Ser-54. This result suggests that this ion plays a structural role in forming the ligand binding pocket.

Dimeric Structure in Crystal—Four molecules in the asymmetric unit were attributable to two pairs of dimers (Fig. 2B). The dimerization of human β-Gal in solution was confirmed by gel-filtration chromatography (Fig. 2C). Fig. 2B shows that the two protomers of each dimer are related by a noncrystallographic 2-fold symmetry axis running parallel to the depth direction of the β-Gal monomer. The dimer had approximate dimensions of 75 × 100 × 30 Å. The dimerization interface was composed of three regions: residues 364–374 in the TIM-β1 loop, residues 559–567 in β-domain 2 (β2 loop), and residues 63–70 in the first helix in the TIM barrel domain (α1). These three regions formed a concave surface in the lateral face of the protomer. The β2 loop of one protomer inserted into the dimerization interface of the other protomer through extensive van der Waals interactions and vice versa. The interface was rich in hydrophobic residues, especially Pro residues, including Pro-559, Pro-563, and Pro-566 in the β2 loop and Pro-364, Pro-366, Pro-367, and Pro-370 in the TIM-β1 loop (Fig. 2D). The surface area in the interface was 904 Å², which is slightly small for a dimerization interface.

Domain Organization of Human β-Gal—Among the reported crystal structures of β-Gal, BT β-Gal, Tr β-Gal, and...
FIGURE 1. Overall architecture of human β-Gal. A, domain organization of β-Gal is shown. N-Glycosylation sites are indicated with arrows. Schematic representations of mature form of β-Gal and of trypsinized recombinant β-GAL are shown below. B, mass spectrometry data of the 20-kDa fragment is shown. The observed peak is in good agreement with the theoretical value. C, sequence alignment and secondary structure assignment of β-Gal are shown. The secondary structures of human β-Gal are shown above the aligned sequences. The α helices and β strands are indicated with cylinders and orange arrows, respectively. The alignment was carried out with ClustalW (49). Residues are colored red for the highest similarity, then green, blue, and black for the lowest.

Crystal Structure of Human β-Galactosidase
JANUARY 13, 2012• VOLUME 287 • NUMBER 3
JOURNAL OF BIOLOGICAL CHEMISTRY
1805
Crystal Structure of Human β-Galactosidase

(A) β-domain 1

(B) β-domain 2

(C) TIM barrel domain

(D) β-domain 2

TIM barrel domain

β-domain 1

NAG2(N464)
NAG3(N498)
NAG4(N555)
TIM barrel domain

β2 loop

β-domain 1 loop

TIM-β1 loop

Galactose

β-domain 2

TIM barrel domain

β-domain 2

440K 134K 67K

(mL)
Psp β-Gal belong to the GH35 family. Human β-Gal showed 32% sequence identity with BT β-Gal. The structure of human β-Gal resembled that of BT β-Gal (779 amino acids, monomer) (PDB code 3D3A), with an r.m.s.d. of 1.2 Å for 474 C atoms. Both proteins showed similar domain organizations, with a TIM barrel domain and two β-domains. Psp β-Gal (1011 amino acids, monomer) and Tr β-Gal (1023 amino acids, monomer) had 23 and 16% overall sequence similarities with human β-Gal, respectively. Psp β-Gal and Tr β-Gal also showed domain organizations similar to that of human β-Gal but possessed two or three additional β-domains (29, 30) (Fig. 3) whose functional roles remain to be elucidated.

Other bacterial β-Gal structures differed considerably in their oligomerization state and domain organization, and their similarity with human β-Gal was limited mainly to the TIM barrel domain. C221 β-Gal (1,023 amino acids, hexamer) belongs to the GH2 family. Its active site opens to the central cavity of the hexamer and is connected by eight channels with exterior solvent (26). EC β-Gal (1,024 amino acids, tetramer) belongs to the GH2 family, and it is the most structurally and biochemically well characterized compound among the β-Gals. Each active site is complemented by a loop of neighboring molecules that is inserted into the active site (27). In human β-Gal, the loop region of β-domain 2 (residues 482–491) plays the same role as the complementation loop of EC β-Gal.

A4 β-Gal (645 amino acids, trimer) belongs to the GH42 family. Each monomer of A4 β-Gal has an active site located inside a large central tunnel formed by the interface of the trimer (28). The active site of human β-Gal is easily accessible from the bulk solvent, in contrast to the active sites of A4 β-Gal, C221 β-Gal, and EC β-Gal, which are segregated from the bulk solvent.

**Ligand Recognition and Catalytic Mechanism**—The catalytic residues, Glu-268 and Glu-188, were located in a deep well in the TIM barrel domain. The electron densities corresponding to the bound ligands in the β-Gal-galactose and β-Gal-DGJ complexes were clear enough to generate unambiguous models. Fig. 4A shows that a galactose molecule is bound to each monomer in the chair conformation, and its 1-OH group has the β-anomer configuration. All of the hydroxyl groups of the ligands made hydrogen bonds with β-Gal; 10 and 8 direct hydrogen bonds were formed between galactose and β-Gal and between DGJ and β-Gal, respectively (Fig. 4. A–D, supplemental Table S1). Additional hydrogen bonds mediated by a couple of water molecules were formed between the ligands and β-Gal. Moreover, aromatic and hydrophobic residues in the active site contributed to ligand recognition via extensive van der Waals interactions (Fig. 4B). These interactions explain how β-Gal specifically recognizes the terminal galactose molecule.

In the β-Gal-galactose complex, the 1-OH group was hydrogen-bonded to the carboxylic group of Glu-188 and, concurrently, to a water molecule. Previously, Glu-268 was reported to be a nucleophile and Glu-188 to be an acid/base catalyst. The reactions proceed with a double-displacement reaction (22). In the crystal structure the configuration of the active site residues around the galactose was suitable for the nucleophilic attack of the Glu-268 carboxylate to the C1 position of galactose. The side chain of Glu-268 was oriented correctly against the substrate by forming hydrogen bonds to Tyr-270, Arg-121, and Tyr-83. In the course of the nucleophilic attack of Glu-268, Glu-188 acts as an acid catalyst. Subsequently, Glu-188 acts as a base, activating a water molecule that, in turn, attacks the C1 position of galactose and leaves the galactose residue.

Glycoside hydrolases are classified as retaining or inverting enzymes, depending on the conservation of the conformation of the anomeric position through the reaction (41). The difference between these classifications is determined by the distance between the oxygens of the two catalytic carboxylates; this distance ranges 4.5–6.5 Å in the retaining enzyme and 9.0–9.5 Å in the inverting enzyme (42). The retaining enzyme forms a covalent intermediate with its substrate, whereas the inverting enzyme activates water molecules to hydrolyze the substrate. In the structure of human β-Gal, the average value of the distance for the four protomers was 5.1 Å, confirming that human β-Gal hydrolyzes its substrate in a retaining manner (13).

**DISCUSSION**

In this study, we determined the first crystal structure of human β-d-galactosidase. Our structural study revealed that human β-Gal showed domain organization that is distinct from those of previously reported β-Gal structures.

**Homodimer of Human β-Gal**—Several groups have reported that mature β-Gal isolated from mammalian tissues forms a dimer (43–45) or a monomer (46). β-Gal has also been isolated as a dimer or tetramer (16). These discrepancies could be due to...
differences in the experimental procedures employed and in particular to differences in pH. Porcine β-Gal exhibits pH-dependent oligomerization; it adopts a monomeric form at neutral pH (pH 7.0) and reversibly associates to form a dimer at acidic pH (43). Considering that the optimal pH for activity is around 4.5, the active form is most probably a dimer. Norden et al. (46) reported that β-Gal is monomeric, but the experiment was performed at neutral pH.

From the structure, enzyme treatments such as with glycosidase and trypsin would be unlikely to affect oligomerization because the glycosylation sites and cleavage site is located far from the dimerization region. Taken together, these reports strongly suggest that β-Gal would form a dimer under physiological conditions, but a definitive demonstration of its oligomeric structure must await further investigation.

**Substrate Binding Mode**—Domain analysis showed that β-Gal has no known protein motifs other than the catalytic domain. A structural similarity search using DALI (47) revealed that both β-domains share structural homology with β-domains of other galactosidases with high homology scores (>7). Interestingly, β-domains show a weak homology to the carbohydrate binding module 35 (CBM35) (48). The comparison of β-domain 1 and CBM35 (PDB code 2W87) yielded a Z-score of 5.3, an r.m.s.d. value of 2.8 Å for 82/138 residues, and a sequence identity of 7%. The comparison of β-domain 2 and CBM35 (PDB code 2W87) yielded a Z-score of 4.9, an r.m.s.d value of 2.6 Å for 74/138 residues, and a sequence identity of 12%. In addition to the low homology, the topology of both structures is different, and the Trp residue that is important for ligand recognition in CBM35 is not conserved in the β-domains of human β-Gal. Therefore, the β-domains of human β-Gal are unlikely to have an affinity for sugar.

**Mutations Causing GM1 Ganglioside and Morquio B Disease**—Based on the structure, we can demonstrate structural insights into lysosomal storage diseases mutations. Nearly 100 gene mutations causing GM1 ganglioside and Morquio B disease are reported in the UniProt data base (40). We mapped these mutations onto the structure of human β-Gal to elucidate relationships between the location of the mutation and its effect on the activity of β-Gal (Fig. 5, A and B, Table 2). The mutations, which were scattered throughout the β-Gal structure, were classified as mutations directly affecting the ligand recognition, mutations inside the protein core, or mutations located in the protein surface (Fig. 5B).
Crystal Structure of Human β-Galactosidase

Mutations Directly Affecting Ligand Recognition—Among the disease-related mutations, only Tyr-270 (Y270D), Tyr-83 (Y83C, Y83H), Trp-273 (W273L), and Tyr-333 (Y333H) directly contributed to the ligand recognition (Fig. 6A). The side chain of Trp-273 was found to lie in the entrance of the ligand binding pocket, tuning the shape of the pocket and partially covering the pocket by interacting with the ligand. When Trp-273 was mutated to a residue with a smaller side chain, such as His, it could not effectively cover the ligand in the proper position. The side chain of Tyr-83 was almost buried in the bottom of the ligand binding pocket, with its phenol ring stacked on the side chains of Arg-121, His-56, and Ile-126. Moreover, the OH group of the side chain was hydrogen-bonded with Arg-121, Ile-126O, Glu-268, Asn-187, and O3 of galactose. The side chain of Tyr-333 formed part of the lateral face of the ligand binding pocket, and its OH group interacted with O6 of galactose. Tyr-270 made a water-mediated interaction with O1 of galactose and also hydrogen-bonded with Glu-268, thereby fixing the catalytic carboxylate of Glu-268 to the appropriate position for the catalytic reactions. Collectively, mutations at these sites disrupted the shape of the ligand binding pocket. Such disruptions caused reduced affinity and/or catalytic efficiency, leading to dysfunction of the enzyme.

Mutations Inside Protein Core—Mutations inside the protein core, which account for 60% of all mutations, can be subdivided into mutations within the domain and mutations located in the domain interface.

Mutations of Arg-59 to Cys or His are known to cause the infantile forms of G\textsubscript{M1} gangliosidosis. Arg-59 was almost totally buried inside the protein core, as shown by an ASA of 3.1 Å\(^2\) near the bottom of the ligand binding pocket (7 Å to the ligand). Its side chain formed hydrogen bonds with His-56 and Asn-313 and with the carbonyl oxygen atoms of Ala-128 and Glu-129 (Fig. 6B). Moreover, the guanidino moiety of Arg-59 was stacked between the side chains of Trp-130 and Tyr-331. These interactions stabilized the structure of the ligand binding pocket. When mutated to Cys or His, which have smaller side chains, these stabilizing interactions were disrupted, and the shape of the ligand binding pocket was changed.

Arg-121 (R121S, infantile G\textsubscript{M1} gangliosidosis) was also buried near the ligand binding pocket (5 Å to the ligand), where it played a structural role in maintaining the pocket. Arg-590 (R590C, infantile G\textsubscript{M1} gangliosidosis, ASA of 5.4 Å\(^2\)) in β-domain 2 linked three domains of β-Gal by making contacts with Glu-478 in β-domain 1, Asn-318 in the TIM barrel domain, and the carbonyl oxygen atom of Asn-479 in β-domain 1 (Fig. 6C). Tyr-316, Asn-318, Thr-329, Arg-442, Arg-482, Lys-578, and Tyr-591 also made interdomain interactions. Therefore, when these residues were mutated, the appropriate domain organization was disrupted, which destabilized the mutant proteins and/or lost the surface to the neuraminidase/cathepsin to form multiprotein complex to transport properly to the lysosomes. These observations support the idea that the three-domain organization of human β-Gal is indispensable for it to exert its normal function in lysosomes.

Mutations Located in Protein Surface—In contrast to mutations in the protein core region, which cause large structural rearrangements of the protein, mutations located in the protein surface had milder effects on the protein; the effect of such mutations could be balanced by local rearrangement of the nearby residues. For example, Arg-201 (R201C, infantile G\textsubscript{M1} gangliosidosis, ASA of 118.9 Å\(^2\)) was located on the lateral face of the TIM barrel domain, which is far from the ligand binding pocket (22 Å to the ligand). The side chain of Arg201 was completely exposed to the solvent and formed a salt-bridge to Asp-198, which is a known cause of the Morquio B mutation (D198Y) (Fig. 6D). No large structural rearrangement occurred in either mutation except for loss of the salt bridge.

R208C is also a known G\textsubscript{M1} gangliosidosis mutation. In this case, the side chain of Arg-208 hydrogen-bonded to the carbonyl oxygen atoms of Gly-212, Val-215, and two water molecules. These interactions stabilized the loop structure, connecting an α4 helix and β5 strand (Fig. 6D). Such muta-
tions in the solvent-exposed regions caused local structural rearrangements and changed the surface properties of the protein, such as the electrostatic potentials and interface to other proteins.

Rationale for Effect of Mutations—In general, we observed that the residues found in the G M1 gangliosidosis and Morquio B diseases do not tend to be very solvent-exposed. Most mutations in the protein core region would cause drastic effects on the overall protein structure and stability, especially for the TIM barrel domain that is responsible for catalysis. Indeed, infantile GM1 ganglioside mutations, which are mutations associated with the most detrimental phenotype, were concentrated in the protein core regions. Mutations associated with milder phenotypes, such as juvenile or adult GM1 gangliosidosis, tended to be exposed to the solvent. The distribution of the mutations associated with Morquio B disease were somewhat biased toward the vicinity of the ligand binding pocket and α-domain 2. However, the effects of some mutations with the most detrimental phenotype were difficult to explain because they were located on the surface and probably caused little structural change. With the structure of human β-Gal presented here, it may become possible to locate the region of the defect causing the development of disease symptoms and to predict, to some extent, the severity of the disease from the three-dimensional protein structure.

Acknowledgments—We thank the staff for data collection at the photon factory. We thank Dr. H. Sakuraba for kindly providing the human β-Gal cDNA.

REFERENCES
1. Alpers, D. H. (1969) Separation and isolation of rat and human intestinal β-galactosidases. J. Biol. Chem. 244, 1238–1246
2. Asp, N. G., and Dahlqvist, A. (1972) Human small intestine galactosidases. Specific assay of three different enzymes. Anal. Biochem. 47, 527–538
3. Distler, J. I., and Jourdian, G. W. (1973) The purification and properties of β-galactosidase from bovine testes. J. Biol. Chem. 248, 6772–6780
4. Oshima, A., Yoshida, K., Shimmoto, M., Fukuhara, Y., Sakuraba, H., and Suzuki, Y. (1991) Human β-galactosidase gene mutations in Morquio B disease. Am. J. Hum. Genet. 49, 1091–1093
5. Yoshida, K., Oshima, A., Shimmoto, M., Fukuhara, Y., Sakuraba, H.,

FIGURE 6. Details of the representative disease-related mutations. Hydrogen bonds are shown as dashed lines. The color scheme is the same as that in Fig. 5A, except for C, where the color scheme is the same as that in Fig. 2B, and the mutations associated with disease are indicated with colored boxes to show the domain interface. A, shown are mutations directly interacting with ligand. B, Arg-59 in the protein core in the TIM barrel domain is shown. C, mutations in the protein core in the domain interface are shown. D, mutations in the protein surface are shown.
Crystal Structure of Human β-Galactosidase

Yanagisawa, N., and Suzuki, Y. (1991) Human β-galactosidase gene mutations in Gm1-gangliosidoses. A common mutation among Japanese adult/chronic cases. *Am. J. Hum. Genet.* 49, 435–442

6. Callahan, J. W. (1999) Molecular basis of Gm1 gangliosidosis and Morquio disease, type B. Structure-function studies of lysosomal β-galactosidase and the non-lysosomal β-galactosidase-like protein. *Biochim. Biophys. Acta* 1455, 85–103

7. Suzuki, Y., Oshima, A., and Namba, E. (2001) β-Galactosidase deficiency (β-galactosidosis). Gm1 gangliosidosis and Morquio B disease in *The Metabolic and Molecular Bases of Inherited Disease* (Scrivar, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., and eds) 8th Ed., pp. 3775–3809, McGraw-Hill Inc., New York

8. Caciotti, A., Garman, S. C., Rivera-Colón, Y., Procopio, E., Prodi, C., Brunetti, M., and Scaglia, F. (2008) GM1 gangliosidosis. Review of *Can. J. Biochem. Cell Biol.* 62, 529–534

11. Oshima, A., Tsuji, A., Nagao, Y., Sakuraba, H., and Suzuki, Y. (1988) Cloning, sequencing, and expression of cDNA for human B-galactosidase. *Biochim. Biophys. Acta* 138, 238–244

12. Yamamoto, Y., Hake, C. A., Martin, B. M., Kretz, K. A., Ahern-Rindell, S. G., and Callahan, J. W. (1994) Kinetic mechanism and characterization of the human B-galactosidase precursor secreted by permanently transfected Chinese hamster ovary cells. *Biochem. J.* 304, 281–288

15. Hoogeveen, A. T., Verheijen, F. W., and Galjaard, H. (1983) The relation between human lysosomal β-galactosidase and its protective protein. *J. Biol. Chem.* 258, 12143–12146

18. Yamamoto, Y., Hake, C. A., Martin, B. M., Kretz, K. A., Ahern-Rindell, A. J., Naylor, S. L., Mudd, M., and O’Brien, J. S. (1990) Isolation, characterization, and mapping of a human β-galactosidase cDNA. *DNA Cell Biol.* 9, 119–127

19. Zhang, S., McCarter, J. D., Okamura-Oho, Y., Yaghi, F., Hinek, A., Withers, S. G., and Callahan, J. W. (1994) Kinetic mechanism and characterization of human β-galactosidase precursor secreted by permanently transfected Chinese hamster ovary cells. *Biochem. J.* 304, 281–288

22. McCarter, J. D., Burgoyne, D. L., Miao, S., Zhang, S., Callahan, J. W., and Withers, S. G. (1997) Identification of Glu-268 as the catalytic nucleophile of B-galactosidase. The C-terminal precursor fragment is an essential domain of the mature enzyme. *J. Biol. Chem.* 272, 10035–10040

24. Morreau, H., Galjart, N., de Vrij, R., Willemsen, R., van der Horst, G. T., and d’Azzo, A. (1989) Alternative splicing of β-galactosidase mRNA generates the classic lysosomal enzyme and a β-galactosidase-related protein. *J. Biol. Chem.* 264, 20655–20663

25. Privitera, S., Prody, C. A., Callahan, J. W., and Hinek, A. (1998) The 67-kDa enzymatically inactive alternatively spliced variant of β-galactosidase is identical to the elastin/laminin-binding protein. *J. Biol. Chem.* 273, 6319–6326

27. Jacobson, R. H., Zhang, X. J., DuBose, R. F., and Matthews, B. W. (1994) Three-dimensional structure of β-galactosidase from *E. coli*. *Nature* 369, 761–766

28. Usui, K., Ohto, U., Ochi, T., Shimizu, T., and Satow, Y. (2011) Crystallography of the glycoside hydrolyase family 42 β-galactosidase from *Thermus thermophilus* A4 and the structure of its complex with galactose. *J. Biol. Chem.* 332, 79–91

34. Aguilar, C. F., Sanderson, I., Moracci, M., Caramella, M., Nucci, R., Rossi, M., and Pearl, L. H. (1997) Crystal structure of the β-glycosidase from the hyperthermophilic archean *Sulfilobus solfataricus*. Resilience as a key factor in thermostability. *J. Mol. Biol.* 271, 789–802

39. Usui, K., Ohto, U., Ochi, T., Shimizu, T., and Satow, Y. (2011) *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, in press

42. McCarter, J. D., and Withers, S. G. (1994) Mechanisms of enzymatic glycoside hydrolysis. *J. Mol. Biol.* 240–255

43. Yamamoto, Y., Fujie, M., and Nishimura, K. (1982) The interrelation between high and low molecular weight forms of GM1-galactosidase. *Biochem. J.* 20655–20663

46. 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, 760–763

47. Ringo, J. A., and Pelayo, A. J. (1997) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 30, 1022–1025

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

49. 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

50. Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* 26, 283–291

51. Delano, W. L. (2010) *The PyMOL Molecular Graphics System*, Version 1.3r1, Schrodinger, LLC, New York

52. Apte, A., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O’Donovan, C., Redaschi, N., and Yeh, L. S. (2004) UniProt, the Universal Protein knowledgebase. *Nucleic Acids Res.* 32, D115–D119

53. Henrisitat, B., and Davies, G. (1997) Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* 7, 637–644

54. McCarter, J. D., and Withers, S. G. (1994) Mechanisms of enzymatic glycoside hydrolysis. *Curr. Opin. Struct. Biol.* 4, 885–892

55. Yamamoto, Y., Fujie, M., and Nishimura, K. (1982) The interrelation between high and low molecular weight forms of Gm1-galactosidase purified from porcine spleen. *J. Biochem.* 92, 13–21

57. Jones, C. S., Mahuran, D., Lowden, J. A., and Callahan, J. W. (1984) Human placental β-galactosidase. Structural and immunological observations. *Can. J. Biochem. Cell Biol.* 62, 529–534

58. Hubbes, M., D’Agrosa, R. M., and Callahan, J. W. (1992) Human placental β-galactosidase. Characterization of the dimer and complex forms of the enzyme. *Biochem. J.* 285, 827–831
Crystal Structure of Human β-Galactosidase

46. Norden, A. G., Tennant, L. L., and O’Brien, J. S. (1974) G₄M₄ ganglioside β-galactosidase. A. Purification and studies of the enzyme from human liver. J. Biol. Chem. 249, 7969–7976
47. Holm, L., and Sander, C. (1995) Dali, a network tool for protein structure comparison. Trends Biochem. Sci. 20, 478–480
48. Montanier, C., van Bueren, A. L., Dumon, C., Flint, J. E., Correia, M. A., Prates, J. A., Firbank, S. J., Lewis, R. J., Grondin, G. G., Ghinet, M. G., Gloster, T. M., Herve, C., Knox, J. P., Talbot, B. G., Turkenburg, J. P., Kerovuo, J., Brzezinski, R., Fontes, C. M., Davies, G. J., Boraston, A. B., and Gilbert, H. J. (2009) Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. Proc. Natl. Acad. Sci. U.S.A. 106, 3065–3070
49. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) ClustalW. Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and weight matrix choice. Nucleic Acids Res. 22, 4673–4680